Evolution of PHDs as Oxygen Sensors: Mechanistic and Structural Studies of the PHD of *Trichoplax adhaerens*, the Simplest Animal, and Mechanistic Studies of a PHD-like Enzyme of the Protist *Monosiga brevicollis*

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



Anna Boleininger
Magdalen College
Trinity Term 2012

Anna Boleininger, Magdalen College
D. Phil. Thesis in Biochemistry, Trinity Term 2012

Evolution of PHDs as Oxygen Sensors: Mechanistic and Structural Studies of the PHD of *Trichoplax adhaerens*, the Simplest Animal, and Mechanistic Studies of a PHD-like Enzyme of the Protist *Monosiga brevicollis*

Abstract

This work aimed to investigate the evolutionary origin of the involvement of the HIF Prolyl Hydroxylases (PHDs) in oxygen sensing. The α/β-heterodimer HIF (Hypoxia Inducible Factor) is a master regulator of oxygen homoeostasis in metazoans. In the nucleus, HIF binds to the Hypoxia Responsive Elements and forms a transcriptional complex that activates the transcription of a multitude of downstream genes. Under normoxic conditions, the Fe(II)- and oxygen-dependent PHDs catalyse 4*R*-prolyl-hydroxylation of the HIF α-subunit, which subsequently leads to its degradation.

It had previously been proposed that the evolution of the HIF-pathway, shared by all metazoans but not found in other organisms, is linked to the rapid diversification of multicellular life during the Cambrian Explosion. This work investigates the structural and biochemical properties of a PHD of the basal metazoan *Trichoplax adhaerens* (taPHD), and a PHD-like enzyme of the protist *Monosiga brevicollis* (mbP4H).

Two crystal structures of taPHD were obtained (1.2-1.3 Å), one containing a *Trichoplax adhaerens* HIFα subunit peptide (taODD). Comparison with crystal structures of human PHD2 showed a high degree of conservation of structural features and enzyme-substrate interactions. The prolyl-residue of taODD, shown to be hydroxylated by taPHD, is occupying the C^4-*endo* conformation in the crystal structure, supporting the previously proposed mechanism of HIFα hydroxylation by PHD2 in humans. A conservation of biochemical properties with human PHD2, such as the formation of a stable enzyme-Fe(II)-2OG complex, was observed and could therefore be key to oxygen sensing by the PHDs.

mbP4H was shown to catalyse 4*R*-prolyl-hydroxylation of taODD. It was proposed that the native substrate of mbP4H is a protein containing a prolyl-hydroxylation site similar to taODD, possibly with a YXXLAP motif. The study of biochemical properties and substrate selectivity of mbP4H suggests that the precursor of PHDs may have had similar properties to mbP4H. Further work on mbP4H could therefore yield clues about the evolutionary origin of HIF-prolyl hydroxylases in oxygen sensing and probe the previously proposed connection between metazoan life and HIF–mediated oxygen sensing.

Acknowledgements

First and foremost, I would like to thank Chris Schofield for being a great scientist and supervisor. The amount of time and attention I have seen him devote to me and other group members are key to his unique leadership style, which shall be an inspiration for many years to come.

Mike McDonough deserves special thanks. Without his skills and expertise, none of the crystallography work would have been possible. Additionally, Mike has taken a great amount of time explaining concepts that were completely new to me at the time in great detail. His guidance and advice were crucial to the work presented in this thesis. Another person who certainly deserves special thanks is Monica Mantri. Being my bench neighbour, she was the first point of call for practical problems and considerations, and the many discussions we had have been a great help on a daily basis. Zhihong Zhang's demonstration of how to carry out streak seeding was key to this work and was very appreciated. James McCullagh's and Nikita Loik's work and advice on mass spectrometry experiments has also been crucial and I am very grateful to them. I would like to thank Adam Hardy for carrying out ESI-MS experiments, for the extensive thesis proofreads and, most importantly, for always being there when help and advice were needed. Richard Hopkinson has been very helpful whenever problems with the automated peptide synthesiser arose and I am greatly indebted to his commitment. Emily Flashman had always an open ear for questions and has greatly contributed to this work in form of discussions and thesis proofreads. Hanna Tarhonskaya deserves many thanks for her advice on kinetic studies and proofreads. Wei Shen Aik has spent a considerable amount of time and patience demonstrating me the in-gel tryptic digest and Western blotting and I am very grateful for this. Ruben Comez-Castellanos deserves special thanks for all the invaluable advice on protein purification techniques. Andrea Szollossi has helped me greatly by letting me use her NMR-buffers. I would also like to thank Inga Pfeffer for providing me with her expression vector. Discussions of crystal structures and PHD2 variants with Rashed Chowdhury have been very helpful, as has been his advice on crystallographic structure refinement. I am very grateful to Akane Kawamura and Louise Walport for thesis proofreads. Ivanhoe Leung has been a great help when it came to NMR experiments. Discussions with Refaat Hamed have always been inspirational and uplifting. Wendy Sobey deserves special thanks as she has always taken her time to repair or replace broken equipment and give general advice. Thanks to Wei Ge for giving me her last mince pie when I had a bad day. It has long been eaten but shall not be forgotten.

I am also very grateful to the BBSRC, who gave me this unique opportunity and provided support throughout my degree.

More importantly, however, I would like to express my gratitude to everyone from the Schofield group who supported me as a friend. These people made my experience truly unforgettable. I wish all the best to all of you on your way and hope to see many of you again in the future. Last but not least, I would like to thank Alastair for all of his advice, support, patience and the countless lifts to and from the lab at random hours.

Table of Contents

Page number

Contributors i
Abbreviations ii

Introduction

1. The HIF Pathway
1.1. The HIF pathway: Discovery and molecular mechanism 1
1.2. Clinical relevance of the HIF pathway 5
1.3. The role of PHDs 5

2. Evolution of the HIF Pathway
2.1. Ancient atmospheres and metazoan evolution 9
2.2. The HIF pathway in other organisms 11
2.3. The origin of the HIF pathway 12

3. *Trichoplax adhaerens*, the Simplest of all Extant Animals
3.1. Habitat and morphology of *Trichoplax adhaerens* 15
3.2. Phylogenetic assignment of placozoans 18
3.3. The HIF pathway in *Trichoplax adhaerens* 20

4. *Monosiga brevicollis*, a Close Unicellular Relative of the Metazoans
4.1. Habitat and morphology of *Monosiga brevicollis* 23
4.2. Phylogenetic assignment of choanoflagellates 24

5. References 26

Part A: Crystal Structures of the *Trichoplax Adhaerens* HIF Prolyl Hydroxylase

A.1. Insights Gained from the Crystal Structures of Human PHD2
A.1.1. Available crystal structures of human PHD2 37
A.1.2. Key structural features of hsPHD2 39
A.1.3. The structure of hsPHD2 in context of other 2OG-oxygenases 43
A.1.4. Summary 46

A.2. *Trichoplax adhaerens* PHD: Purification, Activity Assessment and Crystallisation
A.2.1. Sequence alignments 48
A.2.2. Expression and purification of a suitable taPHD construct 50

A.2.3. MS/MS analysis of purified taPHD 53
A.2.4. Detection of 2OG turnover by taPHD using ^{1}H-NMR 56
A.2.5. taPHD hydroxylates a *Trichoplax adhaerens* HIFα sequence in a 2OG- and ascorbate-dependent manner 61
A.2.6. Choice of inhibitors for co-crystallisation with taPHD 62
A.2.7. Choice of initial conditions for crystallization trials 65
A.2.8. Evaluation of initial hits and optimisation 65
A.2.9. Data collection and structure solution 67
A.2.10. Summary 69

A.3. Crystal Structure of taPHD with taHIFα (477-497)
A.3.1. Binding of taODD, hsCODD and hsNODD peptides to taPHD 71
A.3.2. Co-crystallisation of taPHD with taODD 73
A.3.4. Summary 76

A.4. Refinement and Comparison of the taPHD Crystal Structures to the Crystal Structures of hsPHD2
A.4.1. A brief introduction to crystallographic structure refinement 77
A.4.2. Crystallographic refinement of the crystal structures of taPHD 78
A.4.3. Comparison of the structures taPHD and taPHD.taODD 71
A.4.4. Comparison of the structures taPHD.taODD and hsPHD2.hsCODD 95
A.4.5. Conclusions drawn from the comparison of taPHD.taODD and hsPHD2.hsCODD 98
A.4.6. Comparison of taPHD to other 2OG-oxygenases 99

A.5. Summary of Section A 100
A.6. References 102

Part B: Mechanistic Studies of taPHD

B.1. Hydroxylation of taODD by taPHD: Stereochemistry and Coupling to 2OG Turnover
B.1.1. taPHD forms a stable complex with Fe(II) and 2OG 106
B.1.2. Coupling between substrate hydroxylation and 2OG turnover 107
B.1.3. Investigation of stereochemical requirements for hydroxylation 110
B.1.4. Amino acid analysis of hydroxylated taODD 113

B.2. The Kinetics of 2OG and Peptide Turnover by taPHD
B.2.1. K_m and k_{cat} values in context 115
B.2.2. Determination of K_m and k_{cat} values of taPHD for 2OG, taODD and hsCODD 117
B.2.3. Summary 123

B.3. Investigation of Enzyme-Substrate Interactions of taPHD and hsPHD2
B.3.1. HIFα peptides of various metazoan species 124
B.3.2. Determination of background +16 Da peaks by MALDI-TOF MS 123
B.3.3. Peptide hydroxylation time courses using MALDI-TOF MS 129
B.3.4. Apparent K_m and k_{cat} values of CODD-homologues for taPHD and hsPHD2 131
B.3.5. Summary 133

B.4. Kinetic Studies in the Context of the Crystal Structures and Phylogenetic Relationships
B.4.1. Rationalisation of observed trends using enzyme-substrate interactions identified in the crystal structures 134
B.4.2. Consideration of phylogenetic relationships 141
B.4.3. Summary 144

B.5. Investigating the Binding of NODD by Hydroxylation Assays with taPHD Variants
B.5.1. Interaction between $R281_{hsPHD2}$ and $D412_{hsHIF\text{-}1\alpha}$ of NODD 146
B.5.2. The clinically observed hsPHD2 mutation P317R 147

B.6. Summary of Section B 150
B.7. References 153

Part C: Studies of a PHD-like Enzyme from the Protist *Monosiga brevicollis*

C.1. Previously Described PHD-like Enzymes and mbP4H
C.1.1. *Dictyostelium discoideum* Skp1 hydroxylase ddP4H 156
C.1.2. *Perkinsus olseni* HPH 157
C.1.3. mbP4H, a PHD-like enzyme of *Monosiga brevicollis* 157
C.1.4. Sequence alignments of PHDs and PHD-like sequences 158
C.1.5. Summary 163

C.2. 2OG Turnover Assays and Substrate Screening with mbP4H
C.2.1. His-Tag Purification 164
C.2.2. Uncoupled 2OG-turnover by NMR 165
C.2.3. Substrate screening by MALDI-TOF MS 168
C.2.4. 2OG turnover in the presence of ODD peptides 170
C.2.5. Quantitative comparison of taODD turnover to succinate production 172
C.2.6. Attempts at observing the mbP4H.Fe(II).2OG complex 174
C.2.7. Summary and Discussion 174

C.3. Position and Stereochemistry of Hydroxylation by mbP4H
C.3.1. Hydroxylations carried out by 2OG oxygenases 176
C.3.2. MS/MS Analysis of modified taODD peptide 176
C.3.3. Amino Acid Analysis 180
C.3.4. Summary 182

C.4. Modification of mbP4H Substrate Range Through Residue Substitution
C.4.1. mbP4H Variant C359V 183
C.4.2. mbP4H Variants P239D and G256R 184
C.4.3. Screening results for mbP4H variants 187
C.4.4. The significance of the HXD motif 189
C.4.5. Summary 190

C.5. Crystallisation of mbP4H
C.5.1. Selection of screening conditions 192
C.5.2. Initial hits and optimization 193
C.5.3. Attempts at solving the structure: Halide and heavy metal soaks and preparation of seleno-methionine mbP4H 195
C.5.4. Summary 198

C.6. Strategies for Identifying the Native Substrate of mbP4H
C.6.1. HIF-pathway components in *Monosiga brevicollis* 199
C.6.2. Possible strategies 200
C.6.3. ODD-like sequences of mbP4H found by HMM searches 200
C.6.4. Preparation of FLAG-tagged mbP4H construct for pull-down experiments 201
C.6.5. Future work for substrate identification 202

C.7. Summary of Section C 204
C.8. References 207

Materials and Methods

1. Molecular Cloning Techniques
1.1. PCR for introduction of new restriction enzyme sites or for preparation of *N* - terminally truncated constructs 209
1.2. Visualisation of PCR products on agarose gel 210
1.3. Extraction of PCR bands from agarose gel 211
1.4. Quantification of purified DNA 211
1.5. DNA sequencing 211
1.6. Ligation of PCR products into a plasmid vector 211
1.7. Sub-cloning of an insert into a different expression vector using the same restriction sites 212
1.8. Site-directed mutagenesis 213

2. Protein Expression in *E. coli*
2.1. Expression trials 214
2.2. Large-scale protein expression 214

3. Protein Purification
3.1. Preparation of cell lysate for FPLC 215
3.2. His-affinity purification 215
3.3. GST-tag purification 216
3.4. Cation exchange 216
3.5. Size exclusion chromatography 217
3.6. SDS-PAGE analysis of purification fractions 217

4. In-Gel Tryptic Digest 218
5. Activity Assay Conditions 219
6. Quantification of 2OG Turnover by ^{1}H-NMR 220
7. Quantification of Peptide Hydroxylation by MALDI-TOF MS 221
8. Non-Denaturing Mass Spectrometry 222
9. Electrospray Ionisation Mass Spectrometry 223
10. Crystallisation Trials 224
11. Harvesting of Crystals, Data Collection and Data Processing 224

12. Peptide Synthesis and Purification
12.1. Synthesis 225
12.2. HPLC purification of peptides 226

13. MS/MS Analysis of Peptides 227
15. Determination of K_m Values and Initial Rates
15.1. K_m values 228
15.2. Initial rates 229

16. Preparation of Seleno-Methionine Substituted Protein 230
17. Soaking Crystals in Heavy Atom Solutions 230
18. Western Blotting 230

Appendix

Appendix 1: Sequences, Vectors and Primers 232
Appendix 2: Agarose and SDS-PAGE Gels 238
Appendix 3: Complementary Data 241
Appendix 4: Hydroxylated MS/MS Fragments 245
Appendix 5: Further Crystallisation Attempts 246

Contributors

Unless otherwise stated, all contributors are members of the Christopher Schofield group. The list below gives the names of people who have carried out experiments presented in this thesis. The corresponding experiments are listed in the right hand column.

Contributors: Experiments

Dr Michael A. McDonough	*All crystal harvesting, crystallographic data collection, structure solution and heavy atom soaking*
Dr Mukram Mackeen[1]	*MS/MS analysis of taPHD*
Nikita Loik	*Amino acid analysis of the peptide taODD*
Dr James S. O. McCullagh[2]	*Tandem mass spectrometry of taODD*
Dr Adam Hardy	*ESI-MS analysis of taPHD and mbP4H*

The list below gives the names of people who have prepared materials that were used in this work. The corresponding materials are listed in the right hand column.

Contributors: Materials

Christoph Loenarz	*Preparation of the constructs taPHD t64 and mbP4H in pET-28a(+)* *Synthesis of the peptides DANRE, STRPU, TRICA, PALPU, NEMVE, ANOGA and 4R/S-fluoro-prolyl CODD*
Inga Pfeffer	*Preparation of the vector FLAG-pET-28a(+)*
Marina Demetriades	*Synthesis of the inhibitor 'UN9'*

1. Benedikt Kessler group, Centre for Cellular and Molecular Physiology
Nuffield Department of Clinical Medicine, University of Oxford

2. Mass Spectrometry Facilities, Department of Chemistry, University of Oxford

Abbreviations

(v/v)	Volume to volume ratio
(w/v)	Weight to volume ratio in g/mL
2,4/2,5-PDCA	Pyridine-2,4/2,5-dicarboxylic acid
2OG	2-Oxoglutarate
2TY	2× Tryptone Yeast (bacterial growth medium)
ANOGA	ODD sequence of *Anopheles gambiae*
AQC	6-Amino-quinolyl-*N*-hydroxysuccinimidyl carbamate
BBOX1	ϒ-butyrobetaine hydroxylase 1
BIS-TRIS	(Bis(2-hydroxyethyl)-amino-tris(hydroxymethyl)-methane)
BLAST	Basic local alignment search tool
bp	Base pair
C	Celsius
CAD	*C*-Terminal transactivation domain
CBP	Cyclic AMP-response element binding protein binding protein
CODD	*C*-terminal oxygen-dependent degradation domain
CV	Colum Volume
Da	Dalton
DANRE	ODD sequence of Danio rerio
DAOCS	Deacetoxycephalosporin C synthase
DCM	Dichloromethane
DMF	Dimethylformamide
DMSO	Dimethylsulfoxide
DNPH	2,4-Dinitrophenyl hydrazine
dNTP	Deoxyribonucleotide triphosphate
DSBH	Double-stranded β-helix
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
ESI-MS	Electrospray-ionisation mass spectrometry
FIH	Factor inhibiting HIF
FPLC	Fast protein liquid chromatography
GST	Glutathione-S-transferase
HEPES	(*N*-[2-Hydroxyethyl]piperazine-*N*-[2-ethanesulfonic acid])
HIF	Hypoxia-inducible factor
HLH	Helix-loop-helix
HMM	Hidden Markov model
HPLC	High performance liquid chromatography
HRE	Hypoxia responsive elements
hsPHD2	PHD2 of *Homo sapiens*
HyP	Hydroxy-proline
IPTG	Isopropyl-β-D-thiogalactoside
JGI	Joint Genome Institute
kcat	Turnover number
kDa	kilo Dalton
Km	Michaelis Menten constant

L*C*-MS	Liquid chromatography coupled with mass spectrometry
MAD	Multi-wavelength anomalous diffraction
MALDI-TOF	Matrix-assisted laser desorption/ionisation time of flight
mbP4H	Prolyl-4-hydroxylase of *Monosiga brevicollis*
MES	2-(*N*-Morpholino)ethanesulfonic acid
MS	Mass spectrometry
MYND	Myeloid translocation protein 8, Nervy, and DEAF1
NEMVE	ODD sequence of *Nematostella vectensis*
NMR	Nuclear magnetic resonance
NODD	*N*-terminal oxygen-dependent degradation domain
NOG	*N*-Oxalylglycine
OD_{600}	Optical density at 600 nm
ODD	Oxygen-dependent degradation domain
PALPU	ODD sequence of *Palaemonetes pugio*
PAS	See PER-ARNT-SIM
PAX1	Paired box protein 1
PCR	Polymerase chain reaction
PDB	Protein data bank
PER-ARNT-SIM	(Period circadian protein)-(Aryl hydrocarbon receptor nuclear translocator)-(Single-minded protein)
Pfu	*Pyrococcus furiosus*
PHD	Prolyl-hydroxylase domain
poHPH	HIF-prolyl hydroxylase of *Perkinsus olseni*
ppm	Parts per million
pVHL	Von-Hippel-Lindau protein
Q-TOF	Quadrupole time-of-flight
RMSD	Root mean square deviation
rpm	Rotations per minute
s	Second
SAD	Single-wavelength anomalous dispersion
SDS-PAGE	(Sodium dodecylsulfate)-(polyacrylamide gel electrophoresis)
STRPU	ODD sequence of *Strongylocentrotus purpuratum*
TAE	TRIS-HCl/acetic acid/EDTA
taODD	ODD sequence of *Trichoplax adhaerens*
taPHD	PHD of *Trichoplax adhaerens*
Taq	*Thermus aquaticus*
TCEP	(Tris(2-carboxyethyl)phosphine)
TEMED	*N*,*N*,*N*,'*N*'-Tetramethylethylene-diamine
TRICA	ODD sequence of *Tribolium castaneum*
TRIS	Tris(hydroxymethylethyl)amino-methane
UV	Ultra violet
V	Volts
VCB	pVHL-elongin B - elongin C
VHL	Von-Hippel-Lindau
V_{max}	Maximum turnover rate
WT	*Wild-type*

Introduction

1. The HIF Pathway

All aerobic organisms ranging from mycobacteria to humans require molecular oxygen to produce energy. Deprivation of this key resource causes major cellular stress. To overcome the challenge of low oxygen levels, also referred to as hypoxia, aerobic organisms have evolved oxygen sensing mechanisms that trigger a number of adaptive responses if required. Hypoxia Inducible Factor (HIF), initially identified as a regulator of erythropoietin biosynthesis in human cells (Semenza 1991), is considered to be a master regulator of oxygen homoeostasis in metazoan organisms (Elvidge et al. 2006).

1.1. The HIF pathway: Discovery and molecular mechanism

Although it was observed in the late nineteenth century that red blood cell production in humans increases with travel to high altitudes (Viault 1890), the molecular mechanism behind this observation remained unclear until about a hundred years later in 1991, when HIF was identified as a transcription factor regulating the synthesis of erythropoietin (Semenza 1991). Erythropoietin is a glycoprotein hormone that controls red blood cell production. Once it was discovered that the transcriptional enhancers binding HIF, the Hypoxia Responsive Elements (HREs), were widely abundant in cells that do not produce erythropoietin, it became apparent that the HIF mechanism is involved in the control of other physiologically relevant processes (Maxwell et al. 1993). Since then, substantial research effort has been undertaken to gain in-depth knowledge of the proteins mediating the HIF pathway in humans.

HIFs are heterodimeric, i.e. proteins consisting of α- and β-subunits, both of which are expressed constitutively (see Figure 1.1. for domain architecture of the HIF subunits) (Wang et al. 1995).

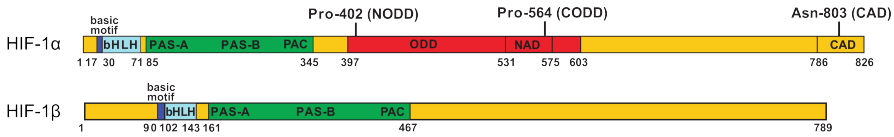


Figure 1.1. Domain architecture of human HIF-1α and HIF-1β. Abbreviations and functionality: Basic motif – DNA-binding sequence; bHLH – basic helix-loop-helix (allows for dimerisation of HIF subunits); PAS-A/B – PER-ARNT-SIM (thought to be involved in HIFα -HIFβ recognition that promotes HLH-dimerisation and DNA binding); PAC – motif *C*-terminal to PAS (part of PAS-domain); ODD – oxygen-dependent degradation domain (region recognised and hydroxylated by PHDs); NODD – *N*-terminal ODD; CODD – *C*-terminal ODD; NAD – *N*-terminal transactivation domain (involved in co-factor recruitment); CAD – *C*-terminal transactivation domain (involved in co-factor recruitment, contains site for asparaginyl hydroxylation by FIH1). The domain architecture of human HIF-2α closely resembles the architecture of HIF-1α, containing the same functional domains. HIF-3 α, in contrast, is markedly different as it probably lacks both NODD and CAD sequences (Ema et al. 1997; Gu et al. 1998; Masson et al. 2001).

In the nucleus, HIF binds to an HRE consensus sequence in DNA (5'-RCGTG-3') associated with HIF target genes. HIF recruits further transcriptional co-activators to from a complex that activates the transcription of over 100 human target genes downstream of the HREs (see Table 1.2.) (Arany et al. 1996; Ke & Costa 2006). In what could constitute a feedback mechanism, one of the genes upregulated by HIF is *PHD2*. Under normoxic conditions, the constitutively expressed HIF α-subunit (HIFα) is hydroxylated by the Fe(II) and oxygen-dependent enzymes HIF-prolyl hydroxylases (PHDs) and Factor-Inhibiting-HIF (FIH), a HIFα asparaginyl hydroxylase (Chowdhury et al. 2008; Kaelin & Ratcliffe 2008). Both the HIF-prolyl and HIF-asparaginyl hydroxylations disrupt the mode of action of the HIF heterodimer. Prolyl-hydroxylation marks HIFα for ubiquitination co-catalysed by an E3 ubiquitin ligase in which the von Hippel-Lindau protein (pVHL) acts as the recognition element, forming the VCB complex with elongins B and C (Maxwell et al. 1999). This leads to subsequent degradation of HIFα. Asparaginyl hydroxylation, in contrast, inhibits the recruitment of transcriptional co-factors such as CBP/P300 (see Abbreviations), thereby inhibiting the formation of a transcriptional complex (Freedman et al. 2002; Lando et al. 2002; Sang et al. 2002).

Therefore, both the stability and the transcriptional activity of HIF are negatively regulated by dioxygen-dependent hydroxylases.

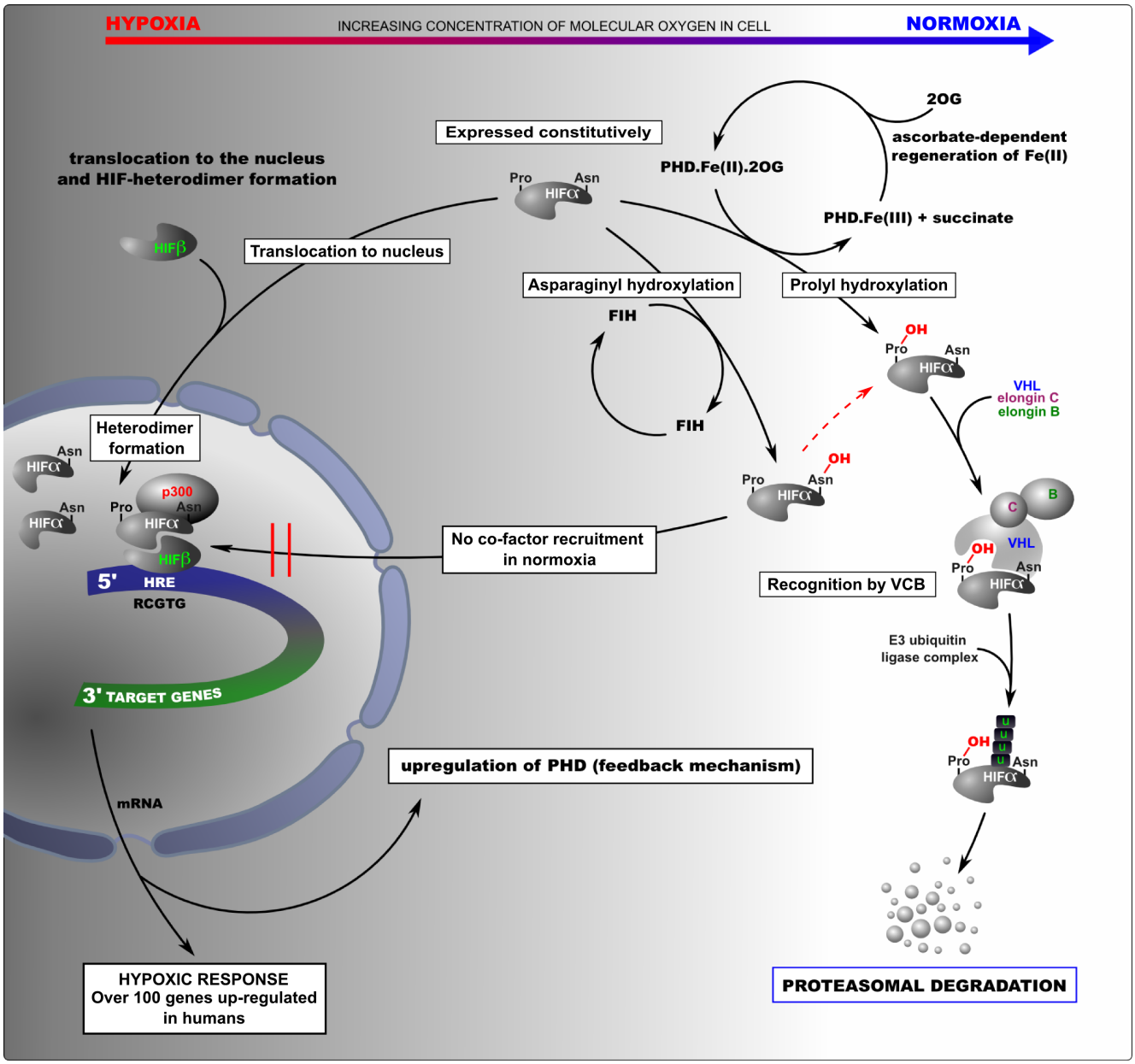


Figure 1.2. Some key components of the HIF pathway in mammals. Processes depicted on the right hand side of the scheme take place in normoxia and processes depicted on the left hand side increasingly take place as the concentration of molecular oxygen in the cell decreases. In normoxia, the constitutively expressed HIFα is hydroxylated at prolyl residues by the PHDs, enabling binding to the VCB complex. Subsequent ubiquitinylation of HIFα marks it for proteasomal degradation, resulting in very low levels of HIFα protein in most tissues in normoxia. In hypoxia, hydroxylation of HIFα by the oxygen-dependent PHDs cannot take place and HIFα accumulates, forms a heterodimer with the abundant HIFβ and recruits transcriptional co-factors in the nucleus. The HIF transcriptional complex binds to HREs within the DNA and thereby regulates the expression of a multitude of genes. Asparaginyl hydroxylation by FIH prevents the recruitment of transcriptional co-factors and thereby provides an alternative route of suppressing the hypoxic response in normoxia.

In humans, three HIFα isoforms are known, all of which can interact with the β-subunit but are encoded by distinct loci and have different functions (see 1.1.) (Ema et al. 1997; Gu et al. 1998).

HIFα isoform	Expression tissue	Preferred PHD	Function
HIF-1α	Ubiquitous in all cells	PHD2	Responsible for broad range hypoxic response and gene regulation, e.g. regulation of glycolytic pathway
HIF-2α	Endothelium, kidney, lung, heart, small intestine and others	PHD1, PHD3	It has been suggested that HIF-2α has different target genes from HIF-1α, e.g. regulation of erythropoietin in liver
HIF-3α	Endothelium, heart, lung, kidney and others [HIF-3α] but generally poorly studied	CODD hydroxylated by PHD2 but further details unknown	Least well understood; recognises the same DNA sequences as HIF-1α and HIF-2α heterodimers and may compete with them; some splice variants form transcriptionally inactive heterodimers with HIF-1α that could potentially bias the transcriptional response towards HIF-2α

Table 1.1. Comparison of the three human HIFα isoforms. Current findings indicate that, at least to some extent, both HIF-1α and HIF-2α control cell metabolism and redox homeostasis through non-overlapping functions (Majmundar et al. 2010).

COMPONENT REGULATED BY HIF	
Gene	Protein
Hormones	
EPO	Erythropoietin
LEP	Leptin
Energy metabolism	
GLUT1	Glucose transporter 1
HK1/2	Hexokinase 1/2
PFKL	6-Phosphofructo-1-kinase
GAPDH	Glyceraldehyde-3-P dehydrogenase
ALDA/ALDOC	Aldolase A/C
ENO1	Enolase 1
PGK1	Phosphoglycerate kinase 1
LDHA	Lactate dehydrogenase A
PKM	Pyruvate kinase M
ADK3	Adenylate kinase 3
TGM2	Tissue transglutaminase
CA9	Carbonic anhydrase 9
Growth and apoptosis	
NIP3	Nineteen kDa interacting protein 3
NIX	Pro-apoptotic factor
ENG	Endoglin
WT1	Wilms' tumour suppressor
AFP	α-Fetoprotein
Cell proliferation	
TGF-α/β3	Transforming growth factors α/β3
IGF2	Insulin-like growth factor 2
IGFBP3	Insulin-like growth factor binding protein 3
CIP1/WAF1	Cellular cyclin-dependent kinase inhibitor
Angiogenetic signalling	
VEGF	Vascular endothelial growth factor A
EG-VEGF	Endothelial gland derived VEGF
VEGFR1	VEGF receptor
PAI1	Plasminogen-activator inhibitor 1
UPAR	Urokinase plasminogen activator receptor

COMPONENT REGULATED BY HIF	
Gene	Protein
Cell migration	
CXCR4	Chemokine receptor 4
AMF/GPI	Autocrine motility factor
LRP1	LDL receptor related protein 1
Transcriptional regulation	
DEC1/2	Differentiated embryo chondrocyte 1/2
ETS1	Avian erythroblastosis virus E2 oncogene homologue 1
CITED2/p35srj	CBP/p300-interacting transactivator
NUR77	Nuclear orphan receptor
Transport	
TF	Transferrin
TFRC	Transferrin receptor
CP	Ceruloplasmin
ABCB1	Multidrug-resistance P-glycoprotein
Cytoskeleton formation	
KRT14	Keratin 14/18/19
Vasomotor regulation	
ET1	Endothelin 1
ADM	Adrenomedullin
INOS	Inducible nitric oxide synthase
ENOS	Endothelial nitric oxide synthase
HMOX1	Heme oxygenase 1
NPPA	Atrial natriuretic peptide
Matrix and barrier functions	
TFF	Intestinal trefoil factor
CATHD	Cathepsin D
FN1	Fibronectin 1
MMP2	Matrix metalloproteinase 2
P4HA1	Pro-collagen prolyl hydroxylase α(I)
COL5A1	Collagen type V (α1)
CD73	Ecto-5'-nucleotidase
Feedback loop	
EGLN1/3	HIF-prolyl hydroxylases 2/3

Table 1.2. Some of the genes found to be directly activated by HIF using functional delineation of HIF binding to HREs (Schofield & Ratcliffe 2004; Semenza 2007; Elvidge et al. 2006). Large-scale gene expression arrays suggest that the number of genes positively or negatively regulated by hypoxia may be of several hundreds (Koong et al. 2000; Wykoff et al. 2000; Jiang et al. 2003). The large number of target genes and the variety of processes they are involved in mirrors the complexity of oxygen homeostasis in complex animals (Schofield & Ratcliffe 2004). One key group of HIF-regulated genes are the glycolytic enzymes, e.g. aldolase and phosphoglycerate kinase, which up-regulates glycolysis to counteract the reduced ATP production caused by hypoxia (Semenza 2007).

1.2. Clinical relevance of the HIF pathway

Since its discovery, HIF has been linked to various human health conditions. Examples include ischaemic heart disease, pulmonary hypertension, arthritis and cancer (Semenza 2001a; Semenza 2001b). A brief summary of the most thoroughly studied diseases linked to the HIF pathway is given in 1.3.

Disease	Ischaemic heart disease	Anaemia	Rheumatoid arthritis	Polycythemia	Cancer
Role of HIF	Up-regulation of the HIF pathway do adapt to local hypoxia in diseased tissue	Therapeutic up-regulation of the HIF-pathway could lead to increased production of red blood cells	Findings suggest that inhibition of the HIF target gene *VEGF* may be an effective treatment by suppressing angiogenesis in affected tissue	Several mutations found to increase HIF-stability: C598T in VHL; R371H and P317R in PHD2 and G1609T in HIF-1α	Many genes playing key roles are regulated by HIF; examples include *MYC* (proliferation), *VEGF* (angiogenesis), *NDRG2* (apoptosis), *PDK1* (metabolism of hypoxic tumours), *CXCR4* (cell migration)

Table 1.3. The genes regulated by HIF have been linked to many diseases and the Table above contains only a handful of dozens of examples found throughout literature. While up-regulation of certain genes by HIF has been shown to protect tissue from damage during strokes and heart attacks (Majmundar et al. 2010), a similar protective effect is suspected to aid the survival of hypoxic tumour cells. Accordingly, in many cancer types, increased expression of HIFα has been shown to correlate with poor clinical prognosis (Semenza 2007; Majmundar et al. 2010).

Up-regulating or down-regulating the HIF response for therapeutic applications is complicated by the presence of negative feedback loops, which requires careful consideration. One example of such a negative feedback loop is the up-regulation of two PHD enzymes, PHD1 and PHD3, by HIF (Wei & Yu 2007; Pan et al. 2007). Further complexity arises from the presence of the different HIFα isoforms, the exact interplay between which still remains to be determined (Webb et al. 2009).

1.3. The role of PHDs

The PHDs (from Prolyl Hydroxylase Domain containing protein) belong to the family of 2-oxoglutarate (2OG) dependent oxygenases. These enzymes employ dioxygen for substrate hydroxylation while simultaneously decarboxylating 2OG to give succinate and carbon dioxide (Webb et al. 2009). Key structural features of 2OG-oxygenases include an active site formed by an eight-stranded β-barrel or 'jelly roll' motif. Within the active site, Fe(II) is generally chelated by two histidines and a carboxylate group (a

more detailed description of the structural features of human PHD2 can be found in Part A) (Elkins et al. 2003; McDonough et al. 2006).

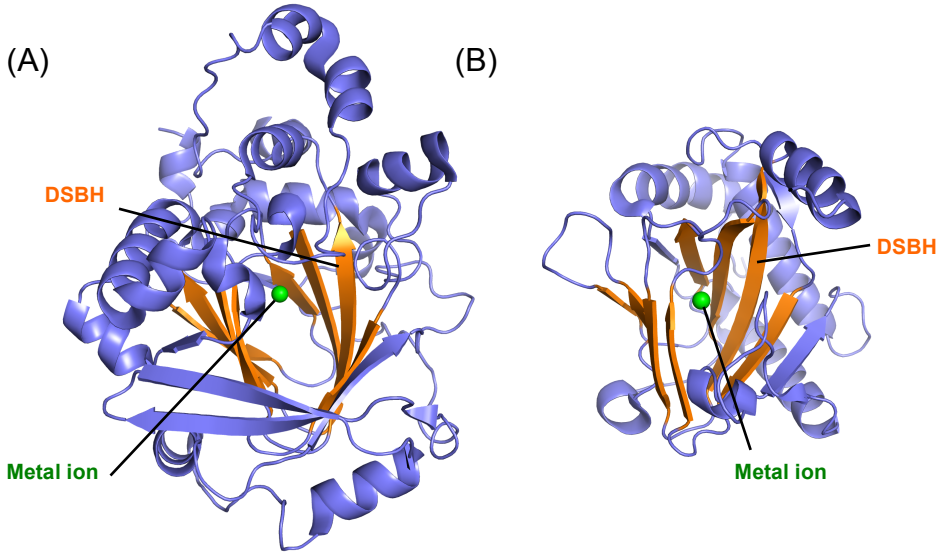


Figure 1.3. Cartoon representations of crystal structures of the human 2OG-oxygenases PHD2 (A; PDB ID=3HQR) and JMJD2A (B; PDB ID=2OQ7), illustrating the eight-β-sheet-stranded jelly roll motif (DSBH; orange) and the position of the active site metal (green).

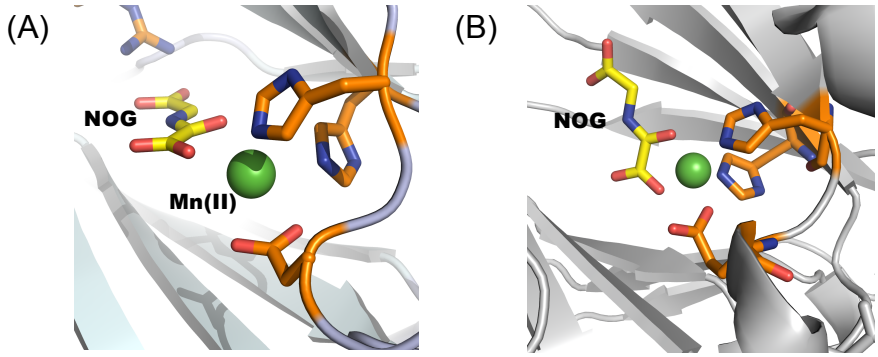


Figure 1.4. Active sites of human PHD2 (A; PDB ID=3HQR) and JMJD2A (B; PDB ID=2OQ7). The metal ion (green) is bound by two histidine residues and a carboxylic acid residue (orange), a feature conserved across 2OG-oxygenases. The ligand in both cases is *N*-oxalylglycine (yellow), a 2OG mimic and a generic 2OG-oxygenase inhibitor. In the human PHD2 structure, Mn(II) is substituted for Fe(II).

The dependence of PHDs on oxygen for HIFα hydroxylation made them the 'prime suspects' for molecular oxygen sensing. This proposal is supported by estimates for their K_m values for oxygen, which, despite variations arising from different experimental methods, are significantly higher than intracellular partial pressure of dioxygen (Webb

et al. 2009). A high K_m value for oxygen implies that PHD activity is exquisitely dependent on oxygen levels across the entire physiological range (Majmundar, Wong, & Simon, 2010). Of the three PHDs characterised in humans (PHDs 1-3), PHD2 is the most abundant and is thought to be the dominant oxygen sensing enzyme in most tissues (Berra et al. 2003; Kaelin & Peter J Ratcliffe 2008). The other two PHDs are less well studied, partially because of their low abundance in most tissues and difficulties associated with production of active recombinant enzyme (Aprelikova et al. 2004; Berra et al. 2003; D'Angelo et al. 2003). PHD1 and PHD3 are thought to contribute more towards the regulation of HIF-2α than PHD2 (see Table 1.1.) (Appelhoff et al. 2004).

Given their suggested role as molecular oxygen sensors, it is inevitable that the activity of the PHDs would be of clinical interest in relation to diseases linked to the HIF pathway. On one hand, loss-of-function mutations of PHD2 have been observed in polycythemia, a condition in which excess red blood cells are produced (see Table 1.3.) (McDonough et al. 2006; Percy et al. 2006). Cancer cell lines, often characterised by genetic instability and dependence on the HIF pathway for tumour growth and survival, are also very likely to favour PHD mutants of lower activity (Webb et al. 2009). On the other hand, however, for conditions such as ischaemic heart disease and anaemia, inhibiting the PHDs and thereby provoking a hypoxic response would be of therapeutic benefit (see Table 1.3.). Most studies reported to date have made use of generic 2OG-oxygenase inhibitors based on 2-oxoglutarate analogues (see Figure 1.5.).

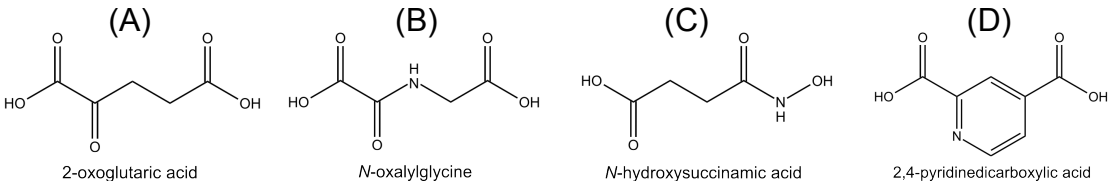


Figure 1.5. 2OG (A) and three examples of generic 2OG-oxygenase inhibitors: (B) *N*-oxalylglycine (NOG), an *N*-oxalyl amino acid; (C) *N*-hydroxysuccinamic acid, a hydroxamic acid; (D) 2,4-pyridinecarboxylate, a cyclic 2OG analogue. The main mode of action of generic 2OG-oxygenase inhibitors is chelation of the active site iron and competition with 2OG and possibly oxygen.

However, undesired inhibition of closely related enzymes could lead to toxic side effects. Given the occurrence of at least 60 2OG-oxygenases in humans, the design of enzyme-specific inhibitors presents a major challenge in the development of drugs targeting the PHDs (Hewitson et al. 2005).

2. Evolution of the HIF Pathway

The composition of the terrestrial atmosphere has varied greatly throughout the 4.5 billion years of the Earth's history. These variations have played a major role in the evolution of life and it has been proposed that the HIF pathway evolved during the Cambrian explosion as a key pathway to regulate oxygen homeostasis in multicellular organisms (Loenarz et al. 2011).

2.1. Ancient atmospheres and metazoan evolution

Careful studies of fossils suggest that many key events in the evolution of life, including the rapid diversification of metazoans, mass extinction events and gigantism, coincided with changes in the composition of the terrestrial atmosphere (Berner 1999; Dudley 1998; Falkowski et al. 2005; McElwain et al. 2009). At the same time, the Earth's atmosphere was shaped by the life forms inhabiting it. Around 3 billion years ago, photosynthetic cyanobacteria in the oceans are thought to have started producing larger amounts of oxygen, which initially was trapped out by oxidation of inorganic minerals on the ocean ground (Nisbet & Sleep 2001). It was not until 1.5 billion years ago that the Earth's reserves of readily oxidised minerals such as iron ran out and oxygen started accumulating in the oceans and atmosphere, slowly initially and then rapidly, peaking at 15%-20% at around 540 million years ago (see Figure 2.1.) (Taylor & McElwain 2010). Although highly toxic for most organisms of the pre-Cambrian, the utilisation of dioxygen to generate ATP proved to be a very efficient energy source for eukaryotes (Rich 2003). The rich abundance of metazoan fossils from the Cambrian era compared to the pre-Cambrian suggests the onset of a rapid evolution and diversification of metazoan phyla at around 540 million years ago, often referred to as the Cambrian explosion (Budd 2008; Morris 2006).

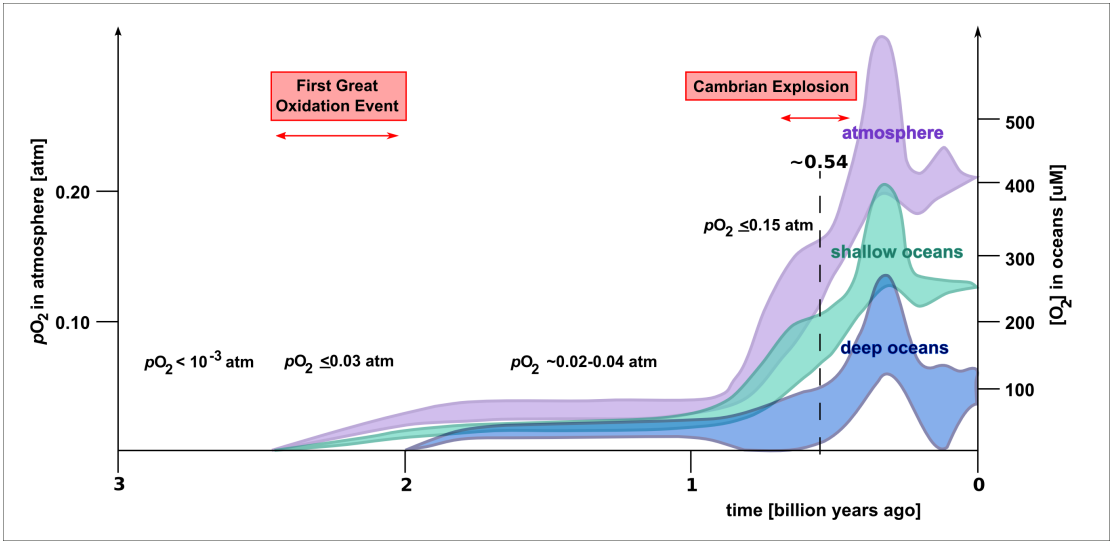


Figure 2.1. Estimated changes in oxygen levels in the deep and shallow oceans and the atmosphere throughout the last 3 billion years (figure adapted from Holland, 2008). The filled areas represent the ranges of current estimates. The First Great Oxidation Event marks the appearance of atmospheric dioxygen of biological origin around 2.4 billion years ago. While cyanobacteria had been producing dioxygen for a longer time period, chemical capture through readily oxidised material such as iron had buffered any effects on the atmosphere. The First Great Oxidation Event marks the saturation of oxygen scavengers. The Cambrian explosion, ca. 540 million years before our time, refers to the sudden and abrupt appearance of metazoan fossils, rather than an oxygenation event. Its temporal coincidence with rapid increases in atmospheric and oceanic oxygen levels, possibly caused by land-based plants, has fuelled many speculations over the correlation between oxygen availability and metazoan evolution (Knoll 1999; Falkowski et al. 2005).

The challenge of slow oxygen diffusion through multiple cell layers is proposed to have lead to the development of sophisticated hypoxia sensing and response mechanisms in the most ancient metazoans (Loenarz et al. 2011). While oxygen sensing mechanisms have been identified in several unicellular organisms and *Saccharomyces cerevisiae* (see Box 2) (Gilles-Gonzalez et al. 1994; Bunn & R. Poyton 1996; Krude et al. 1997), to date no evidence has been found to suggest that the HIF pathway extends beyond the metazoans (Rytkönen et al. 2011). It is therefore likely that the HIF pathway was the key to the success of the metazoan species in its possession (Taylor & McElwain 2010).

2.2. The HIF pathway in other organisms

Orthologs of the minimally required components of the HIF pathway, PHD, HIFα and HIFβ, pVHL and the co-activators CBP/P300 have been observed in metazoans. In the nematode *Caenorhabditis elegans*, the PHD2 ortholog EGL9 has been identified and shown to be involved in regulating the worm's metabolism in an oxygen-dependent manner (Epstein et al. 2001; Hampton-Smith & Peet 2009; Powell-Coffman 2010). Fatiga, a *Drosophila melanogaster* ortholog of PHD2, was found to be involved in oxygen sensing and regulates tracheal branching (Hampton-Smith & Peet 2009).

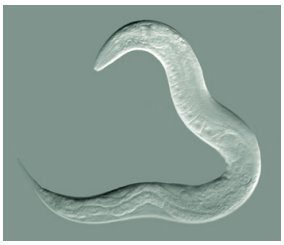
(A)  (B) 

Figure 2.2. (A) *Caenorhabditis elegans,* hermaphrodite (image by Bob Goldstein, UNC Chapel Hill); (B) male *Drosophila melanogaster* (image by André Karwath).

Unlike humans and vertebrates, however, both *Drosophila melanogaster* and *Caenorhabditis elegans* possess only one PHD enzyme, most similar to human PHD2 ('hsPHD2'), and only one HIFα isoform, resembling the vertebrate HIF-1α (Epstein et al. 2001; Centanin et al. 2005). Analyses over a wide range of animal phyla (see Table 2.1.) suggest that the HIF pathway developed a higher degree of complexity with multiple HIFα isoforms and up to three PHDs in larger animals with respiratory and circulatory systems (Rytkönen et al. 2011).

Species	HIF	PHD	VHL	FIH
Monosiga brevicollis (choanozoan)	None identified	PHD2-like enzyme	None identified	FIH-like enzymes
Caenorhabditis elegans (nematode)	HIF-1α and HIFβ	PHD2 (EGL9)		None identified
Drososphila melanogaster (arthropod)	HIF-1α and HIFβ	PHD2 (Fatiga)		None identified
Danio rerio (chordate/fish)	HIF-1α, HIF-2α and HIFβ	PHD2 and PHD3	Conserved	FIH1
Mus musculus (chordate/mammal)	HIF-1α, HIF-2α, HIF-3α and HIFβ	PHD2, PHD3 and PHD1		FIH1
Homo sapiens (chordate/mammal)	HIF-1α, HIF-2α, HIF-3α and HIFβ	PHD2, PHD3 and PHD1		FIH1

Table 2.1. Components of the HIF pathway in representatives of choanozoans (protists), nematodes, arthropods, fish and mammals (Rytkönen et al. 2011). HIF-1α is conserved throughout the metazoans and seems to be the earliest HIFα isoform, with HIF-2α appearing in fish and reptiles and HIF-3α in mammals. Similarly, the evolution of PHD3 and PHD1 appears to be associated with increasing complexity. HIF-1α itself, although conserved, has gained complexity over the course of evolution by gaining the NODD, and the CAD where FIH is present (Hampton-Smith & Peet 2009). While VHL, another crucial component of the HIF pathway, is conserved throughout the metazoan kingdom, the presence of FIH seems more erratic in early animals. Functional FIH homologs as present in chordates are lacking in some arthropods and nematodes, but can be found again in other life forms like the flour beetle *Tribolium castaneum* and the more ancient starlet sea anemone *Nematostella vectensis* (Hampton-Smith & Peet 2009).

2.3. The origin of the HIF pathway

Despite its apparent conservation throughout the metazoans, little is known about the origin of the HIF pathway. HIFα-like homologs have not been identified in the yeast *Saccharomyces cerevisiae*, or in *Monosiga brevicollis*, both considered closely related to metazoans (Rytkönen et al. 2011). Bioinformatic analyses of the prevalence of the HIF-binding HRE regions (5'-RCGTG-3') across more than 60 eukaryote genomes indicated that an enrichment of HRE-motifs closely upstream of genes is present in metazoans and plants but not other eukaryotes (Loenarz et al. 2011). However, as no other putative HIF pathway components were reported in plants, the observation of HRE enrichments does not necessarily mean that a HIF pathway is present because the HRE motif is rather short.

The components of the HIF pathway that are apparently easiest to trace beyond the metazoan kingdom are the highly conserved PHDs. Several PHD-like enzymes in protists and bacteria have been described (Clifton et al. 2006). An example is the parasitic marine protist *Perkinsus olseni* (Leite et al. 2008). Although it has been shown that the

Perkinsus HPH (HIF-Prolyl-Hydroxylase) is up-regulated by hypoxia mimics in a way similar to the up-regulation of hsPHD2 by hypoxia in human cells, prolyl hydroxylation by this enzyme has not been demonstrated (Leite et al. 2008). Alignments of the *Perkinsus olseni* HPH with metazoan PHDs show that it lacks a highly conserved histidine residue involved in Fe(II)-binding, casting doubts on the claims that this PHD-like enzyme may be a 'true' PHD.

In contrast, a PHD-like enzyme found in the socially living soil amoeba *Dictyostelium discoideum* has been proposed to be involved in oxygen sensing, although in a HIF-independent manner (van der Wel et al. 2005). *Dictyostelium* goes through several unicellular and multicellular life stages, two of which involve a motile slug-like shape and an immobile fruiting body (Tyler 2000). The transition between these two stages requires oxygen levels above 10% and is dependent on the *Dictyostelium discoideum* prolyl-4-hydroxylase (ddP4H). This PHD-like enzyme hydroxylates Skp1, an adaptor of the SCF-class of E3 ubiquitin ligases (Zheng et al. 2002). Since the SCF-class is considered to be evolutionary related to VHL (van der Wel et al. 2005), these findings are of interest with regards to the evolution of the HIF pathway and suggest that the Skp1 pathway and the HIF pathway could share a common ancestor mechanism. It is likely that such PHD-like enzymes were recruited to hydroxylate HIF early in the evolution of metazoans (Rytkönen et al. 2011).

The most basal animal experimentally demonstrated to employ the HIF pathway is the placozoan *Trichoplax adhaerens* (Loenarz et al. 2011), while one of the supposedly closest living relative of metazoans, the choanozoan *Monosiga brevicollis*, possesses PHD-like enzymes (Loenarz 2010). The following two Sections (Sections 3 and 4) introduce the habitat, morphology and phylogenetic relations in these two evolutionary interesting organisms.

Examples of oxygen sensing mechanisms in bacteria

Escherichia coli has several direct and indirect oxygen sensing mechanisms that it employs to regulate gene expression of various components of the cellular metabolism. A direct oxygen sensing mechanism is the **fumarate and nitrate reduction (FNR) pathway**. FNR is a transcriptional regulator that is inactive under normoxic conditions. Under hypoxia, FNR forms the active homodimeric complex and down-regulates the transcription of components required for aerobic metabolism and up-regulates components of the anaerobic metabolism (Kiley & Beinert 2003).

FNR activity is regulated directly by molecular oxygen concentration and is mediated by a structurally dynamic cluster of iron molecules, [4Fe-4S]. Under normoxic conditions, the cluster is predominantly present in the oxidised $[2Fe\text{-}2S]_2$ state, which inhibits dimerisation by inducing structural changes within the FNR monomer. At low oxygen concentrations, the $[4Fe\text{-}2S]_2$ cluster is prevalent and dimerisation of FNR can occur (Bailey-Serres & Chang 2005).

ArcAB is another pathway that regulates transcription as a function of oxygen levels in *Escherichia coli*. The ArcAB system is composed of the membrane bound histidine kinase ArcB and its phosphorylation target ArcA. ArcB can undergo auto-phosphorylation in response to a reduced electron flow through the respiratory chain, thereby acting as an indirect oxygen sensor (Alexeeva et al. 2003). Auto-phoshorylation of ArcB activates phosphorylation of ArcA, which in return regulates carbon catabolism and cellular redox status (Unden & Bongaerts 1997; Alexeeva et al. 2003).

Another example of a bacterial direct oxygen sensing pathway is **FixL/FixJ** of *Rhizobia* and *Bradyrhizobium* genera (Fischer 1994). The activity of the oxygen and heme binding histidine kinase FixL is directly dependent on molecular oxygen levels. At low oxygen levels, oxygen is released from the heme binding site of FixL, which triggers auto-phosphorylation of FixL and phosphorylation of the transcription factor FixJ (Bailey-Serres & Chang 2005).

Oxygen sensing by heme-binding proteins in yeast

Saccharomyces cerevisiae also has multiple pathways that respond to low oxygen level (Kwast et al. 1998; R. O. Poyton 1999). One of the predominant pathways, the **Hap1 pathway**, is indirect and regulated by rates of heme synthesis, which are proportional to oxygen levels (Hon et al. 2003). Hap1 regulates the expression of genes for aerobic metabolism. At low oxygen levels, heme production is slow and correspondingly, levels of active Hap1 decrease.

Saccharomyces cerevisiae also possesses the heme-binding tetrameric transcription factor **Hap2/3/4/5**, which similarly to Hap1, which acts in a way similar to Hap1 (Bailey-Serres & Chang 2005).

The Skp1-pathway of *Dictyostelium discoideum*

The slime mold *Dictyostelium discoideum* possesses a prolyl-hydroxylase (ddP4H1, see Part C) that appears to be an ortholog of hsPHD2 (West et al. 2004). However, no HIFα-orthologs were identified in *Dictyostelium discoideum* (West et al. 2004). The substrate of ddP4H1 was found to be Skp1, a subunit of the SCF class of E3-Ub ligases (Lando et al. 2003; Knowles et al. 2003). E3-Ub ligases are involved in polyubiquitination and degradation of numerous other proteins, including regulators of the cell cycle and transcription factors. Prolyl-4-hydroxylation of Skp1 by ddP4H1 leads to glycosylation of Skp1, thereby modulating its activity (van der Wel et al. 2005). It has been proposed that Skp1 is involved the in the regulation of culmination of *Dictyostelium discoideum*, which is dependent on environmental oxygen levels (van der Wel et al. 2005).

Oxygen sensing in plants

A clear understanding of oxygen sensing pathways in plants is not present to date (Drew 1997; Geigenberger 2003). As for other aerobic organisms, the rate of oxygen uptake by plants needs to be adjusted to the rate of oxygen consumption within cells. Within different plant tissues and organs, plants can experience different oxygen concentrations and anoxic regions may occur despite surrounding cells having normal oxygen levels (Geigenberger 2003). Anoxic plant cells reduce aerobic respiration and energetic processes (Dongen et al. 2004). Whether the underlying regulation of gene expression is regulated by direct or indirect oxygen sensing pathways is currently unclear (Bailey-Serres & Chang 2005). A recent study has shown that that one branch of the ubiquitin-dependent *N*-end rule pathway for protein degradation, present in both animals and plants, functions as an oxygen-sensing mechanism in *Arabidopsis thaliana* (Licausi 2011).

Box 1. HIF-Independent Oxygen Sensing Pathways: A Brief Introduction

3. *Trichoplax adhaerens*, the Simplest of all Extant Animals

3.1. Habitat and morphology of *Trichoplax adhaerens*

Consisting of only five cell types and having a minimalist 98-megabase genome, *Trichoplax adhaerens*, the only named representative of the placozoan phylum, is the simplest extant (non-extinct) animal known (Miller & Ball 2008).

The first recorded discovery of the placozoans dates back to a late nineteenth century aquarium. Inspired by their shape, resembling small disks of a diameter of 1-2 mm, they were named placozoans, literally meaning "flat animals" (Schulze 1883). Misinterpreted as hydra larvae (Syed & B Schierwater 2002), they were dismissed until found in the wild more than a hundred years later and classified as an own phylum (Pearse 1989; Maruyama 2004).



Figure 3.1. Light microscope image of a *Trichoplax adhaerens* animal of about 0.5 mm diameter in size. *Trichoplax* are usually shapeless in appearance, constantly changing their contours as they move across a surface (image by Oliver Voigt).

The *Trichoplax adhaerens* bauplan is composed of two epithelial layers sandwiching a layer of multi-nucleate fibre cells (Grell 1971; Bernd Schierwater 2005). *Trichoplax adhaerens* lacks any symmetry, with the only spatially distinguishing features being top and bottom. While the upper epithelium seems to form a protection layer, the lower epithelium is used for attachment to and ciliar motion across a surface, which could be a

rock in shallow waters. The lower epithelium is also equipped with gland cells and used for encapsulating and digesting prey, usually algae (Grell & Benwitz 1971; Grell & Benwitz 1981; Grell & Ruthmann 1991). The fibre cell layer seems to be responsible for contractile motion that defines the shape of the animal (Srivastava et al. 2008). While *Trichoplax adhaerens* attached to a surface appear as flat disks, they arch their bodies when encapsulating prey and roll into tubes when forced to detach from a surface (Schulze 1883; Grell & Benwitz 1981; Ruthmann et al. 1986; Grell & Ruthmann 1991).

The commonly observed mode of propagation of *Trichoplax adhaerens* in the laboratory is vegetative reproduction by binary fission. Production of buds referred to as "swarmers" has also been observed (Schütze et al. 1999; Srivastava et al. 2008; Wainright et al. 1993). Although sexual reproduction has never been definitely observed, *Trichoplax adhaerens* has been documented to form female oocytes and male gemocytes, suggesting that in the wild, there may be a second, sexual mode of reproduction (Grell 1981; Aleshin & Petrov 2002; Giribet 2002).

Given the nearly invisible appearance of *Trichoplax adhaerens*, a full picture of the ecological distribution of this small animal is difficult to obtain. Some conclusions can be drawn from their collection sites, which are commonly warm tropical and subtropical marine water in the latitudinal band between 30° North and 30° South (Eitel & Schierwater 2010). Specimens have been found on mangrove tree roots, reefs, boat docks and stony beaches (Fraschetti et al. 2008).

Although initially described as only one species based on their morphology, the analysis of nuclear DNA suggest considerable genetic diversity (Srivastava et al. 2008; Miller & Ball 2008). Surprisingly, the *Trichoplax adhaerens* genome revealed a very high level of genetic complexity for an animal with such a simple morphology, leading to the theory

that the placozoans are a secondarily simplified phylum (Miller & Ball 2008), i.e. that it evolved from a later branching phylum through secondary simplification. The *Trichoplax adhaerens* genome is one of the smallest known animal genomes with 98 megabases and contains about 11,500 protein-coding genes (Srivastava et al. 2008). Nevertheless, many genes found in *Trichoplax adhaerens* have homologues involved in complex processes of bilaterian animals, including components of many developmentally regulated signalling pathways and animal-specific transcription factors (Miller & Ball 2008). Examples of bilaterian pathways predicted in *Trichoplax adhaerens* are TGF-β, Wnt, Notch and JAK-STAT (although some are incomplete). Members of the Pax, SOX, T-box and Fox families of transcription factors have been found to match sponges in their level of complexity (see Box 2) (Miller & Ball 2008; Srivastava et al. 2008).

TGF-β: Transforming growth factor beta; protein controlling proliferation and cellular differentiation, among others, and plays a key role in immunity, cancer, heart disease, diabetes, Marfan syndrome (genetic disorder of the connective tissue); antiproliferative factor in normal epithelial cells and early stages of oncogenesis (Elliott & Blobe 2005; Hill et al. 2009).

Wnt signaling pathway network: group of secreted lipid-modified signalling proteins involved in embryogenesis, cell differentiation and cell polarity differentiation (Logan & Nusse 2004).

Notch signalling pathway: highly conserved cell signalling system present in most multicellular organisms; promoting proliferative signalling during neurogenesis (Kopan & Ilagan 2009).

JAK-STAT signalling pathway: alternative signalling pathway to the second messenger system; conserved in metazoans; transmits information form chemical signals outside the cells through the cell membrane into gene promotors in the cell nucleus; composed of three main components: a receptor, the Janus kinase (JAK) and the Signal Transducer and Activator of Transcription (STAT) (Murray 2007).

Pax: Paired box proteins; tissue specific transcription factors containing a paired domain and usually a partial or complete homeodomain; important in early animal development for tissue specification and epimorphic limb regeneration (Lang et al. 2007).

SOX genes: Sry-related HGM box genes; encode transcription factors that bind to the minor groove in DNA; contain the highly conserved high mobility group (HMG) box of SRY, a gene involved in sex determination; important in different aspects of development, such as sex determinatin and neuronal development (Kamachi et al. 2000).

T-box proteins: transcription factors involved in limb and heart development (Naiche et al. 2005).

Fox: Forkhead box proteins; transcription factors regulating the expression of genes involved in cell growth, proliferation, differentiation, and longevity; 'forkhead box' is a DNA binding motif, also known as the 'winged helix' (Myatt & Lam 2007).

Box 2. Selected metazoan genes involved in complex signalling processes of bilaterians.

Consistent with a secondarily simplified organism, the *Trichoplax adhaerens* genome also contains components of the post-synaptic scaffold and, unlike in sponges, among those are also channel and receptor proteins (Miller & Ball 2008; Srivastava et al. 2008). The presence of such genes in an organism morphologically as simple as the *Trichoplax adhaerens* remains an enigma. It has, however, to be considered that cellular morphology can be rather deceptive and what appears to be one cell type could indeed be functionally distinct types (Monteiro et al. 2006; Martinelli & Spring 2003; Adamska et al. 2007). Some of the transcription and signalling factors may play a role in processes such as chemotaxis, phototaxis, fission, swarming, and possibly sexual reproduction (Srivastava et al. 2008).

3.2. Phylogenetic assignment of placozoans

The simple morphology of *Trichoplax adhaerens* has attracted much attention because it resembles the 'Urmetazoan' (see Box 3), a hypothetical organism which represents the ancestor of all metazoan life as predicted by Buetschli's placula hypothesis (DeSalle & Bernd Schierwater 2008; Blackstone 2009; Philippe et al. 2009; Siddall 2010). However, long-standing questions about the phylogenetic position of this enigmatic creature have only been partially answered through the sequencing of its nuclear genome as the knowledge about its life cycle and biology is far from complete (Eitel et al. 2011).

While undoubtedly the simplest extant animal known, *Trichoplax adhaerens* occupies a disputed position phylogenetically (DeSalle & Schierwater 2008; Miller & Ball 2008; Blackstone 2009; Schierwater et al. 2009; Philippe et al. 2009; Siddall 2010; Hejnol et al. 2009; de Jong et al. 2009). Early studies on a small number of genes suggested that placozoans could be secondarily simplified cnidarians (Ender & Schierwater 2003). Analysis of small subunit rRNA, in contrast, suggested that placozoans were eumetazoans and either a sister group of bilaterians or the earliest eumetazoan branch

(Silva et al. 2007; Signorovitch et al. 2007). Complete mitochondrial genome analyses revealed that placozoans could be the earliest branching basal metazoan phylum (Putnam et al. 2007).

'Phylogenetics' is derived from Greek
phylon (φῦλον) "race" and *genetikos* (γενετικός) "relative to birth"] and stands for the study of evolutionary relatedness using molecular sequencing data.

'Eumetazoa' is derived from Greek *eu* (εῦ) "well", *metá* (μετά) "after" and zóon (ζῷον) "animal"] and describes a clade comprising all metazoans apart from sponges. The basis for this distinction is that eumetazoans go through an embryonic stage known as the gastrula, which leads to the development of tissues and/or organs. Whether placozoans are eumetazoans is disputed.

'Bilateria' collectively describes all animals with bilaterian symmetry, meaning they have distinct sites that can be assigned to be the front, the back, the top and the bottom.

'Urmetazoan' refers to a hypothetical last common ancestor of all metazoans.
At least five different hypotheses about the form of such an organism have been proposed and will remain in conflict with each other as long as the phylogenetic relationships at the base of the metazoan phyla have not been fully resolved. Within the scope of this work, the most notable hypothesis states that the last common ancestor had the form of a placula, an amorphous and asymmetric organism bearing close resemblance with the placozoans.

'Sister group' is a term used for a group that is more closely related to the group in question than any other group.

Box 3. A brief introduction to evolutionary biology terms used in the main text.

Analysis of the complete nuclear genome led to the proposal that placozoans are basal eumetazoans, diverging after sponges and before cnidarians and bilaterians (see Figure 3.2.) (Srivastava et al. 2008). Although these results cannot exclude a more basal position, they argue against the idea that placozoans are secondarily simplified cnidarians or bilaterians (Srivastava et al. 2008).

While some argue that the lack of symmetry axes in many adult sponges is an indication that the latest common metazoan ancestor did not possess any symmetry either, it has to be noted that the some adult sponges such as the calcisponges as well as larval organisation of sponges have axial symmetry (Brusca & Brusca 2003; Manuel 2009). Therefore, absence of symmetry in some sponges and in *Trichoplax adhaerens* could be derived rather than ancestral (Philippe et al. 2004).

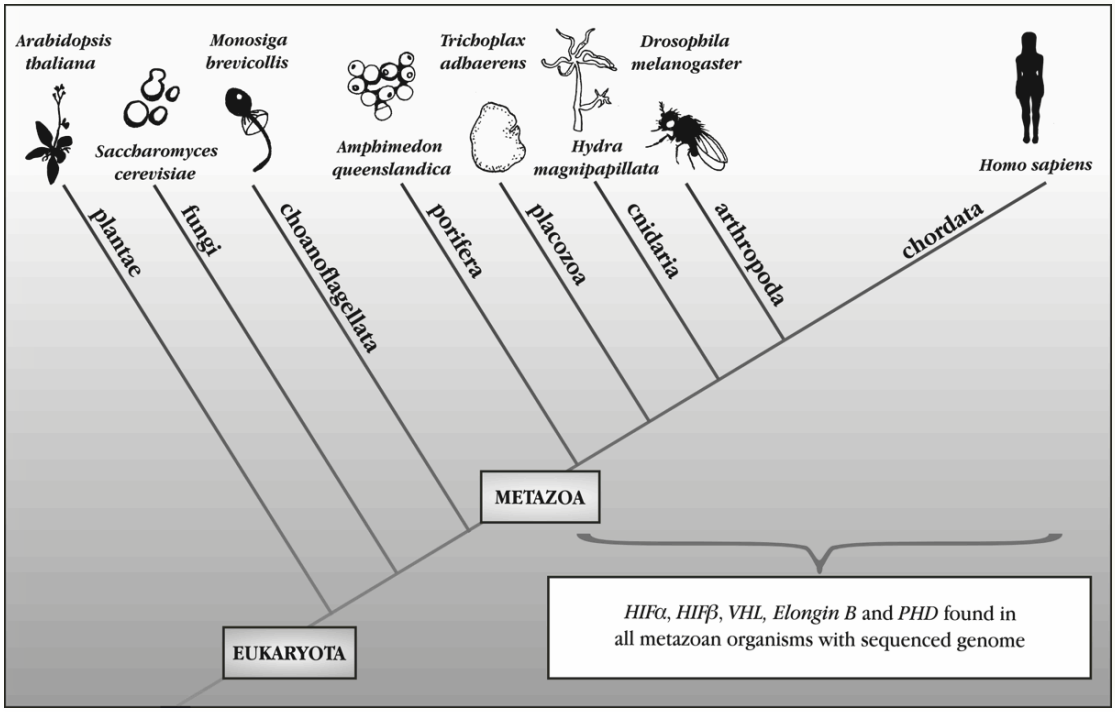


Figure 3.2. Phylogenetic tree, illustrating the relative positions of *Trichoplax adhaerens* and *Monosiga brevicollis* according to the results yielded from the analysis of their genomic DNA (Srivastava et al. 2008). While these studies suggest that, rather than being the most basal phylum, placozoans diverged after the sponges (porifera), choanoflagellates are considered to be the closest living relatives of metazoans.

3.3. The HIF pathway in *Trichoplax adhaerens*

A recently reported study to which I contributed (Loenarz et al. 2011) indicates that *Trichoplax adhaerens* employs the HIF pathway to respond to hypoxia in a way rather similar to the human mechanism. The genes *taPHD*, *taHIFα*, *taHIFβ* and *taVHL*, which are homologues of the human genes *PHD2*, *HIF-1α*, *HIF-1β* and *VHL,* were identified in *Trichoplax adhaerens.*

The predicted *Trichoplax adhaerens* HIFα contains the conserved basic helix-loop-helix (bHLH) and the PAS motives. The 548 amino acid protein also contains a *C*-terminal oxygen dependent degradation domain (CODD), but unlike human HIF-1α is lacking both the *N*-terminal equivalent (NODD) and the *C*-terminal transactivation domain

(CAD), modified by FIH1 in humans. No FIH homolog has been identified in *Trichoplax adhaerens* (Loenarz et al. 2011).

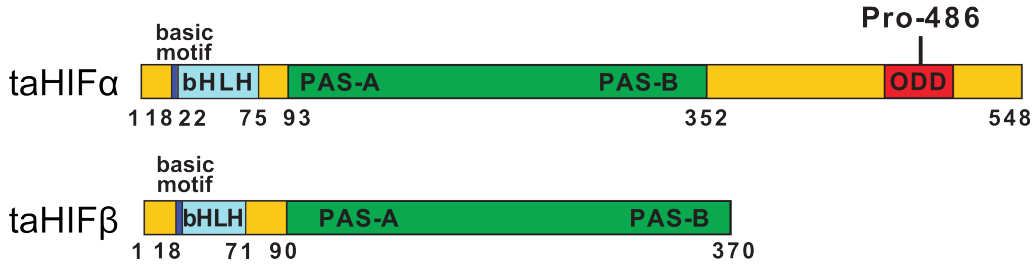


Figure 3.3. Predicted domain architectures of *Trichoplax adhaerens* HIFα and HIFβ. Abbreviations and functionality: Basic motif – DNA-binding sequence; bHLH – basic helix-loop-helix (allows for dimerisation of HIF subunits); PAS-A/B – PER-ARNT-SIM (thought to be involved in HIFα -HIFβ recognition that promotes HLH-dimerisation and DNA binding); ODD – oxygen-dependent degradation domain (region recognised and hydroxylated by PHDs). The domain architecture of taHIFα closely resembles hsHIF-2α, but is more minimalistic, containing only one hydroxylation site rather than three. The taODD resembles human CODD more than human NODD, suggesting that NODD evolved after CODD. The lack of an FIH hydroxylation site in taHIFα may also imply that the FIH-pathway evolved after the PHD-pathway.

Recombinant taPHD expressed in *E. coli* has been shown to hydroxylate a human CODD sequence (HIF-1$\alpha_{556-575}$) and the *Trichoplax adhaerens* ODD sequence (taHIF$\alpha_{477-497}$) *in vitro* at the evolutionary conserved prolyl residues Pro-564 and Pro-486, respectively. Further evidence for the involvement of taPHD in oxygen sensing was obtained from expressing taPHD in human 293T cells, which led to a reduction of human HIF-1α levels in both normoxia and hypoxia (Loenarz et al. 2011).

The employment of a VHL homolog in the degradation of taHIFα was demonstrated indirectly by observing binding of prolyl-hydroxylated taODD to human VHL by non-denaturing mass spectrometry.

Several genes with human homologs known to be up-regulated by HIF have also been up-regulated in *Trichoplax adhaerens* exposed to 2% atmospheric oxygen levels for 16 hours (see Table 3.1.).

Gene	Enzyme	Role of human homolog	Up-regulation
taCDO	Cysteine dioxygenase	Cysteine catabolism	2.5-fold
taPGK	Phospoglycerate kinase	Glycolysis	5.3-fold
taPHD	HIF-prolyl hydroxylase	HIF signalling	5.5-fold
taPLOD	Procollagen lysine-5-hydroxylase	Structural	5.5-fold
taIDE	Insulin-degrading enzyme	Glucose metabolism	7.3-fold
taALDO	Fructose-biphosphate aldolase	Glycolysis	16.8-fold
taPDK	Pyruvate dehydrogenase kinase	Glycolysis/citric acid cycle	24.3-fold

Table 3.1. The up-regulation of *Trichoplax* genes homologous to human HIF targets was induced by incubation of *Trichoplax* specimen at 2% oxygen levels for 16 hours and determined by RT-q-PCR. Normalised to β–*actin*; n=3; RNA extracted from 20-100 individuals (Loenarz et al. 2011).

These findings provide solid evidence for the existence of a fully functioning HIF pathway in *Trichoplax adhaerens*, making it the simplest organism known to employ PHD, HIF and pVHL for oxygen-dependent gene regulation.

4. *Monosiga brevicollis*, a Close Unicellular Relative of the Metazoans

Monosiga brevicollis is a unicellular protist belonging to the family of choanoflagellates (see Figure 3.2). Due to their morphological similarity to the feeding cells of sponges (choanocytes), choanoflagellates have long been considered to be candidates for the closest living relatives of metazoans (James-Clark, 1868). Despite this interest, relatively little is known about the cell biology and behaviour of choanoflagellates to this date. The recent completion of the full nuclear genome of *Monosiga brevicollis* has opened new doors to studies on the evolution of metazoan life and the involved signalling pathways (King et al. 2008). Inspection of the nuclear genome of *Monosiga brevicollis* suggested the presence of a PHD-like enzyme ('mbP4H'), the study of which is described in Section C.

4.1. Habitat and morphology of *Monosiga brevicollis*

Over 125 species of choanoflagellates have been identified. They are globally distributed and widely abundant in marine and freshwater environments (Buck & Garrison 1988; Thomsen & Larsen 1992). *Monosiga brevicollis* displays the typical choanoflagellate morphology, consisting of a main cell body, a single apical flagellum used for motion and a ring of actin-filled protrusions (microvilli), forming a collar around the flagellum for capture and ingestion of bacteria (see Figure 4.1.) (King et al. 2008).

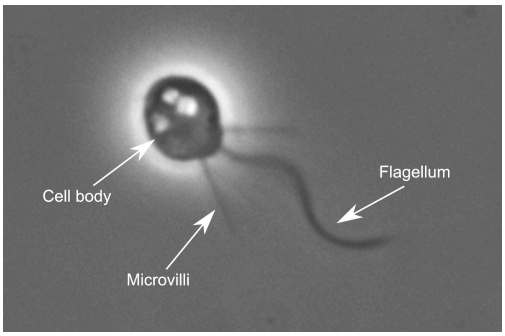


Figure 4.1. Phase contrast image of *Monosiga brevicollis*, highlighting the key choanoflagellate morphological features (image by Stephen Fairclough).

Choanoflagellates have a key role in the microbial food web by linking bacteria to higher trophic levels (positions in the food chain) (Boenigk & Arndt 2002). Choanoflagellates are free-swimming in water or adhering to a surface (Leadbeater 1983). There is little detailed knowledge into their life histories. Although *Monosiga brevicollis* is strictly solitary, coloniality has been observed with several other choanoflagellates, for example of the genus *Proterospongia*. Nevertheless, even the colonial species go through a solitary life phase lead (Leadbeater 1983). It is yet unclear whether choanoflagellates undergo sexual reproduction. While such behaviour has not been observed under laboratory conditions, the identification of key genes involved in meiosis suggests that a sexual phase in their life cycle is a plausible scenario (Carr et al. 2010).

4.2. Phylogenetic assignment of choanoflagellates

Morphological comparisons over a hundred years ago led to the proposal that choanoflagellates are a sister group of metazoans (James-Clark 1866). Analysis of mitochondrial genomes and expression sequence tags (ESTs) from a variety of protists supported this theory and established choanoflagellates as the closest unicellular relatives of metazoan (Lang et al. 2002; Ruiz-Trillo et al. 2008). The last common ancestor of choanoflagellates, placozoa, sponges and eumetazoa shared in the late pre-Cambrian is likely to have been a unicellular organism capable of forming simple colonies (Peterson & Butterfield 2005). Studying *Monosiga brevicollis* might provide us with crucial clues about the molecular biology and life cycle of this ancestor organism.

With around 42 megabases, the genome of *Monosiga brevicollis* is comparable in size to the genome of filamentous fungi and small diatoms (King et al. 2008). Annotation of the genome using protein domain prediction databases revealed 76 proteins domains that have previously been thought to be exclusive to metazoans (King et al. 2008). This is of particular interest as many of these domains are involved in cell signalling and adhesion

in metazoans (King et al. 2008). It is, however, unclear whether metazoan features missing in *Monosiga brevicollis* have evolved later or were subsequently lost. Consequently, with respect to metazoan evolution, it remains unclear which genes were adapted from ones pre-existing in unicellular organisms and which originated in the early metazoans (Ruiz-Trillo et al. 2008).

5. References

Adamska, M. et al., 2007. Wnt and TGF-beta expression in the sponge Amphimedon queenslandica and the origin of metazoan embryonic patterning. *PloS one*, 2(10), p.e1031. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2000352&tool=pmcentrez&rendertype=abstract [Accessed March 1, 2012].

Aleshin, V. & Petrov, N., 2002. Molecular evidence of regression in evolution of metazoa. *Zh Obshch Biol.*, 63(3), pp.195-208.

Alexeeva, S. et al., 2003. Requirement of ArcA for Redox Regulation in Escherichia coli under Microaerobic but Not Anaerobic or Aerobic Conditions. *Journal of Bacteriology*, 185(1), pp.204-209.

Appelhoff, R.J. et al., 2004. Differential function of the prolyl hydroxylases PHD1, PHD2, and PHD3 in the regulation of hypoxia-inducible factor. *The Journal of biological chemistry*, 279(37), pp.38458-65. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15247232 [Accessed March 6, 2012].

Aprelikova, O. et al., 2004. Regulation of HIF prolyl hydroxylases by hypoxia-inducible factors. *Journal of cellular biochemistry*, 92(3), pp.491-501. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15156561 [Accessed April 6, 2012].

Arany, Z. et al., 1996. An essential role for p300/CBP in the cellular response to hypoxia. *Proceedings of the National Academy of Sciences of the United States of America*, 93(23), pp.12969-73. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=24030&tool=pmcentrez&rendertype=abstract.

Bailey-Serres, J. & Chang, R., 2005. Sensing and signalling in response to oxygen deprivation in plants and other organisms. *Annals of botany*, 96(4), pp.507-18. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16051633 [Accessed March 17, 2012].

Berner, R.A., 1999. Ancient atmospheres and the evolution of oxygen sensing via the hypoxia-inducible factor in metazoans. *Proc. Natl. Acad. Sci USA*, 96(September), pp.10955-10957.

Berra, E. et al., 2003. HIF prolyl-hydroxylase 2 is the key oxygen sensor setting low steady-state levels of HIF-1alpha in normoxia. *The EMBO journal*, 22(16), pp.4082-90. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=175782&tool=pmcentrez&rendertype=abstract.

Blackstone, N.W., 2009. A new look at some old animals. *PLoS biology*, 7(1), p.e7. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2631055&tool=pmcentrez&rendertype=abstract [Accessed March 9, 2012].

Boenigk, J. & Arndt, H., 2002. Bacterivory by heterotrophic flagellates: community structure and feeding strategies. *Antonie van Leeuwenhoek*, 81(1-4), pp.465-80. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12448743.

Brusca, R. & Brusca, G., 2003. *Invertebrates* Second Edi., Elsevier, Inc.

Buck, K. & Garrison, D., 1988. Distribution and abundance of choanoflagellates (Acanthoecidae) across the ice-edge zone in the Weddell Sea, Antarctica. *Marine Biology*, 269(March 1986), pp.263-269.

Budd, G.E., 2008. The earliest fossil record of the animals and its significance. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 363(1496), pp.1425-34. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2614223&tool=pmcentrez&rendertype=abstract [Accessed March 16, 2012].

Bunn, H. & Poyton, R., 1996. Oxygen Sensing and Molecular to Hypoxia. *Physiological Reviews*, 76(3).

Carr, M., Leadbeater, B. S. C. & Baldauf, S.L., 2010. Conserved meiotic genes point to sex in the choanoflagellates. *The Journal of eukaryotic microbiology*, 57(1), pp.56-62. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20015185 [Accessed March 8, 2012].

Centanin, L., Ratcliffe, P. J. & Wappner, P., 2005. Reversion of lethality and growth defects in Fatiga oxygen-sensor mutant flies by loss of hypoxia-inducible factor-alpha/Sima. *EMBO reports*, 6(11), pp.1070-5. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1371028&tool=pmcentrez&rendertype=abstract [Accessed March 31, 2012].

Chowdhury, R., Hardy, A. & Schofield, C., 2008. The human oxygen sensing machinery and its manipulation. *Chemical Society reviews*, 37(7), pp.1281-1452.

Clifton, I.J. et al., 2006. Structural studies on 2-oxoglutarate oxygenases and related double-stranded beta-helix fold proteins. *Journal of inorganic biochemistry*, 100(4), pp.644-69. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16513174 [Accessed April 28, 2012].

Conway Morris, S., 2006. Darwin's dilemma: the realities of the Cambrian "explosion". *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 361(1470), pp.1069-83. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1578734&tool=pmcentrez&rendertype=abstract [Accessed March 13, 2012].

DeSalle, R. & Schierwater, B., 2008. An even "newer" animal phylogeny. *BioEssays : news and reviews in molecular, cellular and developmental biology*, 30(11-12), pp.1043-7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18937347 [Accessed March 16, 2012].

Dongen, J.T.V. et al., 2004. Phloem Import and Storage Metabolism Are Highly Coordinated by the Low Oxygen Concentrations within Developing Wheat Seeds 1. *Plant Physiology*, 135(July), pp.1809-1821.

Drew, M.C., 1997. OXYGEN DEFICIENCY AND ROOT METABOLISM: Injury and Acclimation Under Hypoxia and Anoxia. *Annual review of plant physiology and plant molecular biology*, 48, pp.223-250. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15012263.

Dudley, R., 1998. Atmospheric oxygen, giant Paleozoic insects and the evolution of aerial locomotor performance. *The Journal of experimental biology*, 201(Pt 8), pp.1043-50. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9510518.

D'Angelo, G. et al., 2003. Hypoxia up-regulates prolyl hydroxylase activity: a feedback mechanism that limits HIF-1 responses during reoxygenation. *The Journal of biological chemistry*, 278(40), pp.38183-7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12876291 [Accessed April 13, 2012].

Eitel, M & Schierwater, B, 2010. The phylogeography of the Placozoa suggests a taxon-rich phylum in tropical and subtropical waters. *Molecular ecology*, 19(11), pp.2315-27. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20604867 [Accessed March 19, 2012].

Eitel, Michael et al., 2011. New insights into placozoan sexual reproduction and development. *PloS one*, 6(5), p.e19639. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3098260&tool=pmcentrez&rendertype=abstract [Accessed March 7, 2012].

Elkins, J.M. et al., 2003. Structure of factor-inhibiting hypoxia-inducible factor (HIF) reveals mechanism of oxidative modification of HIF-1 alpha. *The Journal of biological chemistry*, 278(3), pp.1802-6. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12446723 [Accessed April 2, 2012].

Elliott, R.L. & Blobe, G.C., 2005. Role of transforming growth factor Beta in human cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 23(9), pp.2078-93. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15774796 [Accessed March 10, 2012].

Elvidge, G.P. et al., 2006. Concordant regulation of gene expression by hypoxia and 2-oxoglutarate-dependent dioxygenase inhibition: the role of HIF-1alpha, HIF-2alpha, and other pathways. *The Journal of biological chemistry*, 281(22), pp.15215-26. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16565084 [Accessed March 28, 2012].

Ema, M. et al., 1997. A novel bHLH-PAS factor with close sequence similarity to hypoxia-inducible factor 1 regulates the VEGF expression and is potentially involved in lung and vascular development. *Proc. Natl. Acad. Sci USA*, 94(April), pp.4273-4278.

Ender, A. & Schierwater, B., 2003. Placozoa Are Not Derived Cnidarians: Evidence from Molecular Morphology. *Molecular Biology and Evolution*, 20(1), pp.130-134. Available at: http://mbe.oxfordjournals.org/cgi/doi/10.1093/molbev/msg018 [Accessed March 12, 2012].

Epstein, A. C. et al., 2001. C. elegans EGL-9 and mammalian homologs define a family of dioxygenases that regulate HIF by prolyl hydroxylation. *Cell*, 107(1), pp.43-54. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11595184.

Falkowski, P.G. et al., 2005. The rise of oxygen over the past 205 million years and the evolution of large placental mammals. *Science (New York, N.Y.)*, 309(5744), pp.2202-4. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16195457 [Accessed March 1, 2012].

Fischer, H.M., 1994. Genetic regulation of nitrogen fixation in rhizobia. *Microbiological reviews*, 58(3), pp.352-86. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=372973&tool=pmcentrez&rendertype=abstract.

Fraschetti, S., Terlizzi, A. & Boero, F., 2008. How many habitats are there in the sea (and where)? *Journal of Experimental Marine Biology and Ecology*, 366(1-2), pp.109-115. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0022098108003353 [Accessed April 23, 2012].

Freedman, S. et al., 2002. Structural basis for recruitment of CBP$\overline{p}$300 by hypoxia-inducible factor-1 *Biophysics*, 99(8), pp.5367-5372.

Geigenberger, P., 2003. Response of plant metabolism to too little oxygen. *Current Opinion in Plant Biology*, 6(3), pp.247-256. Available at: http://linkinghub.elsevier.com/retrieve/pii/S1369526603000384 [Accessed March 20, 2012].

Gilles-Gonzalez, M. a et al., 1994. Heme-based sensors, exemplified by the kinase FixL, are a new class of heme protein with distinctive ligand binding and autoxidation. *Biochemistry*, 33(26), pp.8067-73. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8025112.

Giribet, G., 2002. Relationships among metazoan phyla as inferred from 18S rRNA sequence data: a methodological approach. *EXS*, 92, pp.85-101.

Grell, K., 1981. Trichoplax adhaerens and the origin of Metazoa. *Accademia nazionale dei Lincei*, pp.107–121. Available at: http://onlinelibrary.wiley.com/doi/10.1002/cbdv.200490137/abstract [Accessed April 29, 2012].

Grell, K., 1971. Trichoplax adhaerens, F.E. Schulze und die Entstehung der Metazoen F. Harrison & J. Westfall, eds. *Naturwiss. Rundsch.*, 24, pp.160-161.

Grell, K. & Benwitz, G., 1971. Die Ultrastruktur von Trichoplax adhaerens. *Cytobiologie*, 4, pp.216–240.

Grell, K. & Benwitz, G., 1981. Ergänzende untersuchungen zur ultrastruktur von Trichoplax adhaerens. *Zoomorphology*, 98,, pp.47–67.

Grell, K. & Ruthmann, A., 1991. Placozoa. In F. Harrison & J. Westfall, eds. *Microscopic Anatomy of Invertebrates, Placozoa, Porifera, Cnidaria, and Ctenophorn*. New York: Elsevier, Inc., pp. 13–28.

Gu, Y. et al., 1998. Molecular characterization and chromosomal localization of a third alpha-class hypoxia inducible factor subunit, HIF3alpha. *Gene Expression*, 7(3), pp.205-213.

Hampton-Smith, R.J. & Peet, D.J., 2009. From polyps to people: a highly familiar response to hypoxia. *Annals of the New York Academy of Sciences*, 1177, pp.19-29. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19845603 [Accessed April 28, 2012].

Hejnol, A. et al., 2009. Assessing the root of bilaterian animals with scalable phylogenomic methods. *Proceedings. Biological sciences / The Royal Society*, 276(1677), pp.4261-70. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2817096&tool=pmcentrez&rendertype=abstract [Accessed March 10, 2012].

Hewitson, K.S. et al., 2005. Oxidation by 2-oxoglutarate oxygenases: non-haem iron systems in catalysis and signalling. *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, 363(1829), pp.807-28; discussion 1035-40. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15901537 [Accessed March 15, 2012].

Hill, J.J. et al., 2009. Glycoproteomic analysis of two mouse mammary cell lines during transforming growth factor (TGF)-beta induced epithelial to mesenchymal transition. *Proteome Science*, 7(1), p.2. Available at: http://www.proteomesci.com/content/7/1/2 [Accessed May 5, 2012].

Hon, T. et al., 2003. A mechanism of oxygen sensing in yeast. Multiple oxygen-responsive steps in the heme biosynthetic pathway affect Hap1 activity. *The Journal of biological chemistry*, 278(50), pp.50771-80. Available at: http://www.ncbi.nlm.nih.gov/pubmed/14512429 [Accessed April 16, 2012].

James-Clark, H., 1866. Note on the infusoria flagellata and the spongiae ciliatae. *Am. J. Sci*, 1, pp.113–114.

James-Clark, H., 1868. On the spongiae ciliatae as infusoria flagellata; or observations on the structure, animality, and relationship of Leucosolenia botryoides. *Ann. Mag. Nat. His.*, 1, pp.133–142, 188–215. 250–264.

Jiang, Y. et al., 2003. Gene Expression Profiling in a Renal Cell Carcinoma Cell Line: Dissecting VHL and Hypoxia-Dependent Pathways. *Molecular Cancer Research*, (1), pp.453-462.

de Jong, D. et al., 2009. Multiple dicer genes in the early-diverging metazoa. *Molecular biology and evolution*, 26(6), pp.1333-40. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19276153 [Accessed March 1, 2012].

Kaelin, W.G. & Ratcliffe, Peter J, 2008. Oxygen sensing by metazoans: the central role of the HIF hydroxylase pathway. *Molecular cell*, 30(4), pp.393-402. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18498744 [Accessed March 2, 2012].

Kamachi, Y. et al., 2000. Pairing SOX off with partners in the regulatin of embryonic development. *TIG*, 9525(99), pp.2-7.

Ke, Q. & Costa, M., 2006. Hypoxia-Inducible Factor-1 (HIF-1). *Molecular Pharmacology*, 70(5), pp.1469-1480.

Kiley, P.J. & Beinert, H., 2003. The role of Fe–S proteins in sensing and regulation in bacteria. *Current Opinion in Microbiology*, 6(2), pp.181-185. Available at: http://linkinghub.elsevier.com/retrieve/pii/S1369527403000390 [Accessed March 12, 2012].

King, N. et al., 2008. The genome of the choanoflagellate Monosiga brevicollis and the origin of metazoans. *Nature*, 451(7180), pp.783-8. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2562698&tool=pmcentrez&rendertype=abstract [Accessed March 5, 2012].

Knoll, a. H., 1999. Early Animal Evolution: Emerging Views from Comparative Biology and Geology. *Science*, 284(5423), pp.2129-2137. Available at: http://www.sciencemag.org/cgi/doi/10.1126/science.284.5423.2129 [Accessed March 6, 2012].

Knowles, H.J. et al., 2003. Effect of Ascorbate on the Activity of Hypoxia-inducible Factor in Cancer Cells. *Cancer Research*, 63, pp.1764-1768.

Koong, A.C. et al., 2000. Hypoxia-inducible factors and the response to hypoxic stress. *Cancer Research*, 60, pp.883-887.

Kopan, R. & Ilagan, M.X.G., 2009. The canonical Notch signaling pathway: unfolding the activation mechanism. *Cell*, 137(2), pp.216-33. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2827930&tool=pmcentrez&rendertype=abstract [Accessed March 2, 2012].

Krude, T. et al., 1997. Cyclin/Cdk-dependent initiation of DNA replication in a human cell-free system. *Cell*, 88(1), pp.109-19. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9019396.

Kwast, K.E., Burke, P.V. & Poyton, R.O., 1998. Oxygen sensing and the transcriptional regulation of oxygen-responsive genes in yeast. *The Journal of experimental biology*, 201(Pt 8), pp.1177-95. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9510529.

Lando, D. et al., 2002. Asparagine Hydroxylation of the HIF Transactivation Domain : A Hypoxic Switch. *Science*, 295(February), pp.858-861.

Lando, D. et al., 2003. Oxygen-dependent regulation of hypoxia-inducible factors by prolyl and asparaginyl hydroxylation. *European Journal of Biochemistry*, 270(5), pp.781-790. Available at: http://doi.wiley.com/10.1046/j.1432-1033.2003.03445.x [Accessed April 29, 2012].

Lang, B.F. et al., 2002. The closest unicellular relatives of animals. *Current biology : CB*, 12(20), pp.1773-8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18717747.

Lang, D. et al., 2007. PAX genes: roles in development, pathophysiology, and cancer. *Biochemical pharmacology*, 73(1), pp.1-14. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16904651 [Accessed March 6, 2012].

Leadbeater, B. S. C., 1983. Life-history and ultrastructure of a new marine species of Proterospongia (Choanoflagellida). *Journal of the Marine Biological Association of the United Kingdom*, 63, pp.135-160. Available at: http://www.journals.cambridge.org/abstract_S0025315400049857.

Leite, R.B., Brito, A.B. & Cancela, M.L., 2008. An oxygen molecular sensor, the HIF prolyl 4-hydroxylase, in the marine protist Perkinsus olseni. *Protist*, 159(3), pp.355-68. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18539525 [Accessed March 22, 2012].

Loenarz, C., 2010. *Mechanistic and Evolutionary Studies on Oxygen Sensing in Animals*. Elsevier, Inc.

Loenarz, C. et al., 2011. The hypoxia-inducible transcription factor pathway regulates oxygen sensing in the simplest animal, Trichoplax adhaerens. *EMBO reports*, 12(1), pp.63-70. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21109780 [Accessed April 19, 2012].

Logan, C.Y. & Nusse, R., 2004. The Wnt signaling pathway in development and disease. *Annual review of cell and developmental biology*, 20, pp.781-810. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15473860 [Accessed March 2, 2012].

Majmundar, A.J., Wong, W.J. & Simon, M.C., 2010. Hypoxia-inducible factors and the response to hypoxic stress. *Molecular cell*, 40(2), pp.294-309. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3143508&tool=pmcentrez&rendertype=abstract [Accessed March 7, 2012].

Manuel, M., 2009. Early evolution of symmetry and polarity in metazoan body plans. *Comptes rendus biologies*, 332(2-3), pp.184-209. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19281951 [Accessed March 5, 2012].

Martinelli, C. & Spring, J., 2003. Distinct expression patterns of the two T-box homologues Brachyury and Tbx2/3 in the placozoan Trichoplax adhaerens. *Development genes and evolution*, 213(10), pp.492-9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/13680223 [Accessed March 7, 2012].

Maruyama, Y.K., 2004. Occurrence in the field of a long-term, year-round, stable population of placozoans. *The Biological bulletin*, 206(1), pp.55-60. Available at: http://www.ncbi.nlm.nih.gov/pubmed/14977730.

Masson, N. et al., 2001. Independent function of two destruction domains in hypoxia-inducible factor-α chains activated by prolyl hydroxylation. *EMBO Journal*, 20(18), pp.5197-5206. Available at: http://www.nature.com/doifinder/10.1093/emboj/20.18.5197.

Maxwell, P.H. et al., 1999. The tumour suppressor protein VHL targets hypoxia-inducible factors for oxygen-dependent proteolysis. *Nature*, 399(6733), pp.271-5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10353251.

Maxwell, P.H., Pugh, C W & Ratcliffe, P J, 1993. Inducible operation of the erythropoietin 3' enhancer in multiple cell lines: evidence for a widespread oxygen-sensing mechanism. *Proceedings of the National Academy of Sciences of the United States of America*, 90(6), pp.2423-7. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=46099&tool=pmcentrez&rendertype=abstract.

McDonough, M. A. et al., 2006. Cellular oxygen sensing: Crystal structure of hypoxia-inducible factor prolyl hydroxylase (PHD2). *Proceedings of the National Academy of Sciences of the United States of America*, 103(26), pp.9814-9. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1502536&tool=pmcentrez&rendertype=abstract.

McElwain, J.C., Wagner, P.J. & Hesselbo, S.P., 2009. Fossil plant relative abundances indicate sudden loss of Late Triassic biodiversity in East Greenland. *Science (New York, N.Y.)*, 324(5934), pp.1554-6. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19541995 [Accessed March 6, 2012].

Miller, D.J. & Ball, E.E., 2008. Animal evolution: Trichoplax, trees, and taxonomic turmoil. *Current biology : CB*, 18(21), pp.R1003-5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19000794.

Monteiro, A. S. et al., 2006. A low diversity of ANTP class homeobox genes in Placozoa. *Evolution & development*, 8(2), pp.174-82. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16509895.

Murray, P.J., 2007. The JAK-STAT Signaling Pathway: Input and Output Integration. *Journal of Immunology*, 178, pp.2623-2629.

Myatt, S.S. & Lam, E.W.-F., 2007. The emerging roles of forkhead box (Fox) proteins in cancer. *Nature reviews. Cancer*, 7(11), pp.847-59. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17943136 [Accessed March 1, 2012].

Naiche, L. a et al., 2005. T-box genes in vertebrate development. *Annual review of genetics*, 39, pp.219-39. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16285859 [Accessed March 1, 2012].

Nisbet, E.G. & Sleep, N.H., 2001. The habitat and nature of early life. *Nature*, 409(6823), pp.1083-91. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11234022.

Pan, Y. et al., 2007. Multiple factors affecting cellular redox status and energy metabolism modulate hypoxia-inducible factor prolyl hydroxylase activity in vivo and in vitro. *Molecular and cellular biology*, 27(3), pp.912-25. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1800695&tool=pmcentrez&rendertype=abstract [Accessed March 24, 2012].

Pearse, V., 1989. Growth and behavior of Trichoplax adhaerens : First record of the phylum Placozoa in Hawaii. , pp.117-121.

Percy, M.J. et al., 2006. A family with erythrocytosis establishes a role for prolyl hydroxylase domain protein 2 in oxygen homeostasis.

Peterson, K.J. & Butterfield, N.J., 2005. Origin of the Eumetazoa: testing ecological predictions of molecular clocks against the Proterozoic fossil record. *Proceedings of the National Academy of Sciences of the United States of America*, 102(27), pp.9547-52. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1172262&tool=pmcentrez&rendertype=abstract.

Philippe, H. et al., 2004. Phylogenomics of eukaryotes: impact of missing data on large alignments. *Molecular biology and evolution*, 21(9), pp.1740-52. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15175415 [Accessed March 12, 2012].

Philippe, H. et al., 2009. Phylogenomics revives traditional views on deep animal relationships. *Current biology : CB*, 19(8), pp.706-12. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19345102 [Accessed March 9, 2012].

Powell-Coffman, J.A., 2010. Hypoxia signaling and resistance in C. elegans. *Trends in endocrinology and metabolism: TEM*, 21(7), pp.435-40. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20335046 [Accessed March 17, 2012].

Poyton, R.O., 1999. Models for oxygen sensing in yeast: implications for oxygen-regulated gene expression in higher eucaryotes. *Respiration physiology*, 115(2), pp.119-33. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10385027.

Putnam, N.H. et al., 2007. Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science (New York, N.Y.)*, 317(5834), pp.86-94. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17615350 [Accessed March 9, 2012].

Rich, P., 2003. The molecular machinery of Keilin's respiratory chain. *Biochemical Society Transactions*, 31, pp.1095–1105.

Ruiz-Trillo, I. et al., 2008. A phylogenomic investigation into the origin of metazoa. *Molecular biology and evolution*, 25(4), pp.664-72. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18184723 [Accessed March 9, 2012].

Ruthmann, A., Behrendt, G. & Wahl, R., 1986. The ventral epithelium of Trichoplax adhaerens (Placozoa). *Zoomorphology*, 106,, pp.115–122.

Rytkönen, K.T. et al., 2011. Molecular evolution of the metazoan PHD-HIF oxygen-sensing system. *Molecular biology and evolution*, 28(6), pp.1913-26. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21228399 [Accessed April 28, 2012].

Sang, N. et al., 2002. Carboxyl-Terminal Transactivation Activity of Hypoxia-Inducible Factor 1alpha Is Governed by a von Hippel-Lindau Protein-Independent, Hydroxylation-Regulated Association with p300/CBP. *Molecular and Cellular Biology*, 22(9), pp.2984-2992.

Kent, W., *A Manual of the Infusoria*, London: Elsevier, Inc.

Schierwater, Bernd, 2005. My favorite animal, Trichoplax adhaerens. *BioEssays : news and reviews in molecular, cellular and developmental biology*, 27(12), pp.1294-302. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16299758 [Accessed March 9, 2012].

Schierwater, Bernd, de Jong, D. & Desalle, R., 2009. Placozoa and the evolution of Metazoa and intrasomatic cell differentiation. *The international journal of biochemistry & cell biology*, 41(2), pp.370-9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18935972 [Accessed March 9, 2012].

Schofield, C.J. & Ratcliffe, Peter J, 2004. Oxygen sensing by HIF hydroxylases. *Nature reviews. Molecular cell biology*, 5(5), pp.343-54. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15122348 [Accessed March 12, 2012].

Schulze, F., 1883. Trichoplax adhaerens. *Zool. Anz.*, 6, pp.92-97.

Schütze, J. et al., 1999. Evolutionary relationships of Metazoa within the eukaryotes based on molecular data from Porifera. *Proceedings. Biological sciences / The Royal Society*, 266(1414), pp.63-73. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1689648&tool=pmcentrez&rendertype=abstract.

Semenza, G. L., 2001a. Hypoxia-inducible factor 1: control of oxygen homeostasis in health and disease. *Pediatric research*, 49(5), pp.614-7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11328942.

Semenza, G. L., 2001b. Hypoxia-inducible factor 1: oxygen homeostasis and disease pathophysiology. *Trends in molecular medicine*, 7(8), pp.345-50. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11516994.

Semenza, G. L., 1991. Hypoxia-inducible nuclear factors bind. , 88(July), pp.5680-5684.

Semenza, G. L., 2007. Oxygen-dependent regulation of mitochondrial respiration by hypoxia-inducible factor 1. *The Biochemical journal*, 405(1), pp.1-9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17555402 [Accessed March 6, 2012].

Siddall, M.E., 2010. Cladistics Unringing a bell : metazoan phylogenomics and the partition bootstrap. *Cladistics*, 26, pp.444-452.

Signorovitch, A.Y., Buss, L.W. & Dellaporta, S.L., 2007. Comparative genomics of large mitochondria in placozoans. *PLoS genetics*, 3(1), p.e13. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1781491&tool=pmcentrez&rendertype=abstract [Accessed April 29, 2012].

Silva, F.B.D., Muschner, V.C. & Bonatto, S.L., 2007. Phylogenetic position of Placozoa based on large subunit (LSU) and small subunit (SSU) rRNA genes. *Genetics and Molecular Biology*, 30(1), pp.127-132. Available at: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1415-47572007000100022&lng=en&nrm=iso&tlng=en.

Srivastava, M. et al., 2008. The Trichoplax genome and the nature of placozoans. *Nature*, 454(7207), pp.955-60. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18719581.

Syed, T. & Schierwater, B, 2002. TRICHOPLAX ADHAERENS: DISCOVERED AS A MISSING LINK , FORGOTTEN AS A HYDROZOAN , RE-DISCOVERED AS A KEY TO METAZOAN EVOLUTION. *Vie Milieu*, 52(4), pp.177-187.

Taylor, C.T. & McElwain, J.C., 2010. Ancient atmospheres and the evolution of oxygen sensing via the hypoxia-inducible factor in metazoans. *Physiology (Bethesda, Md.)*, 25(5), pp.272-9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20940432 [Accessed April 28, 2012].

Thomsen, H. & Larsen, J., 1992. Loricate choanoflagellates of the Southern Ocean with new observations on cell division in Bicosta spinifera (Throndsen, 1970) from Antarctica and Saroeca attenuata Thomsen, 1979, from the Baltic Sea. *Polar Biology*, 168, pp.53-63.

Tyler, M., 2000. *Developmental Biology: A guide for experimental study* 2nd Editio. M. Sunderland, ed., Elsevier, Inc.

Unden, G. & Bongaerts, J., 1997. Alternative respiratory pathways of Escherichia coli: energetics and transcriptional regulation in response to electron acceptors. *Biochimica et biophysica acta*, 1320(3), pp.217-34. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9230919.

Viault, F., 1890. Sur l'augemtation considerable du nombre glubules rouges das le sang chez les habitants des hauts plateaux de l'Amerique du Sud. *C.R. Acad. Sci. Paris*, 111, pp.917-918.

Wainright, P. et al., 1993. Monophyletic origins of the metazoa: an evolutionary link with fungi. *Science*, 260, pp.340-342.

Wang, G.L. et al., 1995. Hypoxia-inducible factor 1 is a basic-helix-loop-helix-PAS heterodimer regulated by cellular O2 tension. *Proceedings of the National Academy of Sciences of the United States of America*, 92(12), pp.5510-4. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=41725&tool=pmcentrez&rendertype=abstract.

Webb, J.D., Coleman, M.L. & Pugh, Christopher W, 2009. Hypoxia, hypoxia-inducible factors (HIF), HIF hydroxylases and oxygen sensing. *Cellular and molecular life sciences : CMLS*, 66(22), pp.3539-54. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19756382 [Accessed March 5, 2012].

Wei, W. & Yu, X.D., 2007. Hypoxia-inducible factors: crosstalk between their protein stability and protein degradation. *Cancer letters*, 257(2), pp.145-56. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17920193 [Accessed April 20, 2012].

van der Wel, H., Ercan, A. & West, C.M., 2005. The Skp1 prolyl hydroxylase from Dictyostelium is related to the hypoxia-inducible factor-alpha class of animal prolyl 4-hydroxylases. *The Journal of biological chemistry*, 280(15), pp.14645-55. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15705570 [Accessed April 28, 2012].

West, C.M. et al., 2004. Cytoplasmic glycosylation of protein-hydroxyproline and its relationship to other glycosylation pathways. *Biochimica et biophysica acta*, 1673(1-2), pp.29-44. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15238247 [Accessed April 26, 2012].

Wykoff, C.C. et al., 2000. Identification of novel hypoxia dependent and independent target genes of the von Hippel-Lindau (VHL) tumour suppressor by mRNA dierential expression profiling. *Oncogene*, 19, pp.6297-6305.

Zheng, N. et al., 2002. Structure of the Cul1-Rbx1-Skp1-F boxSkp2 SCF ubiquitin ligase complex. *Nature*, 416(6882), pp.703-9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11961546.

6. Aims of my work

The objective of my work was to compare the structural and biochemical properties of human PHD2 with taPHD, the PHD2-homolog of the most simple extant metazoan *Trichoplax adhaerens,* and mbP4H, a PHD-like enzyme from a species belonging to the choanoflagellates, thought to be the closest living unicellular relatives of metazoans. The aim of the comparison of hsPHD2, taPHD and mbP4H was to increase our understanding of the evolutionary origin of the involvement of PHDs in oxygen sensing.

Because of the proposed basal phylogenetic position of *Trichoplax adhaerens*, taPHD possibly represents the most ancient version of a PHD known to date. The structural and biochemical features common to both hsPHD2 and taPHD therefore might be essential to their function as molecular oxygen sensors rather than later adaptations.

The *Monosiga brevicollis* enzyme mbP4H has not been studied previously. The objective of studying mbP4H was to investigate the biochemical properties of this 2OG-oxygenase, compare them to the properties of taPHD and hsPHD2 and subsequently assess whether it might be involved in oxygen sensing.

Part A: Crystal Structures of the *Trichoplax Adhaerens* HIF Prolyl Hydroxylase

A.1. Insights Gained from the Crystal Structures of Human PHD2

A.1.1. Available crystal structures of human PHD2

To this date, crystal structures of human PHD2 have been obtained with 2OG and a range of inhibitor ligands in the presence of a metal ion, either Fe(II) or Mn(II) (see Table A.1.1.) .

PDB ID	Structure	Resolution	Year
2Y34	*S*-Nitrosylated hsPHD2 in complex with Fe(II) and UN9	2.01 Å	2010
2Y33	*S*-Nitrosylated hsPHD2 in complex with Zn(II) and UN9	2.00 Å	2010
3OUJ	hsPHD2 in complex with 2-oxoglutarate	2.30 Å	2010
3OUI	hsPHD2-R717 with Fe(II) and 24Z	1.70 Å	2010
3OUH	hsPHD2-R127 with Fe(II) and 014	2.51 Å	2010
3HQR	hsPHD2 in complex with Mn(II), *N*-oxalylglycine and HIF-1α substrate	2.00 Å	2009
3HQU	hsPHD2 in complex with Fe(II), UN9 and partial HIF-1α substrate	2.30 Å	2009
2HBT	hsPHD2 in complex with Fe(II) and UN9	1.85 Å	2006
2HBU	hsPHD2 in complex with Fe(II) and UN9	1.60 Å	2006
2G19	hsPHD2 in complex with Fe(II) and 4HG	1.70 Å	2006
2G1M	hsPHD2 in complex with Fe(II) and 4HG	2.20 Å	2006

Table A.1.1. Currently deposited crystal structures of human PHD2 (March 2012), identified by their Protein Data Bank (PDB) ID codes. The structure containing PHD2 complexed with a HIF-1α substrate HIF-1α (556-574) is highlighted in green. The highest resolution structure is highlighted in orange (see Figure A.1.1. for ligand structures).

In all of the cases, a truncated version of PHD2 lacking the PHD-MYND domain was used. The use of 2OG and Fe(II) requires work under anaerobic conditions to prevent 2OG turnover and oxidation of Fe(II) to Fe(III), making the experimental procedures for obtaining and handling crystals difficult. Only two structures are available with a substrate peptide (PDB ID 3HQR & 3HQU). Of these two structures, only one (3HQR) contains the substrate, a 19mer corresponding to human HIF-1α (556-574), bound within the active site. This is possibly the most interesting structure in the context of

evolution of PHDs because it yields extensive information about the types of interactions involved between human PHD2 and its substrate.

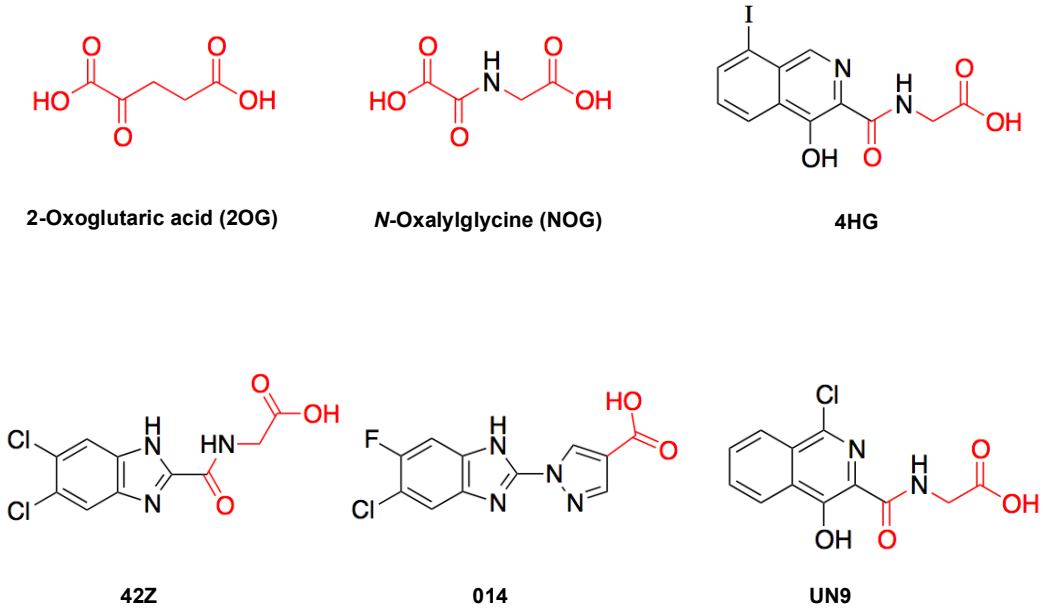


Figure A.1.1. Ligands present in currently deposited human PHD2 (hsPHD2) structures. All of the inhibitors mimic 2OG. NOG binds in a way similar to 2OG but cannot be turned over to succinate. The bicyclic inhibitors bind to the 2OG binding site through the acidic group and coordinate the active site metal with the amide oxygen (benzimidazole nitrogen for 014). The bicyclic aromatic ring system interacts with Y310 of hsPHD2 through π-stacking and perturbs the intramolecular hydrogen bond between D254 and R252 (see Figure A.1.3.).

The highest resolution structure available for hsPHD2, 2HBU, has a resolution of 1.60 Å and contains hsPHD2 in complex with a bicyclic ligand that shall be referred to using its PDB ID 'UN9' throughout this work (Hsieh et al. 2007).

Experience suggests that in structures refined to a resolution of 2-1.5 Å, most residues and folds can be fitted correctly, but many small errors involving rotamers and alternative conformations of side chains, ligand positions and orientations and surface waters are common. At resolutions poorer than 2 Å, side chain rotamers can be difficult to fit, whereas resolutions worse than 3 Å can even yield incorrect prediction of the backbone secondary structure and position. Models of high accuracy with respect to backbone position and folds, side chain rotamers and alternative conformations, ligand and surface water positions can be obtained from data collected at a resolution of 1.5 Å and better (Rhodes 2006). Resolution is, therefore, one of the limitations of currently

deposited hsPHD2 structures. A higher resolution structure of hsPHD2 would yield a more accurate model, facilitate the studies of interactions with ligands and the substrate and be useful in inhibitor design. Other limitations of currently available PHD2 structures are the lack of the PHD-MYND domain and the poor density in the mobile region, called the substrate-enclosing β2/β3 finger-loop region (see Section A.1.2.), apparently caused by disorder. Additionally, the substrate used in the structure with the PDB ID 3HQR is 19 amino acids long, only 17 of which are visible. Combined, these limitations impose constraints on the identification of enzyme-substrate interactions, which are important in understanding the mechanism of action of the PHDs and the evolution of their involvement in oxygen sensing.

A.1.2. Key structural features of hsPHD2

Human PHD2 is composed of two domains, an *N*-terminal domain (residues 21-58) that has homology[1] to MYND-type (from Myeloid-Nervy-DEAF-1) zinc fingers, and the catalytic *C*-terminal domain (residues 181-426). Only truncated constructs lacking the *N*-terminal domain have been crystallised successfully. Therefore, structural information about the MYND domain in PHDs is unavailable. The catalytic domain has homology to other 2OG-oxygenases and contains the conserved double-stranded β-helix (DSBH) fold (see Figure A.1.2.). This motif is composed of eight β-strands (βI to βVIII) (Stirk et al. 1992) that form the major (larger) β-sheet (βI, βVIII, βIII, βVI) and the minor (smaller) β-sheet (βII, βVII, βIV, βV). The DSBH is stabilised by three α-helices (α1-3) that pack along the major β-sheet. hsPHD2 also contains an *N*-terminal helix-strand-helix-strand motif conserved in DAOCS (Valegard et al. 1998; Clifton et al. 2006). The substrate peptide binds along a well-defined groove. Superimposition of the hsPHD2 structures with and without substrate peptide show that while the conformations of the major and minor β-sheets as well as of the α-helices α1-3 are very similar when the

1. When referring to proteins, *homology* means that two proteins, referred to as *homologues*, are derived from the same ancestor protein (Koonin, 2005).

substrate is bound, there is evidence that the region between β2 and β3 (residues 237-254) forms a mobile element.

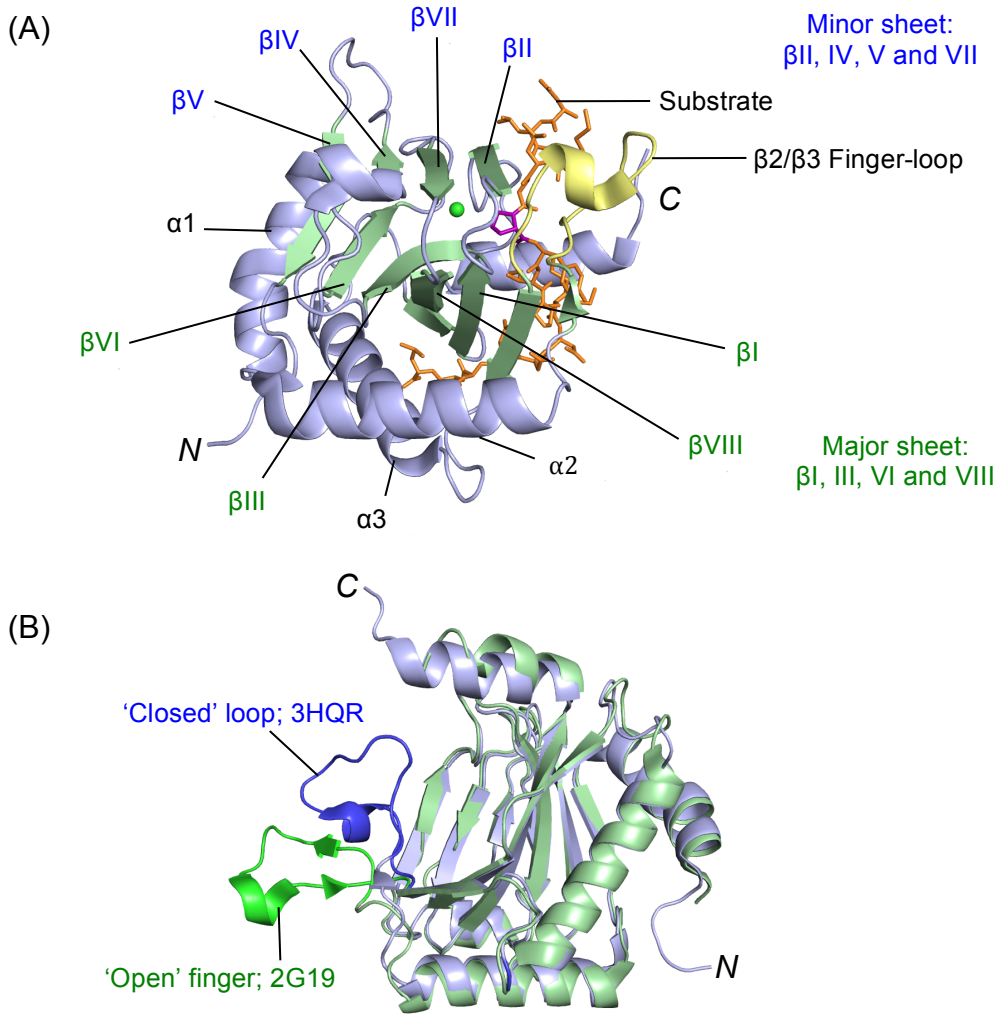


Figure A.1.2. Views from crystal structures of human PHD2, (A) showing the DSBH motif (pale green) and the positions of the α-helices (gray), the β2/β3 finger-loop (yellow), the catalytic site metal ion (bright green) and the substrate (orange sticks with purple P564) in 3HQR; (B) Superimposition of 3HQR (gray) and 2G19 (light green), illustrating the β2/β3 region in 'open' finger (bright green) and 'closed' loop (blue) conformations (substrate has been omitted from 3HQR for clarity).

It was suggested that without substrate, this region adopts a finger-like conformation, while in the presence of bound peptide, it adopts a loop conformation, enclosing the substrate within the active site (Chowdhury et al. 2009).

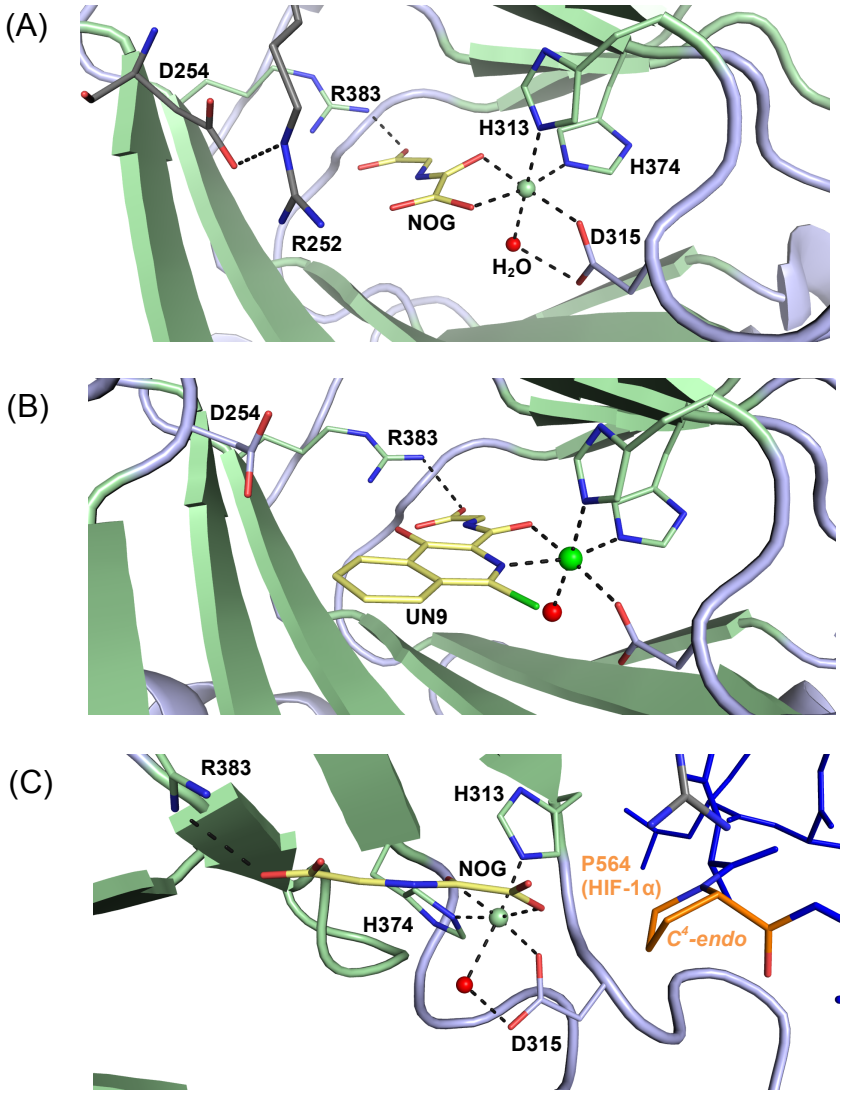


Figure A.1.3. Close-up views of active site of human PHD2. (A) Octahedral coordination around the active site metal and intramolecular hydrogen bond D254-R252 (PDB ID=3HQR); (B) Binding of bicyclic inhibitor UN9 in the structure with PDB ID=2HBT. The intramolecular hydrogen bond D254-R252 is disrupted by the inhibitor; (C) *C^4-endo* conformation adopted by $P564_{HIF\text{-}1\alpha}$ (orange, PDB ID=3HQR).

The active site of human PHD2 is located between the major and minor β-sheets, lying rather deep within the structure in comparison to most other 2OG-oxygenases (Chowdhury et al. 2009). A metal ion, Fe(II) or its surrogate Mn(II), is octahedrally coordinated by H313, D315, H374, the ligand (2OG, NOG or one of the bicyclic inhibitors) and a water molecule (see Figure A.1.3.) (McDonough et al. 2006). While the hsPHD2 residues form a facial triad, 2OG or the inhibitors bind in a bidentate fashion (McDonough et al. 2006). In the substrate structure, the metal-coordinating water

molecule appears to be positioned to form a hydrogen bond to the non-coordinating oxygen of D315. It is suggested that this arrangement secures the Fe(II) in a 6-coordinate state and hinders oxygen binding to Fe(II), thereby preventing turnover of 2OG in the absence of a substrate (Neidig et al. 2007).

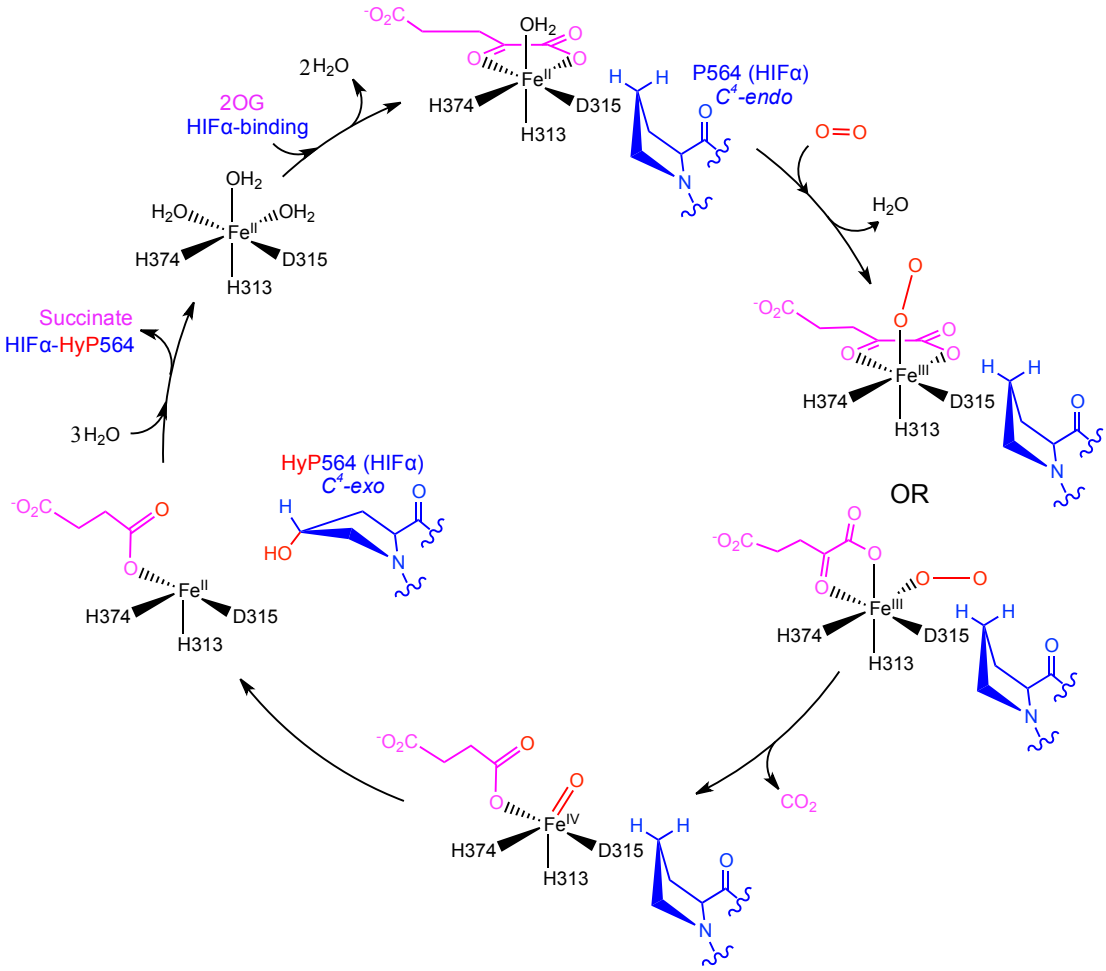


Figure A.1.4. Suggested reaction mechanism of prolyl-hydroxylation by PHDs (Chowdhury et al. 2009). Binding of oxygen can take place *trans* to either histidine ligand. Decarboxylation of 2OG and release of carbon dioxide is followed by substrate hydroxylation and release of succinate. The hydroxylated prolyl-residue rearranges from the *C*⁴-*endo* into the now more favourable *C*⁴-*exo* conformation and exits the active site. Binding of 2OG and another HIFα molecule completes the cycle.

The hydroxylated prolyl-residue P564 is embedded within the active site pocket, with the pro-(4*R*)-carbon-hydrogen bond that is cleaved during the hydroxylation of this residue pointing towards the metal ion. The pyrrolidine ring of P564 adopts the *C*⁴-*endo* conformation. It has been proposed that prolyl-4-hydroxylation by PHDs occurs with the pyrrolidine ring of P564 in a *C*⁴-*endo* conformation. Once hydroxylated, a

stereoelectronic effect in (4*R*)-hydroxy-proline favours the *C^4-exo* conformation and facilitates product release (see Figure A.1.4.) (Mecinović et al. 2009; Gorres et al. 2008). The *C^4-exo* conformation of (4*R*)-hydroxy-P564 has been observed in the structure of VCB with CODDHyP564 (Hon et al. 2002).

Several interactions between the HIFα substrate peptide and hsPHD2 have been observed in the hsPHD2 in complex with a HIFα-peptide (PDB ID=3HQR). For example, the backbone carbonyl oxygen of $A563_{HIF\text{-}1\alpha}$ forms a hydrogen bond to the backbone amine NH of $V241_{PHD2}$ positioned within the β2/β3 finger-loop. The backbone carbonyl of $P564_{HIF\text{-}1\alpha}$ forms a hydrogen bond to the side chain $A322_{PHD2}$. These interactions could help to position $P564_{HIF\text{-}1\alpha}$ productively within the active site (Chowdhury et al. 2009). Hydrophobic interactions are possible between $L559_{HIF\text{-}1\alpha}$ and $V314_{PHD2}$ and $P317_{PHD2}$. $L562_{HIF\text{-}1\alpha}$ can undergo hydrophobic interactions with $I251_{PHD2}$, $V311_{PHD2}$ and $H313_{PHD2}$.

A.1.3. The structure of hsPHD2 in context of other 2OG-oxygenases

A.1.3.1. Structural characteristics of 2OG-oxygenases

2OG-oxygenases are the largest sub-group of the non-heme Fe(II)-dependent oxygenases, a superfamily of enzymes that catalyse a wide range of oxidation reactions (Clifton et al. 2006). All 2OG-oxygenases have been proposed to oxidise their substrates by employing a highly valent ferryl species (Clifton et al. 2006).

IPNS is a non-heme Fe(II)-dependent enzyme that catalyses the formation of isopenicillin N. The crystal structure of Isopenicillin N synthase (IPNS) from *Aspergillus nidulans* in complex with Mn(II) revealed that it contains a DSBH motif (Roach et al. 1995), which forms the core of all 2OG-oxygenases studied to date (Clifton et al. 2006).

DSBH enzymes have been found in different oligomeric forms, with IPNS (Roach et al. 1995), Deacetoxycephalosporin C synthase (DAOCS) (Valegard et al. 1998) and hsPHD2 being monomeric, FIH being dimeric (Elkins et al. 2003; Hewitson et al. 2002; Lancaster et al. 2004) and hydroxypropylphosphonic acid epoxidase (HPPE) existing as a tetramer (Higgins et al. 2005; McLuskey et al. 2005).

The Fe(II) binding site of 2OG-oxygenases is formed by three residues of a conserved HXD/E…H motif. While the HXD/E part is located at the end of the second DSBH strand, the distal H is on the seventh strand of the DSBH. This Fe(II)-coordinating triad is observed in all studied non-heme Fe(II)-dependent oxygenases to date. Variation is seen in the sequential position and type of Fe(II)-coordinating residues.

Current studies and predictions suggest that there are more than 60 human 2OG-oxygenases (Loenarz & C. Schofield 2008), including the PHDs, FIH, collagen hydroxylases and histone demethylases (Mantri et al. 2010). While FIH is part of the Jumonji domain containing (JmjC) subfamily (Hewitson et al. 2002; Klose et al. 2006) the PHDs appear to belong to a discrete subfamily of 2OG-oxygenases, identified by the distinct position of their *C*-terminal helix relative to the active site (Valegard et al. 1998). The same subfamily also includes enzymes involved in the biosynthesis of the cephalosporin family of beta-lactam antibiotics such as DAOCS, the first 2OG-oxygenase for which the crystal structure was solved (Valegard et. al. 1998). Both the PHDs and DAOCS contain a conserved *N*-terminal helix-strand-helix- strand motif. In FIH, the *C*-terminal helix enables dimerisation (Elkins et al. 2003; Dann et al. 2002; Lee et al. 2003), which is not observed with the PHD-subfamily.

A.1.3.2. Mobile elements observed in the structures of 2OG-oxygenases

Some evidence of induced fit upon substrate binding has been observed with IPNS (Roach et al. 1995). Examples include the conformational change of $R279_{IPNS}$ (Roach et al. 1997) and the movement of $Q330_{IPNS}$ away from the metal center when a tripeptide substrate is bound.

Conformational changes upon substrate binding have also been observed with other 2OG-oxygenases such as FIH, where substrate binding is accompanied by conformational changes of active site residue side chains such as $W296_{FIH}$ (Clifton et al. 2006).

However, the conformational changes proposed to be induced by substrate binding for hsPHD2 involving the movement of the β2/β3 finger-loop region are significantly larger than those seen with FIH and IPNS, involving the movement of a whole flexible region with the Cα position of $S242_{hsPHD2}$ differing by ~18 Å between the structures with and without hsCODD (Chowdhury et al. 2009).

A.1.3.3. Comparison of active sites of hsPHD2 and FIH

While the conformations of the two Fe(II)-coordinating histidines of hsPHD2 and FIH are nearly identical, the one of the aspartic acid differs (McDonough et al. 2006).

The structure of FIH with Fe(II), 2OG and substrate peptide (PBD ID=1H2L) reveals that one of the carboxylate oxygens of $D201_{FIH}$ coordinates the iron, while other carboxylate oxygen forms a hydrogen bond to the backbone amide of $N803_{hsHIF\text{-}1\alpha}$ (Elkins et al. 2003). In hsPHD2, in contrast, the second carboxylate oxygen of $D315_{hsPHD2}$ coordinates the active site water that complements the octahedral coordination of the Fe(II), as previously described in Section A.3.4.1.

Differences between hsPHD2 and FIH can also be seen in the 2OG binding site. hsPHD2 binds 2OG through interacting with the 5-carboxylate of 2OG with the basic residue $R383_{hsPHD2}$ (DSBH βVIII) and $Y329_{hsPHD2}$ (DSBH βIII). FIH, in contrast, binds 2OG through $K214_{FIH}$ (DSBH βIV) and $Y145_{FIH}$ (*N*-terminal strand β6). This difference in the 2OG binding sites, among other structural differences (McDonough et al. 2006), sets hsPHD2 and FIH apart as belonging to two distinct sub-families.

A.1.3.4. Structural basis for the formation of a stable 2OG-Fe(II)-complex

The coordination of the second carboxylate oxygen of $D315_{hsPHD2}$ to the metal-bound water in hsPHD2.hsCODD could be key to the formation of an unusually long-lived Fe(II)-2OG complex, as previously observed with hsPHD2 (see Section B.1.1.) (Mecinović et al. 2009). The rationale behind this is that the water stabilises Fe(II) in the 6-coordinate state and inhibits binding of oxygen. Unlike hsPHD2, no formation of an unusually stable Fe(II)-2OG complex was detected with FIH (Chowdhury et al. 2009), in line with the observation that the active site water is not stabilised by the metal coordinating acid $D201_{FIH}$.

A.1.4. Summary

The 11 previously deposited crystal structures of hsPHD2 were introduced briefly. All of these structures contain only the *C*-terminal catalytic domain, lacking the *N*-terminal MYND-domain, the exact function of which remains unclear. The highest resolution deposited is 1.60 Å. Structural studies of enzyme-substrate interactions are, however, limited because only one structure with hsCODD (PDB ID=3HQR) is available at the relatively modest resolution of 2.0 Å, while currently no structure with hsNODD is currently available.

The structural features of the catalytic domain of hsPHD2 were outlined in A1.2. As with other 2OG-oxygenases, the PHD2 catalytic domain contains a conserved double-stranded β-helix (DSBH) fold, stabilised by three α-helices packing along the major β-sheet. The active site is located between the major and minor β-sheets and is deep within the structure in comparison to other 2OG-oxygenases. Two histidines and an aspartic acid residue coordinate the active site Fe(II). The octahedral coordination geometry of the metal ion is completed by the coordination of a water and 2OG, which binds in a bidentate fashion. There is evidence that the β2/β3 region, referred to as the 'finger-loop' region, is mobile and folds down upon the substrate to enclose it within the active site. Structural and mechanistic studies suggest that prolyl-4-hydroxylation by PHDs occurs with the pyrrolidine ring of $P564_{HIF\text{-}1\alpha}$ in the C^4-*endo* conformation.

Comparison of hsPHD2 to structures of other 2OG-oxygenases suggests that the PHDs belong to a discrete subfamily of 2OG-oxygenases, which also contains DAOCS. Although conformation changes upon substrate binding have been observed with other 2OG-oxygenases, the conformational changes upon substrate binding observed for hsPHD2 are significantly larger. Comparison between hsPHD2 and FIH shows differences in the active site aspartic acid and 2OG binding site.

A.2. *Trichoplax adhaerens* PHD: Purification, Activity Assessment and Crystallisation

In order to establish crystallisation trials, a protocol for expression and purification of a suitable form of taPHD needed to be developed. Furthermore, the purified taPHD needed to be active. Inactive protein could be misfolded, yielding misleading structural information. Inhibition and binding experiments with several previously characterised hsPHD2 and 2OG-oxygenase inhibitors aided the choice of ligands for initial crystalliation trials.

A.2.1. Sequence alignments

The predicted sequence of taPHD can be found by BLAST-searching the *Trichoplax adhaerens* genome with the sequence of any of the human PHDs 1-3. The predicted sequence as found in the database, however, contains only 206 amino acids and is missing the *N*-terminal MYND domain present in hsPHD2. This is due to incorrect prediction of intron/exon boundaries. A complete sequence, spanning 300 amino acids and containing both the *N*-terminal MYND and the *C*-terminal catalytic domain was published in reference (Loenarz et al. 2011) (see Appendix 1).

An alignment of the taPHD sequence with those of hsPHDs 1-3 suggests that the highest degree of homology is with hsPHD2, mainly due to the presence of a MYND-domain. Like the hsPHDs 1-3, taPHD contains a sequence homologous to the substrate-enclosing β2/β3 finger-loop. The taPHD finger-loop sequence is two residues shorter than in the human PHD sequences. In hsPHD2, the β2/β3 finger-loop region has been shown to be involved in hsCODD vs. hsNODD selectivity (Flashman et al. 2007). Since taHIFα does not contain a NODD (Loenarz et al. 2011), the taPHD β2/β3 finger-loop region could

represent a more ancient and simpler motif that evolved before the appearance of NODD.

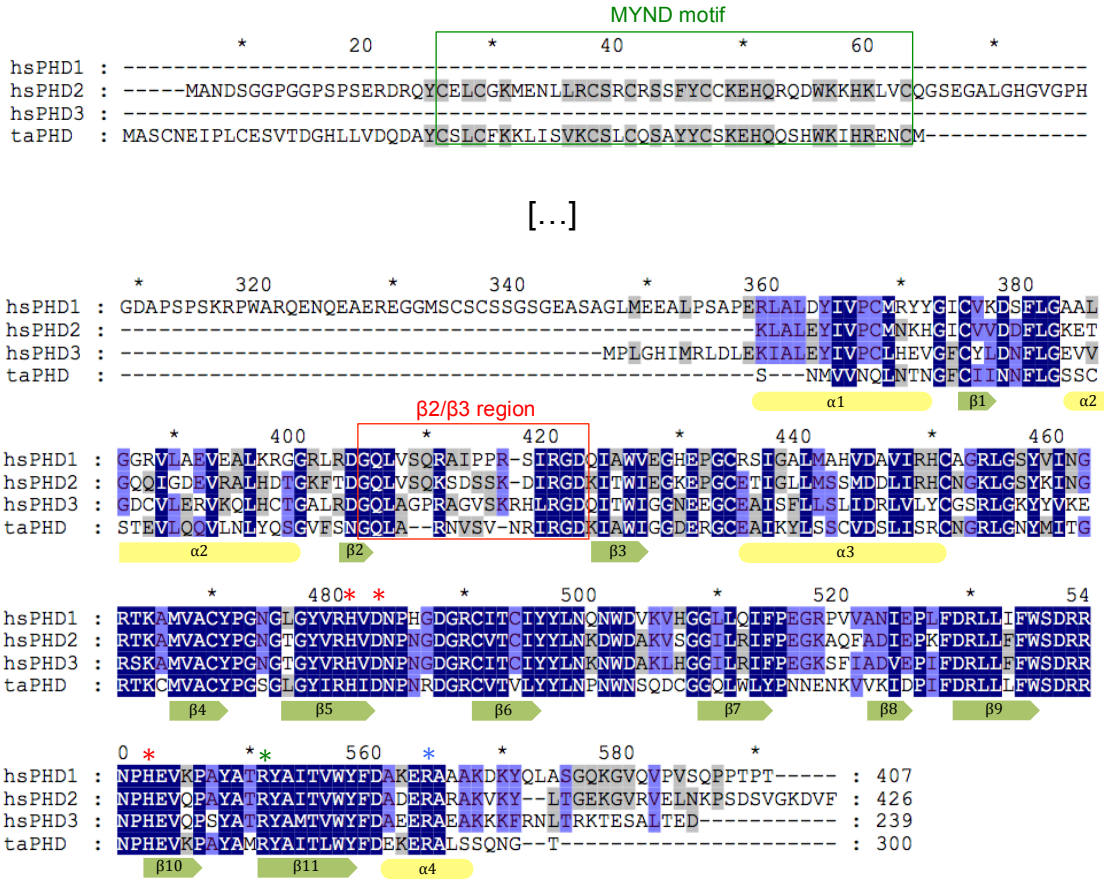


Figure A.2.1. Protein sequence alignment of the hsPHDs 1-3 and taPHD. The MYND-domain is highlighted in green and the β2/β3 finger-loop region is highlighted in red. Fe(II)-coordinating residues are marked with red asterisks, the 2OG-binding arginine is marked with a green asterisk and the substrate-binding arginine is marked with a blue asterisk. While all four sequences contain a conserved β2/β3 finger-loop region and conserved Fe(II)-, 2OG- and substrate-coordinating residues, only hsPHD2 and taPHD possess a MYND-domain. This suggests that taPHD has the highest degree of homology to hsPHD2.

The *C*-terminal catalytic domain is highly conserved (>50% sequence identity with PHDs 1-3 over the residue span corresponding to 300-390 of hsPHD2) and contains the key Fe(II)-binding residues $H209_{taPHD}$, $H270_{taPHD}$ and $D211_{taPHD}$ and the 2OG-binding $R279_{taPHD}$. A residue thought to be key for substrate binding is $R396_{hsPHD2}$ (Chowdhury et al. 2009). In taPHD, this arginine is conserved as R292.

A.2.2. Expression and purification of a suitable taPHD construct

Apart from enzymatic activity, other requirements for a construct to be used for crystallisation usually include reasonable expression levels of soluble protein. One reason for this is that it is optimal to screen for as many conditions as possible with the same purification batch as crystallisation behaviour can differ from batch to batch, for which there is much anecdotal evidence. Several different constructs of taPHD protein were prepared and compared for expression levels of soluble protein in *E. coli* (see Table A.2.1; the details regarding all primers, vector maps and protein sequences used can be found in Appendix 1, while expression and purification protocols, including buffer and column specifications, are given in Materials and Methods 1-3). Two full-length constructs including the *N*-terminal MYND domain were prepared, but expression levels for both proteins were very low. Similar results have been observed with hsPHD2. Several *N*-terminal truncations were prepared based on secondary structure predictions, sequence alignments with hsPHD2 and disorder predictions (www.strubi.ox.ac.uk/RONN). While the 67-end and 72-end (see Table A.2.1.) proteins were found to be insoluble, no expression of the 84-end protein could be detected. Only the 64-end protein, referred to as 't64' throughout this work, yielded reasonable expression levels with and without the 6xHis-tag.

Length	Vector	Tag	Expression
Full-length	pGEX-6p-1	GST	Poor
Full-length	pET-28a(+)	6xHis	Very poor
64-end	pET-28a(+)	6xHis	Medium
67-end	pET-28a(+)	6xHis	Very poor
72-end	pET-28a(+)	6xHis	Poor
84-end	pET-28a(+)	6xHis	Very poor
64-end	pET-24a(+)	None	Medium

Table A.2.1. taPHD expression constructs tested for protein expression levels (see Appendix 1 for primer sequences and vector maps). Protein expression in the soluble fraction of the constructs was tested at 37°C, 30°C, 25°C and 18°C degrees with induction using 0.5 mM IPTG (see Methods 2 for expression trial protocols). Of all constructs prepared, the 64-end construct has shown the highest protein expression levels, which was achieved at 18°C for both the 6xHis and the tag-free constructs (see Appendix 2 for SDS-PAGE gels of expression trials).

t64 taPHD was expressed without a tag in the plasmid vector pET-24a(+) and with an *N*-terminal 6xHis-tag in pET-28a(+). Purification of the resulting protein was carried out using an FPLC system (ÄKTA) with a cation-exchange column for the tag-free construct and a His-affinity column for the 6xHis-tagged construct (both GE HEALTHCARE). Both constructs were further purified by gel filtration (300 mL SP75; GE HEALTHCARE) to give ~95% purity as estimated by SDS-PAGE gels (determined by silver staining and densitometry). The protein yield with both t64 constructs was very low at 5 mg for the 6xHis-tagged construct and 3 mg for the tag-free construct from 8 L bacterial culture after overnight expression at 18°C. The low yields were linked to the fact that a heavy white precipitate was observed in the column purification fractions after the first purification step by His-affinity or cation-exchange, which was attributed to poor protein stability.

The following measures were tried out in order to reduce the protein precipitation problem, but did not lead to improvements:

- Addition of up to 10% glycerol to purification buffers
- Addition of reducing agents to purification buffers (1 mM DTT or 0.2 mM TCEP)
- Supplementing the purification buffers with 2OG
- Pre-loading the fractions with glycerol to yield a final concentration between 25% to 50%
- Reduction of expression time
- Incubation of fractions with 200 mM EDTA disodium salt
- Attempts at re-solubilising the precipitate at either pH=4 or pH=10

It was therefore decided to express the 6xHis-tagged construct covering residues 64-300 and purify the resulting protein using a His-affinity column, followed by gel filtration (see Figure A.2.2.).

Although both full-length constructs (see Table A.2.1.) appeared unsuitable for crystallisation trials and enzyme-intensive kinetic studies as carried out in Part B due to

low levels of soluble protein expression, a small amount of GST-tagged full-length taPHD was prepared for comparative purposes. The full-length construct was active and yielded higher initial rates of peptide turnover compared to t64 truncated taPHD (see Appendix 3). It was nevertheless decided to work with t64 taPHD because larger amounts of enzyme could be obtained, the 6xHis tag was expected to affect protein crystallisation less than the >20 kDa GST tag..

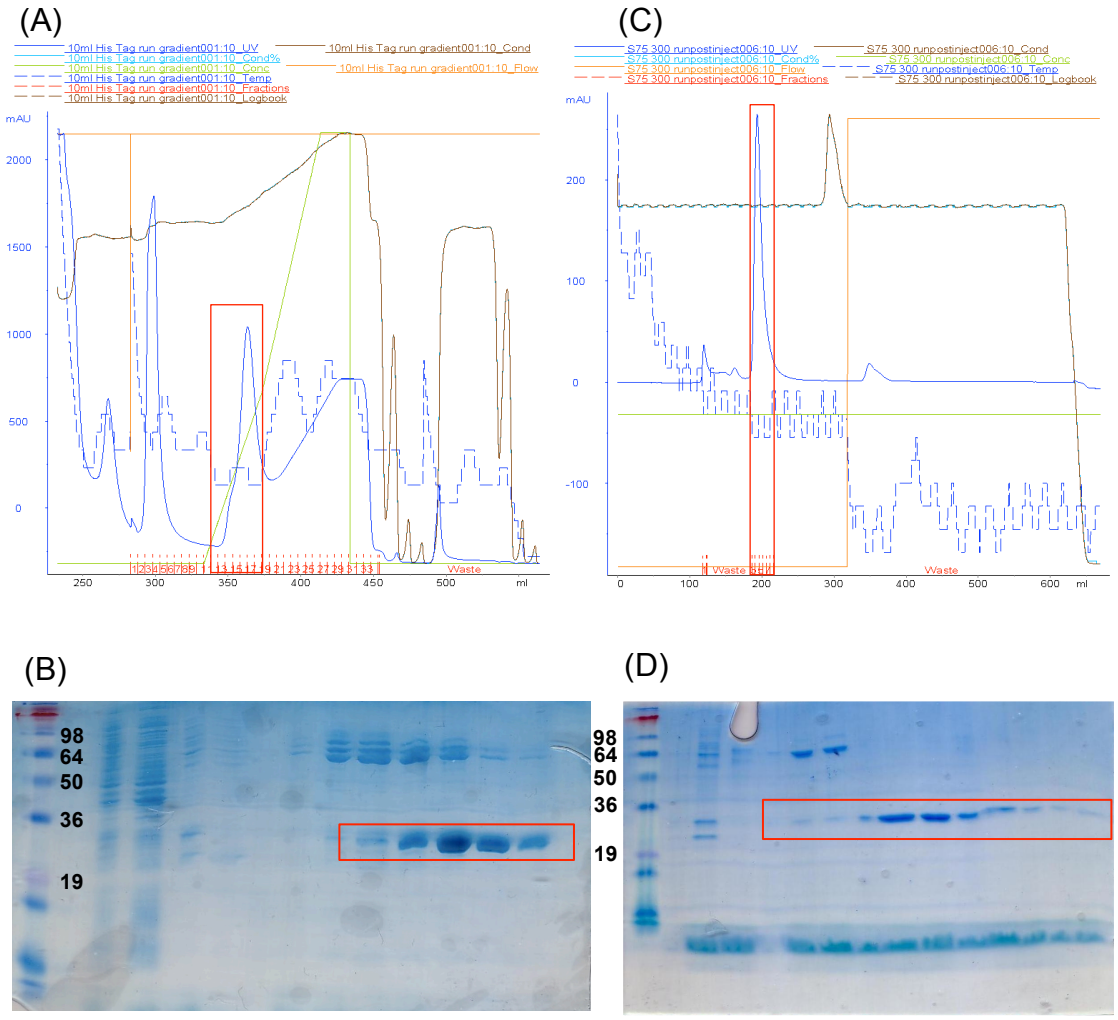


Figure A.2.2. Purification of 6xHis-tagged t64 taPHD. (A) FPLC trace of His-affinity column purification (5 mL His HP, GE HEALHCARE; see Materials and Methods 3 for purification protocols); (B) SDS-PAGE gel of fractions shown in (A); from left to right: protein ladder (masses given in kDa), empty lane, soluble fraction of cell lysate, flowthrough, fractions 9-18; (C) FPLC trace of gel filtration of a mixture of fractions 12-18; (D) SDS-PAGE gel of gel filtration shown in (B); From left to right: protein ladder, empty lane, fractions 1 and 2, empty lane, fractions 3-13. The peaks and bands corresponding to t64 taPHD are highlighted in red. The target construct can be fully separated from the larger impurity (~70 kDa) appearing in most His-purification fractions.

A.2.3. MS/MS analysis of purified taPHD

A small amount of the His-tagged t64 taPHD protein was run on an SDS-PAGE gel. The band was excised, digested in-gel with trypsin and forwarded to collaborators for characterisation by MS/MS (Benedikt Kessler group, Nuffield Department for Clinical Medicine, University of Oxford, see Contributors). The results confirmed the predicted identity of the protein, showed contaminations with *E. coli* proteins and suggested several potential taPHD post-translational modifications (see Figure A.2.5.).

The fragments detected cover 67% of Genbank database entry GI 190584431, a hypothetical *Trichoplax adhaerens* protein. The hypothetical protein sequence matches the sequencing results of the t64 taPHD construct used, but is incomplete. This is because the predicted sequence starts at residue 95 of the actual sequence (Loenarz et al. 2011). Therefore, the MS/MS results cover residues 149-300 of a construct that spans the residues 64-300 (64%).

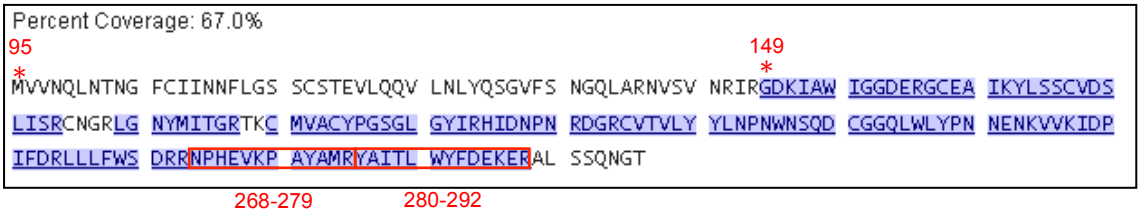


Figure A.2.3. Coverage of hypothetical taPHD sequence GI 190584431 by MS/MS fragments. Two fragments were found to be hydroxylated (framed in red). Note that the protein is too short by prediction, with the predicted sequence starting at residue 95 of the actual sequence (see main text).

Two fragments were identified to be potentially hydroxylated (see Appendix 3 for data). The first fragment spans residues 268-279, with P269 and H270 being potentially hydroxylated residues. The second fragment spans residues 280-292, with Y280 and W285 being potentially hydroxylated residues.

1) 268-279 NP(OH)H(OH)EVKPAYAMR

2a) 280-292 Y(OH)AITLWYFDEKER

2b) 280-292 YAITLW(OH)YFDEKER

Figure A.2.4. Fragments found to be hydroxylation by MS/MS.

While histidyl-, tyrosyl- and tryptophyl-oxidations are common artefacts of the experimental procedure used and could arise from the SDS-PAGE gel and the in-gel digestion (Thiede et al. 2000; Sun & Anderson 2004), prolyl-hydroxylation is less commonly observed and was more likely to represent a biologically relevant modification. To test for self-hydroxylation at position P269, a peptide spanning the residues 260-278 (LLFWSDRRNPHEVKPAYAM) was prepared (see Materials and Methods 8) and tested for modifications by incubation with 3.5 μM taPHD, 300 μM 2OG, 50 μM Fe(II) and 4 mM ascorbate and subsequent analysis using MALDI-TOF MS (see Section A.2.5.). No evidence for modifications could be detected. This suggests that the hydroxy-prolyl residue observed in MS/MS is either an artefact, or that the hydroxylation requires interactions not fulfilled by a short peptide. The fact that no hydroxylation at these sites could be observed in the subsequently obtained crystal structures of taPHD (see Section A.4.) for either of the residues showing apparent hydroxylation by MS/MS analysis, however, suggests that these modifications are artefacts.

Gene ID	Score	Mass/Da	Matches	Sequences	Comments
gi|196005047	969	23867	129 (54)	23 (14)	hypothetical protein TRIADDRAFT_50295 [Trichoplax adhaerens]
gi|15803823	555	23512	42 (25)	18 (13)	30S ribosomal protein S4 [Escherichia coli O157:H7 EDL933]
gi|227115266	1125	26041	59 (31)	20 (10)	30S ribosomal protein S3 [Pectobacterium carotovorum subsp. brasiliensis PBR1692]
gi|281602679	1180	24841	62 (33)	20 (9)	30S ribosomal protein S3 [Shigella flexneri 2002017]
gi|15831679	184	20608	11 (8)	7 (6)	translation initiation factor IF-3 [Escherichia coli O157:H7 str. Sakai]
gi|30061726	309	26812	20 (9)	10 (7)	30S ribosomal protein S2 [Shigella flexneri 2a str. 2457T]
gi|15803719	263	23434	15 (9)	7 (4)	23S rRNA methyltransferase J [Escherichia coli O157:H7 EDL933]
gi|15804574	305	24714	19 (13)	10 (6)	50S ribosomal protein L1 [Escherichia coli O157:H7 EDL933]
gi|253773841	254	24899	13 (7)	7 (5)	23S rRNA pseudouridine synthase E [Escherichia coli BL21-Gold(DE3)pLysS AG]
gi|110805376	244	25934	12 (7)	8 (5)	putative solute/DNA competence effector [Shigella flexneri 5 str. 8401]
gi|15803847	156	22230	8 (5)	6 (4)	50S ribosomal protein L3 [Escherichia coli O157:H7 EDL933]
gi|3201828	71	8852	4 (3)	2 (2)	major outer membrane lipoprotein I [Pseudomonas oleovorans]
gi|15803844	210	29956	11 (6)	8 (5)	50S ribosomal protein L2 [Escherichia coli O157:H7 EDL933]

Figure A.2.5. Database hits for MS/MS fragments of taPHD sample submitted for analysis (only main hits shown). taPHD is highlighted in green, while *E. coli* proteins are highlighted in yellow.

A.2.4. Detection of 2OG turnover by taPHD using ^{1}H-NMR

The mechanistic scheme in Figure A.1.3. suggests that the decarboxylation of 2OG is taking place prior to substrate hydroxylation, which implies that the turnover of 2OG does not have to be followed by substrate turnover. Indeed, 2OG turnover in the absence of a substrate has been observed with 2OG-oxygenases, including FIH and hsPHD2 (Flashman et al. 2010) and can be used to assess the activity of an enzyme when a substrate is not available.

The turnover of 2OG to succinate by taPHD was monitored using ^{1}H-NMR (see Figure A.2.6.). The two β-protons of 2OG are detected as a triplet centered at 2.368 ppm, while the four chemically equivalent protons of succinate can be seen as a singlet centered at 2.338 ppm (exact chemical shifts dependent on buffer pH).

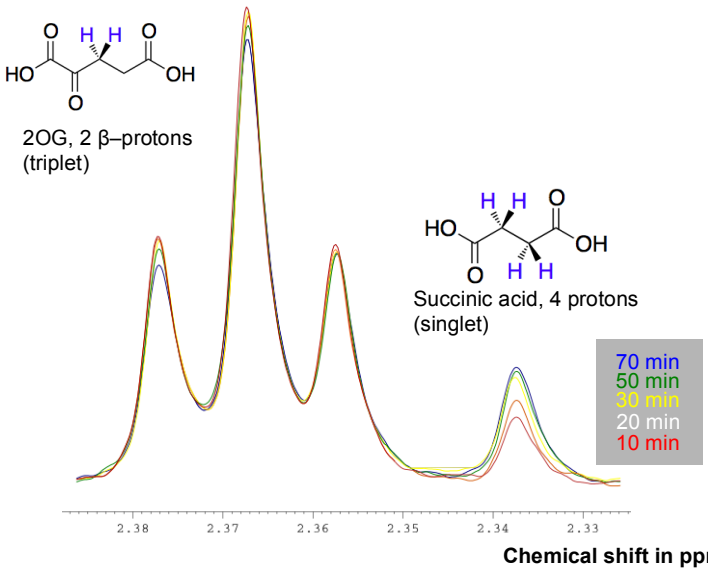


Figure A.2.6. Superimposition of ^{1}H-NMR spectra of a solution containing taPHD (3.5 µM), 2OG (300 µM), Fe(II) (50 µM) and ascorbate (4 mM), only selection region shown. The triplet is caused by the β-protons of 2OG and the singlet by succinate protons. Spectra were taken at defined time intervals from the addition of enzyme to the solution. While the signal of the β-protons of 2OG is decreasing, the succinate signal is increasing, consistent with turnover of 2OG to succinate.

Initially, it needed to be established whether the concentrations of 2OG and succinate could be determined reliably using ^{1}H-NMR. A series of solutions containing varying

amounts of 2OG and succinate were prepared. While the sum of the concentrations of 2OG and succinate was kept constant at 300 µM, the individual concentrations of 2OG and succinate varied from 5 µM to 300 µM. ^{1}H-NMR spectra of the solutions were taken and the ratios of 2OG to succinate were determined by peak integration.

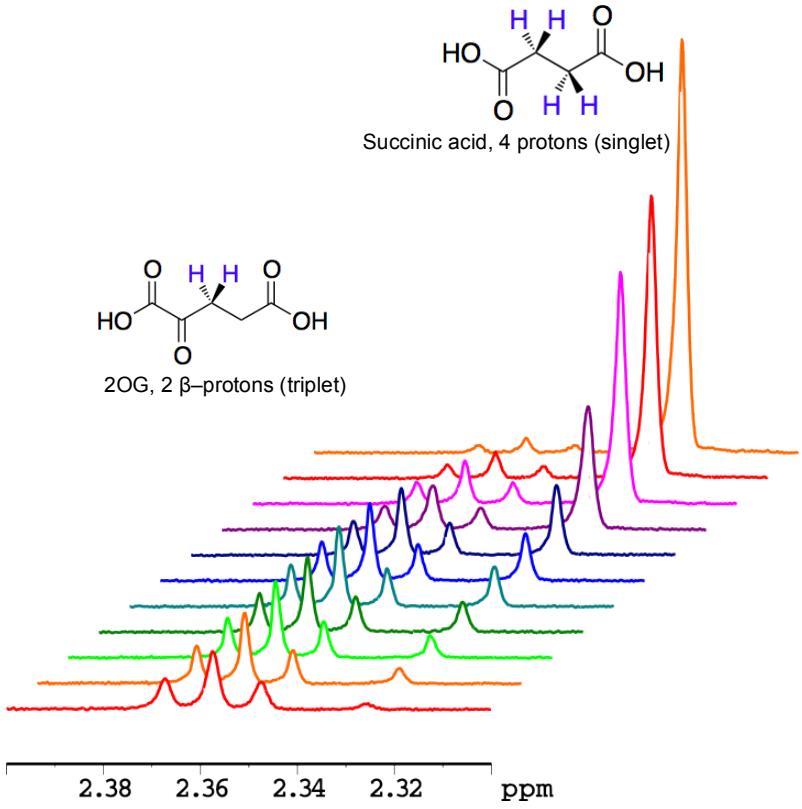


Figure A.2.7. ^{1}H-NMR spectra of solutions containing 2OG and succinate at different ratios and a constant total concentration of 300 µM in deuterated TRIS buffer and D_2O at pH=7.5. 4 mM ascorbate and 50 µM Fe(II) were added to emulate assay conditions.

The concentrations of 2OG and succinate were then calculated assuming a total concentration of 300 µM and taking into consideration that the succinate integral needs to be divided by two (4H vs. 2H for 2OG β-protons). The calculated values were compared to the theoretical values. The deviations of the observed concentrations and the theoretical concentration were less than 10 µM for all of the measurements and averaged to 3.9 µM. It was concluded that this method could be used to determine the concentrations of 2OG and succinate in solution with an average error of ±3.9 µM within

the assumption that the total concentration of both species is constant (see Appendix 3.2. for justification of assumption and discussion).

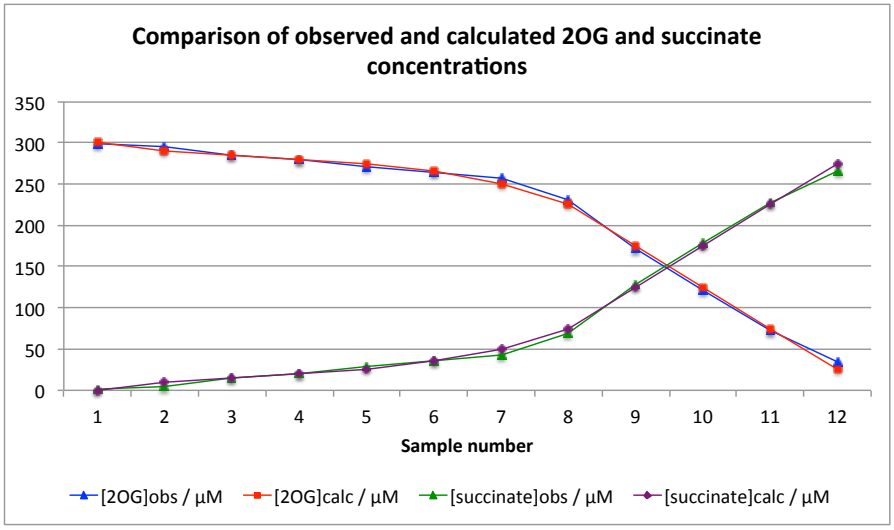


Figure A.2.8. Comparison of calculated (prepared) and observed (determined by ^{1}H-NMR) concentrations of 2OG and succinate. The concentrations of 2OG and succinate were calculated using the ratios of the peak integrals, taking into consideration the numbers of hydrogen atoms giving rise to each signal, and the assumption that [2OG]+[succinate]=300 µM (see Appendix 3.2. for justification of assumption and discussion). The error corresponds to the deviation of the observed concentration from the calculated one and is of identical magnitude but opposite sign for 2OG and succinate. The average error is ±3.9 µM.

Time-dependent 2OG turnover experiments contained a mixture of *Trichoplax adhaerens* PHD (3.5 µM t64 truncated construct), 300 µM 2OG, 4 mM ascorbate and 50 µM Fe(II) in deuterated TRIS (pH=7.5) and D_2O. ^{1}H-NMR spectra were recorded at regular time intervals for a time period of 70 minutes (see Materials and Methods 6). The signals for 2OG and succinate were integrated and the succinate integral was divided by two (4H in succinate singlet vs. 2H in 2OG triplet). The total amount of succinate and 2OG was assumed to be 300 µM at any point in time, corresponding to the initial 2OG concentration.

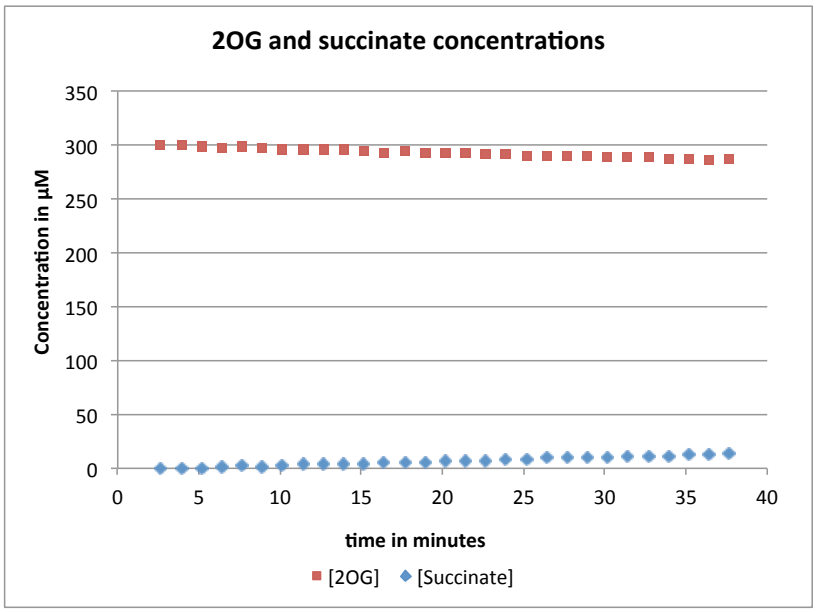


Figure A.2.9. Concentrations of 2OG and succinate as a function of time in an assay containing 3.5 µM taPHD, 300 µM 2OG, 50 µM Fe(II) and 4 mM ascorbate (see main text for calculation of concentrations). While the concentration of 2OG decreases slowly, a small build-up of succinate can be seen.

A small but clearly detectable build-up of succinate levels over the timecourse time period was observed (see Figure A.2.9.). Varying the enzyme concentration from 3.5 µM to 1.8 µM and omitting the enzyme altogether has demonstrated that the formation of succinate depends on the concentration of enzyme (Figure A.2.10.). The data was fitted with exponential functions of the type $y=(y_0-y_\infty)\times\exp(-K\times x)+y_\infty$. The y-intercept y_0 was set to be equal to zero, as the results presented in Table A.2.2. demonstrate that the detected amount of succinate at $t=0$ is negligible. The fitting function used describes the variation of concentration with time for a reactant in a reaction with first-order kinetics $[A]=[A]_0\times\exp(-k\times t)$, where $[A]_0$ is the initial concentration of A and k is the first-order rate constant. In the case of enzymatic turnover of 2OG to succinate, the real reaction order is not known and the use of a first-order fitting curve is a mere approximation. The value of K represents the rate constant within the first-order approximation and is also dependent on the enzyme concentration. This is because at an excess of substrate

over enzyme, the rate of turnover depends on the availability of free enzyme binding sites.

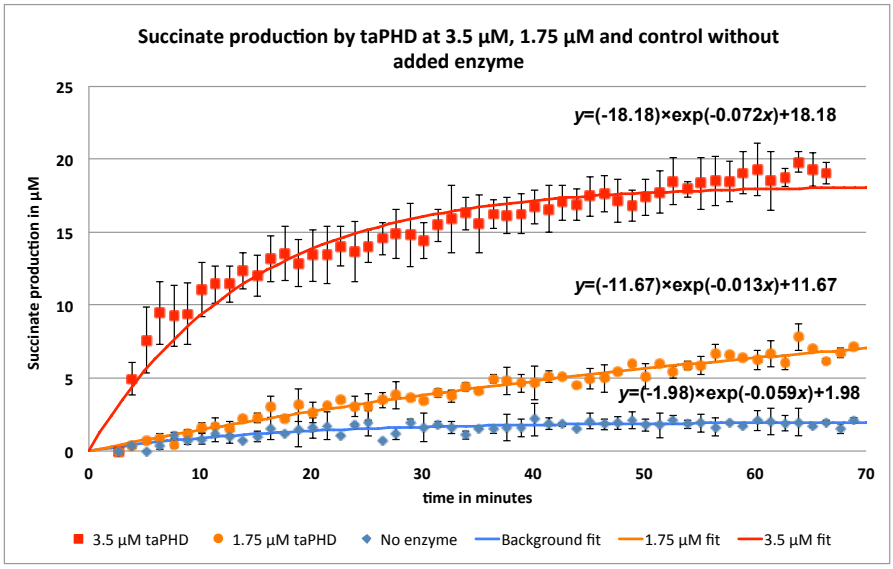


Figure A.2.10. Formation of succinate at two different taPHD concentrations and in the absence of enzyme. The concentrations were calculated from the integrals of ^{1}H-NMR signals as described in the main text and Materials and Methods 6. The data was fitted with the exponential function $y=(y_0-y_\infty)\times exp(-K\times x)+y_\infty$. Note that this first-order kinetics fit is only an approximation and does not make claims about the real reaction order.

Initial rates of succinate production were estimated by fitting a straight line through the first time points until t=5 min and the origin and determining the linear gradient.

[taPHD] / µM	r_i / µM/min
0	0.0282
1.75	0.0993
3.5	1.19

Table A.2.2. Initial rates of succinate production, determined by linear fitting of time points in the range between t=0 and t=5 min.

Comparison of the estimated initial rates r_i shows that doubling the concentration of taPHD to 3.5 µM from 1.75 µM increases the initial rate by more than ten fold. At an enzyme concentration of 3.5 µM, the initial rate is 42-fold the background rate in the absence of enzyme. These results show that the observed turnover of 2OG to succinate depends on the presence and concentration of enzyme.

The determination of uncoupled 2OG turnover by taPHD confirmed that the purified construct displayed enzymatic activity and was therefore suitable for substrate screening and crystallisation trials.

A.2.5. taPHD hydroxylates a *Trichoplax adhaerens* HIFα sequence in a 2OG- and ascorbate-dependent manner

Substrate turnover assays were carried out by adding 100 μM of a peptide covering the residues 477-497 of the ODD of *Trichoplax adhaerens* HIFα ('taODD'; see Figure A.2.11; to a solution containing 3.5 μM taPHD, 300 μM 2OG, 4 mM ascorbate and 50 μM Fe(II) in TRIS-buffer (50 mM, pH=7.5).

taHIFα 477-497 'taODD' EKEDYDDLAPFVPPPSFDNRL

Figure A.2.11. Sequence of taHIFα 477-497 peptide, referred to as 'taODD', with the evolutionarily conserved prolyl-residue P486 highlighted in red. Experimental evidence for the position and stereochemistry of P486 hydroxylation by taPHD is outlined in (Loenarz et al. 2011) and Section C.3.2.

The reaction was incubated at 25°C for 1 hour (see Materials and Methods 5 for assay protocols). The peptide mass was then determined using MALDI-TOF MS (see Methods 7) and compared to the 'no-enzyme' control, an equivalent assay without added enzyme. The presence of a new peak at +16 Da in the enzyme-treated sample indicated that the *Trichoplax adhaerens* HIFα (taODD) sequence was heavier by the mass of one oxygen atom, consistent with prolyl-hydroxylation. Amino acid analysis of the modified taODD peptide, described in Section C.3.2, confirmed that the +16 Da peak was caused by (4*R*)-prolyl-hydroxylation. Repeating the peptide hydroxylation assay in the absence of the co-factors has shown that taODD hydroxylation by taPHD does not proceed in the absence of 2OG and ascorbate, consistent with taPHD being a 2OG-oxygenase.

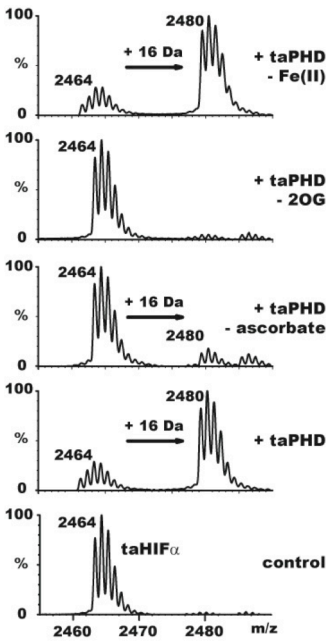


Figure A.2.12. Hydroxylation of taHIFα 477-497 ('taODD') (Loenarz et al. 2011), determined by MALDI-TOF MS after 1 hour incubation at room temperature. The mass of the taODD peptide is 2464 Da. Accordingly, the hydroxylated peptide is detected at 2480 Da (+16 Da). All samples contained 100 µM taODD, 3.5 µM taPHD, 4 mM ascorbate, 50 µM Fe(II) and 300 µM 2OG minus one reagent as stated on the right hand side of the spectra: No Fe(II) in '+taPHD - Fe(II)'; no 2OG in '+taPHD -2OG'; no ascorbate in '+taPHD - ascorbate'; all reagents in '+taPHD' and no taPHD in 'control'. The absence of the +16 Da peak in 'control', '+taPHD -ascorbate' and '+taPHD -2OG' suggests that the turnover is dependent on 2OG and ascorbate. A +16 Da peak can be seen in the '+taPHD -Fe(II)' sample. Non-denaturing mass spectra of taPHD suggest that this may be because the enzyme co-purifies with Fe(II) (see Section A.2.6.). The dependence of enzymatic activity on 2OG, ascorbate and Fe(II) is characteristic of a 2OG-oxygenase and has previously been demonstrated with hsPHD2, FIH and other 2OG oxygenases (Mahon et al. 2001; Vissers et al. 2007; Hirsilä et al. 2003; Koivunen et al. 2004; C. J. Schofield & Ratcliffe 2004).

Leaving out Fe(II) did not to have any effect on the intensity of the +16 Da peak. The most likely reason for this observation is that taPHD co-purifies with Fe(II) in the active site, similarly to hsPHD2 (Mecinović et al. 2009). This proposal is supported by non-denaturing mass spectrometry of taPHD as described in Section A.2.6.

A.2.6. Choice of inhibitors for co-crystallisation with taPHD

It is not uncommon for a protein to crystallise only in the presence of a ligand. Such a ligand is often an inhibitor that stabilises the active site by tight binding, thereby reducing the flexibility of the protein and facilitating crystallisation. Therefore, it was decided to set up crystallisation plates of taPHD with various ligands in order to increase the likelihood of obtaining protein crystals. Potential ligands were selected based on previous crystallisation experience with hsPHD2 and other 2OG-oxygenases. Inhibition assays were carried out to test for the potency of the selected inhibitors (see Figure A.2.13.), while non-denaturing mass spectrometry was used to assess the binding

of the potential ligands to taPHD (see Figure A.2.14). The taODD turnover assay results suggest that NOG, 2,4-PDCA and UN9 are relatively potent inhibitors of taPHD (see A.2.13.). This correlates with *in vitro* results obtained with hsPHD2, which yielded the IC_{50} values of 18.52 µM for NOG, 1.92 µM for 2,4-PDCA and 0.073 µM for UN9 (Stubbs et al. 2009).

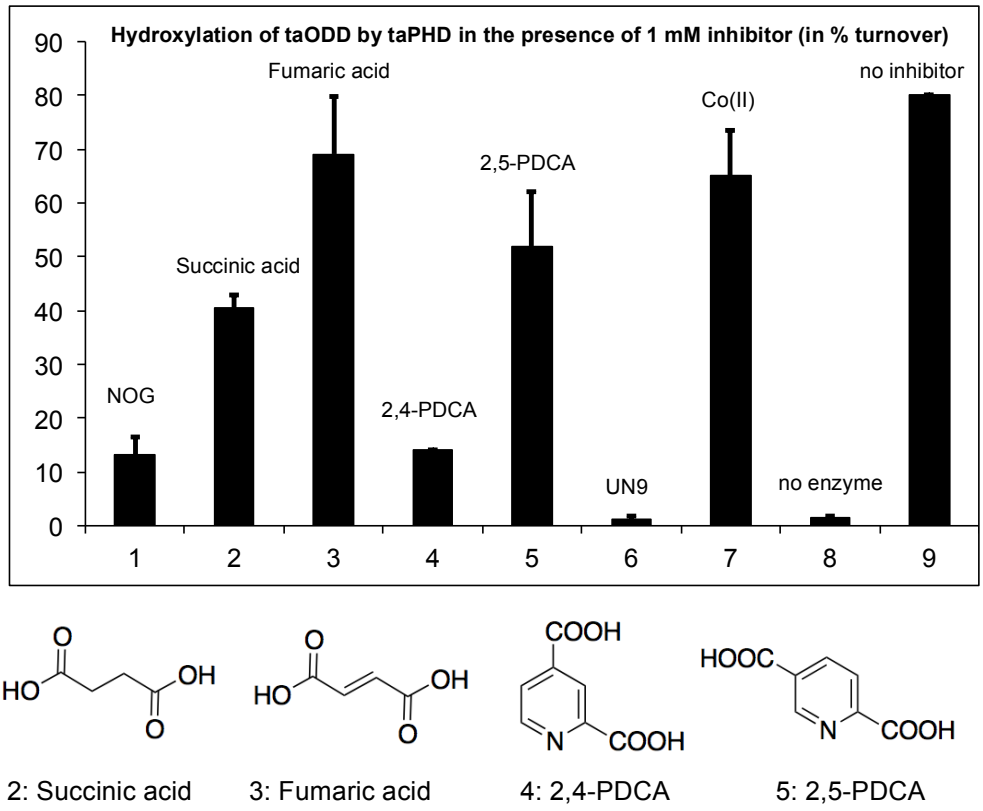


Figure A.2.13. Percentage of taODD hydroxylation by taPHD in the presence of inhibitor, as determined by MALDI-TOF MS after 20 minutes incubation at 25°C in presence of ascorbate, Fe(II) and 2OG (concentrations as in A.2.5.). Apart from Co(II), all compounds have been co-crystallised with 2OG-oxygenases previously. NOG is a 'generic' 2OG-oxygenase inhibitor and has been co-crystallised with hsPHD2 (Chowdhury et al. 2009), JMJD2A (Ng et al. 2007) and FIH1 (Yang et al. 2011). Succinic acid is the product of 2OG turnover, an intermediate in the citric acid cycle and the pre-cursor of fumaric acid. Both acids have previously been identified to be hsPHD2 inhibitors *in vitro* and *in vivo*, thereby linking the regulation of the HIF-pathway to the citric acid cycle (Koivunen et al. 2007). FIH1 has been co-crystallised with succinic and fumaric acid (Hewitson et al. 2007). 2,4-PDCA and 2,5-PDCA (pyridine-2,4/5-dicarboxylic acid) are well-characterised inhibitors of 2OG-oxygenases and have also been shown to inhibit hsPHD2 (Tschank et al. 1987; Rose et al. 2011). Crystal structures with 2,4-PDCA are available for FIH1 (Conejo-Garcia et al. 2010) and JMJD2A (PDB ID=2VD7; not published). Co(II) is a hypoxia-mimic, displacing Fe(II) from the binding site. It has been shown to inhibit both hsPHD2 and FIH1 *in vitro*. UN9 is a potent inhibitor of hsPHD2 and has yielded several structures, e.g. PDB ID 2HBT and 2HBU (not published).

The compounds tested for inhibition of taPHD were also investigated for their ability to bind to taPHD (see Figure A.2.14.). A tightly bound compound is more likely to stabilise the active site than a compound binding weakly.

Non-denaturing mass spectra taken at cone voltages of 50 V, 80 V and 120 V show that the compounds bind with the order of binding strength 2,4-PDCA > 2,5-PDCA > NOG > UN9 (see Figure A.2.14. and Materials and Methods 8 for protocol). No significant binding of succinate or fumarate could be detected at 80 V. These results demonstrate that while 2,4- and 2,5-PDCA are not as potent inhibitors as UN9, they can bind to taPHD. It was therefore decided to use all three of the above as well as NOG in initial crystallisation attempts.

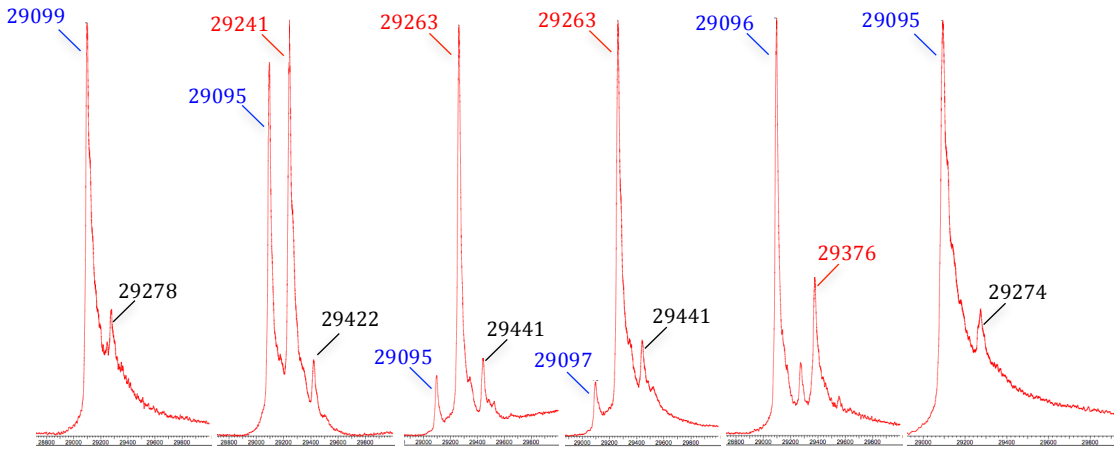


Figure A.2.14. Non-denaturing mass spectra of taPHD with inhibitors and 2OG taken at 80 V cone voltage (see Appendix 3 for spectra taken at other cone voltages). The enzyme peak can be seen at 29095-29099 Da and is labelled in blue with the observed mass. The calculated mass of taPHD is 29042 Da. As no metal was added to the enzyme in this experiment, the mass difference of 57-53 Da suggests that taPHD co-purifies with Fe(II), which has an average mass of 55.85 Da. This hypothesis is supported by the findings in Section A.2.5. and similar observations with hsPHD2 (Mecinović et al. 2009). The red peak labels correspond to the masses of the enzyme-inhibitor complexes: taPHD.NOG at 29241 Da (+146 Da); taPHD.2,4-PDCA at 29263 Da (+168 Da); taPHD.2,5-PDCA (+168 Da); taPHD.UN9 at 29376 Da (+280 Da). The peaks at +180-182 Da, labelled in black, are thought to arise from a sugar molecule bound to taPHD (e.g. MW of glucose = 180.16 Da). These spectra show that although 2,4- and 2,5-PDCA were found to be weak inhibitors of taPHD, they bind more strongly to the enzyme (87% and 89% bound at 80 V) than UN9 (28% bound at 80 V). It is possible that they bind at an enzyme site other than the active site. NOG was also found to bind relatively tightly (53% bound at 80 V).

A.2.7. Choice of initial conditions for crystallization trials

Based on the results described in the previous chapters, a taPHD construct and a range of ligands were chosen for initial crystallisation trials. The construct was truncated 6xHis-tagged taPHD (residues 64-301), purified using a His-affinity column followed by gel filtration (see Materials and Methods 3 for protocols). As the protein yield for this construct was relatively low at around 5 mg from 8 L medium, it was decided not to cleave the 6xHis-tag in order to minimise protein losses throughout the purification procedure. Following gel filtration, the protein was concentrated to 20 mg/mL. Observations from previous purifications had suggested that the protein precipitated at higher concentrations.

The commercial screen JCSG-Plus (MOLECULAR DIMENSIONS) was chosen for initial screens because it covers a wide range of different conditions, including precipitants such as polyethylene glycol, inorganic salts, polyalcohols, organic acids and a pH-range covering pH=4 for pH=10. Ten sitting-drop plates were set up using an automated dispensing robot (PHOENIX). Six plates contained one of the four ligands selected above at 1 mM each and two did not contain an added ligand. Mn(II) was added to all screens to a final concentration of 2 mM as a stable divalent metal ion to substitute for the active site Fe(II), which is subject to oxidation in aerobic conditions. Four plates were incubated at 4°C and four at 20°C (see Table A.2.3.).

A.2.8. Evaluation of initial hits and optimisation

After five days, small crystals were observed in two plates (see Table A.2.3.). All hits contained 20-25% w/v polyethylene glycol 3350.

Plate number	Enzyme concentration	Ligand	Metal	Screen	T/°C	Hits	Comment
taPHD.1	20 mg/mL	None	Mn(II)	JCSG-Plus	20	None	
taPHD.2		NOG				None	
taPHD.3		UN9				H.8.2; H.11.1	Clusters of plates
taPHD.4		2,4-PDCA				None	
taPHD.5		2,5-PDCA				None	
taPHD.6		None			4	None	
taPHD.7		NOG				None	
taPHD.8		UN9				A.8.1 /2; A.9.1; H.10.1	Stacks of needles
taPHD.9		2,4-PDCA				None	
taPHD.10		2,5-PDCA				None	

Table A.2.3. Sitting-drop plates set up with taPHD with description of initial hits. The commercial screen JCSG-Plus (HAMPTON RESEARCH) was used for all plates. Mn(II) was added to the taPHD solution. From experience with other 2OG-oxygenases, Mn(II) was found to stabilise taPHD. The temperature of the plates changes the equilibration rates in the drops and can affect the growth rate and consequently the shape of crystals. After five days, hits were recorded in five wells. While none of the hits were harvestable due to their small size (~50 µm at largest dimension), the initial hits yielded a starting point for optimised crystallisation attempts.

JCSG-Plus at 20°C with UN9

H.8.2 (0.2 M sodium chloride, 25% w/v polyethylene glycol 3350, 0.1 M bis-tris pH=5.5)
H.11.1 (0.2 M magnesium chloride, 25% w/v polyethylene glycol 3350, 0.1 M bis-tris pH=5.5)

JCSG-Plus at 4°C with UN9

A.8.1 (0.2 M ammonium formate, 20% w/v polyethylene glycol 3350)
A.8.2 (0.2 M ammonium formate, 20% w/v polyethylene glycol 3350)
A.9.1 (0.2 M ammonium chloride, 20% w/v polyethylene glycol 3350)
H.10.1 (0.2 M ammonium acetate, 25% w/v polyethylene glycol 3350, 0.1 M bis-tris pH=5.5)

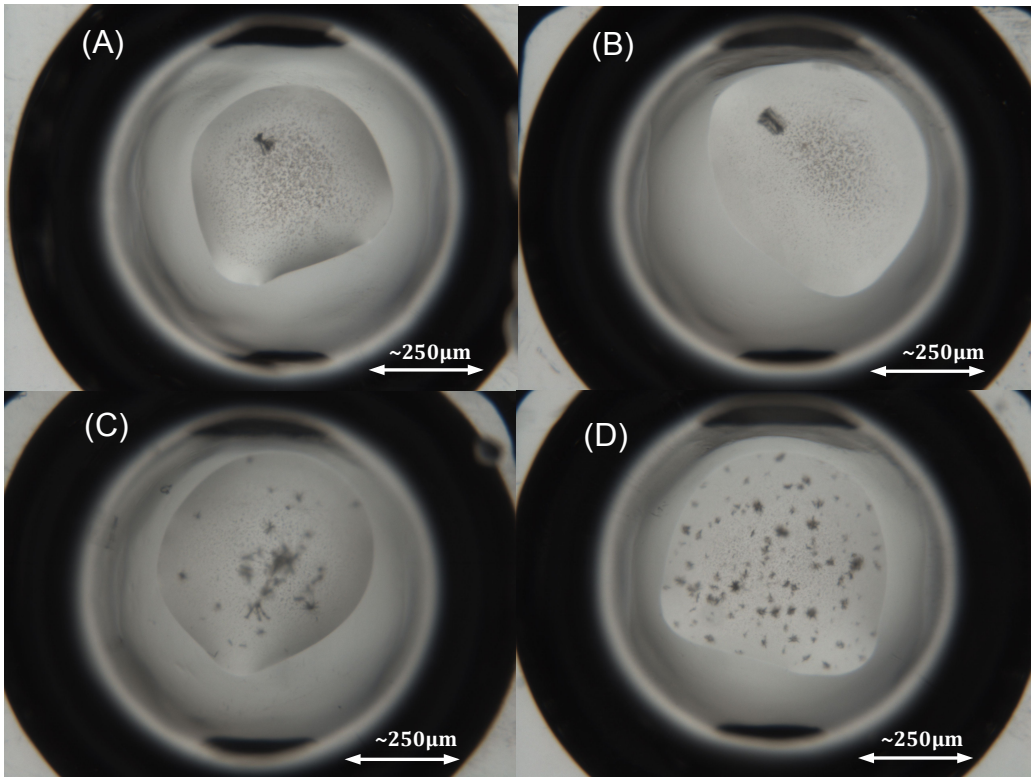


Figure A.2.15. Initial hits with taPHD. Clusters of flat plates observed in H.8. (A) and in H.11. at 20°C (B). Numerous small overnucleated clusters in A.8. (C) and H.10. at 4°C (D).

Initial X-ray diffraction data collected in-house (RIGAKU; see Materials and Methods 11) has shown that the clusters of plates as found in plate taPHD.3 were protein crystals rather than inorganic salt. Diffraction at a resolution better than 2 Å was observed in-house. It was therefore decided to work towards improving the size of the initial crystals using the initial hits as seed crystals (see Materials and Methods 11 for protocol). A hanging drop set-up was used and a crystal harvested from well H.11. of plate taPHD.3 was ground up using a commercial seeding kit and used for seeding (seeding kit and plates from HAMPTON RESEARCH). Well solutions were prepared to match the screen well solutions. All hit conditions were attempted both at 4°C and at 20°C. While the conditions H.8. and A.8. did not produce any crystals, conditions A.9. and H.11. produced small, overnucleated crystals at 4°C. Condition H.10. produced a large crystal cluster 12 hours after seeding (see Figure A.2.16.). These large clusters of plates were found to be easily reproducible when using the purification batch that yielded the initial hits.

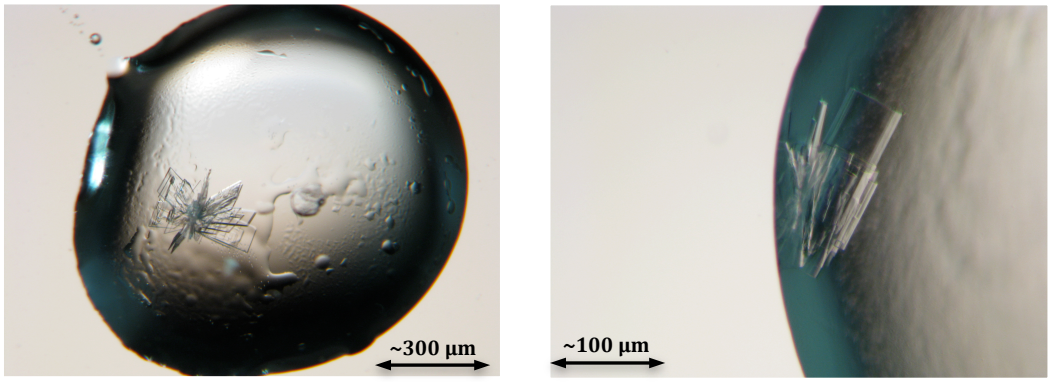


Figure A.2.16. Example of crystals of taPHD t64 with UN9 and Mn(II), obtained through streak seeding at conditions equivalent to JCSG-Plus well H.10. The plates were carefully separated from each other and frozen in liquid nitrogen immersed in 75% well solution and 25% glycerol. The glycerol acts as a cryoprotectant and reduces the formation of ice crystals in the vicinity of the protein crystal.

A.2.9. Data collection and structure solution

Crystals of taPHD produced through streak seeding in a well solution containing 0.2 M ammonium acetate, 25% w/v PEG 3350 and 0.1 M bis-tris pH 5.5 (well H.10. of commercial screen JCSG-Plus from HAMPTON RESEARCH; see Section A.2.8.) were

harvested by carefully separating the small plates from the main cluster, following by freezing in liquid nitrogen in 75% well solution and 25% glycerol. X-ray diffraction data at a resolution of 1.2 Å was collected using a synchrotron source at a wavelength of 0.97950 Å (Diamond beam line I04). The data were processed with HKL-2000 (see Figure A.2.17. for example of image file).

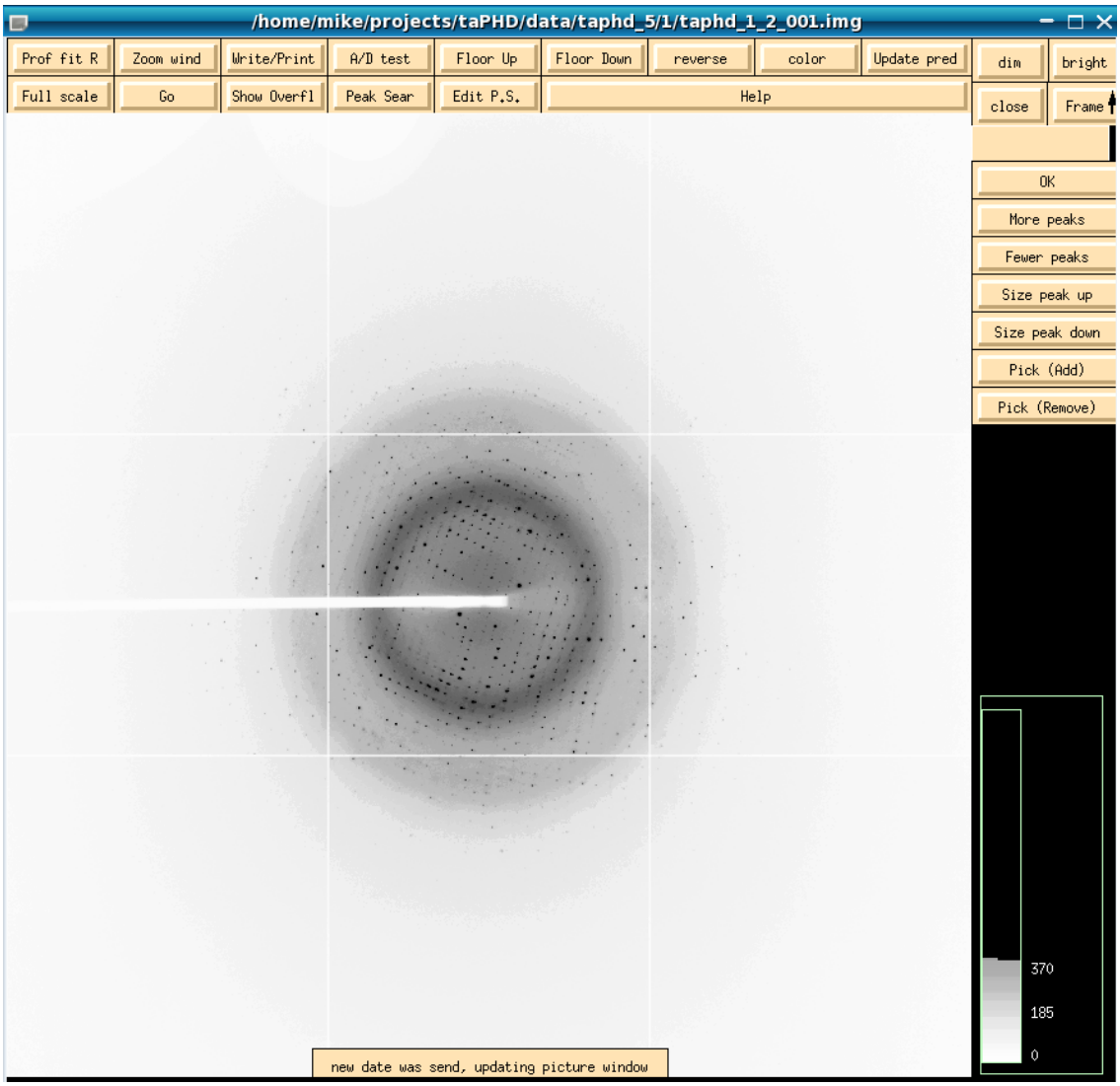


Figure A.2.17. Example diffraction image, viewed in HKL-2000 (http://www.hkl-xray.com/).

Unique reflections		74209
Space group		P 1 21 1
Molecules/ASU		1
Unit cell parameters		
	a	40.579 Å
	b	59.413 Å
	c	51.202 Å
	α	90 °
	β	100.742 °
	γ	90 °
Resolution range		39.86–1.20 Å
Completeness in resolution range		0.99984
Mean I/σ(I)		3.06 (1.39)
Wilson B value		13.03 Å^2

Table A.2.4. Data collected at a synchrotron source (DIAMOND I04) with taPHD crystals (see Section A.2.8.). One molecule per asymmetric unit was detected. The value in brackets for mean I/σ(I) represents the highest resolution shell. See Section A.4 for refinement and structure discussion.

The structure was solved by Dr Michael McDonough by molecular replacement using the hsPHD2 structure 2HBU as the search model using PHASER (McCoy et al. 2007). Refinement was carried out using COOT (Emsley & Cowtan 2004) and PHENIX (Adams et al. 2010) and is described in Section A.4. The final values obtained for R_{free} and R_w were 0.1537 and 0.1374, respectively (see Section A.4.1. for a brief introduction for crystallographic refinement).

A.2.10. Summary

Alignments of the sequence of taPHD with hsPHDs 1-3 show that taPHD contains a MYND domain and therefore most closely resembles hsPHD2. The catalytic domain is highly conserved between taPHD and hsPHD2 and contains the key Fe(II)- and 2OG-binding residues. A sequence homologous to the mobile β2/β3 finger-loop of hsPHD2 is present in taPHD.

The catalytic domain of taPHD was recombinantly expressed in *E. coli* and purified using a His-affinity column followed by gel filtration. The identity of the protein was confirmed by MS/MS analysis. Activity of the purified construct was assessed by

monitoring 2OG turnover by taPHD in the absence of substrate using ^{1}H-NMR. The 2OG turnover to succinate was shown to be caused enzymatically. A 21-amino-acid peptide covering the ODD sequence of *Trichoplax adhaerens* HIFα (taODD) was shown to be hydroxylated by taPHD to give the (4*R*)-stereochemistry.

Several characterised 2OG-oxygenase inhibitors were tested for inhibition of taODD hydroxylation and binding to taPHD using non-denaturing mass spectrometry. *N*-oxalylglycine, a bicyclic inhibitor with the PDB ID UN9, and pyridine-2,4/5-dicarboxylic acids were chosen as ligands for crystallisation trials. Mn(II) was used as a surrogate for Fe(II). Small crystals were obtained using the commercial screen JCSG-Plus and optimised by streak seeding. X-ray diffraction data at a resolution of 1.2 Å were collected using a synchrotron source. The structure was solved by molecular replacement.

A.3. Crystal Structure of taPHD with taHIFα (477-497)

Obtaining a crystal structure of taPHD in complex with its substrate taHIFα was of great interest because it would allow for the comparison of the enzyme-substrate interactions of taPHD to observations made with the crystal structure of hsPHD2 with CODD.

A.3.1. Binding of taODD, hsCODD and hsNODD peptides to taPHD

The binding of taODD (taHIFα 477-497), hsCODD (hsHIF-1α 556-574) and hsNODD (hsHIF-1α 394-412) to taPHD was investigated using non-denaturing mass spectrometry.

taHIFα 477-497 'taODD'	EKEDYDDLA**P**FVPPPSFDNRL
hsHIF-1α 556-574 'hsCODD'	DLDLEMLA**P**YIPMDDDFQL
hsHIF-1α 394-412 'hsNODD'	PDALTLLA**P**AAGDTIISLD

Figure A.3.1. Amino acid sequences of the 19mer peptides 'hsCODD' and 'hsNODD' and the 21mer 'taODD'. hsCODD and hsNODD were obtained from a commercial supplier (GSL CHINA). taODD was synthesised in-house using an automated peptide synthesiser (CS BIO) and purified by HPLC (see Section 12 for methods). The prolyl-residues hydroxylated by hsPHD2 and taPHD, P486 in taHIFα, and P564/P402 in hsHIF-1α, are highlighted in red.

Because taODD and hsCODD are good substrates of taPHD, displaying 70-80% turnover within one hour (Loenarz et al. 2011) (see Section B.2.2.1.), both peptides were also potential candidates for co-crystallisation trials. Turnover of hsNODD, in contrast, is very slow by taPHD, with hydroxylation levels below 20% after an overnight incubation (see Section B.5.3.), consistent with the findings that taHIFα does not contain a NODD (Loenarz et al. 2011).

taODD was chosen as the substrate for co-crystallisation because it exhibited stronger binding than hsCODD (see Figure A.3.2.) and because it was of greater interest, being the

natural substrate of taPHD. Furthermore, taODD was found to be much more soluble in aqueous solution than hsCODD and hsNODD (data not shown). Therefore, a final peptide concentration of 10 mM taODD could be used in the crystallisation solution, whereas the solubility of the hsCODD 19mer limits the final concentration to several mM only. This is important because a large excess of substrate often increases the chance of co-crystallisation.

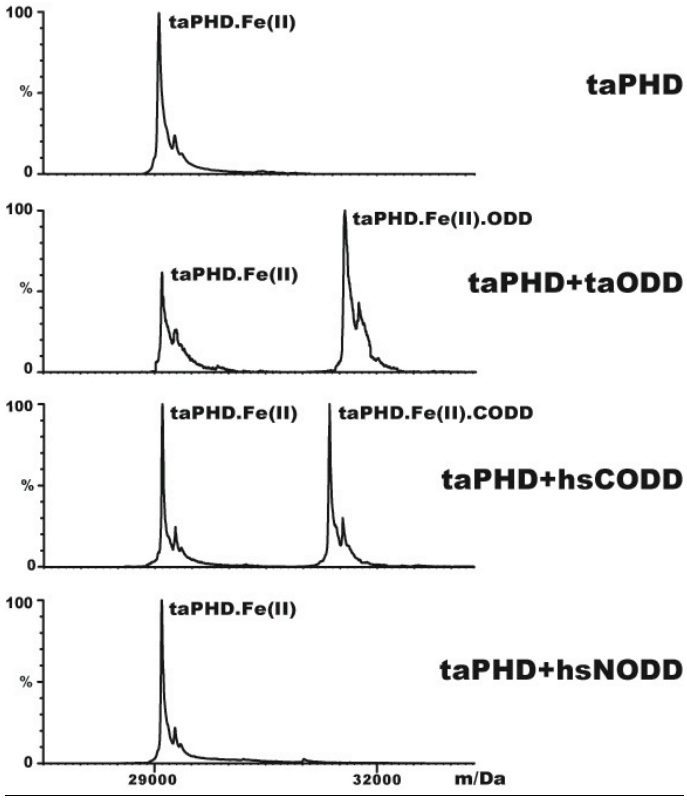


Figure A.3.2. Non-denaturing mass spectrometry data, illustrating the binding of taODD, hsCODD and hsNODD to taPHD at a cone voltage of 120 V. No binding of hsNODD was observed. taODD and hsCODD were found to be bound to 62% and 51% of the enzyme, respectively. The strength of binding of these peptides to taPHD is reflected in their hydroxylation rates as described in Section B.3.3, with initial rates of 6.5 µM/min calculated for taODD and 2.8 µM/min for hsCODD, whereas hsNODD is turned over only very slowly with low hydroxylation levels of < 20% detectable after an overnight incubation only.

A.3.2. Co-crystallisation of taPHD with taODD

The crystallisation condition that yielded the crystals used for collection of the 1.2 Å data in Section A.2.9. was based on the conditions of well H.10. of the commercial screen JCGS-Plus (HAMPTON RESEARCH).

JCSG-Plus H.10.
H.10. (0.2 M ammonium acetate, 25% w/v polyethylene glycol 3350, 0.1 M bis-tris pH=5.5)

Although it is not possible to predict the crystallisation behaviour of a protein upon addition of an extra component, such as a substrate peptide, experience suggests that conditions that have previously yielded crystals are sometimes a good starting point. Therefore, an optimization screen based on the JCSG-Plus screen well H.10. with varying concentrations of ammonium acetate and polyethylene glycol was designed.

JCSG-Plus H.10. optimisation screen
0.08-0.4 M ammonium acetate, 15-25% w/v polyethylene glycol 3350, 0.1 M bis-tris pH=5.5

An initial test of the optimisation screen with taPHD, Mn(II) and UN9 yielded 15 hits across the plate at 20°C and 4 considerably smaller hits at 4°C. All of the hits were of the same morphology as the initial hits observed with this condition, consisting of small clusters of needles. These observations confirmed that the optimisation screen based on H.10. could be used for generating taPHD crystals in the presence of UN9 in a reproducible manner. This was an important step because it was desirable to avoid broad screening with the substrate peptide taODD in order to preserve both enzyme and peptide. The same taPHD construct as used in Section A.3.1, truncated at the *N*-terminus until residue 64 and carrying an *N*-terminal 6xHis tag, was used for co-crystallisation with taODD. Mn(II) was added to a final concentration of 1 mM to stabilise the metal binding site. Rather than UN9, NOG was used at 2 mM as a 2OG analogue ligand. The reason for using NOG was that being rather bulky compared to 2OG, UN9 could prevent correct positioning of $P486_{taHIF\alpha}$ in the active site of taPHD. The protein-metal-ligand mix

was incubated at room temperature for 10 minutes. A stock solution of taODD in 50 mM TRIS at pH=7.5 was added to yield a final peptide concentration of 10 mM. The resulting mixture was incubated for further 10 minutes prior to addition to the crystallisation plates to allow for equilibration. Two plates with identical components, taPHD.11. and taPHD.12. were set up. The plate taPHD.11. was incubated at 20°C and the plate taPHD.12. was incubated at 4°C.

Plate	[Enzyme]	Ligand	Metal	[taODD]	Screen	T/°C	Hits	Comment
taPHD.11	20 mg/mL	NOG	Mn(II)	10 mM	H.11/JCSG-Plus optimisation	20	E.10; E.11; G.10; G.11; F.9; F.10; F.11	Rectangular rods
taPHD.12						4	D.11; E.10; G.10; G.11; F.10; F.11	Stacks of needles

Table A.3.1. Two plates were set up for co-crystallisation of taPHD with taODD, a 19mer substrate peptide. The protein and Mn(II) were used at the same concentrations that yielded the 1.2 Å structure described in Section A.2. The stable 2OG analogue NOG was used at a concentration of 2 mM. The plates were set up with an optimisation screen based on the conditions H.11. of the commercial screen JCSG-Plus (HAMTON RESEARCH). Several hits were recorded in both plates. While at 4°C, only small stacks of needles were observed, the 20°C plate contained numerous rectangular rods of a typical size of 120 µm × 30 µm, sometimes also surrounded by needle stacks as illustrated in Figure A.3.3.

The plate taPHD.11. generated several hits after two weeks, including crystals in the shape of rectangular rods of a typical size of 120 µm × 30 µm (see Figure A.3.3.).

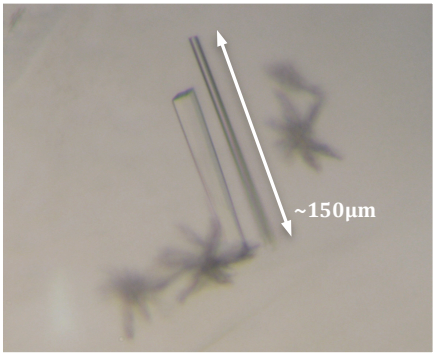


Figure A.3.3. taPHD co-crystallised with taODD in well F.11. of plate taPHD.11.

A.3.3. Data collection and structure solution (Diamond)

Crystals from the wells F.11. (0.31 M ammonium acetate, 24% w/v polyethylene glycol 3350, 0.1 M bis-tris pH=5.5) and E.10. (0.26 M ammonium acetate, 23% w/v polyethylene glycol 3350, 0.1 M bis-tris pH=5.5) were harvested and frozen in well solution and 25% glycerol. X-ray diffraction data at a resolution of 1.3 Å was collected using a synchrotron source at a wavelength of 0.97950 Å (Diamond beam line I02). HKL-2000 (Otwinowski & Minor 1997) was used for data processing. One molecule per asymmetric unit in the space group P1 (triclinic) was detected. The structure was solved by molecular replacement using the hsPHD2 structure with the PDB ID 3HQR as the search model (PHASER). Refinement was carried out using COOT and PHENIX and is described in Section A.4.1. The final values obtained for R_{free} and R_w were 0.1705 and 0.1447, respectively.

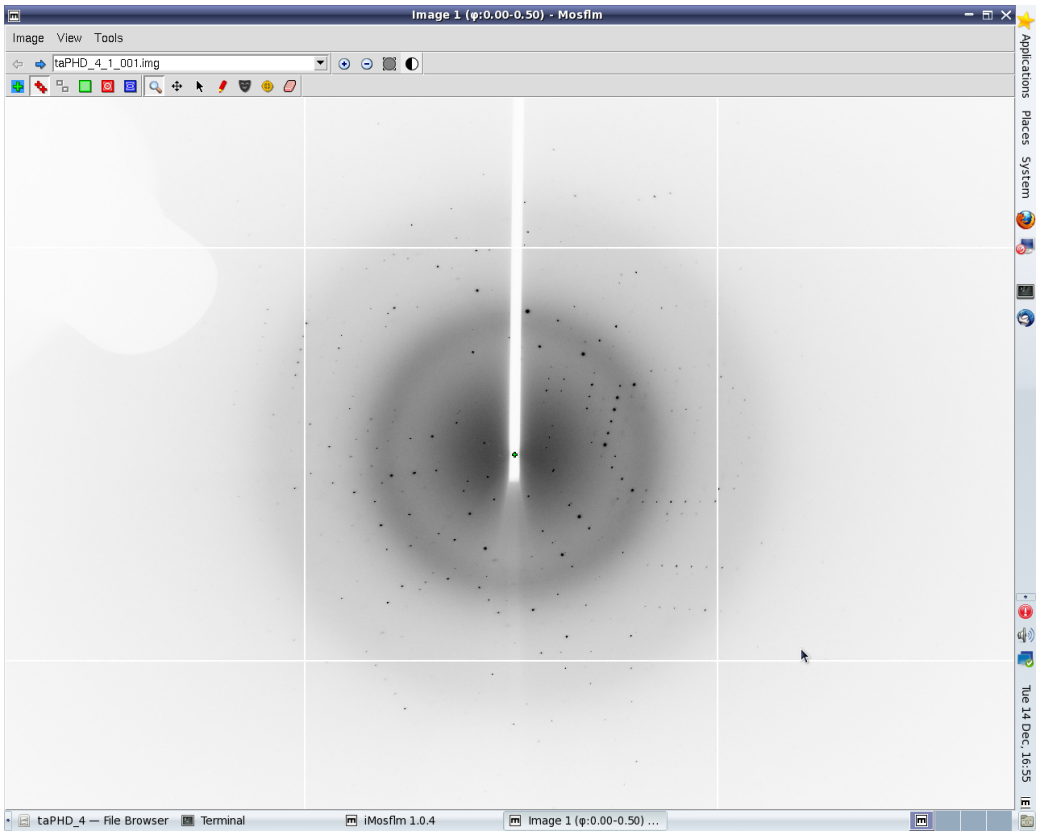


Figure A.3.4. Example diffraction image collected for taPHD in complex with the taODD substrate, viewed in MOSFILM (www.mrc-lmb.cam.ac.uk/harry/mosflm/).

Unique reflections		55428
Space group		P 1
Molecules/ASU		1
Unit cell parameters		
	a	41.321 Å
	b	40.839 Å
	c	42.076 Å
	α	95.665 °
	β	114.064 °
	γ	103.69 °
Resolution range		37.48–1.30 Å
Completeness in resolution range		0.948
Mean I/σ(I)		3.95 (1.37)
Wilson B value		14.57 Å^2

Table A.3.2. Data collected at a synchrotron source (DIAMOND I02) with taPHD crystals. One molecule per asymmetric unit was detected.

A.3.4. Summary

The binding of taODD, hsCODD and hsNODD to taPHD was investigated using non-denaturing mass spectrometry. taODD was chosen for co-crystallisation trials because it exhibited strongest binding of the three peptides and because, being the native substrate, it was of greater interest.

taPHD.taODD crystals were obtained in optimisation plates based on the condition JCSG-Plus H.11. X-ray diffraction data at a resolution of 1.3 Å was collected using a synchrotron source. The structure was solved by molecular replacement using the hsPHD2.CODD structure (PDB ID=3HQR) as the search model.

A.4. Refinement and Comparison of the taPHD Crystal Structures to the Crystal Structures of hsPHD2

A.4.1. A brief introduction to crystallographic structure refinement

Once the initial phases have been obtained from molecular replacement or other solution techniques, an initial model of the protein can be built. The atomic positions in this initial model can be refined to fit the observed data better and yield a better set of phases, which can then in return be used to calculate an improved electron density map. A new model can then be fitted to this improved map, and so forth. Ideally, the refinement is continued until the agreement between the model and the diffraction data cannot be improved further. The agreement is measured by the *R*-factor, the ratio of the difference of the observed and calculated structure factors of each reflection and the sum of the measured structure factors of all reflections:

$$R = \frac{\sum_{\text{all reflections}} |F_o - F_c|}{\sum_{\text{all reflections}} |F_o|} \qquad (1)$$

The *R*-factor should be minimised throughout a successful refinement procedure. A commonly used but different quality criterion is R_{free}. It is calculated as in Equation 1 above but is using only a subset of reflections, called a 'test set', which is not included in the structure refinement and usually comprises of 5-10% of all reflections. While *R* describes how a particular model predicts the entire data set produced by this model, R_{free} describes how well the current model predicts the 'test set' not used in refinement, thereby avoiding the circularity apparent in the calculation of *R*. Generally, both *R*-values are monitored in refinement and the *R* factor is called R_{w} ('w' for working set of reflections) to avoid confusion. For R_{free}, a rule of thumb is that the optimised factor can

be as low as the numerical value of the resolution in ångström divided by 10. R_{free} is commonly expressed in percent. Such low values, however, are rarely achieved in practice because of ambiguities in the model building process. Refinement is usually carried out by manually adjusting the model and can be time consuming for a high-resolution structure due to the large amount of detail requiring refinement.

A.4.2. Crystallographic refinement of the crystal structures of taPHD

Refinement of the crystal structures of taPHD and taPHD.taODD was carried out using PHENIX with modifications by COOT between rounds of refinement. The general strategy outlined in Table A.4.1. was followed for both structures.

Rounds	Refined parameters
1	Rigid body refinement of initial model
2-4	Isotropic refinement of atomic *xyz* coordinates and displacement parameters; sidechain rotamers adjusted
5-7	Addition of riding hydrogens; clashing sidechains adjusted
8-12	Solvent picking and refinement
13-20	Building and refinement of termini and gaps
21	Introduction of anisotropic refinement parameters for non-hydrogen atoms
22-25	Modelling of alternative sidechain conformations
26-30	Further building of gaps and termini
31	Refinement of weights

Table A.4.1. Refinement strategy for the crystal structures taPHD and taPHD.taODD.

Unless otherwise stated, the default parameters of PHENIX were used. The initial model built by PHASER was refined as a rigid body. Metal ion coordination and ligand restraints were automatically generated in PHENIX and used in all refinement rounds after manual inspection. In all subsequent rounds, the *xyz* coordinates, individual B-factors of all non-hydrogen atoms and the bulk solvent model were refined. Incorrect molecular geometries and clashes were manually edited using a molecular graphics workstation of the COOT software between each round.

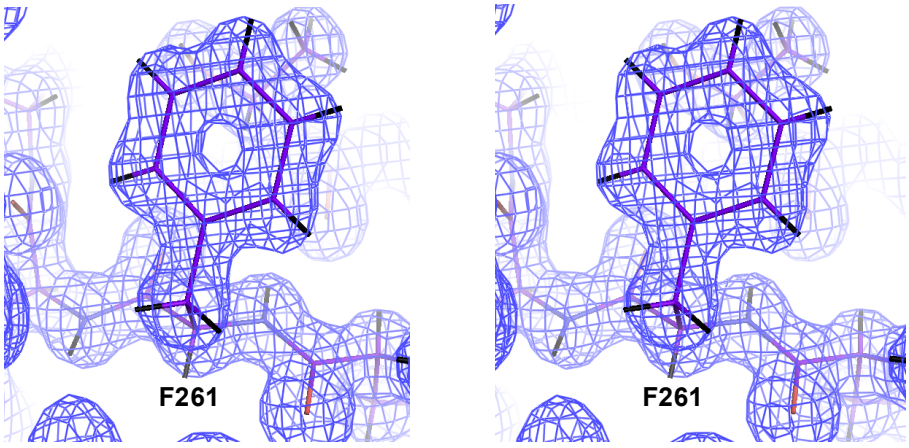


Figure A.4.1. Stereo view of stick representation of the 1.2 Å taPHD crystal structure model (violet) and mesh representation of the corresponding calculated $|2F_o\text{-}F_c|$ density map contoured to 1 σ (blue). Riding hydrogen atoms (see main text) are shown in black.

Riding hydrogens were added to the models after 5-7 rounds of coordinate refinement. In the riding hydrogen model within PHENIX, the addition of hydrogens does not add additional refinement parameters or contributes to structure factor calculation. The position of each hydrogen atom is calculated from the atoms it is bound to ('parent atom'). The riding hydrogens inherit the occupancies of the parent atom and are assigned isotropic B-factors derived from the parent atom B-factor. The developers of PHENIX recommend the use of the riding hydrogen model at all resolutions above 1 Å in order to account for clashes between different parts of the protein without the introduction of additional parameters, thereby reducing the risk of overfitting by essentially providing anti-bumping restraints[1] to improve the geometry of the model (instructions from www.phenix.org). Adding riding hydrogens yields a drop in R_{free} of 0.7-1.2 percentage points.

Water molecules were added to $|F_o\text{-}F_c|$ electron density peaks > 3 σ and within hydrogen bonding distance to hydrogen bond donors and acceptors of the protein, refined and checked manually every round. Water molecule occupancy was fixed as 1. Water molecules closer than 1.8 Å to residues or other waters were removed, as were waters with closest hydrogen bond contacts larger than 3.2 Å. Introduction of water molecules

1. In protein crystallography, *anti-bumping restraints* refer to restraints imposed on the refinement of atomic displacement parameters that take into account predicted steric clashes between non-connected atoms (www.phenix.org).

yielded an R_{free} drop of 2.7-3.1 percentage points. Since the coordinates and intensities of each reflection are dependent on all atoms in the unit cell, modelling solvent is an important part of structure refinement.

The polypeptide termini and gaps were manually built as the model improved between refinement rounds until no further improvement of the R_{free} could be achieved. Atoms with B-factors above 50 Å^2 and/or poor map agreement (correlation coefficient 'CC' < 0.75) were removed. The main disordered regions for both taPHD structures were the *N*- and *C*-termini, neither of which are fully visible in the electron density maps of the structures, the β2/β3 finger-loop region (residues 136-150), a major part of which (135-146) is only visible in the taPHD.taODD structure, and the βIV/βV loop (residues 242-249), which contains a gap of several residues (245-247) that could not be built in either of the structures. After completion of the builds, all non-hydrogen atoms were refined with anisotropic B-factors, leading to a drop in R_{free} ranging from 1.4-2.3 percentage points. Refinement of anisotropic parameters typically only leads to improvements of the model when the resolution of the data is equal to or better than 1.5 Å (Rhodes 2006). Going through the whole model manually every round allowed for the identification and fitting of alternative conformers.

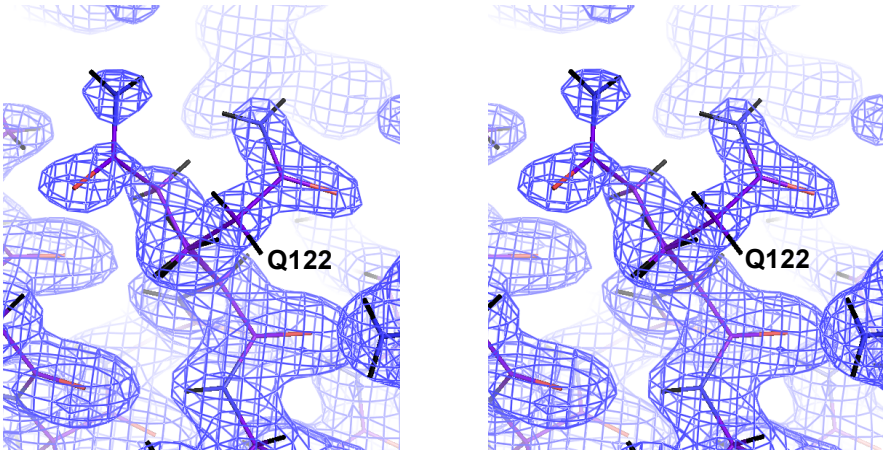


Figure A.4.2. Stereo view of alternative side chain conformations of Q122 of the 1.2 Å taPHD crystal structure. Same colour scheme as in Figure A.4.1. applies.

The final refinement rounds were carried out with optimisation of the X-ray and stereochemistry/atomic displacement parameter weighting, which has yielded an R_{free} decrease of 0.5-1 percentage points.

A.4.3. Comparison of the structures taPHD and taPHD.taODD

A.4.3.1. Secondary structure, gaps and termini

The two crystal structures taPHD and taPHD.taODD are very similar in terms of the secondary structure (RMSD of 0.121 for all Cα), side chain rotamers (see Table A.4.3.) and active site metal coordination.

	taPHD	taPHD.taODD
Resolution	1.2 Å	1.3 Å
R_{free}/R_w	0.1537/0.1374	0.1705/0.1447
Rmsd bonds	0.013 Å	0.0098 Å
Rmsd angles	1.384°	1.188°
Number of amino acids	200	217 (protein) + 17 (peptide)
Ligand/metal	None/Mn(II)	NOG/Mn(II)
Number of waters	156	171
$<B_{factor}>$ protein	19.96 Å^2	22.1 Å^2
$<B_{factor}>$ Mn(II)	10.96 Å^2	8.55 Å^2
$<B_{factor}>$ NOG	-	13.48 Å^2
$<B_{factor}>$ water	32.83 Å^2	32.90 Å^2
***N*-terminal start residue**	80	73
***C*-terminal end residue**	295	296
Gaps	135-146; 245-248	76-80; 140; 245-247
Side chain alt. conf.	17	9

Table A.4.2. Comparison of the refined crystal structures taPHD and taPHD.taODD.

Residues with sidechains refined in two alternative conformations	
taPHD	**taPHD.taODD**
THR83	-
-	GLU90
SER93	-
-	MET95
SER115	-
CYS116	CYS116
GLU119	-
GLN122	GLN122
SER130	-
-	ILE152
ASP158	-
SER169	-
SER174	SER174
SER177	SER177
SER202	-
ARG208	-
-	ILE210
LEU239	-
GLU271	-
-	LEU284
PHE287	-
ARG292	-

Table A.4.3. Amino acid residues refined with two alternative side chain conformations.

The main differences between the two structures are found in the number of alternative conformations modelled, the number of residues visible and the presence of NOG and taODD in taPHD.taODD, whereas the taPHD structure was refined without a ligand coordinating to the active site metal (see below).

Of the 32 additional amino acids refined in the taPHD.taODD structure, 17 are accounted for by the substrate peptide. Three additional *N*-terminal residues (73-75) and one additional residue in the βIV/βV gap (248) are observed (see Figure A.2.5.), while the remaining 11 amino acids missing are part of the β2/β3 finger-loop. The β2/β3 finger-loop residues, in contrast, appear to be highly disordered in the taPHD structure. It is therefore proposed that in the absence of peptide, the β2/β3 finger-loop of taPHD is flexible.

The *N*-terminal residues 73-75 observed in the taPHD.taODD structure are held in place by a disulfide bridge formed between C73 and C116. Since the side chain of C116 is present in two alternative conformations in both crystal structures, the residues 73-75 were all refined in the taPHD.taODD structure as part of the same conformer as the bridge-forming conformer of C116 with an occupancy of 0.5. While K75 has been refined without its side chain due to low correlation coefficient (CC) values, some residual electron density can be seen in the maps. Residues 76-80, however, are disordered. The exact roles of the disulphide bridges in hsPHD2 and taPHD are unclear. It is possible that they support the overall enzyme structure or assist the folding process during translation.

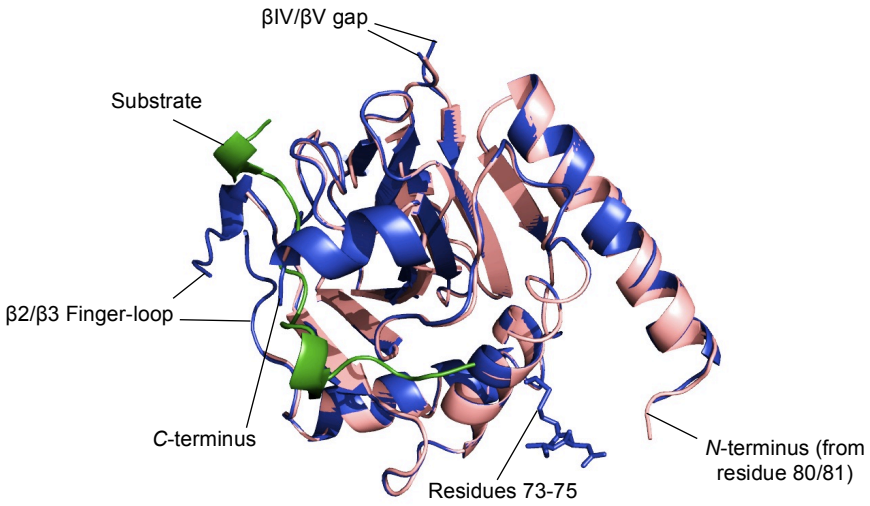


Figure A.4.3. Overlay of the crystal structures taPHD (salmon) and taPHD.taODD (blue with green substrate peptide), illustrating differences in the β2/β3 finger-loop region and the *N*- and *C*-termini. While only residue 140 is missing in the β2/β3 finger-loop region the taPHD.taODD structure, the taPHD structure is missing 11 residues (see Table A.4.2.). Residues 73-75 with the disulfide bond between C73 and C116 were refined only in taPHD.taODD and are shown as sticks.

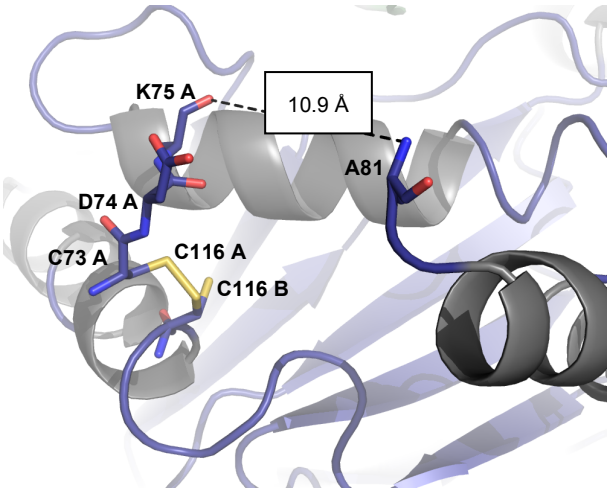


Figure A.4.4. View of *N*-terminal residues of taPHD.taODD. Residues 73-75 were refined as part of alternative conformer A, held in place by a disulfide bond between C73 and alternative conformer A of C116.

The flexibility observed in the βIV/βV DSBH region is interesting in the light of other 2OG-oxygenases (see Section A.1.3.1.) and is discussed further in Section A.4.6.

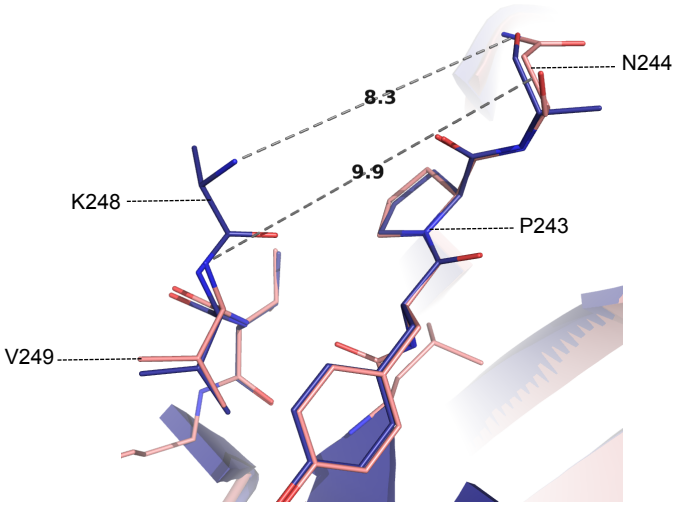


Figure A.4.5. Overlay of stick representation of the βIV/βV DSBH region of taPHD (salmon) and taPHD.taODD (blue). The minor differences in the models that can be seen are comprised of the side chain of N244 in taPHD and residue K248 in taPHD.taODD.

A.4.3.2. The β2/β3 finger-loop region

Interestingly, in the taPHD structure, the β2/β3 finger-loop region is not observed for residues between N134 and G149. For the corresponding residues in the taPHD.taODD structure, the β2/β3 finger-loop region is observed and turns towards the substrate. These observations support the previously suggested hypothesis based on crystal structures of hsPHD2, which proposes that the β2/β3 finger-loop occupies an extended 'finger'- like conformation, pointing out in the absence of substrate, while bending to a 'closed hand' conformation upon substrate binding, thereby enclosing it in the active site. The suggested 'hinge points' for this conformational change are residues G238 and G253 in hsPHD2, which are equivalent to G135 and G149 in taPHD (see alignment in Section A.2.1.).

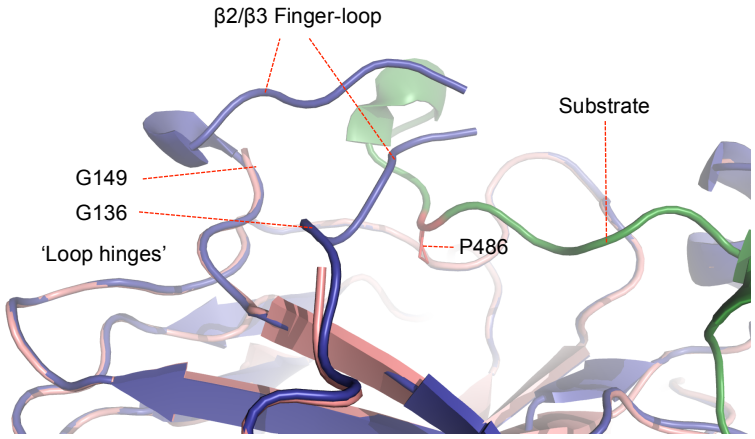


Figure A.4.6. Overlay of the crystal structures of taPHD (salmon) and taPHD.taODD (blue with green substrate peptide), showing the β2/β3 finger-loop region (residues 136-149) forming a loop that encloses the substrate in the active site. Note that the β2/β3 finger-loop region of taPHD is only visible until just before G135 and after G149, the equivalents of which have previously been proposed to represent flexible 'hinge points' in the β2/β3 finger-loop of hsPHD2 (Chowdhury et al. 2009).

Inspection of the β2/β3 finger-loop region of the taPHD.taODD structure led to the identification of 6 polar interactions between loop residues and two polar interactions between residues of the loop and residues of the substrate peptide (see Table A.4.4. and Figure A.4.7.).

Residues taPHD	Type of interaction (loop-loop)	Label in Figure A.4.7.
S134-I152	Backbone-backbone H-bond x2	A
N135-K151	Side chain-side chain H-bond	B
G136-D150	Backbone-backbone H-bond	C
L138-V142	Backbone-backbone H-bond	D
L138-N141	Backbone-backbone H-bond	E
R148-D150	Side chain-side chain salt bridge	F
Residues taPHD-taODD	**Type of interaction (loop-substrate)**	
A139-A485	Backbone-backbone H-bond	G
Q137-F487	Backbone-backbone H-bond	H

Table A.4.4. Interactions identified in the β2/β3 finger-loop region of the taPHD.taODD structure.

While the intramolecular interactions between loop residues (see Table A.4.4.) could explain how the β2/β3 finger-loop region is stabilised where it is essentially folded onto itself, intermolecular interactions between loop residues and substrate could stabilise the substrate in the correct position within the active site. This proposal is supported by the observation that the two identified backbone interactions engaging $A485_{taHIF\alpha}$ and $F487_{taHIF\alpha}$ are on either side of the hydroxylated prolyl residue $P486_{taHIF\alpha}$.

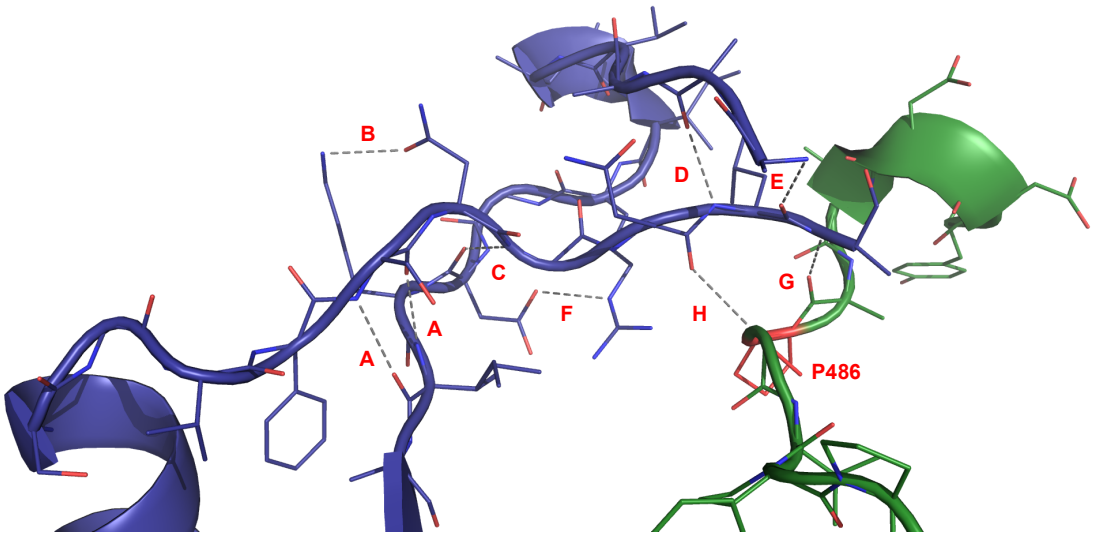


Figure A.4.7. β2/β3 finger-loop region (blue) and the substrate peptide taODD (green) of taPHD.taODD. The hydroxylated prolyl residue $P486_{taHIF\alpha}$ is shown in red. Observed hydrogen bonds are shown in grey (see of interactions in Table A.4.4.).

A.4.3.3. Active site

Inspection of the electron density maps of the active site region of the taPHD structure revealed that the ligand UN9 was not present in the active site (see Figure A.4.8.).

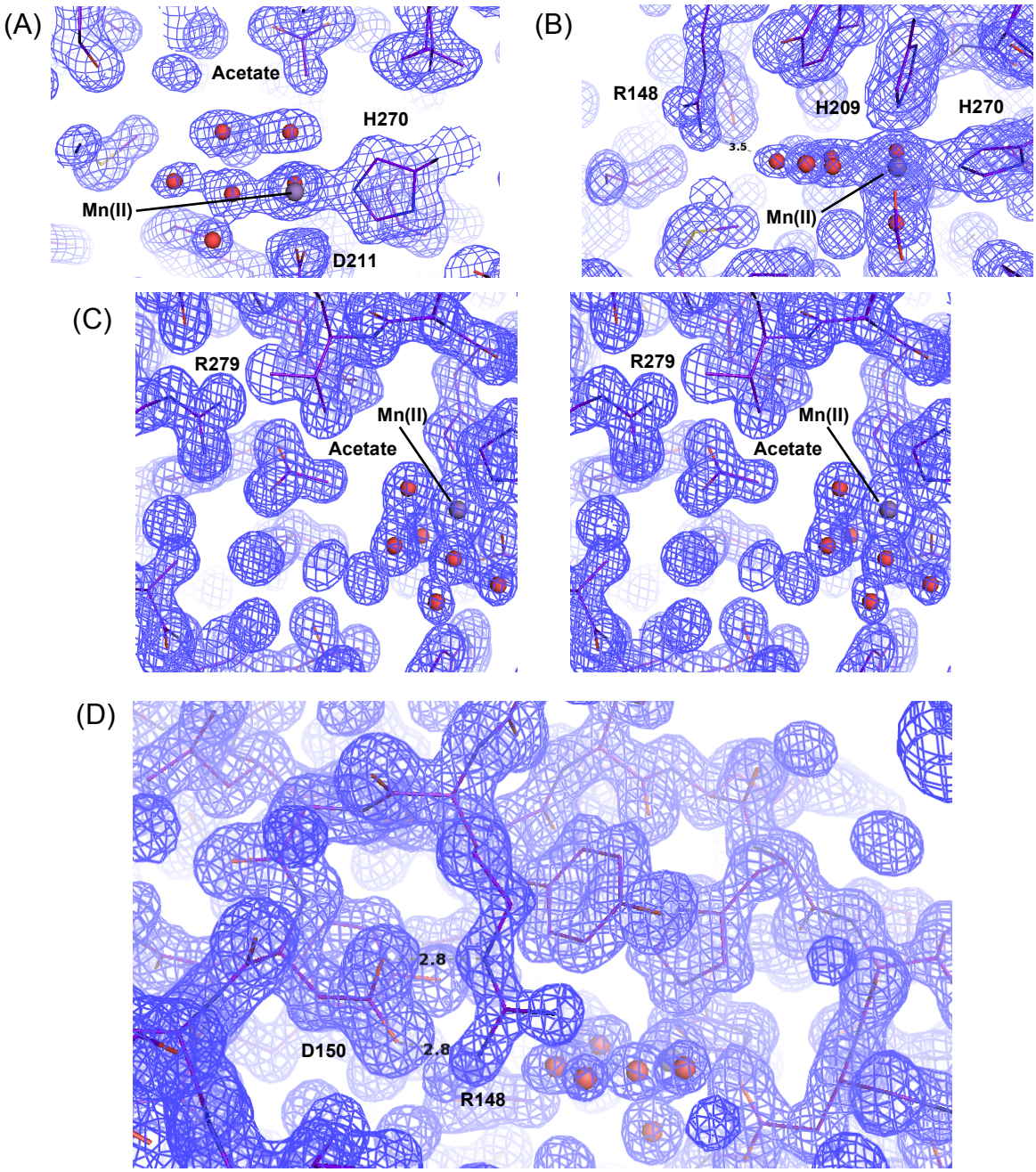


Figure A.4.8. Views of active site region of the taPHD structure, shown with the $|2F_o\text{-}F_c|$ electron density map contoured to 1 σ. View (B) is the same as view (A) but rotated by 90° clockwise around the short edge of the paper. Although the ligand UN9 (see Section A.1.1.) was added to the crystallisation mixture, the active site density does not match the shape of UN9. Therefore, the structure of taPHD was refined with water molecules (shown in red) coordinating to Mn(II) (grey) and an acetate (present in crystallisation buffer, see Section A.2.8.) bound to $R279_{taPHD}$, shown in stereoview (C). View (D) shows the salt bridge between $R148_{taPHD}$ and $D150_{taPHD}$, seen in the taPHD structure but not observed in the crystal structure of hsPHD2 with UN9 because it is disturbed by the aromatic system of UN9 (see Figure A.1.3.).

Although electron density for a carboxylic acid can be clearly seen in proximity of $R279_{taPHD}$ (see Figure A.4.8.C), the electron density surrounding the metal coordination site is of the wrong size and shape for a bicyclic ligand (see Figures A.4.8.A. and A.4.8.B.). These conclusions are supported by the findings that the salt bridge between $R148_{taPHD}$ and $D150_{taPHD}$ is intact (see Figure A.4.8.D). Attempts at refining a UN9 molecule in the active site at a low occupancy not only resulted in poor difference maps, but also in severe steric clashes between the side chain of $R148_{taPHD}$ and UN9, leading to out-of-plane distortion of the aromatic system of UN9. Therefore, the structure of taPHD was refined with an acetate molecule, present in the crystallisation conditions (see Section A.2.8.), binding to $R279_{taPHD}$ and water molecules occupying the active site electron density. Despite being a potent inhibitor of taPHD, UN9 was shown to bind to taPHD only weakly compared to other taPHD inhibitors by non-denaturing mass spectrometry (Section A.2.6.). It is therefore very likely that UN9 is not observed in the crystal structure of taPHD for these reasons. Nevertheless, it is somewhat striking that the active site electron density appears planar (see Figure A.4.8.B). It is therefore speculated that the active site may be occupied by a mixture of water molecules and acetate molecules from the crystallisation condition.

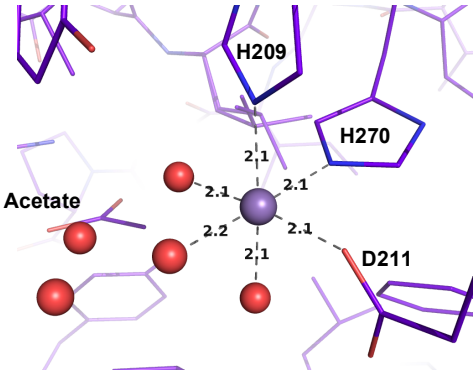


Figure A.4.9. View of active site of the taPHD structure, showing octahedral coordination geometry around the active site Mn(II) and coordination bond lengths in Å. The bond lengths are similar to the ones determined in the hsPHD2 structures with the PDB IDs 3HQR (Mn(II) in active site) and 2G19 (Fe(II) in active site) (Chowdhury et al. 2009).

In contrast to the taPHD structure, NOG is clearly visible in the active site of the taPHD.taODD structure, biding to $R279_{taPHD}$ with its carboxylate and coordinating the metal ion in a bidentate fashion (see Figure A.4.10.).

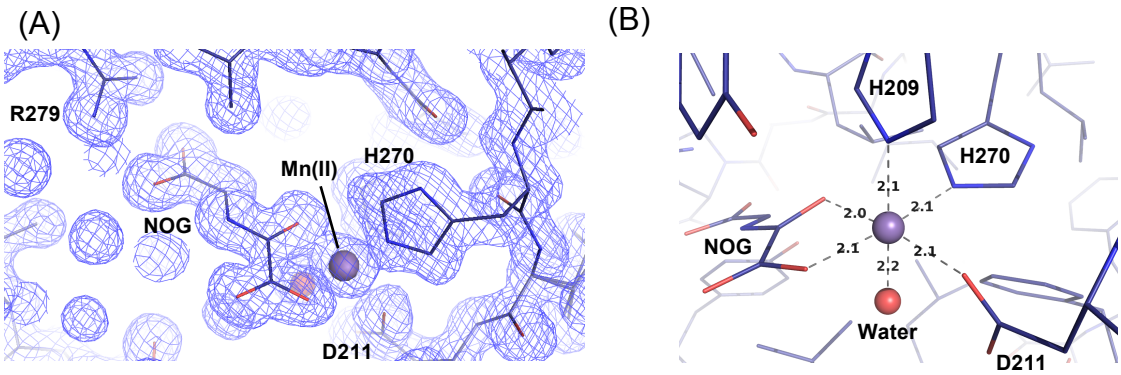


Figure A.4.10. (A) View of active site of the taPHD.taODD structure, shown with $|2F_o\text{-}F_c|$ electron density map contoured to 1 σ. The NOG is clearly distinguishable and can be seen coordinating to the Mn(II) and R279; (B) The Mn(II) ion is octahedrally coordinated with bond length values similar to the ones determined in the taPHD structure.

A.4.3.4. Binding mode of taODD

17 of the 21 substrate peptide amino acids are observed in the taPHD.taODD structure, missing the *N*-terminal fragment EKED (residues 477-480). All observed residues have low B-factors (average ~27 Å^2) and good CC values (>0.8), suggesting that this region of the substrate is rather rigid when bound to taPHD. Several polar and hydrophobic interactions between the taODD substrate peptide and the protein have been identified and are discussed in detail in Section A.4.4, where they are compared to previously identified interactions between hsCODD and hsPHD2. The hydroxylated prolyl-reside $P486_{taHIF\alpha}$ is occupying the *C*^{4}*-endo* conformation, in agreement with the suggested mechanism of hsCODD hydroxylation with hsPHD2 (see Section A.1.2.).

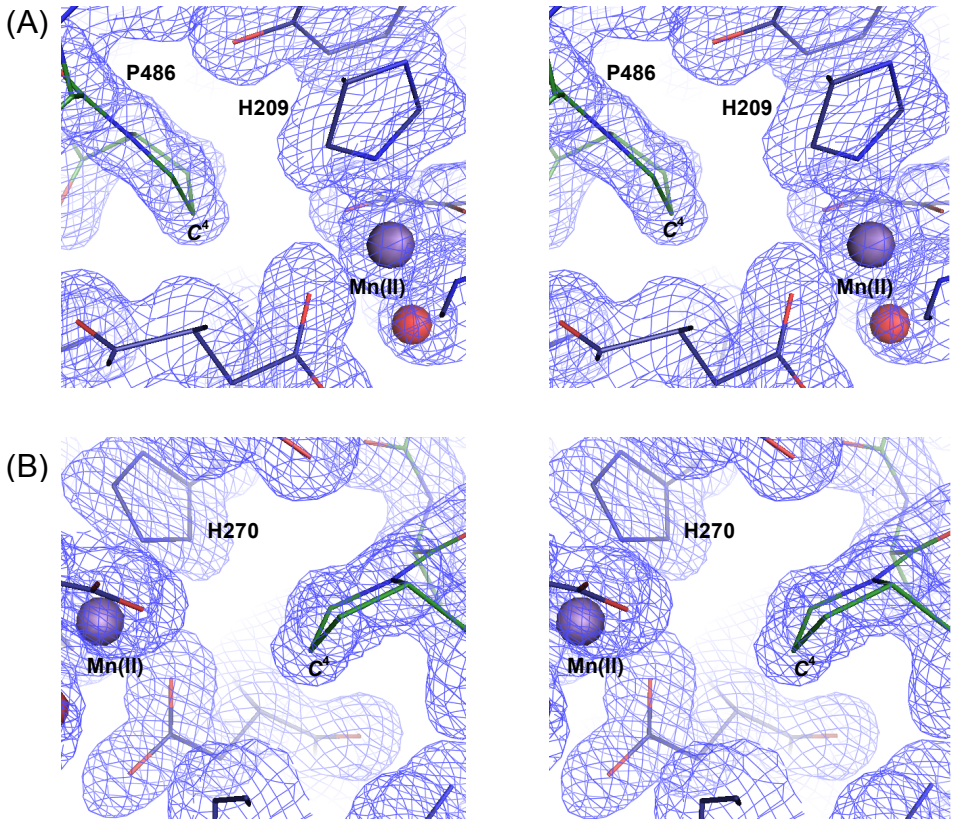


Figure A.4.11. Stereoviews showing that the taODD residue P486$_{taHIF\alpha}$, which is being hydroxylated by taPHD at the *C*^{4}-position, is observed in the *C*^{4}-*endo*-prolyl conformation. View (B) is the same as view (A) but rotated by 180° around axis following long edge of paper (B).

The taODD substrate peptide is apparently bound to the enzyme within the active site groove by both polar and hydrophobic interactions (see Tables A.4.5. and A.4.6.). Two side chain-side chain polar interactions, two side chain-backbone polar interactions and four backbone-backbone polar interactions were identified (see Figure A.4.12. and Table A.4.5.). Additionally, residues Y481$_{taHIF\alpha}$, L484$_{taHIF\alpha}$, P487$_{taHIF\alpha}$ and L497$_{taHIF\alpha}$ were found to be positioned within hydrophobic pockets formed by taPHD side chains (see Figure A.4.13. and Table A.4.6.). These 12 interactions could be key in holding the peptide in place throughout the catalytic cycle.

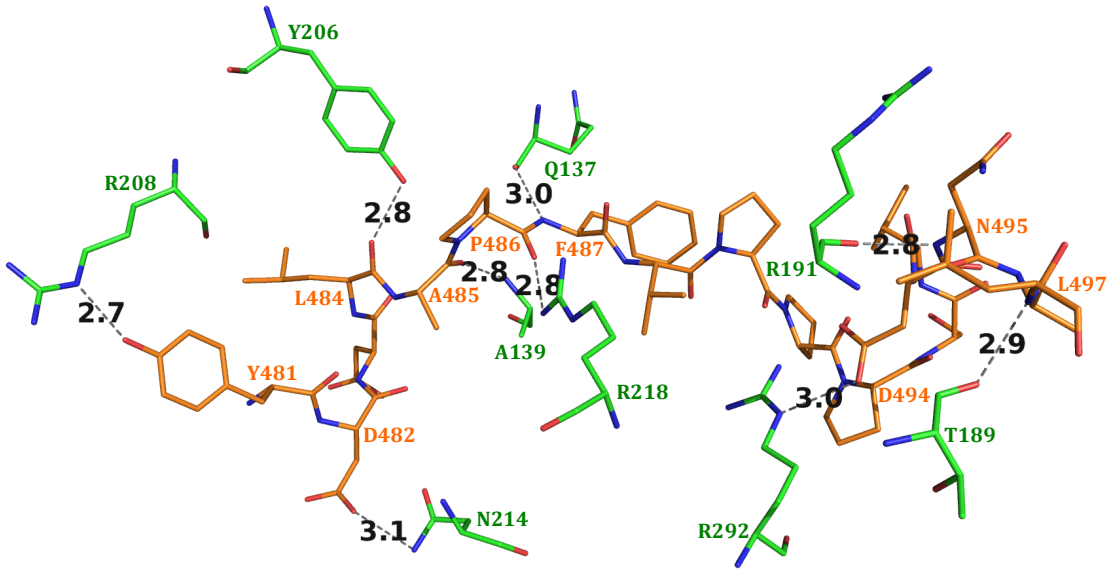


Figure A.4.12. Polar interactions between taODD (orange) and taPHD (green, only interacting residues shown). The interacting residues are labelled near the α-carbon in green (taPHD) and orange (taODD).

Residues taPHD-taODD	Type of polar interaction	Atoms taPHD-taODD	Distance/Å
R208-Y481	Side chain-side chain H-bond	NE-OH	2.7
R292-D494	Side chain-side chain salt bridge	O-NH2	3.0
Y206-L484	Backbone-side chain H-bond	OH-O	2.8
R218-P486	Backbone-side chain H-bond	NH2-O	2.8
Q137-F487	Backbone-backbone H-bond	O-N	3.0
A139-A485	Backbone-backbone H-bond	N-O	2.8
T189-L497	Backbone-backbone H-bond	O-N	2.9
R191-N495	Backbone-backbone H-bond	O-N	2.8

Table A.4.5. List of polar enzyme-substrate interactions between taPHD and taODD.

Residues (taPHD)-taODD	Type of interaction
(I210; P213)-Y481 (I147; I207)-L484 (W154; F493/taODD)-P489 (I176; I188; Y286)-L497	Formation of hydrophobic pocket

Table A.4.6. List of hydrophobic enzyme-substrate interactions between taPHD and taODD.

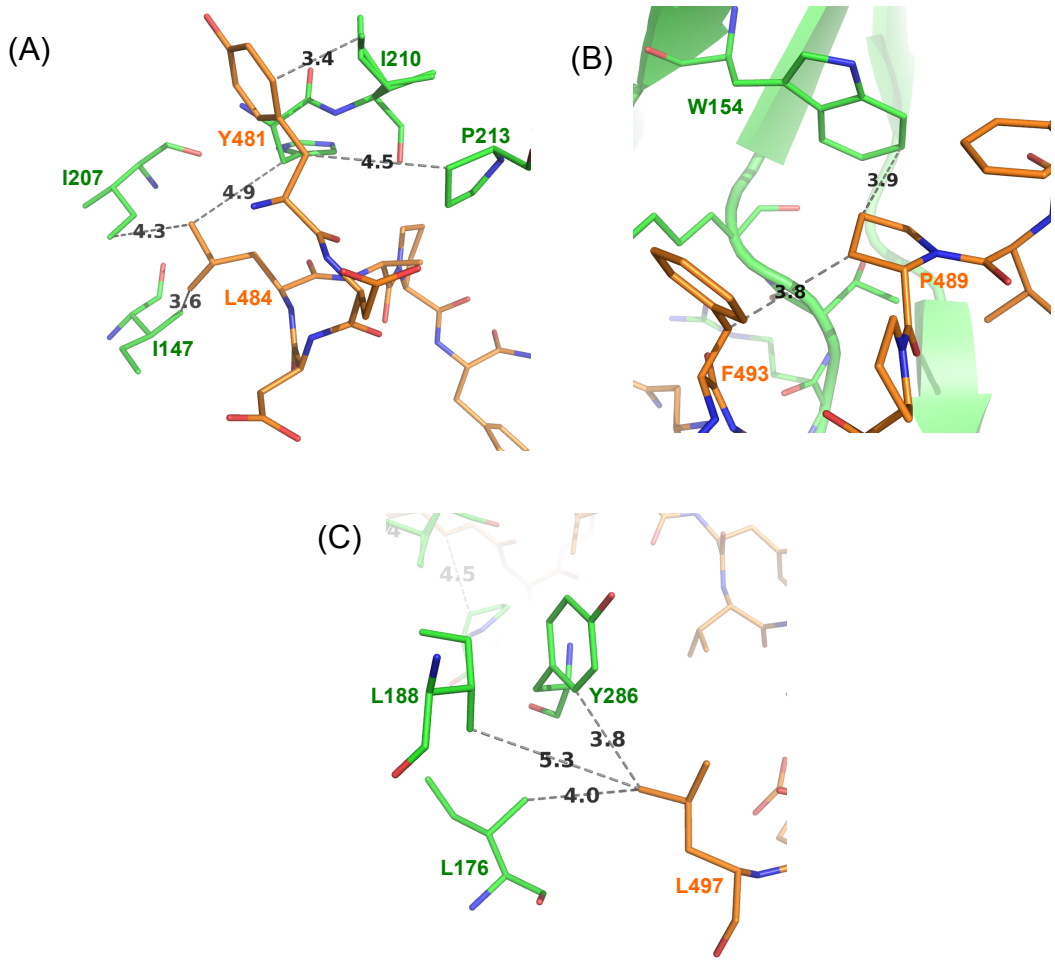


Figure A.4.13. Apparent hydrophobic interactions between taODD (orange) and taPHD (green). (A) Showing taODD residues Y481$_{taHIF\alpha}$ and L484$_{taHIF\alpha}$; (B) showing residue P489$_{taHIF\alpha}$ and (C) showing taODD residue L497$_{taHIF\alpha}$. The interacting residues are labelled in green (taPHD) and orange (taODD) near the α-carbon.

4.3.5. Comparison of temperature factors

The average temperature factors (B-factor) for each residue of the structures taPHD and taPHD.taODD were plotted against residue number (see Figures A.4.14. and A.4.15.). Inspection of the plots suggests that there are two flexible regions: the β2/β3 finger-loop region (residues 136-150) and the βIV/βV region (residues 242-250). The average B-factor over the residues 136-150 and 242-150 are around 35% and 55% higher than the average B-factor for the whole structure, respectively. Comparison of B-factor plots of taPHD and taPHD.taODD suggests that flexibility in the region covering residues 122-147 is reduced upon substrate binding. This residue span includes most of the β2/β3

finger-loop (136-150), which in hsPHD2 has been previously proposed to change conformation upon substrate binding (Chowdhury et al. 2009). Residues 122-131 are part of the α–helix α2 (residues 117-131). The lower B-factors observed in α2 of taPHD.taODD compared taPHD might imply that α2 is stabilised by taODD binding, which could be caused by a conformational change in the β2/β3 finger-loop region.

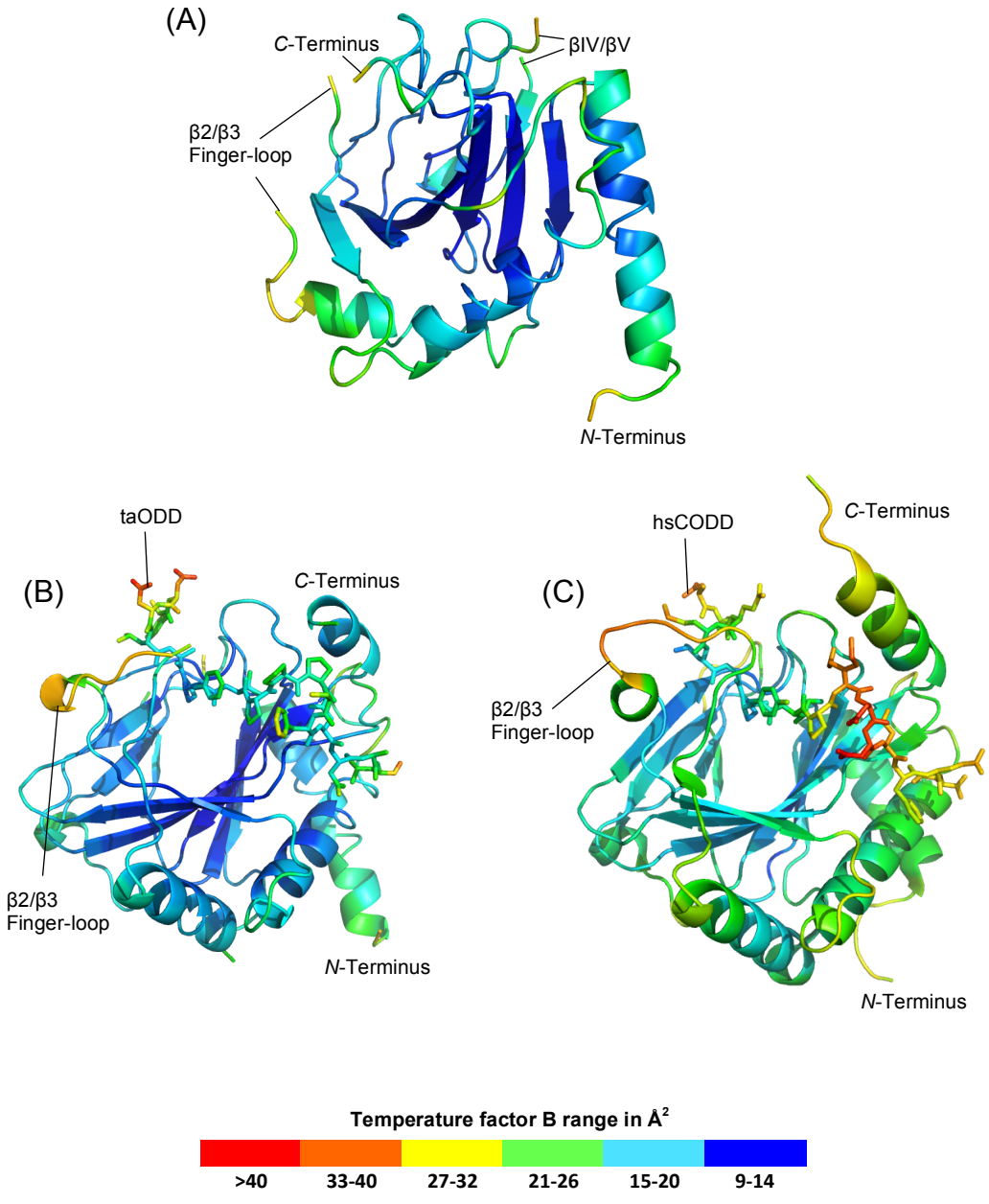


Figure A.4.14. Views from crystal structures taPHD (A), taPHD.taODD (B) and hsPHD2.hsCODD (C), coloured by the temperature factor B. The most flexible regions of taPHD are the *C*- and *N*-termini of the protein and the taODD peptide, the βIV/βV region and the β2/β3 finger-loop region. In hsPHD.hsCODD, in contrast, the most flexible region spans residues 568-570$_{\text{hsHIF-1α}}$

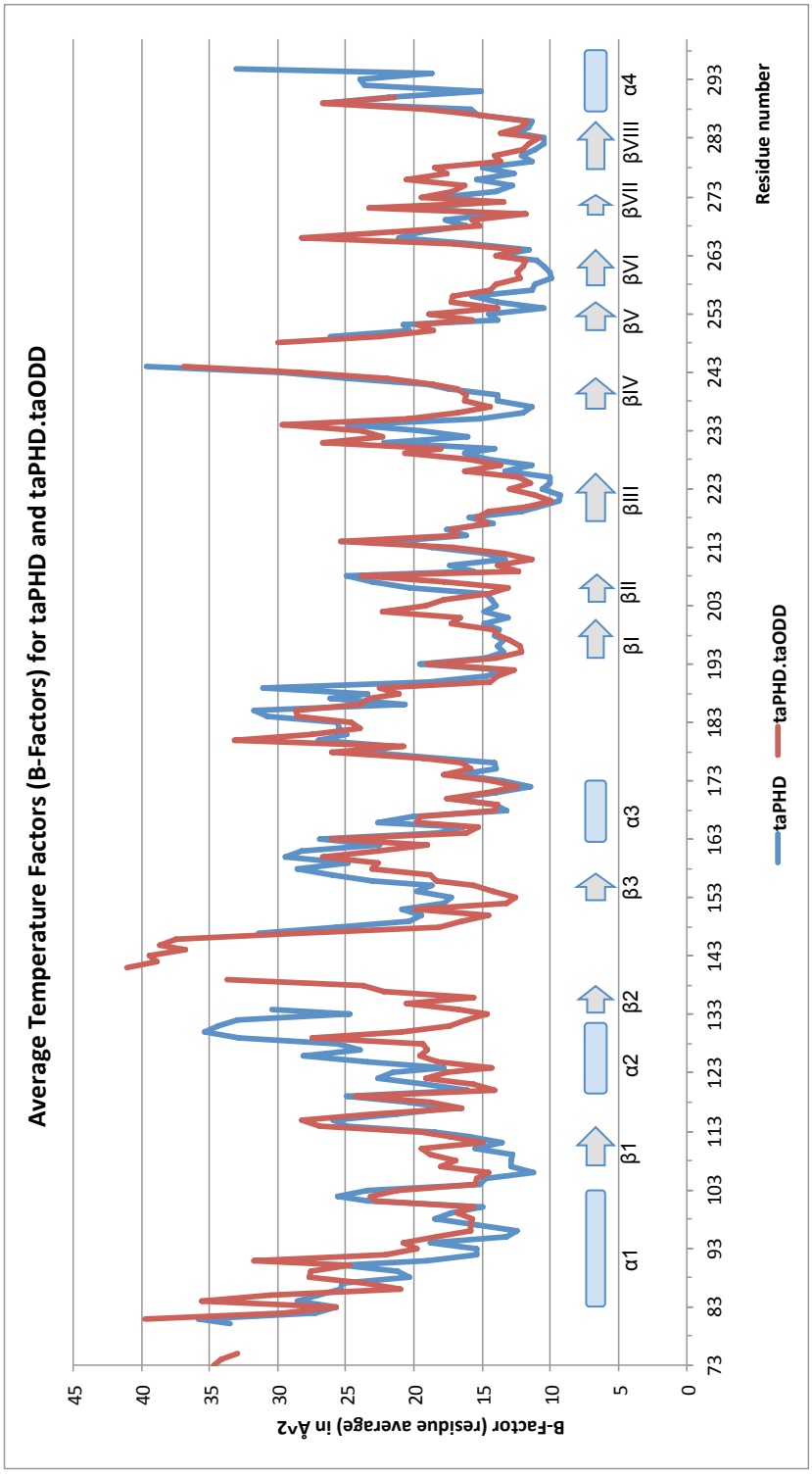


Figure A.4.15. A plot of residue averaged temperature factors against residue number for the structures taPHD and taPHD.taODD. Residues 122-147 appear to be stabilised in the presence of taODD.

A.4.4. Comparison of the structures taPHD.taODD and hsPHD2.hsCODD

The structure of taPHD with its substrate taODD was compared to the structure of hsPHD2 with hsCODD (PDB ID=3HQR). This particular structure of hsPHD2 was chosen because like taPHD.taODD, it contains NOG, Mn(II) and a substrate peptide. The secondary structure of taPHD closely resembles the secondary structure of hsPHD2 (see Figure A.4.16; RMSD=0.531 for all Cα). The DSBH, comprised of a minor and a major β-sheet, is conserved, with three α-helices α1-3 packing along the major β-sheet (see Section A.1.2.). Like in hsPHD2, the active site of taPHD is lying deep within the enzyme.

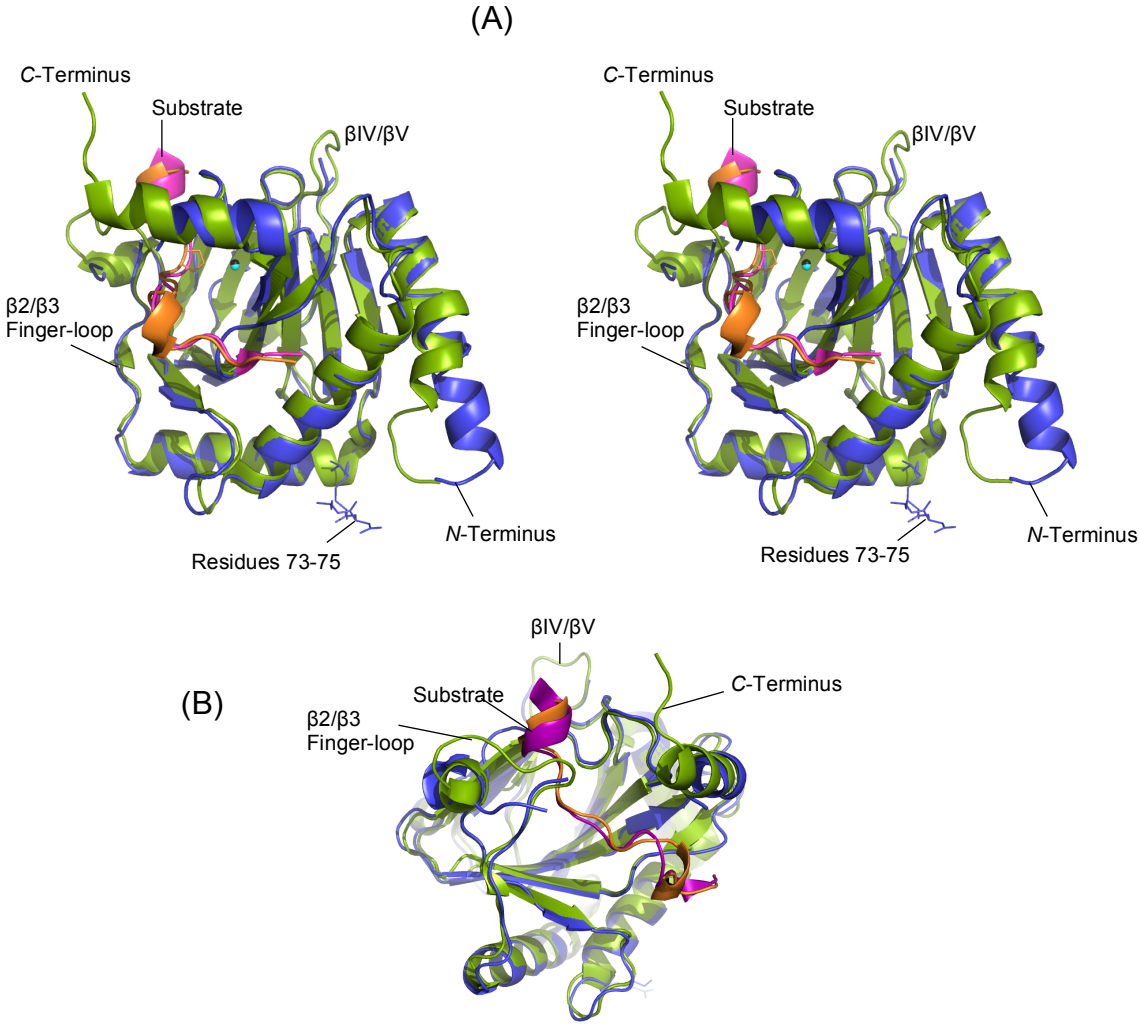


Figure A.4.16. Views of overlay of the crystal structures taPHD.taODD (blue with orange taODD) and hsPHD2.hsCODD (PDB ID=3HQR; green with magenta CODD); (A) Stereo view; (B) View rotated clockwise by 90° along long edge of paper to show substrate binding groove.

Comparison of the βIV/βV regions shows that residues 244 and 248 of taPHD have geometry similar to their hsPHD2 equivalents.

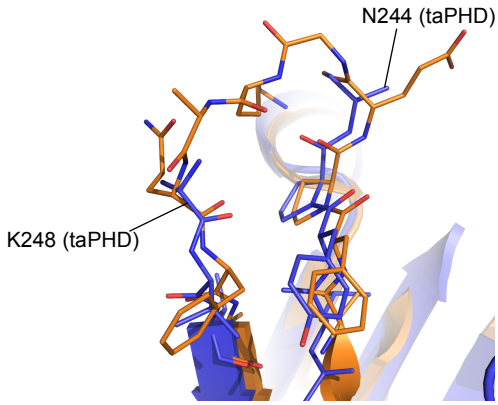


Figure A.4.17. Overlay of the βIV/βV region of taPHD.taODD (blue) and hsPHD2 (3HQR).

The flexibility of the βIV/βV region is interesting in the light of the predicted amino acid sequence of mbP4H, a PHD-like enzyme of *Monosiga brevicollis* (see Part C). mbP4H has a 26 amino acid insert in the βIV/βV region. Since the βIV/βV region lies along the substrate-binding axis, it is possible that it is involved in further enzyme-substrate interactions not observed with a short peptide. Notably, the βIV/βV region has previously been observed to interact with substrates, e.g. in the case of BBOX (Tars et al. 2010).

Another interesting feature of the taPHD.taODD structure is the three *N*-terminal residues 73-75 and the disulfide bonds observed between C73 and C116. Neither residue C116 nor C73 are conserved in hsPHD2. Variants with a serine either in place of C73 or C116 could be prepared or tested for enzymatic activity. The disulfide bond between the hsPHD2 residues C201 and C208, connecting the *C*-terminal helix α1 with β1, is not observed in taPHD. While C208 of hsPHD2 is conserved in taPHD as C106, the amino acid equivalent to hsPHD2 C201 is a glutamine in taPHD (Q99).

The β2/β3 finger-loop region of hsPHD2 is longer by 2 residues compared to taPHD. An overlay (Figure A.4.18.) shows that the general conformations are very similar, with both loops 'hinging' at the same positions and 'reaching' over to enclose the substrate within the active site.

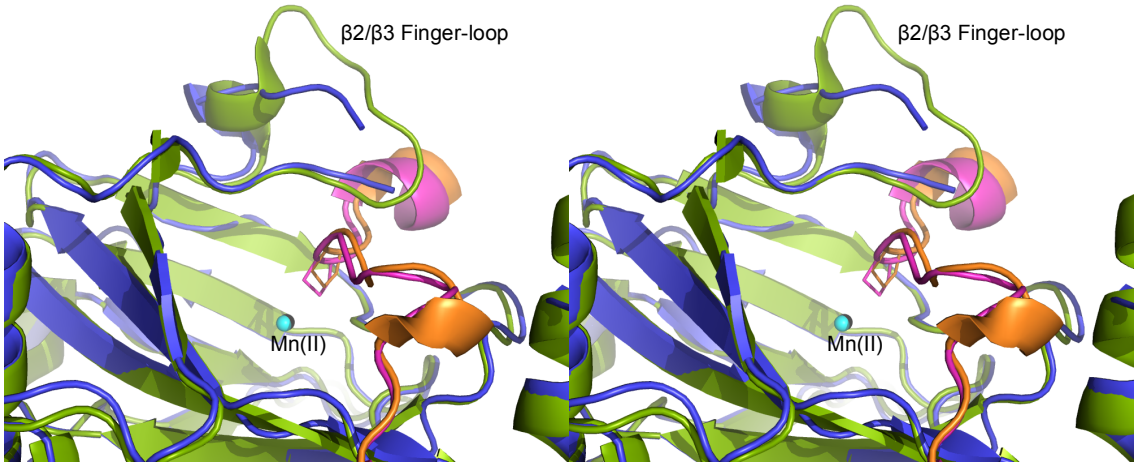


Figure A.4.18. Stereo view of overlay of the β2/β3 finger-loop region of taPHD.taODD (blue with orange taODD) and hsPHD2.hsCODD (green with magenta hsCODD). The Mn(II) ion is shown in cyan.

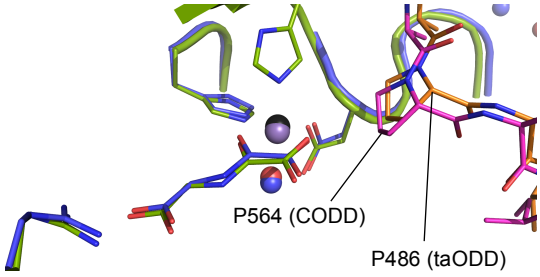


Figure A.4.19. Overlay of the active site regions of taPHD (blue with orange taODD) and hsPHD2 (green with magenta hsCODD). The Mn(II)-ion is shown in grey/black. The Mn(II)-coordinating water is shown in red/blue.

Some of the interactions between taPHD and taODD revealed by the structure taPHD.taODD have previously been observed in the hsPHD2.hsCODD structure (PDB

ID=3HQR). The interactions between hsPHD2 and hsCODD and between taPHD and taODD are compared in Table A.4.7.

Residues (hsPHD2-CODD)	Type of interaction	Residues (taPHD-taODD)
V241-A563	Backbone-backbone H-bond	A139-A485
R322-P564	Backbone-side chain H-bond	R218-P486
(V314; P317)-L559	Hydrophobic pocket formed by enzyme residues	(P269; I210; P213)-Y481
(I251; V311)-L562		(I147; I207; H209)-L484
(I280; I292; Y390)-L574		-
R396-D571	Side chain-side chain salt bridge	R292-D494

Table A.4.7. Summary of enzyme-substrate interactions identified between hsPHD2 and hsCODD (Chowdhury et al. 2009) and the corresponding interactions between taPHD and taODD.

A.4.5. Conclusions drawn from the comparison of taPHD.taODD and hsPHD2.hsCODD

The comparison of the structures taPHD.taODD and hsPHD2.hsCODD (PDB ID=3HQR) shows that the structural features of taPHD and hsPHD2 closely resemble one another, including the secondary structure of both enzymes, the conformation of the β2/β3 finger-loop region in the presence of substrate and the coordination geometry of the active site. Analogous interactions between hsPHD2 and hsCODD can be found between taPHD and taODD. Notable differences are seen in the βIV/βV region, which is disordered in taPHD and differs in amino-acid sequence compared to hsPHD2. In taPHD, this region might have a functionality not identified in hsPHD2, possibly involving enzyme-substrate interactions. The β2/β3 finger-loop region of hsPHD2 is also longer by two residues. This may reflect on the fact that hsPHD2 is hydroxylating hsHIF-1α at two distinct sites, hsCODD and hsNODD, whereas no NODD has been identified in taPHD (Loenarz et al. 2011).

The high similarity between taPHD and hsPHD, including the enzyme-substrate interactions, suggests that the majority of the functional features of the PHDs have evolved in basal metazoans, prior to the diversification of cnidarians of even sponges,

dependent the exact phylogenetic position of the placozoan phylum (see Introduction). It is therefore plausible that more basic PHD-like enzymes, showing some but not all of the features of the PHDs could exist in more simple unicellular organisms. Tracking down such PHD-like enzymes would help understand the evolutionary origin of the involvement of 2OG-oxygenases in oxygen sensing. A study on mbP4H, one of such PHD-like enzymes from the unicellular eukaryote *Monosiga brevicollis*, is presented in Part C of this work.

A.4.6. Comparison of taPHD to other 2OG-oxygenases

There is high structural similarity between taPHD and hsPHD2. A comparison of the structure of taPHD to available structures of selected 2OG-oxygenases yields very similar key conclusions as ones described in the comparison of hsPHD2 to other 2OG-oxygenases (see Section A.1.3.):

- taPHD belongs the same sub-family of 2OG-oxygenases as hsPHD2 and DAOCS, identified by a conserved *N*-terminal helix-strand-helix-strand motif
- There is some evidence that similarly to hsPHD2, taPHD contains a mobile substrate-binding loop, although an open 'finger' conformation could not be observed with taPHD and the loop appears disordered in the absence of a substrate
- In contrast to hsPHD2, the βIV/βV loop of taPHD is disordered. In the JmjC family proteins FIH and JMJD2A, a βIV/βV insert is positioned to be involved in substrate recognition (Mantri et al. 2010). Whether the βIV/βV of taPHD is involved in interactions with taHIFα is currently unclear.

A.5. Summary of Section A

Previously, it has been shown that hsPHD2 is composed of two domains, an *N*-terminal domain that has homology to MYND-type zinc fingers, and the catalytic *C*-terminal domain, which contains the conserved DSBH fold. A mobile region observed between β2 and β3 is commonly referred to as the 'β2/β3 finger-loop'. Previous work suggests that it adopts a disordered finger-like conformation in the absence of substrate and a loop conformation in the presence of substrate, enclosing the substrate within the active site. A sequence alignment of hsPHDs 1-3 and taPHD suggests the highest degree of homology to hsPHD2. taPHD contains a MYND domain and the *C*-terminal catalytic domain with conserved Fe(II)- and 2OG- coordinating residues.

Truncated taPHD lacking the *N*-terminal MYND domain was recombinantly expressed in *E. coli* and purified by His-affinity tag purification and gel filtration. The activity of the purified construct was confirmed by measurement of turnover of 2OG to succinate in the absence of a substrate and by detecting the hydroxylation of a 21mer taHIF-α substrate peptide ('taODD') by MALDI-TOF MS. Amino acid analysis of the modified taODD peptide confirmed that the reaction catalysed by taPHD is (4*R*)-prolyl-hydroxylation.

Crystal structures of taPHD and taPHD in complex with taODD and NOG were obtained at resolutions of 1.2 Å and 1.3 Å, respectively, and refined to R_{free} values of 0.1537 and 0.1705. Comparison to available structures of hsPHD2 shows that the secondary structure is highly conserved between taPHD and hsPHD2. Significant differences can be seen in the β2/β3 finger-loop region and the βIV/βV region. The taPHD β2/β3 finger-loop region is shorter by two amino compared to the human sequence. Like hsPHD2, the taPHD β2/β3 finger-loop region adopts a loop conformation in the taPHD.taODD structure, enclosing the substrate in the active site. In the taPHD structure

without a substrate, most of the β2/β3 finger-loop region is disordered and is visible up to the positions suggested to act as 'hinge points' in hsPHD2, supporting the suggestion that the β2/β3 finger-loop region is a mobile element in the PHDs.

In the taPHD.taODD structure, the hydroxylated prolyl-reside P486 is occupying the C^4-*endo* conformation, supporting the suggested mechanism of hsCODD hydroxylation with hsPHD2, which involves a change from C^4-*endo* to C^4-*exo* conformations of the prolyl-ring upon hydroxylation (Chowdhury et al. 2009).

Polar and hydrophobic enzyme-substrate interactions between taPHD and taODD were identified. Comparison to the hsPHD2.CODD structure shows that equivalents of most of these interactions are also observed between hsPHD2 and CODD. Overall, a remarkably high structural similarity was observed between taPHD and hsPHD2. The next Section explores the biochemical properties of taPHD and relates them to previous findings on hsPHD2.

A.6. References

Adams, P., Afonine, P., Bunkóczi, G., Chen, V., Davis, I., Echoo, N., Headd, J., et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Cryst*., *D66*, 213-221.

Chowdhury, R., McDonough, M. A, Mecinović, J., Loenarz, C., Flashman, E., Hewitson, K. S., Domene, C., et al. (2009). Structural basis for binding of hypoxia-inducible factor to the oxygen-sensing prolyl hydroxylases. *Structure (London, England: 1993)*, *17*(7), 981-9. Elsevier, Inc. doi:10.1016/j.str.2009.06.002

Clifton, I. J., McDonough, M. A, Ehrismann, D., Kershaw, N. J., Granatino, N., & Schofield, C. J. (2006). Structural studies on 2-oxoglutarate oxygenases and related double-stranded beta-helix fold proteins. *Journal of inorganic biochemistry*, *100*(4), 644-69. Elsevier, Inc. doi:10.1016/j.jinorgbio.2006.01.024

Conejo-Garcia, A., McDonough, M. A., Loenarz, C., McNeill, L. a, Hewitson, K. S., Ge, W., Liénard, B. M., et al. (2010). Structural basis for binding of cyclic 2-oxoglutarate analogues to factor-inhibiting hypoxia-inducible factor. *Bioorganic & medicinal chemistry letters*, *20*(20), 6125-8. doi:10.1016/j.bmcl.2010.08.032

Dann, C. E., Bruick, R. K., & Deisenhofer, J. (2002). Structure of factor-inhibiting hypoxia-inducible factor 1: An asparaginyl hydroxylase involved in the hypoxic response pathway. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(24), 15351-6. Elsevier, Inc. doi:10.1073/pnas.202614999

Elkins, J. M., Hewitson, K. S., McNeill, L. A, Seibel, J. F., Schlemminger, I., Pugh, C. W., Ratcliffe, P. J., et al. (2003). Structure of factor-inhibiting hypoxia-inducible factor (HIF) reveals mechanism of oxidative modification of HIF-1 alpha. *The Journal of biological chemistry*, *278*(3), 1802-6. Elsevier, Inc. doi:10.1074/jbc.C200644200

Flashman, E., Bagg, E. A. L., Chowdhury, R., Mecinovic, J., Loenarz, C., McDonough, M. A., Hewitson, K. S., et al. (2007). Kinetic Rationale for Selectivity toward N- and C-terminal Oxygen-dependent Degradation Domain Substrates Mediated by a Loop Region of Hypoxia-Inducible Factor Prolyl Hydroxylases. *Journal of Biological Chemistry*, *283*(7), 3808-3815. Elsevier, Inc. doi:10.1074/jbc.M707411200

Flashman, Emily, Hoffart, L. M., Hamed, R. B., Bollinger, J. M., Krebs, C., & Schofield, C. J. (2010). Evidence for the slow reaction of hypoxia-inducible factor prolyl hydroxylase 2 with oxygen. *The FEBS journal*, *277*(19), 4089-99. Elsevier, Inc. doi:10.1111/j.1742-4658.2010.07804.x

Gorres, K. L., Edupuganti, R., Krow, G. R., & Raines, R. T. (2008). Conformational preferences of substrates for human prolyl 4-hydroxylase. *Biochemistry*, *47*(36), 9447-55. Elsevier, Inc. doi:10.1021/bi8009373

Hewitson, K. S., Liénard, B. M. R., McDonough, M. A., Clifton, I. J., Butler, D., Soares, A. S., Oldham, N. J., et al. (2007). Structural and mechanistic studies on the inhibition of the hypoxia-inducible transcription factor hydroxylases by tricarboxylic acid cycle intermediates. *The Journal of biological chemistry*, *282*(5), 3293-301. doi:10.1074/jbc.M608337200

Hewitson, K. S., McNeill, L. A., Riordan, M. V., Tian, Y.-M., Bullock, A. N., Welford, R. W., Elkins, J. M., et al. (2002). Hypoxia-inducible factor (HIF) asparagine hydroxylase is identical to factor inhibiting HIF (FIH) and is related to the cupin structural family. *The Journal of biological chemistry*, *277*(29), 26351-5. Elsevier, Inc. doi:10.1074/jbc.C200273200

Higgins, L. J., Yan, F., Liu, P., Liu, H.-wen, & Drennan, C. L. (2005). Structural insight into antibiotic fosfomycin biosynthesis by a mononuclear iron enzyme. *Nature*, *437*(7060), 838-44. Elsevier, Inc. doi:10.1038/nature03924

Hirsilä, M., Koivunen, P., Günzler, V., Kivirikko, K. I., & Myllyharju, J. (2003). Characterization of the human prolyl 4-hydroxylases that modify the hypoxia-inducible factor. *The Journal of biological chemistry*, *278*(33), 30772-80. doi:10.1074/jbc.M304982200

Hon, W.-ching, Wilson, M. I., Harlos, K., Pugh, C. W., Maxwell, P. H., Ratcliffe, P. J., Stuart, D. I., et al. (2002). Structural basis for the recognition of hydroxyproline in HIF-1 a by pVHL. *Nature*, *417*(June), 975-978.

Hsieh, M. M., Linde, N. S., Wynter, A., Metzger, M., Wong, C., Langsetmo, I., Lin, A., et al. (2007). HIF-prolyl hydroxylase inhibition results in endogenous erythropoietin induction, erythrocytosis, and modest fetal hemoglobin expression in rhesus macaques. *Blood*, *110*(6), 2140-2147. Elsevier, Inc. doi:10.1182/blood-2007-02-073254

Klose, R. J., Kallin, E. M., & Zhang, Y. (2006). JmjC-domain-containing proteins and histone demethylation. *Nature reviews. Genetics*, *7*(9), 715-27. Elsevier, Inc. doi:10.1038/nrg1945

Koivunen, P., Hirsilä, M., Günzler, V., Kivirikko, K. I., & Myllyharju, J. (2004). Catalytic properties of the asparaginyl hydroxylase (FIH) in the oxygen sensing pathway are distinct from those of its prolyl 4-hydroxylases. *The Journal of biological chemistry*, *279*(11), 9899-904. doi:10.1074/jbc.M312254200

Koivunen, P., Hirsilä, M., Remes, A. M., Hassinen, I. E., Kivirikko, K. I., & Myllyharju, J. (2007). Inhibition of hypoxia-inducible factor (HIF) hydroxylases by citric acid cycle intermediates: possible links between cell metabolism and stabilization of HIF. *The Journal of biological chemistry*, *282*(7), 4524-32. Elsevier, Inc. doi:10.1074/jbc.M610415200

Lancaster, D. E., McNeill, L. A, McDonough, M. A, Aplin, R. T., Hewitson, K. S., Pugh, C. W., Ratcliffe, P. J., et al. (2004). Disruption of dimerization and substrate phosphorylation inhibit factor inhibiting hypoxia-inducible factor (FIH) activity. *The Biochemical journal*, *383*(Pt. 3), 429-37. Elsevier, Inc. doi:10.1042/BJ20040735

Lee, C., Kim, S. J., Jeong, D. G., Lee, S. M., & Ryu, S. E. (2003). Structure of human FIH-1 reveals a unique active site pocket and interaction sites for HIF-1 and von Hippel-Lindau. *The Journal of biological chemistry*, *278*(9), 7558-63. Elsevier, Inc. doi:10.1074/jbc.M210385200

Loenars, C., Coleman, M. L., Boleininger, A., Schierwater, B., Holland, P. W. H., Ratcliffe, P. J., & Schofield, C. J. (2011). The hypoxia-inducible transcription factor pathway regulates oxygen sensing in the simplest animal, Trichoplax adhaerens. *EMBO reports*, *12*(1), 63-70. Elsevier, Inc. doi:10.1038/embor.2010.170

Loenarz, C., & Schofield, C. (2008). Expanding chemical biology of 2-oxoglutarate oxygenases. *Nature Chemical Biology*, *4*(3), 152-156. Elsevier, Inc.

Mahon, P. C., Hirota, K., & Semenza, G. L. (2001). FIH-1: a novel protein that interacts with HIF-1alpha and VHL to mediate repression of HIF-1 transcriptional activity. *Genes and Development*, *15*, 2675-2686. doi:10.1101/gad.924501.HIF-1

Mantri, M., Krojer, T., Bagg, E. A, Webby, C. A, Butler, D. S., Kochan, G., Kavanagh, K. L., et al. (2010). Crystal Structure of the 2-Oxoglutarate- and Fe(II)-Dependent Lysyl Hydroxylase JMJD6. *Journal of molecular biology*, 211-222. Elsevier, Inc. doi:10.1016/j.jmb.2010.05.054

McDonough, M. a, Li, V., Flashman, E., Chowdhury, R., Mohr, C., Liénard, B. M. R., Zondlo, J., et al. (2006). Cellular oxygen sensing: Crystal structure of hypoxia-inducible factor prolyl hydroxylase

(PHD2). *Proceedings of the National Academy of Sciences of the United States of America*, *103*(26), 9814-9. Elsevier, Inc. doi:10.1073/pnas.0601283103

McLuskey, K., Cameron, S., Hammerschmidt, F., & Hunter, W. N. (2005). Structure and reactivity of hydroxypropylphosphonic acid epoxidase in fosfomycin biosynthesis by a cation- and flavin-dependent mechanism. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(40), 14221-6. Elsevier, Inc. doi:10.1073/pnas.0504314102

Mecinović, J., Loenarz, C., Chowdhury, R., & Schofield, C. J. (2009). 2-Oxoglutarate analogue inhibitors of prolyl hydroxylase domain 2. *Bioorganic & medicinal chemistry letters*, *19*(21), 6192-5. Elsevier, Inc. doi:10.1016/j.bmcl.2009.09.005

Neidig, M. L., Brown, C. D., Light, K. M., Fujimori, D. G., Nolan, E. M., Price, J. C., Barr, E. W., et al. (2007). CD and MCD of CytC3 and taurine dioxygenase: role of the facial triad in alpha-KG-dependent oxygenases. *Journal of the American Chemical Society*, *129*(46), 14224-31. Elsevier, Inc. doi:10.1021/ja074557r

Ng, S. S., Kavanagh, K. L., McDonough, M. A, Butler, D., Pilka, E. S., Lienard, B. M. R., Bray, J. E., et al. (2007). Crystal structures of histone demethylase JMJD2A reveal basis for substrate specificity. *Nature*, *448*(7149), 87-91. doi:10.1038/nature05971

Rhodes, G. (2006). *Crystallography Made Crystal Clear* (3rd Editio.). Elsvier.

Roach, P., Clifton, I. J., Fulop, V., Harlos, K., Barton, G., Hajdu, J., Andersson, I., et al. (1995). Crystal structure of isopenicillin N synthase is the first from a new structural family of enzymes. *Nature*, *375*, 700-704. Elsevier, Inc.

Roach, P. L., Clifton, I. J., Hensgens, C. M., Shibata, N., Schofield, C. J., Hajdu, J., & Baldwin, J. E. (1997). Structure of isopenicillin N synthase complexed with substrate and the mechanism of penicillin formation. *Nature*, *387*(6635), 827-30. Elsevier, Inc. doi:10.1038/42990

Rose, N., McDonough, M. ., King, O., Kawamura, A., & Schofield, C. (2011). Inhibition of 2-oxoglutarate dependent oxygenases. *Chemical Society reviews*, *40*, 4364-4397.

Schofield, C. J., & Ratcliffe, P. J. (2004). Oxygen sensing by HIF hydroxylases. *Nature reviews. Molecular cell biology*, *5*(5), 343-54. Elsevier, Inc. doi:10.1038/nrm1366

Stirk, H., Woolfson, D., Hutchinson, E., & Thornton, J. (1992). Depicting topology and hanoedness in jellyroll structures. *FEBS letters*, *308*(1), 1-3. Elsevier, Inc.

Stubbs, C. J., Loenarz, C., Mecinović, J., Yeoh, K. K., Hindley, N., Liénard, B. M., Sobott, F., et al. (2009). Application of a proteolysis/mass spectrometry method for investigating the effects of inhibitors on hydroxylase structure. *Journal of medicinal chemistry*, *52*(9), 2799-805. Elsevier, Inc. doi:10.1021/jm900285r

Sun, G., & Anderson, V. E. (2004). Prevention of artifactual protein oxidation generated during sodium dodecyl sulfate-gel electrophoresis. *Electrophoresis*, *25*(7-8), 959-65. doi:10.1002/elps.200305800

Tars, K., Rumnieks, J., Zeltins, A., Kazaks, A., Kotelovica, S., Leonciks, A., Sharipo, J., et al. (2010). Crystal structure of human gamma-butyrobetaine hydroxylase. *Biochemical and biophysical research communications*, *398*(4), 634-9. Elsevier, Inc. doi:10.1016/j.bbrc.2010.06.121

Thiede, B., Lamer, S., Mattow, J., Siejak, F., Dimmler, C., Rudel, T., & Jungblut, P. R. (2000). Analysis of missed cleavage sites, tryptophan oxidation and N-terminal pyroglutamylation after in-gel tryptic digestion. *Rapid Communication in Mass Spectrometry*, *14*(6), 496-502. doi:10.1002/(SICI)1097-0231(20000331)14:6<496::AID-RCM899>3.0.CO;2-1

Tschank, G., Raghuntath, M., Guenzler, V., & Hanauske-Abel, H. (1987). Pyridinedicarboxylates, the first mechanism-derived inhibitors for prolyl 4-hydroxylase, selectively suppress cellular hydroxyprolyl biosynthesis. *Biochemistry*, *248*, 625-633. Elsevier, Inc.

Valegard, K., Lloyd, M. D., Hara, T., Ramaswamy, S., Perrakisk, A., Thompsonk, A., Lee, H.-jen, et al. (1998). Structure of a cephalosporin synthase, *394*(805-808), 1948-1951. Elsevier, Inc.

Vissers, M. C. M., Gunningham, S. P., Morrison, M. J., Dachs, G. U., & Currie, M. J. (2007). Modulation of hypoxia-inducible factor-1 alpha in cultured primary cells by intracellular ascorbate. *Free radical biology & medicine*, *42*(6), 765-72. doi:10.1016/j.freeradbiomed.2006.11.023

Yang, M., Ge, W., Chowdhury, R., Claridge, T. D. W., Kramer, H. B., Schmierer, B., McDonough, M. a, et al. (2011). Asparagine and aspartate hydroxylation of the cytoskeletal ankyrin family is catalyzed by factor-inhibiting hypoxia-inducible factor. *The Journal of biological chemistry*, *286*(9), 7648-60. doi:10.1074/jbc.M110.193540

Part B: Mechanistic Studies of taPHD

B.1. Hydroxylation of taODD by taPHD: Stereochemistry and Coupling to 2OG Turnover

B.1.1. taPHD forms a stable complex with Fe(II) and 2OG

If PHDs act as molecular oxygen sensors, their reactivity likely needs to be finely tuned to oxygen levels and their reaction with oxygen in cells must be the limiting factor in a cascade of many enzymatic processes.

A characteristic of hsPHD2, at least *in vitro,* is that it forms an unusually stable complex with Fe(II) and 2OG in the absence of a substrate. This property distinguishes hsPHD2 from other 2OG-oxygenases that catalyse prime substrate uncoupled turnover of 2OG, in particular FIH1 (McNeill et al. 2005). A structural explanation for this unusually stable complex formation could be the relatively deep-lying active site of hsPHD2 in comparison to other structurally studied 2OG-oxygenases (McDonough et al. 2006). A long-lived complex between the enzyme, Fe(II) and 2OG could form in cells and rapidly bind HIFα, and then react with oxygen, so that the determining factor for the turnover rate is the oxygen availability.

The stability of the taPHD.Fe(II).2OG complex was investigated using non-denaturing mass spectrometry at a cone voltage of 50 V (see Materials and Methods 8). Five minutes after the addition of taPHD to a solution containing 50 μM Fe(II) and 300 μM 2OG, two new peaks in the mass region between 29000 Da and 29500 Da were observed (see Figure B.1.1): 29042 Da, which corresponds to taPHD.Fe(II) and has the mass of 6xHis-tagged t64 taPHD complexed with Fe(II) (see Section A.2.6.), and 29188 Da, which is heavier by the mass of one 2OG molecule (calculated mass 146.11 Da) and

corresponds to the taPHD.Fe(II).2OG complex. The intensity ratio between the taPHD.Fe(II) peak and the taPHD.Fe(II).2OG peak was 1:1.1 after five minutes. After incubation of the sample at room temperature for 24 hours, the intensity ratio changed to 1:0.7. This suggests that the half-life of the taPHD.Fe(II).2OG complex exceeds 24 hours under these conditions.

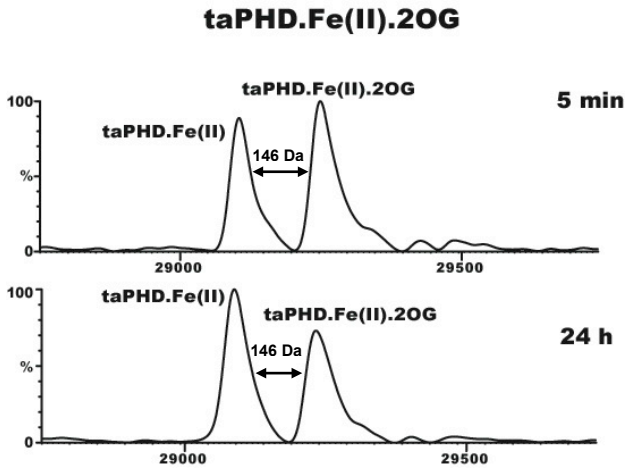


Figure B.1.1. Non-denaturing mass spectra show that taPHD forms a stable complex with Fe(II) and 2OG with a half-life of over 24 hours.

This result is in agreement with previous work on hsPHD2 (Mecinović et al. 2009), suggesting that the findings on hsPHD2 may be applicable to some other if not all PHDs. The formation of an unusually stable enzyme-Fe(II)-2OG complex may therefore be an intrinsic property of the PHDs that is key to oxygen sensing by enabling a slow reaction with oxygen.

B.1.2. Coupling between substrate hydroxylation and 2OG turnover

Uncoupled turnover of 2OG by taPHD was described in Part A and occurs at an initial rate of ~1.2 μM/min as determined by NMR analyses. In the presence of a substrate, turnover of 2OG by 2OG-oxygenases is generally greatly enhanced (Hausinger 2004). The degree of coupling in the presence of a substrate (amount of 2OG turned over compared to amount of substrate turned over) was determined to be 1:1 for hsPHD2 (Flashman et al. 2010). Results for different 2OG-oxygenases are, however, difficult to

compare to the PHDs because their substrates vary greatly, ranging from other proteins such as the transcription factor HIFα to small molecules such as penicillin. The experimental techniques, conditions and assumptions therefore also vary considerably and are presumably different in cells.

To investigate the extent of uncoupled turnover of 2OG by taPHD, the initial rates of succinate production by taPHD in the absence and in the presence of taODD were determined by ^{1}H-NMR as described in Section A.2.4. (see Figures B.1.2. and B.1.3.).

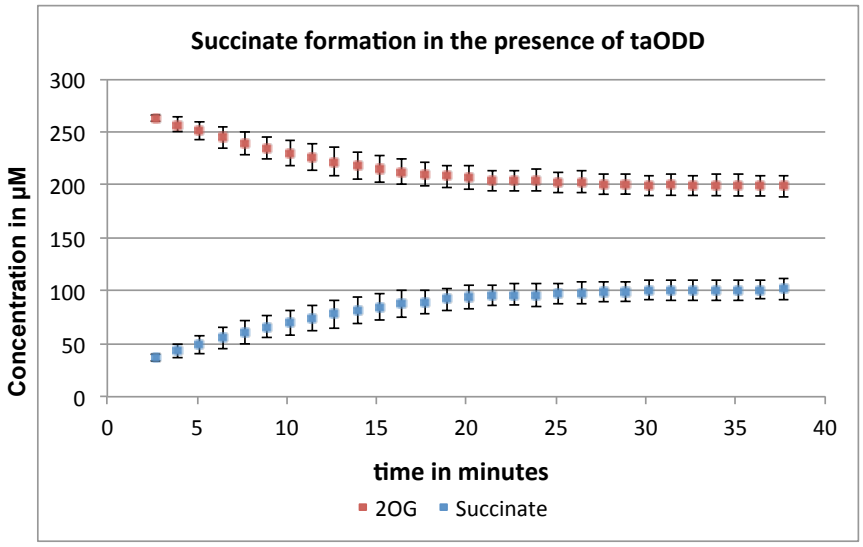


Figure B.1.2. Conversion of 2OG to succinate as a function of time in an assay (avg. of n=3) containing 3.5 µM taPHD, 100 µM taODD, 300 µM 2OG, 4 mM ascorbate and 50 µM Fe(II).

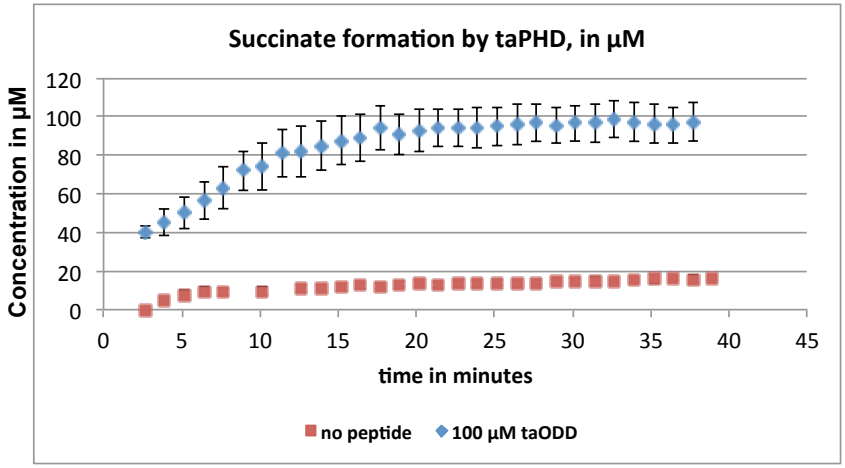


Figure B.1.3. Succinate production by 3.5 µM taPHD in the presence and absence of taODD as a function of time, determined by ^{1}H-NMR (n=3; error bars show standard deviation).

Consistent with the non-denaturing MS results described above, the results show that 2OG turnover in the absence of a substrate is very slow in comparison to the turnover detected in the presence of a substrate. The initial rate in the presence of taODD (11.07 µM/min) is around 10 times the initial rate determined in the absence of taODD (1.19 µM/min; see Section A.2.4.). The coupling between 2OG turnover and substrate hydroxylation in taPHD was investigated by splitting one assay mix into two parts after the addition of enzyme. The first part (160 µL) was transferred to an NMR tube and used for measuring 2OG and succinate by ^{1}H-NMR. Small samples (3 µL each) were taken out of the second part of the assay (100 µL) at regular time intervals, snap-frozen in liquid nitrogen and subsequently analysed by MALDI-TOF MS*. The same concentrations as described in the legend of Figure B.1.2. were used, apart from 2OG, which was used at 100 µM to match the taODD peptide concentration. Initial rates of succinate formation and peptide hydroxylation were determined as described above and yielded 7.84 µM/min for succinate formation and 6.70 µM/min for peptide hydroxylation (see Figure B.1.4.). The ratio of these initial rates is ~1.2:1.

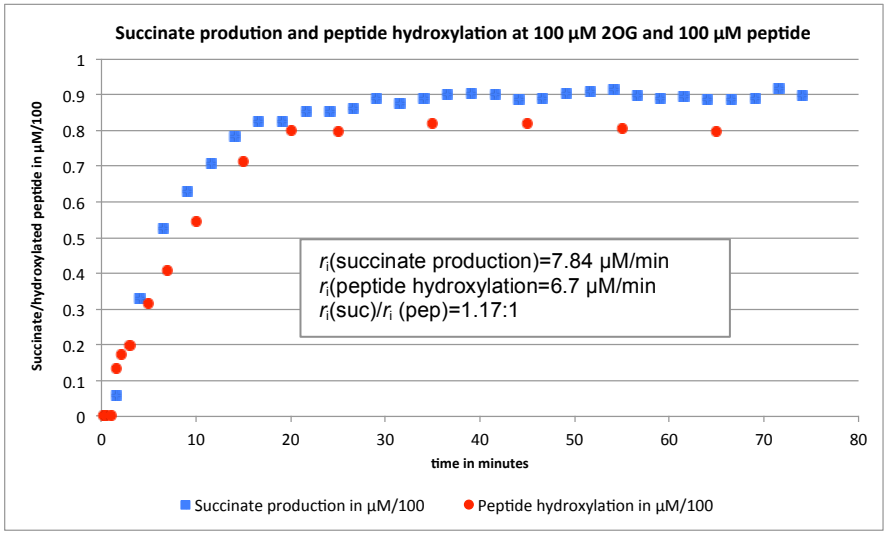


Figure B.1.4. Production of succinate, determined by ^{1}H-NMR, and hydroxylation of taODD, determined by MALDI-TOF MS on the same sample. Visual inspection of the data suggests that the stoichiometry of 2OG turnover and peptide hydroxylation is close to 1, with slightly more succinate produced than peptide hydroxylated. Initial rates are shown in the box and suggest a coupling ratio of ~1.2:1.

*Note: Relative peak intensities were used to determine fractional peptide turnover under the assumption that taODD and HyP-taODD behave similarly at the given MALDI-TOF conditions (also

The calculated coupling ratio implies that the amount of succinate formed is higher than the amount of peptide hydroxylated. This could be due turnover of 2OG not coupled to peptide hydroxylation or a disagreement between the techniques used. The coupling ratio previously determined for hsPHD2 using ^{1}H-NMR for the quantification of 2OG turnover and CODD hydroxylation was 1:1 (Flashman et al. 2010). These results demonstrate further similarities in the biochemical properties of hsPHD2 and taPHD.

B.1.3. Investigation of diastereomeric selectivity of the hydroxylation reaction

A prolyl-residue in a protein or peptide can be hydroxylated at the pro-(4*R*) or the pro-(4*S*) C-H bond to yield either (4*R*)-hydroxy-proline or (4*S*)-hydroxy-proline (see Figure B.1.5).

taPHD was incubated separately overnight with two 19mer diastereomeric hsCODD-analogues, (4*R*)-fluoro-prolyl CODD and (4*S*)-fluoro-prolyl hsCODD (Loenarz et al. 2009)(apart from modifications described, sequence as in Figures A.3.1; see Figure B.1.6. for results). The reaction mixtures were analysed by MALDI-TOF MS for hydroxylation.

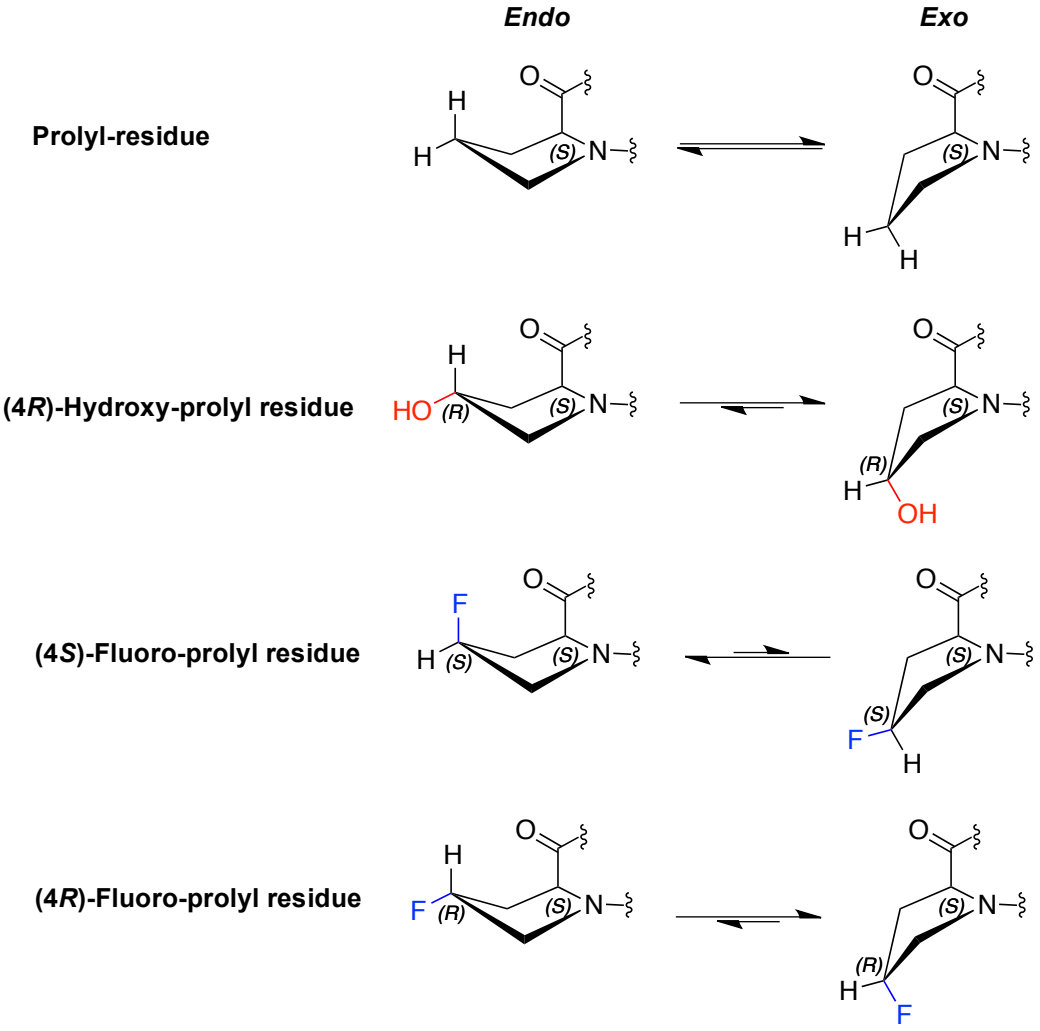


Figure B.1.5. 4-substituted prolyl-derivatives.

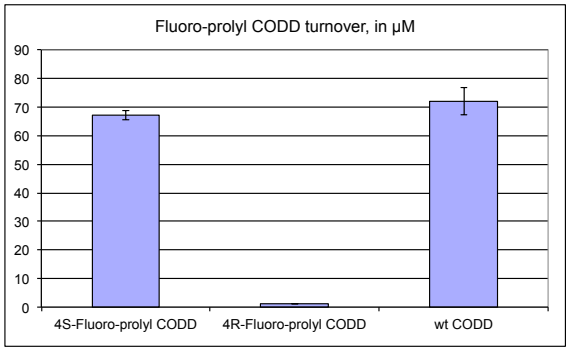


Figure B.1.6. Hydroxylation of fluorinated hsCODD analogues by taPHD (100 µM). Hydroxylation is favoured when a pro-(4*R*) C-H bond equivalent is available, as is the case with (4*S*)-fluoro-prolyl CODD.

While the turnover of (4*R*)-fluoro-prolyl hsCODD was negligible, (4*S*)-fluoro-prolyl hsCODD was hydroxylated to the same extent as *wild-type* hsCODD. These results suggest that prolyl-hydroxylation by taPHD proceeds via a mechanism with a strong diastereomeric preference towards the pro-(4*R*) C-H bond.

Once hydroxylated, the fluorinated hsCODD analogues undergo fluoride elimination, forming a keto-group. This process leads to an overall -4 Da shift as determined by MALDI-TOF MS, rather than a +16 Da shift. The resulting keto-hsCODD derivative can be trapped out with 2,4-dinitrophenyl hydrazine (DNPH; also known as *Brady's reagent*), giving rise to a +180 Da peak.

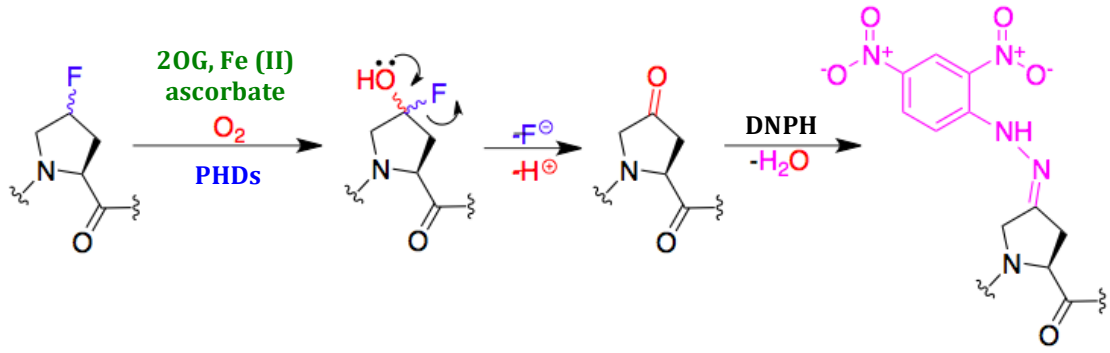


Figure B.1.7. Elimination of fluoride from 4-fluoro-4-hydroxy-prolyl residue and capture of the resulting ketone by DNPH. The condensed product results in a +180 Da mass shift.

The condensation product of the keto-prolyl-peptide and DNPH was detected by MALDI-TOF MS (see Figure B.1.8.). The use of DNPH facilitates the quantification of turnover of the fluorinated hsCODD analogues more easily because a shift of -4 Da is difficult to quantify by MALDI-TOF MS due to overlap of peaks caused by isotope splitting.

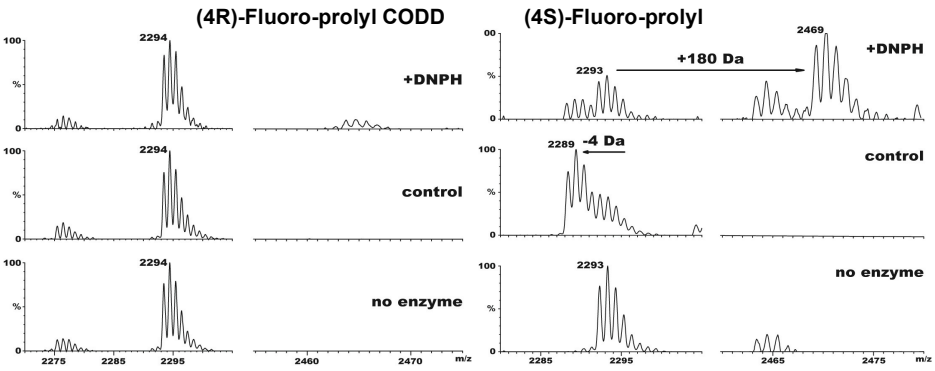


Figure B.1.8. The (4*R*)- and (4*S*)-fluoro-prolyl hsCODD analogues were treated with DNPH after incubation with enzyme to trap out and quantify the ketone formed. The peptide masses are marked above the peaks.

B.1.4. Amino acid analysis of hydroxylated taODD

In order to determine the stereochemistry of hydroxylation of taODD by taPHD, amino acid analysis of hydroxylated taODD peptide was carried out (see Section C.3.3. for detailed description). The peptide was hydrolysed with HCl to single amino acids. These were then derivatised to block the polar amino group, separated by HPLC and detected by mass spectrometry. Comparison of the spectra with hydroxy-proline standards (3*R*/*S*- and 3/*4*-hydroxy-proline) yielded the conclusions that the product of taPHD is (4*R*)-hydroxy-prolyl taODD, similar to (4*R*)-hydroxy-prolyl hsCODD, the product of hsPHD2. Amino acid analysis was carried out by Nikita Loik.

B.1.5. Summary and Conclusions

Several biochemical properties that are proposed as key characteristics of hsPHD2 compared to other 2OG-oxygenases were investigated in taPHD. It was shown that similarly to hsPHD2, taPHD forms an unusually stable enzyme-2OG-Fe(II) complex. The coupling between 2OG turnover and peptide hydroxylation was determined as ~1.2:1, which is similar to the 1:1 coupling ratio previously determined for hsPHD2. Hydroxylation by taPHD was shown to underlie the same stereochemical requirements

as by hsPHD2, only turning over prolyl-residues in the C^4-*endo* conformation. Amino acid analysis has confirmed that the reaction catalysed by taPHD is (4*R*)-prolyl-hydroxylation, which is the same reaction catalysed by hsPHD2. These results suggest that the key biochemical properties of hsPHD2 and taPHD could be universal to all PHDs and play key roles required in their function as molecular oxygen sensors.

B.2. The Kinetics of 2OG and Peptide Turnover by taPHD

Kinetic studies of enzymes are key to gaining insight into the enzyme-substrate interactions throughout the course of a catalytic cycle and form a basis for comparison between similar enzymes and substrates. Under appropriate circumstances, K_m values can be used to compare how tightly a substrate is bound within the active site of an enzyme, while the turnover number k_{cat} allows for the comparison of the maximum number of substrate molecules turned over in a given time span.

B.2.1. K_m and k_{cat} values in context

The K_m value of an enzyme with respect to a certain substrate is often used in literature as a measure of the affinity of the enzyme to the given substrate. The k_{cat} value is used as a measure of the speed of substrate turnover.

$$v = \frac{V_{max}[S]}{K_m + [S]} \tag{2.1}$$

In the equation above, v is the turnover rate and V_{max} represents the maximum substrate turnover rate at saturating substrate concentrations. K_m is equal to the substrate concentration at which the initial reaction rate is half the maximal rate.

k_{cat} is also referred to as the turnover number. It represents the maximum number of substrate molecules turned over by one enzyme molecule throughout a given time span, usually a second. In a kinetic model assuming the reversible formation of an enzyme-substrate intermediate, followed by the irreversible substrate turnover, k_{cat} can be derived as follows:

$$E + S \underset{k_r}{\overset{k_f}{\rightleftharpoons}} ES \xrightarrow{k_{cat}} E + P$$

If $[E] << [S]$ $$v = k_{cat}[E]_0 \frac{[S]}{K_m+[S]} \quad (2.2)$$

Using equation 2.1 $$k_{cat} = V_{max} / [E]_0 \quad (2.3)$$

The kinetic model as described above is based on the assumption that the formation of a substrate-enzyme complex is much faster than product formation. An alternative method, developed by Briggs and Haldane, is known as the "quasi-steady-state" approximation. The underlying assumption of the Briggs-Haldane model is that the concentration of the intermediate complex does not change on the time-scale of product formation:

$$k_f = [E][S] = k_r / [ES] + k_{cat}[ES] \quad (2.4)$$

$$[ES] = \frac{[E]_0[S]}{K_m+[S]} \quad (2.5)$$

$$[K_m] = \frac{k_r + k_{cat}}{k_f} \quad (2.6)$$

The Briggs-Haldane analysis is therefore only valid if $\frac{[E]_0}{K_m+[S]_0} << 1$

Like most models describing biological processes, the Michaelis-Menten model is based on a simplified view of the processes involved and can therefore break down under a variety of circumstances. One example is allosteric binding, which refers to the presence of several binding sites within the same enzyme, with the occupancy of one site affecting

the binding affinity of the other site(s), resulting in negative or positive cooperative binding. Another example is the inhibition of the enzyme by either the substrate or the product when present at high concentrations.

Experimental determination of K_m, V_{max} and k_{cat} is typically carried out by conducting a series of fixed time-point assays at different substrate concentrations and determining the initial rate. The constants are then calculated using non-linear regression techniques. With multi-substrate enzymes such as the PHDs, the concentration of only one of the substrates is varied to obtain an apparent K_m value for a given substrate. The other substrates are used at saturating concentrations so that their concentrations can be assumed constant throughout the initial rate determination. When deciding on the time from starting the assay to data acquisition, both the theoretical concept as well as practical limitations needs to be considered. On one hand, the rate determined needs to be as close to the 'true' initial rate as possible. Enough time needs to have passed for the pre-equilibrium or the steady-state to form, but not enough to deplete the substrate to an extent large enough to have a significant effect on the turnover rate. On the other hand, very short assay times are limited by experimental restraints such as quench or acquisition time and the detection sensitivity of the technique used.

B.2.2. Determination of K_m and k_{cat} values of taPHD for 2OG, taODD and hsCODD

B.2.2.1. Choice of time point for initial rate determination

The apparent K_m values of taPHD for 2OG and a substrate peptide were determined by carrying out a series of end-point assays at varying 2OG or peptide concentrations. Peptide turnover was quantified by MALDI-TOF MS analyses. For comparison, all experiments carried out with taPHD were also repeated with hsPHD2. 2OG turnover was determined indirectly by measuring peptide hydroxylation by MALDI-TOF MS and then assuming a 1.2:1 coupling ratio (see Section B.1.2.).

To determine a suitable time point for initial rate determination, timecourse hydroxylation assays with hsPHD2 and taPHD were carried out using MALDI-TOF MS (see Figure B.2.1.).

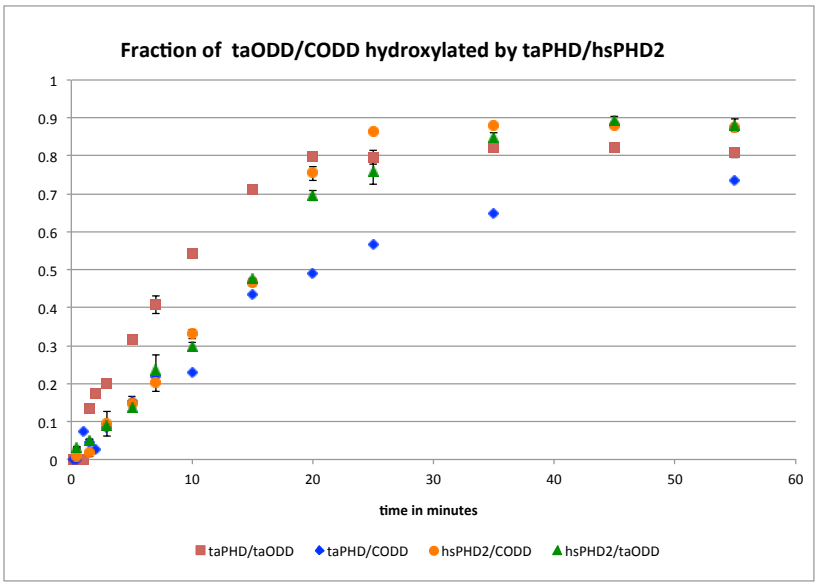


Figure B.2.1. The hydroxylation of taODD and hsCODD by taPHD and hsPHD2 as a function of time was determined by MALDI-TOF MS using 3.5 µM enzyme, 100 µM peptide, 300 µM 2OG, 4 mM ascorbate and 50 µM Fe(II) (n=3; error bars represent standard deviation).

Analysis of the resulting curves suggested that a time point of five minutes was suitable as the curves appear linear in this region (steady-state approximation applies). At the same time, hydroxylation levels are high enough to give a good signal to noise ratio by MALDI-TOF MS. A later time point, such as seven minutes, for example, was expected to yield an inaccurate estimate for the initial rates as the curve had started to flatten in this region.

B.2.2.2. Determination of apparent K_m values

Apparent K_m values with 2OG, taODD and hsCODD (see Figure A.3.1.) were determined for hsPHD2 and taPHD by carrying out end-point hydroxylation assays at t=5 minutes and analysing the data using non-linear fitting.

(A)

		Apparent K_m / µM		Apparent k_{cat} / s^{-1}	
Varying substrate	**Fixed substrate**	**hsPHD2**	**taPHD**	**hsPHD2**	**taPHD**
taODD	300 µM 2OG	22±3	25±5	0.94±0.04	0.89±0.05
CODD	300 µM 2OG	13±3	16±8	1.14±0.06	0.72±0.08
2OG	100 µM taODD	18±3	45±9	0.82±0.02	0.84±0.03
2OG	100 µM CODD	20±4	32±4	0.82±0.02	0.67±0.02

(B)

	Apparent K_m of hsPHD2 in µM		
Method	**2OG**	**CODD**	**Reference**
2OG turnover/crude insect cell lysate	60	7	(Hirsila et al, 2003)
Post-reaction derivatisation fluorescence	25±6	4±1	(McNeill et al, 2005)
Oxygen consumption assay	55±11	21.6±7	(Ehrismann et al, 2007)
Turnover of radiolabelled 2OG	-	36.7±8.4	(Flashmann et al, 2008)
Peptide hydroxylation by MALDI-TOF MS	20±4	13±3	This work

Table B.2.1. (A) Apparent K_m and k_{cat} values of taPHD and hsPHD2 determined by varying the concentration of peptide at constant 2OG concentration or varying 2OG concentration at constant peptide concentration. (B) Comparison to literature data for hsPHD2 and hsCODD.

Comparison of the hsPHD2 K_m values for hsCODD and 2OG with literature data (see Table B2.1.) shows that the apparent K_m values determined with hsPHD2 vary dependent on the experimental technique and substrate used and are likely to be dependent on the purification protocol used for hsPHD2. In the context of the data available throughout the literature, the K_m values determined with hsPHD2 above were therefore deemed reasonable.

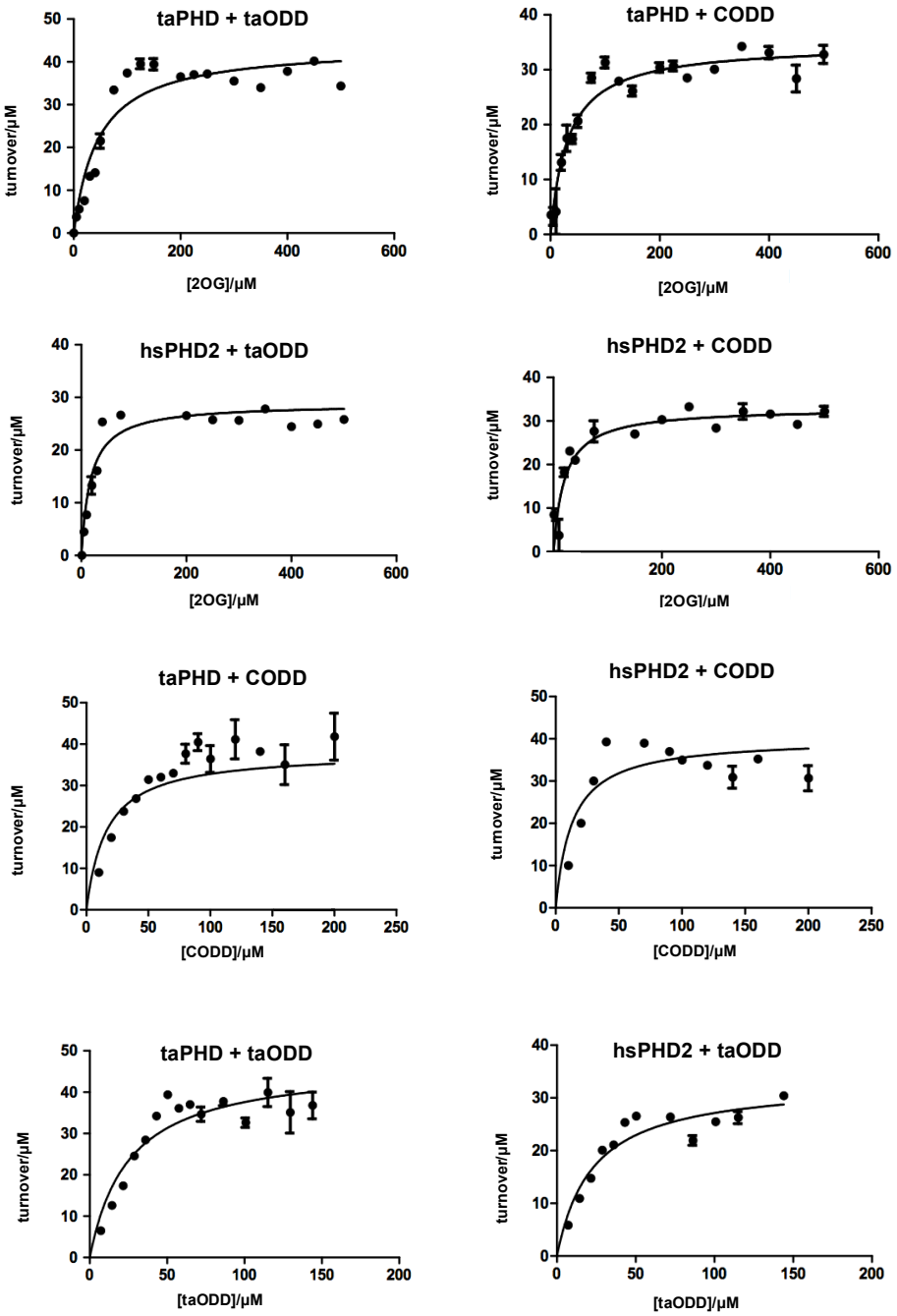


Figure B.2.2. K_m curves for taPHD and hsPHD2 with fixed concentrations of taODD and CODD and varying 2OG concentrations (see Materials and Methods 15). Non-linear fitting was carried out with GRAPH PAD PRISM (www.graphpad.com). The amount of hydroxylated peptide was determined by MALDI-TOF MS after incubation with enzyme at room temperature for 5 minutes (n=3; error bars represent standard deviation).

B.2.2.3. Comparison and discussion of apparent K_m values

The apparent K_m(taODD) for hsPHD2 and for taPHD are the same within error (see Table B.2.1.). So are the apparent K_m(CODD) for hsPHD2 and for taPHD. As can be deduced from equation 2.6, this suggests similar binding of hsCODD by taPHD and hsPHD2, and similar binding of taODD by both enzymes. This is supported by comparisons of the crystal structures (see Section A.4.4.), which has shown that the two enzymes have a high structural similarity.

While the apparent K_m (CODD) and K_m (taODD) are the same within error for taPHD, the apparent K_m(CODD) value is slightly lower than K_m (taODD) for hsPHD2 (Table B.2.1.A.). This suggests that hsPHD2 binds hsCODD more tightly than taODD, whereas taPHD does not distinguish between hsCODD and taODD. This observation could be of biological significance and suggest that the sequence of hsCODD is adapted to the binding site of hsPHD2, whereas the sequence of taODD is not. Since the observed difference in apparent K_m values is rather small, however, this difference may not be biologically relevant.

Most of the K_m curves show a decrease of turnover after the saturating concentration of 2OG has been reached (Figure B2.2.). This suggests substrate inhibition, which becomes apparent at higher substrate concentrations. A consequence of substrate inhibition is an underestimation of V_{max} values and accordingly, k_{cat}. Generally, reactions displaying substrate inhibition do not follow ideal Michaelis-Menten kinetics (Cornish-Bowden 1979). Therefore, the apparent K_m and k_{cat} values need to be interpreted in the context of other experimental observations.

The apparent K_m(2OG) values for hsPHD2 are significantly lower than for taPHD with both hsCODD and taODD being the fixed substrate. While this observation

suggests tighter binding of 2OG by hsPHD2, it could also arise from the incomplete coupling between 2OG turnover and taODD hydroxylation observed with taPHD (Section B.1.2.). At a coupling ratio of ~1.2:1, higher concentrations with taPHD than with hsPHD2 are needed to achieve saturation as determined by measurement of peptide turnover because about one sixth of the 2OG turned over does not contribute towards peptide hydroxylation.

For a given enzyme, the apparent K_m(2OG) values are the same within error with either taODD or hsCODD being the fixed component. While these observations suggest that the effect of the peptide sequence on the apparent K_m(2OG) is not measurable, they need to be interpreted with caution because of the substrate inhibition effect observed with taODD at high 2OG concentrations.

B.2.2.4. Comparison and discussion of apparent k_{cat} values

All k_{cat} values for hsPHD2 are larger or the same within error compared to taPHD. This may have no biological meaning and be caused by a higher concentration of active enzyme in the case of hsPHD2.

Notably, the k_{cat} values for taPHD are higher with taODD compared to hsCODD. This observation is reflected in the initial rate determinations for peptide turnover in Section B.3.3, where the initial rate ratio for hsCODD vs. taODD turnover is ~1:2.3 for taPHD. A possible molecular basis for this differentiation has been identified by inspection of the crystal structures taPHD.taODD and hsPHD2.hsCODD and could be caused by the interactions between $Y481_{taHIF\alpha}$ with hydrophobic pockets of taPHD (P269, I210 and P213), the geometry of which seems to be adapted to taODD. These interactions are discussed in detail in Section B.4.

A comparison of the k_{cat} values for a given enzyme and peptide, determined either by varying the peptide concentration or the concentration of 2OG, shows that the turnover number determined by varying the 2OG concentration is always lower than when varying the peptide concentration. This could be caused from an underestimate of V_{max} at high concentrations of 2OG, caused by substrate inhibition at high 2OG concentrations. This effect is clearly pronounced with taODD as the fixed substrate (see Figure B.2.2.) but may also be present with CODD to a less pronounced extent.

B.2.3. Summary

The apparent K_m and k_{cat} values were determined for hsPHD2 and taPHD with taODD, CODD and 2OG. Analysis of the data yielded the following conclusions:

1. taPHD binds the peptides taODD and hsCODD in a way very similar to the binding of the same two peptides by hsPHD2.
2. hsPHD2 binds hsCODD more tightly than it binds taODD.
3. hsPHD2 binds 2OG more tightly than taPHD.
4. The apparent K_m(2OG) values are not significantly different with taODD and hsCODD as the fixed substrate.
5. The k_{cat} values suggest that taPHD turns over taODD at a higher maximum rate V_{max} than it turns over hsCODD.
6. Substrate inhibition at high 2OG concentrations could indicate a deviation from Michaelis-Menten kinetics, meaning that the K_m and k_{cat} values need to be interpreted with caution and in context of other experimental results.

B.3. Investigation of Enzyme-Substrate Interactions of taPHD and hsPHD2

Crystal structures of taPHD.taODD and hsPHD2.hsCODD lead to the identification of various hydrogen bonds, salt bridges and hydrophobic interactions between the PHDs and their substrates (see Tables A.4.5-7), in the given case peptides corresponding to the HIFα sequences of the *C*-terminal ODD. However, it is unknown whether these interactions are present in solution and if so, what their relative effect on the binding and turnover of the substrate is. It is very plausible that further transient interactions not observed in the crystal structures exist and have a significant effect on reaction rates and mechanism. Therefore, kinetic studies of hsPHD2 and taPHD using a variety of different substrates were carried out (Sections B.3.1.-B.4.). Where possible, the results were related to the crystal structures and previous results with human hsPHD2.

B.3.1. HIFα peptides of various metazoan species

The K_m values determined in Section B.2. suggested that hsPHD2 behaves slightly differently towards taODD than towards hsCODD. This could be caused by the variations in the structures of the enzymes and/or the peptides. Five HIF-1α CODD-homologues (Loenarz et al. 2011) of various metazoan species (see Table B.3.1.) were used for comparative time course assays and determination of initial rates and apparent K_{m} and k_{cat} values to study the effects of a varying substrate peptide sequence and to identify key peptide residues involved in interactions with the taPHD and hsPHD2. It was also of interest whether any observed differences in turnover rates correlated with the evolutionary relationship between the source species of a given peptide sequence and humans/*Trichoplax adhaerens*. The range of species spanned by the available peptides covered the very primitive cnidaria (*Nematostella vectensis*) and echinodermata

(*Strongylocentrotus purpuratum*), arthropoda (*Tribolium castaneum*, *Palaemonetes pugio*) and chordata (*Danio rerio*). All of these peptides have been shown to be substrates of taPHD by MALDI-TOF MS assays (see Appendix 3) and all but TRICA have also been shown to be substrates of hsPHD2 (Loenarz et al. 2011).

Species	Gene ID	Sequence	Abbreviation
Human	gi:4504385	DLDLEMLA**P**YIPMDDDFQL	CODD
Danio rerio	ENSDARG00000044550	LDELDSLA**P**YIPMHGEDFLL	DANRE
Strangylocentrotus purpuratum	gi:115929387	DELAMRA**P**YIPMGEDFDL	STRPU
Tribolium castaneum	gi:189237669	ESDLVAKA**P**YITMNMGDDL	TRICA
Palaemonetes pugio	gi:50261639	LDEFDMRA**P**FIPLSNELLM	PALPU
Nematostella vectensis	JGI:scaffold_26	SNELQNRA**P**YIPPPTGDAAL	NEMVE
Trichoplax adhaerens	JGI:56360	EKEDYDDLA**P**FVPPPSFDNRL	taODD

Table B.3.1. hsCODD, taODD and HIFα-CODD-homologues. Rather than containing CODD and NODD, HIFα of *Nematostella vectensis* contains only one ODD, which resembles the sequence of hsCODD more than of hsNODD (Loenarz et al. 2011). The sequences are listed in the order of the phylogenetic distance of the source species to humans (top of list) and *Trichoplax adhaerens* (bottom of list).

All of the peptides described above were synthesised and purified by HPLC (see Materials and Methods 12) prior to use to ensure that the solutions were free from shorter peptide fragments, which often occur as impurities from incomplete coupling and can inhibit enzymatic activity. However, even pure peptides can yield solutions of concentrations that deviate from the theoretical ones, caused by varying amounts of water retained in the powder form. To circumvent this problem, solutions with a calculated peptide concentration of 1 mM were prepared. ^{1}H-NMR spectra of these solutions were taken and the relative concentrations of the peptides were determined by integration of the aromatic region, adding 100 µM 2OG to each mixture for normalisation (see Figure B.3.1.). The human hsCODD solution was chosen as the standard and its concentration was set at 1 mM. The concentrations of the remaining peptide solutions were adjusted accordingly (see Table B.3.2.). Throughout the assays, the deviations from the calculated concentrations were adjusted by changing the volume of peptide solution added for each peptide.

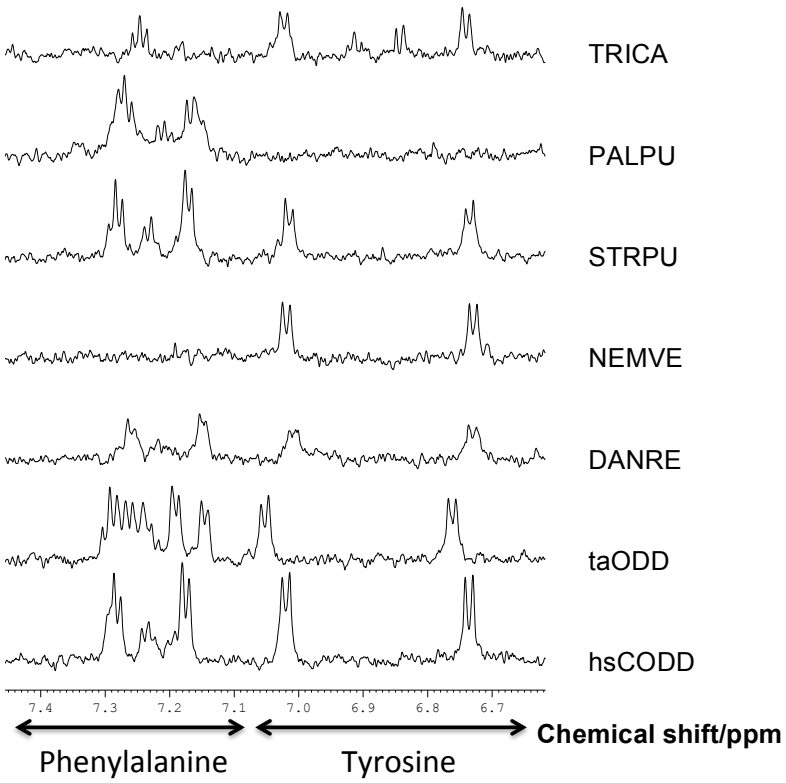


Figure B.3.1. Aromatic region of the ^{1}H-NMR (700 MHz) spectra of hsCODD, taODD and CODD-homologue peptides. The four aromatic protons of tyrosine give rise to two doublets of equal intensity, while the five aromatic protons of phenylalanine yield a more complicated, overlapping pattern. 100 µM 2OG were used for normalisation. The chemical shifts vary slightly for the taODD solution because the peptide was dissolved in d11-Tris rather than D_2O.

Peptide	Correction factor
CODD	1
DANRE	0.51
STRPU	0.72
TRICA	0.53
PALPU	0.7
NEMVE	0.74
taODD	0.72

Table B.3.2. Correction factors for the concentrations of peptide stock solutions relative to the concentration of the hsCODD solution, determined by integrating the ^{1}H-NMR tyrosine and phenylalanine signals shown in Figure B.3.1. and considering the number of aromatic residues in a given peptide. The full peptide sequences are given in Table B.3.1.

B.3.2. Determination of background +16 Da peaks by MALDI-TOF MS

Generally, CODD-like peptides containing a methionine residue show some extent of time dependent +16 Da peak shift by MALDI-TOF MS in the absence of an enzyme. This shift is caused by oxidation of methionine residues to methionine sulfoxide (see Figure B.3.2.). Of all amino acids, the sulfur-containing methionine and cysteine are most prone to oxidation in aqueous solution and oxidised versions of both are common in tissue.

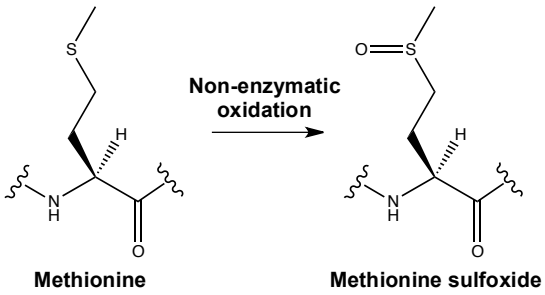


Figure B.3.2. The oxidation of methionine to methionine sulfoxide causes +16 Da peaks in the MALDI-TOF MS spectra of hsCODD and methionine-containing CODD-homologue peptides.

In order to quantify peptide hydroxylation by hsPHD2 and taPHD using MALDI-TOF MS, the time dependence of background +16 Da peaks in the absence of an enzyme was determined.

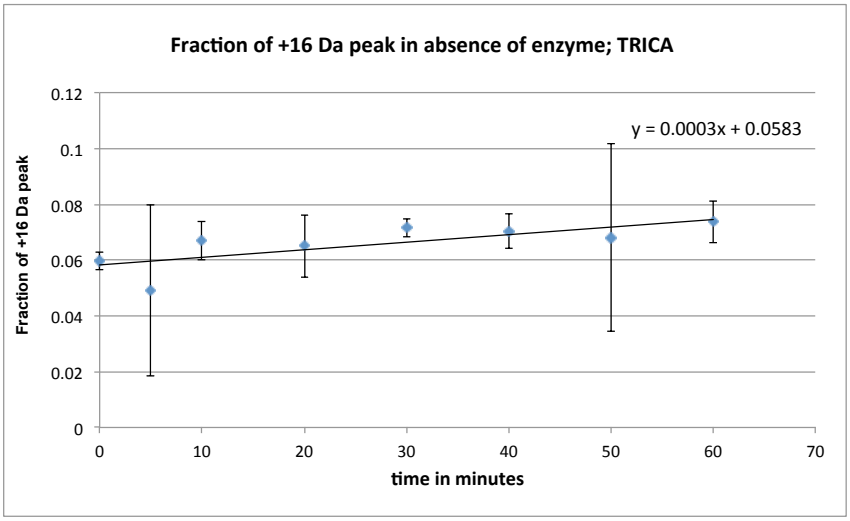


Figure B.3.3. Time course with TRICA peptide in the absence of enzyme. The +16 Da fractions were determined by MALDI-TOF MS. The straight line fit was later subtracted from the '+enzyme' time course as background. This was carried out for all peptides (n=3; error bars represent standard deviation).

All peptides containing methionine residues were found to have some degree of +16 Da shift, with varying degrees of initial amount and time dependence. The background +16 Da peak fractions were fitted with straight lines, the equations of which were used for background corrections in the time course assays in the presence of enzymes.

Peptide	Linear background	Met residues
CODD	y=0.0005x+0.076	2
BRARE	y=0.0018x+0.0617	1
STRPU	y=0.0013x+0.2654	2
TRICA	y=0.0003x+0.0583	2
PALPU	y=0.0002x+0.0998	2
NEMVE	y=0.0836	0
taODD	y=0	0

Table B.3.3. Linear fits of background +16 Da peaks determined by MALDI-TOF MS. While the fitting line is parallel to the *x*-axis for peptides without methionine residues, no further correlation between number of methionine residues and background levels can be seen. The background appears to be specific to each particular sequence. The full peptide sequences are given in Table B.3.1.

As expected, no background +16 Da peak was observed with taODD, which does not contain methionine residues. Although the NEMVE peptide does not contain a methionine, a constant low amount of +16 Da peptide was also detected (8.4%), the relative amount of which was not found to be time dependent. This could be due to a residual contamination from the synthesis process not removed by HPLC purification.

While it is possible that peptides containing methionine sulfoxide are hydroxylated again at the prolyl residue by a PHD enzyme, no +32 Da peaks were observed by MALDI-TOF MS with any of the peptides. This suggests that methionine oxidation of a peptide inhibits prolyl-hydroxylation by the PHDs. Possible reasons for this could be that the oxidised peptide does not fit the binding groove well enough because of the additional

oxygen or that it binds non-specifically to the enzyme through hydrogen bonding with the sulfoxide group.

B.3.3. Peptide hydroxylation time courses using MALDI-TOF MS

For peptide hydroxylation assays, the fraction of hydroxylated peptide was monitored as a function of time using MALDI-TOF MS. The assays were started by adding taPHD or hsPHD2 (final concentration of 3.5 μM) to a solution containing 100 μM peptide, 300 μM 2OG, 50 μM Fe(II) and 4 mM ascorbate in 50 mM Tris-HCl (pH=7.5). At regular intervals, 3 μL were taken from the reaction mixture, transferred to Eppendorf tubes and snap-frozen in liquid nitrogen. After completion of the assay, each time point tube was thawed by adding 3 μL α-cyano-4-hydroxycinnamic acid (CHCA), which quenches the reaction by changing the pH of the mixture and serves as a crystallisation matrix for MALDI-TOF MS. The background for each peptide as determined in Section B.3.1. was subtracted from each time point to account for +16 Da peaks caused by non-enzymatic methionine oxidation. Time course hydroxylation assays were carried out with taPHD and hsPHD2 and all of the peptides described in Section B.3.3.

Hydroxylation of the peptides PALPU and TRICA by taPHD and hsPHD2 was negligible at the chosen time scale of 5 minutes. Initial hydroxylation rates were determined by linear fitting in the time span t=0 to t=5 min (see Table B.3.4.), as previously discussed in Section B.3.2.

	r_i(X)in μM/min	
Peptide	**X=taPHD**	**X=hsPHD2**
CODD	2.8	2.9
DANRE	6.0	4.1
STRPU	5.1	3.0
TRICA	0.0	0.0
PALPU	0.0	0.0
NEMVE	1.6	2.3
taODD	6.5	2.8

Table B.3.4 Initial rates for peptide hydroxylation by taPHD and hsPHD2, calculated by linear fitting of the data shown in Figure B.3.4. in the time span between t=0 and t=5 min.

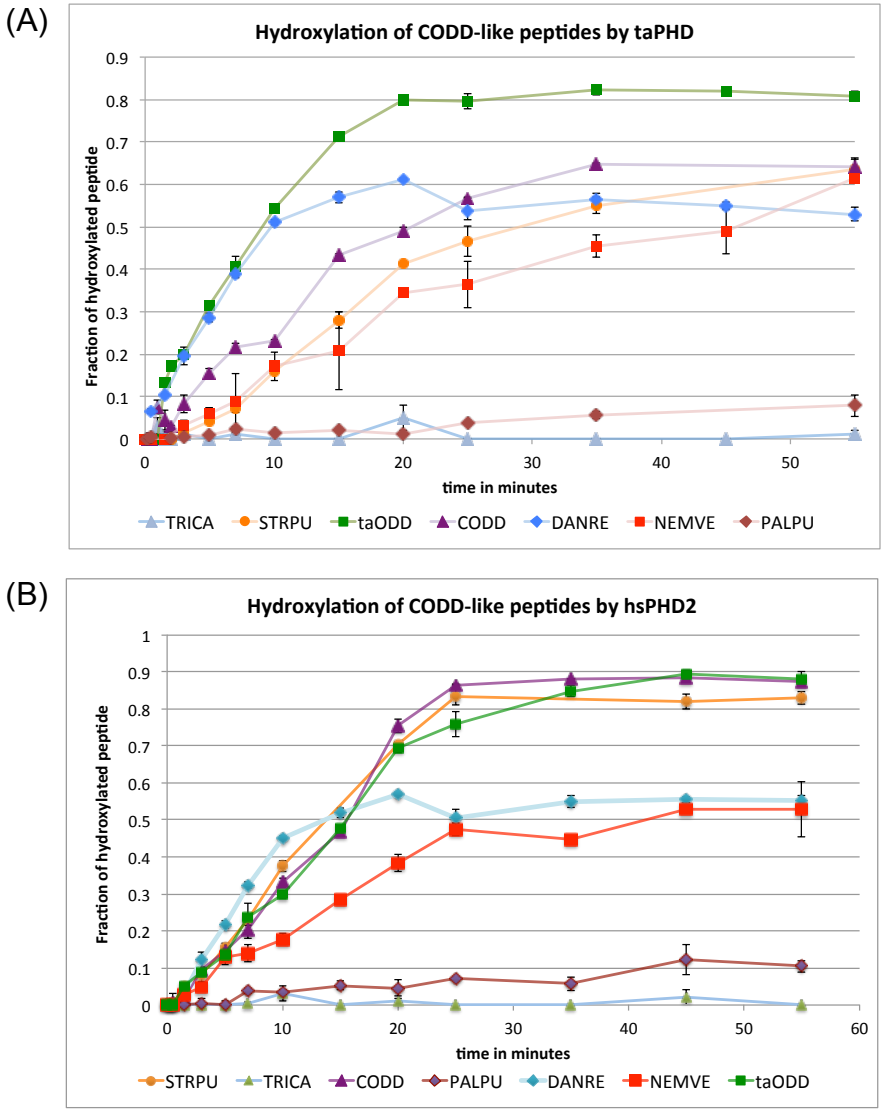


Figure B.3.4. (A) Hydroxylation of CODD-like peptides by taPHD and (B) by hsPHD2 as a function of time, determined by MALDI-TOF MS (n=3; error bars show standard deviation).

While it is difficult to directly compare initial rates between taPHD and hsPHD2 because the precise concentrations of active enzyme are not known, the rates can be compared in relation to the native substrates, taODD and hsCODD. Comparison of the initial rates in Table B.3.4. yielded the following observations:

1. The order of initial rates determined with taPHD is the following:

taODD > DANRE > STRPU > CODD > NEMVE

2. The order of initial rates determined with hsPHD2 is the following:

DANRE > STRPU ≈ CODD ≈ taODD > NEMVE

Interestingly, different relative rate orders are observed for taPHD and hsPHD2, suggesting that they might interact differently with the peptides. While taPHD turns over taODD roughly twice as fast as CODD, the initial rates for taODD and CODD are similar with hsPHD2. The peptide DANRE is turned over fast with both enzymes, at a rate comparable to taODD and is the peptide with the highest turnover rate for hsPHD2. The peptide NEMVE has the lowest initial rates of the five rates recorded for both enzymes. The implications of these results are discussed in Section B.4. in the context of the taPHD.taODD and hsPHD2.hsCODD crystal structures and phylogenetic relationships of the source species.

B.3.4. Apparent K_m and k_{cat} values of CODD-homologues for taPHD and hsPHD2

Apparent K_m and k_{cat} values for the CODD-homologue peptides were determined for taPHD and hsPHD2 as described in Section B.2.2. using 3.5 µM enzyme, 100 µM peptide, 300 µM 2OG, 50 µM Fe(II) and 4 mM ascorbate. The peptides TRICA and PALPU were excluded because of their low initial turnover rates in the time span between t=0 and t=5 minutes (see Section B.3.3.). Using TRICA and PALPU at higher concentrations would have been highly impractical because they are not soluble under assay conditions to more than ~2 mM.

	Apparent K_m(X) in µM		Apparent k_{cat}(X) in s^{-1}	
Peptide	**X=taPHD**	**X=hsPHD2**	**X=taPHD**	**X=hsPHD2**
CODD	16±8	13±3	0.72±0.08	1.14±0.06
DANRE	32±5	66±7	1.43±0.0.09	2.29±0.1
STRPU	49±8	20±4	0.47±0.03	1.23±0.06
TRICA	-	-	-	-
PALPU	-	-	-	-
NEMVE	61±7	28±6	1.09±0.06	1.54±0.04
taODD	25±5	22±3	0.89±0.05	0.94±0.04

Table B.3.5. K_m and k_{cat} values of taPHD and hsPHD2 for hsCODD, taODD and CODD-homologues, determined by MALDI-TOF MS and non-linear fitting (GRAPH PAD PRISM; n=3).

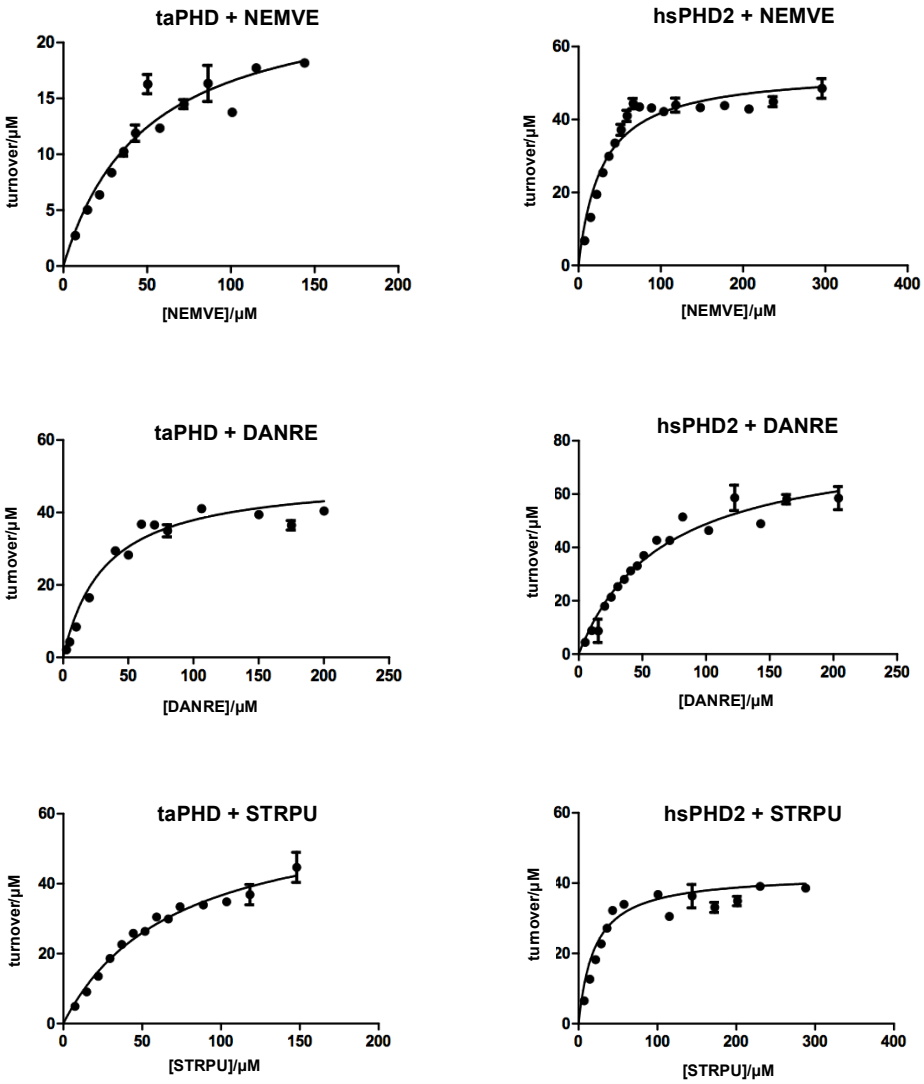


Figure B.3.5. Fitting curves for determining K_m and k_{cat} of CODD-homologue peptides (GRAPH PAD PRISM). The amount of hydroxylated peptide was determined by MALDI-TOF MS after incubation with taPHD or hsPHD2 for 5 minutes (n=3, error bars show standard deviation; see Materials and Methods 15).

Most of the K_m indicate some degree of substrate inhibition. As substrate inhibition implies non-Michaelis-Menten kinetics, the K_m values in Table B.3.5. need to be interpreted with caution as substrate inhibition can distort the values determined for K_m and V_{max}.

For taPHD, K_m values increase in the order:

CODD ≤ taODD ≤ DANRE < STRPU < NEMVE

For hsPHD2, the increasing order is:

CODD ≤ STRPU ≤ taODD ≤ NEMVE < DANRE

The symbol '≤' indicates that the K_m values of the peptides to its left and right are the same within error. The lowest K_m values (interpreted as the tightest binding to the enzyme) are observed for hsCODD, which is the same within error as taODD for taPHD. The peptide NEMVE has high K_m values for both enzymes, whereas DANRE has a medium K_m value with taPHD but a high one with hsPHD2.

The peptide NEMVE has the lowest initial rates for both taPHD and hsPHD2. Since it also has high K_m values compared to other peptides, it is possible that the slow hydroxylation of NEMVE is caused by weak binding to the active site of taPHD and hsPHD2.

B.3.5. Summary

In order to investigate enzyme-substrate interactions important in prolyl-hydroxylation by taPHD and hsPHD2, time course hydroxylation assays were carried out with five peptides corresponding to the CODD-homologous sequences of several metazoan species. Comparison of the initial rates revealed that the trends are different for taPHD and hsPHD2, suggesting that these two PHDs interact differently with the CODD-homologue peptides. The implications of these results are discussed in the next section.

B.4. Kinetic Studies in the Context of the Crystal Structures and Phylogenetic Relationships

The kinetic data for taPHD and hsPHD2 presented in Sections B.2. and B.3. were analysed with respect to two distinct interests. Firstly, it was investigated whether the trends observed in initial rates and K_m values correlate with the sequence of the CODD-homologues and their ability to form enzyme-substrate interactions observed in the crystal structures of taPHD.taODD and hsPHD2.hsCODD (Section B.4.1.). Secondly, the trends were analysed in the light of phylogenetic relationships between *Trichoplax adhaerens*, humans and the source organisms of the CODD-homologues (Section B.4.2.).

B.4.1. Rationalisation of observed trends using enzyme-substrate interactions identified in the crystal structures

Several interactions involving taODD side chains were identified by inspection of the crystal structure of taPHD.taODD (see Section A.4.). Many of these interactions have been previously identified between hsPHD2 and hsCODD (see Table A.4.7.). While equivalents for some of the interacting residues are present in the CODD-homologues used in Sections B.2. and B.3. for initial rate and K_m determinations, others are substituted by amino acids very different in size and character (see Figure B.4.1.).

```
CODD    DLDLEMLAPYIPMDD-DFQL
DANRE   LDELDSLAPYIPMHGEDFLL
STRPU    DELAMRAPYIPMGE-DFDL
TRICA   ESDLVAKAPYITMNMGDD-L
PALPU   LDEFDMRAPFIPLSN-ELLM
NEMVE   SNELQNRAPYIPPPTGDAAL
taODD EKEDYDDLAPFVPPPSFDNRL
```

Figure B.4.1. CODD-homologues used in Sections 2 and 3. The hydroxylated proline is highlighted in red. Residues undergoing side chains-side chain interactions with enzyme are shown in blue.

The significance of the identified interactions was investigated by inspecting the initial rates and K_m values for the peptides that differ significantly from hsCODD and taODD at key positions.

B.4.1.1. Interactions with $Y481_{taHIF\alpha}$ / $L559_{hsHIF\text{-}1\alpha}$

The interactions between taPHD and taODD and hsPHD2 and hsCODD are different for $Y481_{taHIF\alpha}$/$L559_{hsHIF\text{-}1\alpha}$. While both $Y481_{taHIF\alpha}$ and $L559_{hsHIF\text{-}1\alpha}$ are situated in a hydrophobic pocket, $Y481_{taHIF\alpha}$ also forms a hydrogen bond to $R208_{taPHD}$, which is not possible for $L559_{hsHIF\text{-}1\alpha}$.

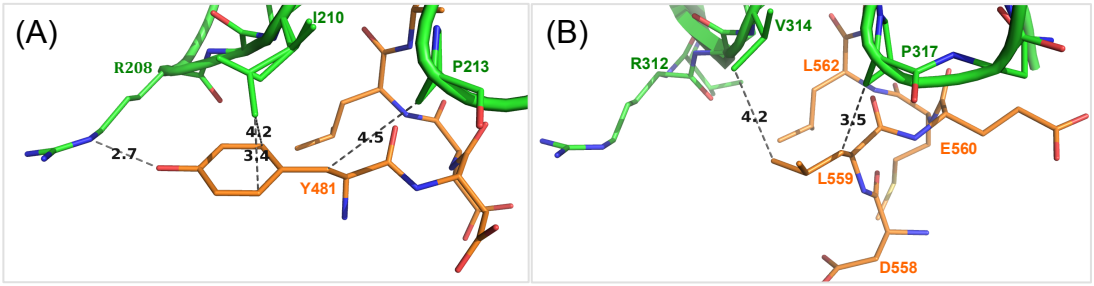


Figure B.4.2. (A) The crystal structure taPHD.taODD shows that $Y481_{taHIF\alpha}$ undergoes hydrophobic interactions with $I210_{taPHD}$ and $P213_{taPHD}$ and may form a hydrogen bond with $R208_{taPHD}$; (B) The crystal structure of hsPHD2.hsCODD (PDB ID=3HQR) shows similar interactions for $L559_{hsHIF\text{-}1\alpha}$, which is situated in a hydrophobic pocket formed by $V314_{hsPHD2}$ and $P317_{hsPHD2}$. In contrast to taPHD.taODD, however, a hydrogen bond is not possible. taODD and hsCODD are shown in orange, while taPHD and hsPHD2 are shown in green.

While $Y481_{taHIF\alpha}$ can form a hydrogen bond with $R208_{taPHD}$ and $R312_{hsPHD2}$, $I210_{taPHD}$ is closer to the hydrophobic ring of $Y481_{taHIF\alpha}$ than would be possible for $V314_{hsPHD2}$, thereby optimising hydrophobic interactions with the aromatic ring (see Figure B.4.2.). These differences between taPHD/taODD and hsPHD2/hsCODD interactions are possibly reflected in the initial rate orders of CODD-homologues for taPHD and hsPHD2 (see Section B.3.). The initial rate of taODD turnover by taPHD (6.5 µM/min) is more than twice the initial rate of hsCODD turnover (2.8 µM/min), while hsPHD2 turns over taODD and hsCODD with similar initial rates (2.8 and 2.9 µM/min, respectively). This is further supported by the observation that the peptide PAPLU, which contains a phenylalanine at the position equivalent to $Y481_{taHIF\alpha}$/$L559_{hsHIF\text{-}1\alpha}$, has a very low initial turnover rate by both enzymes (negligible during first five minutes). Unlike tyrosine, phenylalanine cannot form a hydrogen bond with $R208_{taPHD}$/$R312_{hsPHD2}$ and the planar

aromatic ring is possibly not of optimal geometry for interaction with the hydrophobic pockets of hsPHD2.

To further investigate the different enzyme-substrate interactions for taPHD/taODD and hsPHD2/hsCODD, a taODD variant with a leucine in the position of $Y481_{taHIF\alpha}$ ('taODD*') was prepared:

```
taODD          EKEDYDDLAPFVPPPSFDNRL
taODD*         EKEDLDDLAPFVPPPSFDNRL
```

The initial rates of turnover by taPHD and hsPHD2 (see Figure B.4.1.) were determined for taODD* as described in Section B.3.

	r_i(X)in µM/min	
Peptide	**X=taPHD**	**X=hsPHD2**
CODD	2.8	2.9
taODD	6.5	2.8
taODD*	2.5	2.9

Table B.4.1. Initial rates of taODD* turnover by taPHD and hsPHD2 compared to the rates of CODD, and taODD.

Comparison of the initial rates of taODD* turnover by taPHD and hsPHD2 shows that turnover of taODD* by taPHD is taking place at a rate comparable to hsCODD turnover. hsPHD2, in contrast, turns over taODD, hsCODD and taODD* at the same rates within error. These results suggest that the fast turnover of taODD by taPHD compared to other CODD-homologue peptides is mainly due to the presence of a tyrosine in taODD, which interacts the binding site of taPHD differently than a leucine residue in hsCODD and taODD*. The binding site of hsPHD2, in contrast, does not distinguish between leucine and tyrosine.

B.4.1.2. Hydrophobic interactions with $L484_{taHIF\alpha}$ /$L562_{hsHIF\text{-}1\alpha}$

The hydrophobic pockets surrounding $L484_{taHIF\alpha}$ in taPHD and $L562_{hsHIF\text{-}1\alpha}$ in hsPHD2 are very similar to each other (see Figure B.4.3.).

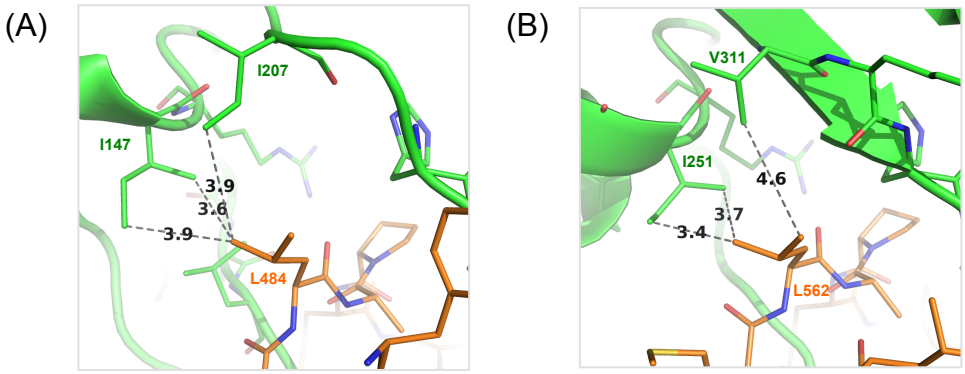


Figure B.4.3. The crystal structures taPHD.taODD (A) and hsPHD.hsCODD (B) show that $L484_{taHIF\alpha}$/$L562_{hsHIF\text{-}1\alpha}$ is situated in a hydrophobic pocket formed by $I147_{taPHD}$ /$I251_{hsPHD2}$ and $I207_{taPHD}$/$V311_{hsPHD2}$. The same colour scheme as in Figure B.4.2. applies.

Of the CODD-homologues tested in Sections B.2. and B.3, DANRE was the only other peptide to contain a leucine at the position equivalent to $L484_{taHIF\alpha}$/$L562_{hsHIF\text{-}1\alpha}$. It is therefore possible that this is the main reason why DANRE has high initial turnover rates for both taPHD (6.0 µM/min, comparable to taODD) and hsPHD2 (4.1 µM/min, highest initial rate for hsPHD2) compared to other CODD-homologues.

The other CODD-homologues contain an arginine at the equivalent position, apart from TRICA, which contains a lysine at this position. Note that both arginine and lysine can also undergo hydrophobic interactions through the aliphatic chain if oriented correctly. In order to investigate whether the presence of a lysine rather than a leucine plays a role in the low initial turnover rates of TRICA, a variant peptide containing a leucine in place of the lysine was prepared ('TRICA*'):

```
TRICA          ESDLVAKAPYITMNMGDDL
TRICA*         ESDLVALAPYITMNMGDDL
```

However, no hydroxylation of TRICA* above background levels was detected (see Appendix 3), suggesting that the presence of a lysine rather than an arginine or leucine is not the only reason for the slow turnover rates of TRICA by taPHD and hsPHD2.

B.4.1.3. Hydrophobic interactions with $P489_{taHIF\alpha}/P567_{hsHIF\text{-}1\alpha}$

$P489_{taHIF\alpha}$ is situated in a hydrophobic pocket formed by $W154_{taPHD}$ and $F493_{taHIF\alpha}$ (see Figure B.4.4.). Hydrophobic interactions between $P567_{hsHIF\text{-}1\alpha}$ and $W258_{hsPHD2}$ are also possible, although a phenylalanine residue is not present in hsCODD and no other hydrophobic residues are in proximity to form a pocket.

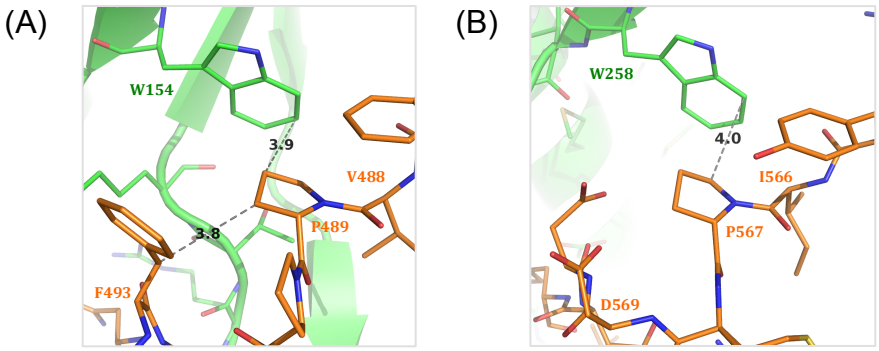


Figure B.4.4. (A) $P489_{taHIF\alpha}$ is situated in a hydrophobic pocket formed by $W154_{taPHD}$ and $F493_{taHIF\alpha}$. (B) $P567_{hsHIF\text{-}1\alpha}$ can interact with $W258_{hsPHD2}$, but a hydrophobic pocket as such is not observed.

Of the CODD-homologue peptides tested, only TRICA does not contain a proline but a threonine at the position equivalent to $P489_{taHIF\alpha}$, which could be another reason for the slow turnover of this peptide by taPHD and hsPHD2.

B.4.1.4. Hydrophobic interactions with $L497_{taHIF\alpha}$ and $L574_{hsHIF\text{-}1\alpha}$

Residues $L497_{taHIF\alpha}$ and $L574_{hsHIF\text{-}1\alpha}$ interact with taPHD/hsPHD2 in a very similar manner, both occupying a hydrophobic pocket formed by a tyrosine and two isoleucines/one isoleucine and a valine (Figure B.4.5.).

Of the CODD-homologues tested in Sections B.2. and B.3, PALPU was the only peptide that does not contain a leucine at the equivalent position. Instead, it contains a methionine.

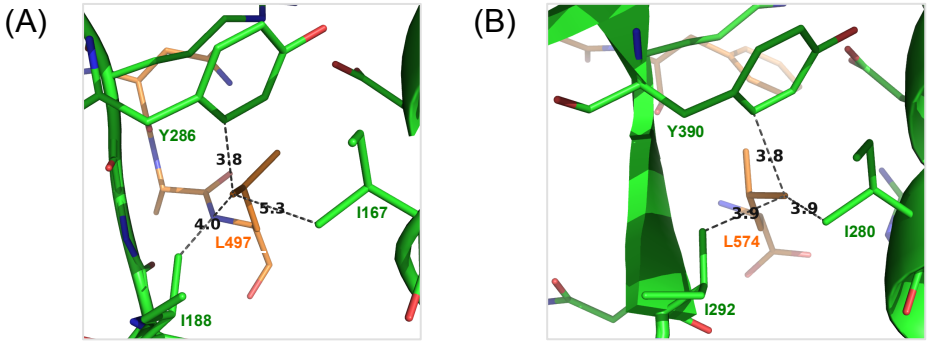


Figure B.4.5. L497$_{taHIF\alpha}$ and L574$_{hsHIF\text{-}1\alpha}$ are both situated in a hydrophobic pocket of taPHD (A) and hsPHD2 (B), formed by two isoleucines in taPHD and a tyrosine, an isoleucine and a valine in hsPHD2.

A variant PALPU peptide, containing a leucine in place of the *C*-terminal methionine (PALPU*) was prepared:

```
PALPU  LDEFDMRAPFIPLSNELLM
PALPU* LDEFDMRAPFIPLSNELLL
```

Initial rate determination for PALPU* with taPHD and hsPHD2 yielded 2.3 μM/min for taPHD and 2.2 μM/min for hsPHD2. Considering that the initial rate of turnover is negligibly small for PALPU under the same conditions, this is a significant enhancement (c.f. initial rate for CODD/taPHD is 2.8 μM/min and 2.9 μM/min for hsCODD/hsPHD2). These results suggest that the hydrophobic interactions at the *C*-terminus of the CODD-homologues are important for peptide hydroxylation.

B.4.1.5. Salt bridge between D494 $_{taHIF\alpha}$/D571$_{hsHIF\text{-}1\alpha}$ and R292$_{taPHD}$/R396$_{hsPHD2}$

A salt bridge between D494$_{taHIF\alpha}$/D571$_{hsHIF\text{-}1\alpha}$ and residue R292$_{taPHD}$/residue R396$_{hsPHD2}$ is observed in the crystal structures of taPHD.taODD and hsPHD2.hsCODD (Figure B.4.6.).

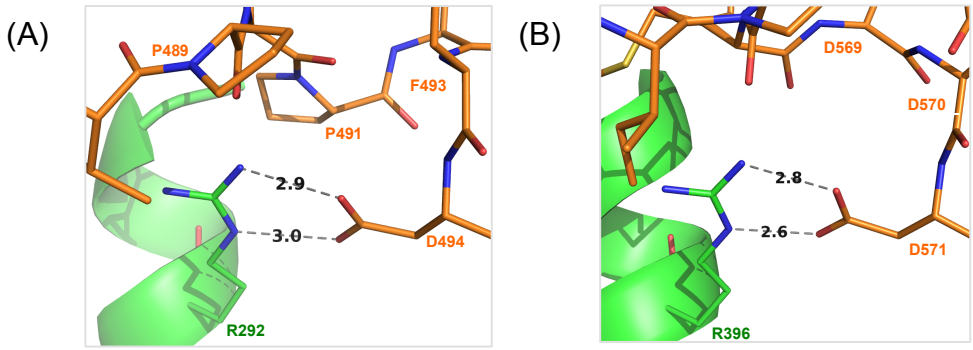


Figure B.4.6. D494$_{taHIF\alpha}$ and D571$_{hsHIF\text{-}1\alpha}$ form a salt bridge with R292$_{taPHD}$ (A) and R396$_{hsPHD2}$ (B).

PALPU contains a glutamic rather than an aspartic acid in the equivalent position, which could be too long to fit appropriately in the binding site provided by taPHD and hsPHD2.

A hsPHD2 variant R396A displayed significantly reduced CODD-activity but unaffected NODD-activity (Chowdhury et al. 2009), suggesting that this interaction is more important for hsCODD than for hsNODD binding. A taPHD R292A variant was prepared by site-directed mutagenesis (see Materials and Methods 1) and purified following the *wild-type* purification protocol (see Materials and Methods 3). Neither taODD nor hsCODD nor any of the CODD-homologue peptides were turned over by the R292A mutant (see Appendix 3), further confirming the importance of this interaction for substrate hydroxylation in taPHD and hsPHD2.

B.4.2. Consideration of phylogenetic relationships

No direct correlation between phylogenetic relationships and initial rate orders for the CODD-homologues was observed, with DANRE, a chordate sequence, having a higher initial rate with taPHD than NEMVE, a sequence from a more closely related cnidarian. Similarly, hsPHD2 turned over hsCODD and taODD at similar initial rates. It is, however, noteworthy that both the arthropod sequences TRICA and PALPU were turned over very slowly compared to other CODD-homologues tested. A third arthropod CODD-homologue, belonging to the mosquito *Anopheles gambiae* (ELDLSMRA**P**YISMSEVDDL), was tested with taPHD and hsPHD2 but did not show any detectable hydroxylation levels after overnight incubation under assay conditions (see Appendix 3).

B.4.2.1. Alignment of PHD2-homologues

A number of residues undergoing side chain-side chain interactions with taODD/hsCODD have been identified previously (Chowdhury et al. 2009). Some of the CODD-homologues tested do not contain the corresponding amino acids to fulfill all of the identified interactions with hsPHD2 and taPHD. To investigate whether the interacting enzyme residues are conserved in the PHDs of the corresponding species, an alignment of PHD2-homologues was prepared (see Figure B.4.7.).

The sequence of *Palaemonetes pugio* PHD was not available because the full genome had not been released (the CODD-homologue sequence was identified by searching an EST library (http://www.marinegenomics.org/). Instead, the sequence of *Anopheles gambiae* PHD was added to the alignment.

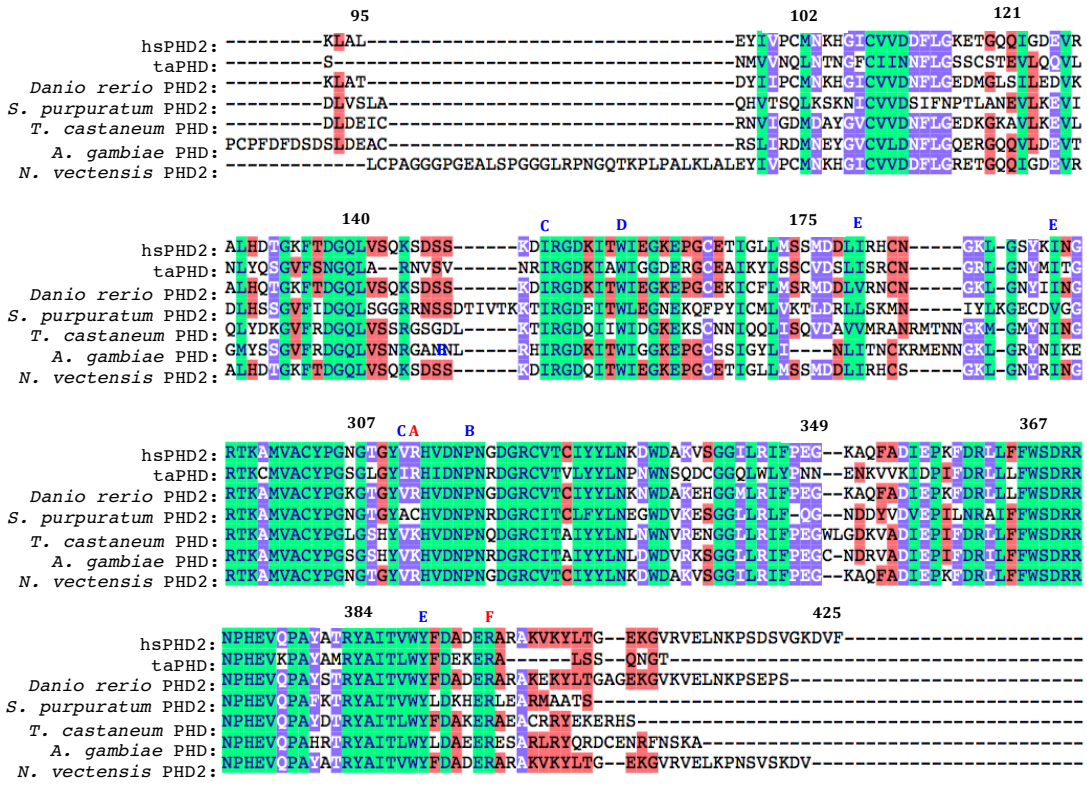


Figure B.4.7. Alignment of PHD2-homologuous sequences (see Table B.3.1. for full name of organisms). Letters mark residues with side chains interacting with taODD/hsCODD in taPHD/hsPHD2, with red letters denoting polar interactions and blue letters hydrophobic interactions (A=hydrogen bonding with Y481$_{taHIF\alpha}$; B=hydrophobic pocked binding Y481$_{taHIF\alpha}$/L559$_{hsHIF\text{-}1\alpha}$; C=hydrophobic pocket binding L484$_{taHIF\alpha}$/L562$_{hsHIF\text{-}1\alpha}$; D=hydrophobic interactions with P489$_{taHIF\alpha}$/P567$_{hsHIF\text{-}1\alpha}$; E=hydrophobic pocket binding L497$_{taHIF\alpha}$/L574$_{hsHIF\text{-}1\alpha}$; F=arginine forming salt bridge with D494 $_{taHIF\alpha}$/D571$_{hsHIF\text{-}1\alpha}$). Numbering refers to hsPHD2 residues.

In line with the high conservation observed for the PHDs (40-60% over catalytic domain for the sequences shown), most of the residues involved in the enzyme-substrate interactions in taPHD and hsPHD2 are conserved throughout the five PHD2-homologues analysed, supporting their importance in substrate binding.

B.4.2.2. Non-conserved residues

Exceptions can be seen in the PHD2 of *Strongylocentrotus purpuratum*, which contains a cysteine at the position equivalent to R208$_{taPHD}$/R312$_{hsPHD2}$ (marked as 'A' in Figure B.4.7.). This substitution, however, is unlikely to be important in substrate-enzyme

interactions because $Y481_{\mathbf{taHIF\alpha}}$ is not conserved in any of the other peptides tested. In agreement with this, only taPHD contains an isoleucine at the position equivalent to $I210_{taPHD}/V314_{hsPHD2}$, allowing for more optimal hydrophobic interaction distances (see Section B.3.1.).

Another noteworthy substitution is the alanine of *Strongylocentrotus purpuratum* PHD2, which occupies the position of $I207_{taPHD}/V311_{hsPHD2}$. In taPHD and hsPHD2, these residues are part of a hydrophobic pocket that interacts with $L484_{taHIF\alpha}/L562_{hsHIF\text{-}1\alpha}$. The peptide STRPU, however, contains an arginine at the equivalent position, for which polar interactions with the backbone and polar side chains are likely to be of greater weight than hydrophobic interactions.

The alignment of PHD2-homologues shows that the PHD-residues interacting with HIFα are highly conserved, whereas the CODD-sequences vary substantially (see Figure B.4.7.). The rapid turnover of taODD by hsPHD2 demonstrates that the LXXLAP motif is not a crucial requirement for peptide turnover. It is, at this stage, difficult to assess whether the YXXLAP motif is an evolutionary precursor of the LXXLAP motif, although experiments with a PHD-like enzyme of *Monosiga brevicollis* presented in Part C may suggest so. Given the poor conservation in the ODD-regions, however, it may be a key clue to identifying an evolutionary precursor of HIFα.

B.4.2.3. β2/β3 Finger-loop region

Compared to the catalytic domain, the middle region of the β2/β3 finger-loop of the PHDs (residues 242-250 of hsPHD2 and 138-146 of taPHD) is poorly conserved between the PHDs aligned in Figure B.4.7. Previous experiments with hsPHD2 have suggested that the loop sequence is important for hsCODD vs. hsNODD selectivity. It is

possible that the loop undergoes transient interactions with the substrate that cannot be observed in the crystal structures.

The poor conservation in the mid-loop region is remarkable in light of the high conservation of the active site region. Different sites in proteins are evolving at different rates, with some being highly conserved and others rapidly changing (Dean & Golding 2000). The mid-loop region could be an evolutionary 'hot spot' region of the PHDs. One of the factors affecting the evolutionary rate of a site is solvent accessibility (Dean & Golding 2000). This is because the side chains buried in the hydrophobic enzyme core are more constrained than the mobile solvent-exposed regions. The β2/β3 finger-loop is such a mobile region situated on the surface of hsPHD2/taPHD (see Figure B.4.7.) and has therefore a high solvent exposure in comparison to the deep-lying active site.

The evolutionary rate has also been observed to be low for regions that interact with the substrate and/or the active site (Dean & Golding 2000). In the crystal structures of hsPHD2.hsCODD and taPHD.taODD, the only mid-loop-substrate interactions observed involved the backbone atoms of the mid-loop regions, in line with the suggestion that this region may be an evolutionary 'hot spot'.

B.4.3. Summary

The enzyme-substrate interactions observed in the crystal structures of taPHD.taODD and hsPHD2.hsCODD were used to rationalise some of the kinetic data trends observed in the Sections B.2. and B.3.

It was proposed that the reason why taPHD hydroxylated taODD much faster than hsCODD is that the hydrophobic pocket of taPHD that interacts with $Y481_{taHIF\alpha}/L559_{hsHIF\text{-}1\alpha}$ is optimised to interact with a tyrosine, as present in taODD but not hsCODD. This

proposal is supported by the observation that taPHD turned over the taODD variant $Y481L_{taHIF\alpha}$ at the same initial rate as hsCODD.

The peptides TRICA and PALPU, both of which were turned over very slowly compared to the other CODD-homologues, were found to lack several residues identified to be involved in enzyme-substrate interactions.

A PALPU variant containing a leucine rather than a methionine at the position equivalent to $L497_{taHIF\alpha}/L574_{hsHIF\text{-}1\alpha}$ was turned over at initial rates similar to the initial rates of the peptide NEMVE, showing a large increase compared to the negligible rates of the *wild-type* sequence under the same conditions. This suggests that one of the reasons for the slow turnover of PALPU is the lack of a leucine to match the hydrophobic pocket of taPHD/hsPHD2 observed to bind $L497_{taHIF\alpha}$ /$L574_{hsHIF\text{-}1\alpha}$.

Interestingly, the two peptides that were turned over slowly, TRICA and PALPU, were both arthropod HIFα sequences. A third arthropod ODD sequence belonging to the *Anopheles gambiae* was not hydroxylated neither by taPHD nor hsPHD2, further suggesting a phylogenetic correlation.

Alignment of PHD2-homologues showed that most of the interacting enzyme residues are conserved, with minor variations in the sequence of the PHD2 of S*trongylocentrotus purpuratum*, which were considered of minor importance. Only taPHD contains an isoleucine in the pocket binding $Y481_{taHIF\alpha}/L559_{hsHIF\text{-}1\alpha}$. Accordingly, only taODD contains a tyrosine at the equivalent position. It may be that the evolutionary pre-cursor of the PHDs may have hydroxylated a YXXLAP rather than a LXXLAP motif.

The low conservation in the β2/β3 finger-loop region compared to the rest of the sequence led to the proposal that this region could be fast-evolving.

B.5. Investigating the Binding of NODD by Hydroxylation Assays with taPHD Variants

A NODD-type sequences was not identified in HIFα of *Trichoplax adhaerens* (Loenarz et al. 2011). *Nematostella vectensis,* the most closely related metazoan organism to *Trichoplax adhaerens* that has a sequenced genome, does not contain a NODD in the HIFα sequence either (Loenarz et al. 2011).

It has been proposed that all vertebrate HIF-1α sequences contain a NODD (Loenarz et al. 2011). Neither the exact function nor the binding mode of NODD in humans are known to date, but its appearance pattern throughout the metazoan species suggests it evolved later than CODD and is required in more complex metazoans.

B.5.1. Interaction between $R281_{hsPHD2}$ and $D412_{hsHIF\text{-}1\alpha}$ of NODD

The residue $R281_{hsPHD2}$ has previously been suggested to interact with $D412_{hsHIF\text{-}1\alpha}$ within hsNODD (Chowdhury 2008). A R281A variant of hsPHD2 shows decreased turnover of hsNODD, but not hsCODD (Chowdhury, unpublished data). This residue is conserved neither in hsPHD3 nor in taPHD, which have $L103_{taPHD}$ and $S177_{taPHD}$ at the equivalent positions, respectively. The observations that hsPHD3 does not turn over hsNODD (Epstein et al. 2001; Hirsilä et al. 2003), and that hsNODD turnover by taPHD is very slow (~20% hydroxylation only detectable in an overnight assay) support this hypothesis. The sequence alignment of hsPHD2-homologues in Figure B.4.7. shows that no arginine is present at the equivalent position in *Strongylocentrotus purpuratum* PHD, *Tribolium castaneum* PHD, and *Anopheles gambiae* PHD, which contain a serine, methionine and threonine at the corresponding positions, respectively. However, analysis of the NODD-sequences of the corresponding species shows that all of them

contain either an aspartic or a glutamic acid at the position equivalent to D412 of human HIF-1α.

```
hsNODD                      PDALTLLAPAAGDTIISLDF
T. castaneum NODD          EPDDLTHLAPVAGDVCVPLDD
A. gambiae NODD            EPDDLTHLAPTAGDACIPLE
S. purpuratum NODD         VEEKLAYLAPTAGDVMIELDP
```

The above suggests that while the interaction between $R281_{hsPHD2}$ and $D412_{hsHIF\text{-}1\alpha}$ is important, the species above may either rely on other interactions for a rapid NODD turnover, or that the hydroxylation of NODD in these species is intrinsically slow.

B.5.2. The clinically observed hsPHD2 mutation P317R

The clinically observed mutation P317R in hsPHD2 is associated with familial erythrocytosis (an inherited disorder also called familial polycythemia, characterised by an increased number of red blood cells) (Percy et al. 2007). The mutation P317R decreases the levels of both hsCODD and hsNODD turnover. It is thought that the reason for this is a narrowing of the entrance to the active site of hsPHD2. *In vitro*, the P317R substitution in hsPHD2 was shown to decrease hsNODD turnover more than hsCODD turnover (Chowdhury, unpublished data). $P317_{hsPHD2}$ is conserved in taPHD as $P213_{taPHD}$.

B.5.3. hsNODD, hsCODD and taODD turnover by the taPHD variants S177R and P213R

The taPHD variants S177R and P213R were prepared to investigate the effect of the positions equivalent to $R281_{hsPHD2}$ and $P317_{hsPHD2}$ on hsNODD, hsCODD and taODD turnover. The constructs were prepared by site-directed mutagenesis, expressed in *E. coli* and purified following the same protocols as for the *wild-type* expression and purification (see Materials and Methods 1-3). End point hydroxylation assays were carried out in 50 mM Tris-HCl at pH=7.5 with 3.5 μM enzyme, 100 μM peptide, 300 μM

2OG, 50 μM Fe(II) and 4 mM ascorbate. The fraction of peptide hydroxylated after 3 hours was determined using MALDI-TOF MS. The data were analysed as described in Section B.3, with background +16 Da peak intensities subtracted from the '+enzyme' sample values.

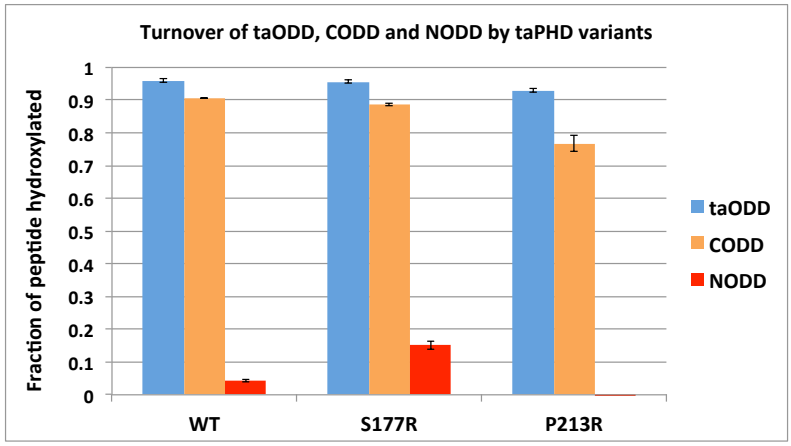


Figure B.5.1. taODD, hsCODD and hsNODD hydroxylation by WT taPHD and taPHD variants after 3 hours, determined by MALDI-TOF MS (n=3; error bards show standard deviation).

hsNODD turnover by *wild-type* taPHD was intrinsically slow and yielded only ~4% of hydroxylated peptide after 3 hours, whereas taODD was turned over 90% and hsCODD to over 90% in the same time span, consistent with the hypothesis that NODD evolved after CODD (Loenarz et al. 2011).

Introducing an arginine into taPHD in the S177R variant increased the hsNODD hydroxylation level by a factor of 3.5 to ~15%, while the level of taODD and hsCODD turnover remained the same within error as with the *wild-type* enzyme under the experimental conditions used. Despite this, the hsNODD turnover of the S177R mutant is still low compared to taODD and hsCODD, suggesting that the binding of hsNODD by arginine is only one of many interactions with hsNODD that taPHD cannot support.

The *Trichoplax adhaerens* equivalent of the clinically observed hsPHD2 mutation did not display any hsNODD hydroxylation above background levels, while the levels of taODD and hsCODD hydroxylation were still in the saturated regime (70%-100% turnover). This result support observations made with hsPHD2 and further confirms that the clinically observed P317R mutation of hsPHD2 is likely to affect hsNODD hydroxylation more significantly than hsCODD hydroxylation.

B.6. Summary of Section B

In this Section, the biochemical properties of taPHD were investigated and compared to hsPHD2. Using non-denaturing mass spectrometry, it was shown that similarly to hsPHD2, taPHD forms an unusually stable Fe(II)-2OG complex, with a half-life greater than 24 hours. These findings support the hypothesis that the formation of an unusually stable Fe(II)-2OG complex is an intrinsic property of the PHDs and is required for their function as oxygen sensors.

The extent of coupling between turnover of 2OG to succinate and peptide hydroxylation was determined. At ~1.2:1, the calculated ratio of the initial rates of succinate production to peptide hydroxylation is very close to the coupling ratio of 1:1 determined for hsPHD2 previously (Flashman et al. 2010).

Amino acid analysis of the hydroxylated taODD has shown that the product of prolyl-hydroxylation by taPHD is (4*R*)-hydroxy-prolyl taODD, homologous to *R*-hydroxy-prolyl hsCODD with hsPHD2.

Apparent K_m values were determined with taODD, hsCODD and five peptides corresponding to metazoan HIFα sequences homologous to hsCODD with taPHD and hsPHD2. The trends observed in the K_m data were analysed together with the initial rates of hydroxylation determined for taODD, hsCODD and the CODD-homologue peptides and rationalised using the crystal structures of taPHD and hsPHD2. Different trends were observed for taPHD and hsPHD2, suggesting that the way these two enzymes interact with substrates is also different to a certain extent. Inspection of the sites of enzyme-substrate interaction identified in Section A yielded plausible explanations for many of the observations. It was proposed that the main reason for the

fast turnover of taODD compared to hsCODD by taPHD but very similar initial rates for both of these peptides by hsPHD2 is due to different interactions in the hydrophobic pocket biding $Y481_{\mathbf{taHIF\alpha}}/L559_{\mathbf{hsHIF\text{-}1\alpha}}$. This suggestion is supported by initial rates of hydroxylation determined for a taODD Y481L variant, which shows a drop in the turnover rate compared to taODD with taPHD, but not with hsPHD2. The zebra fish CODD sequence DANRE was found to be turned over rapidly compared to the other peptides by both taPHD and hsPHD2. It was speculated that this may have to do with the fact that of all CODD-homologues tested, DANRE was the only peptide to contain a leucine at the position equivalent to $L484_{\mathbf{taHIF\alpha}}/L562_{\mathbf{hsHIF\text{-}1\alpha}}$, which interacts with a hydrophobic pocket in taPHD and hsPHD2. Two of the CODD-homologues, TRICA and PALPU, were turned over very slowly by both enzymes. Both of these peptides were found to lack key residues involved in enzyme-substrates identified in Section A. Interestingly, both peptides also belonged to species within the arthropod phylum.

An alignment of hsPHD2 and taPHD sequences to PHD2-homologues from other metazoan species has shown that the middle Section of the β2/β3 finger-loop region is poorly conserved compared to the catalytic domain. It was proposed that this section is an evolutionary 'hot spot'. The differing mid-loop regions of taPHD and hsPHD2 could contribute to the different trends in K_m values and initial rates observed with both enzymes. Although such interactions are not visible in the crystal structures, they could be of transient nature and occur at some stage throughout the catalytic cycle.

The binding mode of hsNODD was investigated using two taPHD variants. taPHD S177R was prepared to test for the importance of R281 in hsPHD2, which has been suggested to bind to D412 of hsNODD but is not conserved in taPHD. Introducing the arginine at the equivalent position in taPHD has increased hsNODD turnover levels by a factor of 3.5. taPHD P213R was prepared to reproduce the effect of the clinical hsPHD2 mutation P317R, which has been shown to decrease the rate of hsNODD hydroxylation more than

hsCODD hydroxylation. No turnover of hsNODD by taPHD P213R was detected, suggesting that the observations made with hsPHD2 may be applicable to other PHDs.

Overall, conservation of biochemical properties between taPHD and hsPHD2 were observed. These properties, such as the formation of a stable Fe(II)-2OG complex, could therefore be key requirements for the function of the PHDs as oxygen sensors. Despite the structural (see Part A) and biochemical conservation between taPHD and hsPHD2, differences were observed in the initial turnover rates and K_m values for taODD, hsCODD and CODD-homologues. Apart from the $Y481_{\mathbf{taHIF\alpha}}$/$L559_{\mathbf{hsHIF\text{-}1\alpha}}$ binding site, the available evidence suggests that the enzyme-substrate interactions between taPHD and taODD and hsPHD2 and hsCODD are very similar. It is therefore proposed that an important basis of different enzyme-substrate interactions could be transient interactions with the mid-loop region, which varies significantly for the PHDs and could represent an evolutionary 'hot spot' within the PHDs.

B.7. References

Choudhary, A., Pua, K. H. & Raines, R. T., 2011. Quantum mechanical origin of the conformational preferences of 4-thiaproline and its S-oxides. *Amino acids*, 41(1), pp.181-6. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2910815&tool=pmcentrez&rendertype=abstract [Accessed March 8, 2012].

Chowdhury, R., 2008. *Structural and Mechanistic Studies on Prolyl Hydroxylases*. Elsevier, Inc.

Chowdhury, R. et al., 2009. Structural basis for binding of hypoxia-inducible factor to the oxygen-sensing prolyl hydroxylases. *Structure (London, England : 1993)*, 17(7), pp.981-9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19604478 [Accessed March 9, 2012].

Cornish-Bowden, A., 1979. *Fundamentals of enzyme kinetics* 3rd ed., Portland Press.

Dean, A.M. & Golding, G.B., 2000. Enzyme Evolut ion Explained (Sort Of) ANTONY M. DEAN. , 17, pp.6-17.

Eberhardt, E. S., Panasik, N. & Raines, R. T., 1996. Inductive Effects on the Energetics of Prolyl Peptide Bond Isomerization: Implications for Collagen Folding and Stability. *Journal of the American Chemical Society*, 118(49), pp.12261-12266. Available at: http://pubs.acs.org/doi/abs/10.1021/ja9623119.

Epstein, A. C. et al., 2001. C. elegans EGL-9 and mammalian homologs define a family of dioxygenases that regulate HIF by prolyl hydroxylation. *Cell*, 107(1), pp.43-54. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11595184.

Flashman, E. et al., 2010. Evidence for the slow reaction of hypoxia-inducible factor prolyl hydroxylase 2 with oxygen. *The FEBS journal*, 277(19), pp.4089-99. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20840591 [Accessed April 28, 2012].

Hausinger, R. P., 2004. *FeII/alpha-ketoglutarate-dependent hydroxylases and related enzymes.*, Available at: http://www.ncbi.nlm.nih.gov/pubmed/15121720 [Accessed March 29, 2012].

Hirsilä, M. et al., 2003. Characterization of the human prolyl 4-hydroxylases that modify the hypoxia-inducible factor. *The Journal of biological chemistry*, 278(33), pp.30772-80. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12788921 [Accessed March 15, 2012].

Hon, W.-Ching et al., 2002. Structural basis for the recognition of hydroxyproline in HIF-1 a by pVHL. *Nature*, 417(June), pp.975-978.

Loenarz, C. et al., 2009. Evidence for a stereoelectronic effect in human oxygen sensing. *Angewandte Chemie (International ed. in English)*, 48(10), pp.1784-7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19180614 [Accessed March 16, 2012].

Loenarz, C. et al., 2011. The hypoxia-inducible transcription factor pathway regulates oxygen sensing in the simplest animal, Trichoplax adhaerens. *EMBO reports*, 12(1), pp.63-70. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21109780 [Accessed April 19, 2012].

McDonough, M. A. et al., 2006. Cellular oxygen sensing: Crystal structure of hypoxia-inducible factor prolyl hydroxylase (PHD2). *Proceedings of the National Academy of Sciences of the United States of America*, 103(26), pp.9814-9. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1502536&tool=pmcentrez&rendertype=abstract.

McNeill, L. A. et al., 2005. Hypoxia-inducible factor prolyl hydroxylase 2 has a high affinity for ferrous iron and 2-oxoglutarate. *Molecular BioSystems*, 1(4), p.321. Available at: http://xlink.rsc.org/?DOI=b511249b [Accessed May 5, 2012].

Mecinović, J. et al., 2009. 2-Oxoglutarate analogue inhibitors of prolyl hydroxylase domain 2. *Bioorganic & medicinal chemistry letters*, 19(21), pp.6192-5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19775891 [Accessed April 28, 2012].

Nic, M., Jirat, J. & Kosata, B., 2006. IUPAC Compendum of Chemical Terminology.

Percy, M. J. et al., 2007. A novel erythrocytosis-associated PHD2 mutation suggests the location of a HIF binding groove. *Blood*, 110(6), pp.2193-6. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1976349&tool=pmcentrez&rendertype=abstract [Accessed April 6, 2012].

Part C: Studies of a PHD-like Enzyme from the Protist *Monosiga brevicollis*

C.1. Previously Described PHD-like Enzymes and mbP4H

C.1.1. *Dictyostelium discoideum* Skp1 hydroxylase ddP4H

The eukaryote *Dictyostelium discoideum*, commonly referred to as slime mold, is an amoeba living in soil and feeding on yeast and bacteria (Raper 1935). Under starvation, the solitary amoebae aggregate to a multicellular slug. This slug then migrates through the soil until it reaches the surface and culminates into a fruiting body west (West 2003; Bonner & Lamont 2005). The spores can get dispersed to distant locations with better food availability. The transition to the culmination stage is triggered by a number of environmental factors, one of which is oxygen concentration. To the slug, an increase in environmental oxygen levels suggests that it has reached the surface, where the spores can be distributed by wind when released (Singleton et al. 2006; West et al. 2007).

One mechanism of oxygen sensing in *Dictyostelium discoideum* is proposed to proceed via hydroxylation of Skp1, an SCF-class subunit of E3 ubiquitin ligases. This hydroxylation is carried out by a Skp1 hydroxylase (ddP4H; from Prolyl-4-Hydroxylase) (Willems et al. 2004; van der Wel et al. 2005). The SCF complex is related to the VCB-type E3 ubiquitin ligase that targets HIFα in animals. ddP4H is homologous to the mammalian HIF-prolyl-hydroxylases and shares many biochemical properties with the PHDs. The sequence identity of ddP4H with the catalytic domain of hsPHD2 (residues 181-426) is 38.7%. Nevertheless, ddP4H hydroxylates neither hsCODD nor hsNODD nor any of the CODD-homologs introduced in Section B.3, which suggest functionality

distinct from the PHDs. It has been proposed that modification of the ubiquitin ligase itself rather than one of its substrates is an ancient precursor to the HIF-pathway (C. M. West et al. 2010).

C.1.2. *Perkinsus olseni* HPH

The protozoan marine parasite *Perkinsus olseni* infects mollusks including clams and oysters. Recently, *HPH2* (from HIF-Prolyl-Hydroxylase), a gene homologous to *EGLN1* (hsPHD2) has been identified in *Perkinsus olseni* (Leite et al. 2008). *HPH2* was found to be upregulated in the presence of the hypoxia mimic cobalt chloride (Leite et al. 2008). It has been suggested that *HPH2* plays a role in the pathogenicity of *Perkinsus olseni.* It has, however, never been shown what the substrate of poHPH is or that the reaction it catalyses is indeed hydroxylation. Given that one of the Fe(II)-coordinating histidine residues is swapped for arginine (see Figure C.1.1.), it is reasonable to doubt that poHPH is a prolyl-hydroxylase, although it would not be impossible.

C.1.3. mbP4H, a PHD-like enzyme of *Monosiga brevicollis*

Monosiga brevicollis is a marine protist of the class choanoflagellata, various species of which are abundant in most geographical regions. Choanoflagellates are considered to be the closest unicellular relatives to metazoans, with some species forming colonies for some or all of their lifespans, while others, including *Monosiga brevicollis*, live a strictly solitary life (see Introduction).

Monosiga brevicollis is particularly interesting following the publication of its nuclear genome (King et al. 2008). Simple BLAST-searches using the hsPHD2 amino acid sequence yield a predicted protein sequence that is closer to the metazoan PHDs than poHPH or ddP4H. In anticipation of homology to the PHDs, this sequence was labelled

P4H (from Prolyl-4-Hydroxylase). As the biology of choanoflagellates is still rather poorly studied, the relevance of oxygen sensing in *Monosiga brevicollis* is unclear. Studying mbP4H may not only reveal more about the choanoflagellates, but importantly may yield crucial hints about the origin of the 2OG oxygenases as oxygen sensors. For this purpose, a synthetic gene covering the residue range 131-485 of the 677-residue predicted protein sequence was obtained (www.invitrogen.com). This sequence was based on sequence alignments, disorder predictions and exon/intron predictions and was thought to represent the true protein sequence more accurately than the predicted 677-residue sequence (Loenarz 2010).

C.1.4. Sequence alignments of PHDs and PHD-like sequences

C.1.4.1. Comparison of the sequences of hsPHD2, poHPH, ddP4H and mbP4H

In order to compare key features of the three PHD-like proteins poHPH, ddP4H and mbP4H, a sequence alignment to hsPHD2 was prepared using GENEDOC (Nicholas et al. 1997).

The conservation between the hsPHD2 and the PHD-like sequences is in general low compared to the conservation between the sequences of PHDs of different animal species (see Section B.4.2.). While poHPH contains a PHD-MYND motif, only one of the two Fe(II)-coordinating histidines is conserved. mbP4H, in contrast, does not possess a MYND sequence but does show considerable conservation with hsPHD2 in the substrate-enclosing β2/β3 finger-loop region and has conserved active site iron-coordinating histidines. It does, however, contain two long inserts (see Section 1.4.2. for details).

The *Dictyostelium discoideum* Skp1 hydroxylase also contains the key active site residues (residues H191, D193, H266) and the 2OG-binding arginine (R276) but shows poor conservation with hsPHD2 and taPHD in the β2/β3 finger-loop region and a shortened *C*-terminus, thereby lacking a key substrate-binding arginine (equivalent to R396 in hsPHD2).

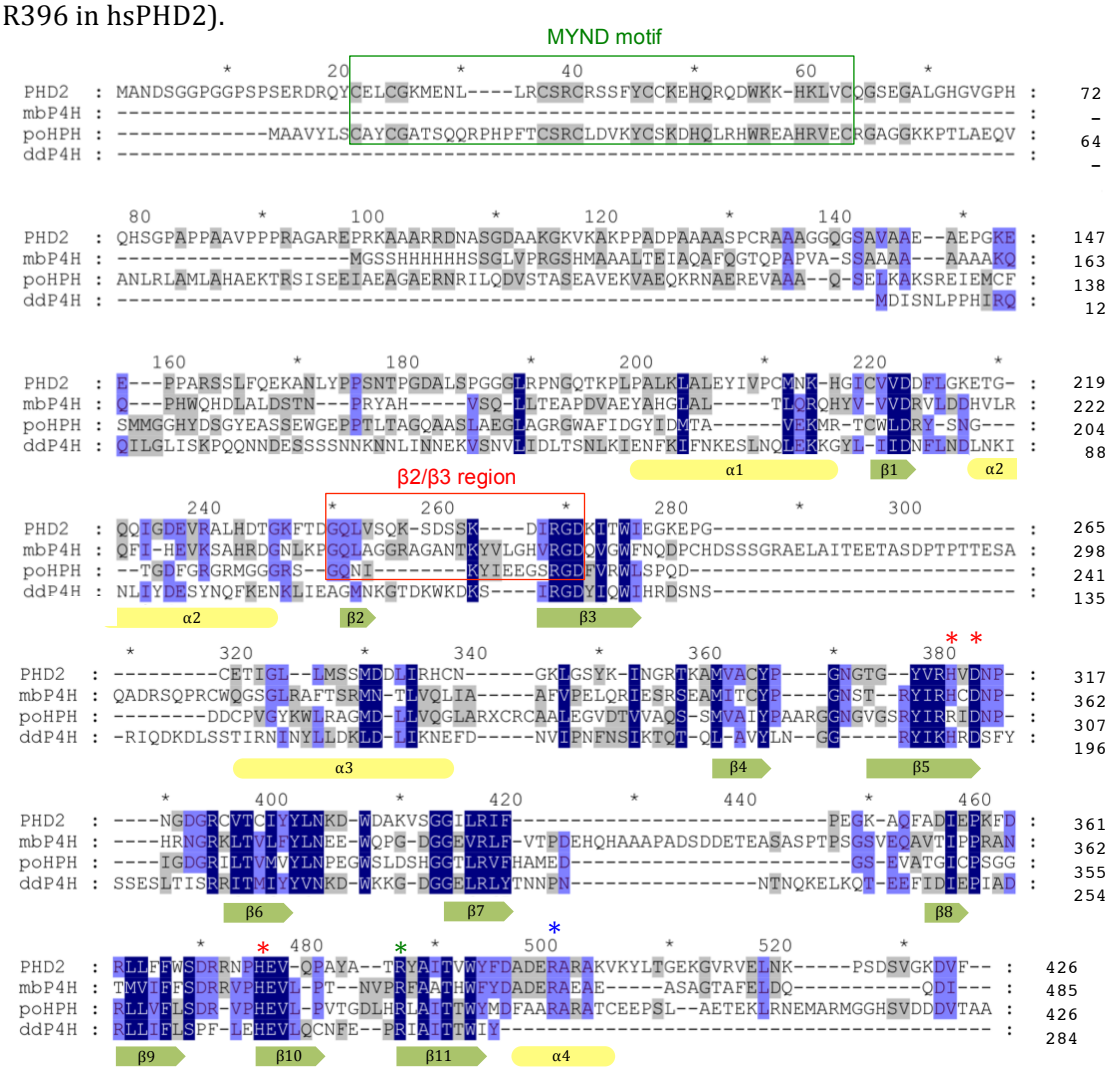


Figure C.1.1. Alignment of the PHD-like sequences mbP4H (JGI 27914), poHPH (GI 154432878) and ddP4H (GI 54660734) to hsPHD2 (GI 13489073). The MYND domain is highlighted in green and the substrate-enclosing loop is highlighted in red. The key Fe(II)-coordinating histidines and aspartic acid are marked with a red asterisk. The 2OG-binding arginine is marked with a green asterisks. The substrate-binding *C*-terminal arginine is marked with a blue asterisk. The alignment shows high conservation of the DSBH β-sheet regions (βI/4-βVIII/11) for all four sequences. Only the mbP4H sequence seems to contain a β2/β3 finger-loop region similar to hsPHD2. In hsPHD2, this region has been suggested to interact with the substrate (Flashman et al. 2007). The MYND domain, the function of which is not known to this date, is only conserved in poHPH. The *C*-terminal arginine demonstrated to interact with the HIFα substrate in hsPHD2 is missing in ddP4H. Later experiments have shown that the *N*-terminal methionine residue of mbP4H is cleaved (see Section C.2.1.).

The hsPHD2 MYND domain, the exact function of which is unknown, is considered to be a conserved feature of the metazoan PHD2-homologues (see Section A.1.). The absence of this domain in the mbP4H sequence therefore clearly distinguishes this enzyme from hsPHD2 and taPHD. The absence of the MYND-domain, however, does not imply that mbP4H is not involved in oxygen sensing, given that truncated taPHD constructs show no difference in enzymatic activity compared to the full length constructs (see Appendix 3). Furthermore, it has been shown that the MYND domain is not essential for hypoxia sensing by PHD in *Caenorhabditis elegans* (Shao et al. 2009). ddP4H also lacks the MYND-domain, but is probably involved in oxygen sensing via a route different to the HIF-pathway (see Introduction). In a way similar to ddP4H, mbP4H could also be involved in oxygen sensing by acting on a non-HIF substrate.

C.1.4.2. Sequence comparison of hsPHD2, taPHD and mbP4H

Interesting features of the mbP4H sequence are the two long inserts compared to hsPHD2 and taPHD (37 and 26 amino acids, residues 273-310 and 391-417, respectively; see Figure C.1.2.). Neither ddP4H nor poHPH have comparable inserts. Secondary structure predictions suggest that these inserts are flexible. A model of mbP4H based on the crystal structures of hsPHD2 and taPHD (see Figure C.1.4; MODELLER) suggests that both inserts form loops that are positioned along the substrate-binding groove.

Although not observed in the PHDs, flexible loop regions between βIV and βV of the jelly-roll motif are present in some other 2OG-oxygenases, for example the human FIH1 (residues 220–259), BBOX1 (residues 233-327) and PAHX (residues 223–233) (McDonough et al. 2005) (see Figure C.1.3.).

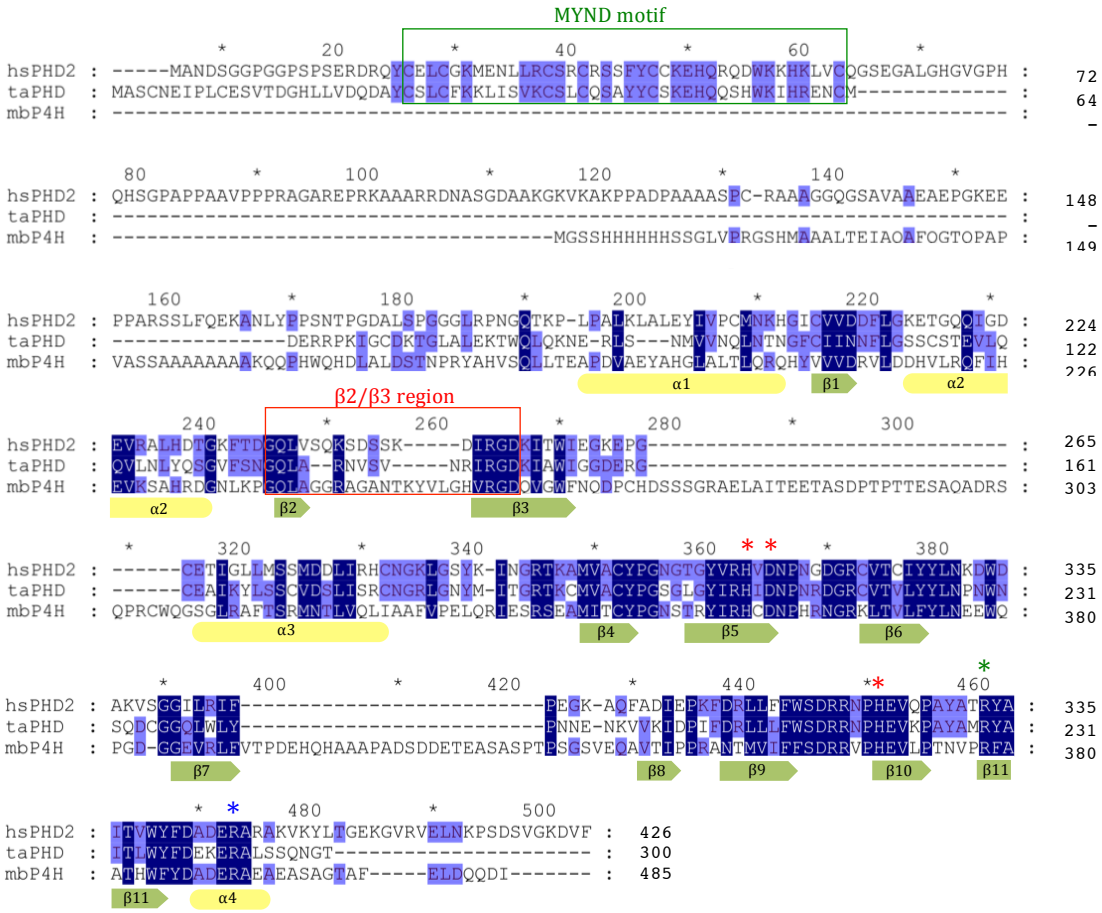


Figure C.1.2. Protein sequence alignment of hsPHD2, taPHD and mbP4H. The same features as in as in Figure C.1.1. are highlighted.

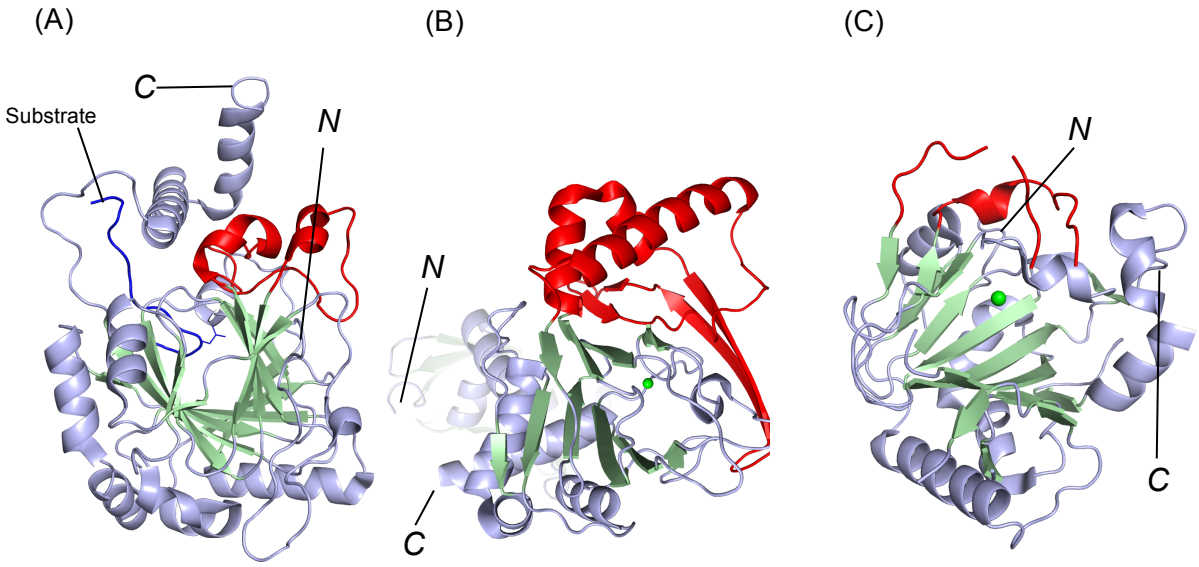


Figure C.1.3. Views from crystal structures of human FIH (A), BBOX1 (B) and PAHX (C). The βIV/βV inserts are highlighted in red.

In mbP4H, the residues between positions 396 and 432 are predicted to correspond to the βIV/βV-region (β6/β7). This region contains the second insert (26 amino acids; see Figure C.1.2.). Both inserts of mbP4H could be involved in substrate recognition and binding in addition to the β2/β3 finger-loop region. Their predicted presence suggests that the interactions between the substrate and the enzyme could be more complex than observed between the PHDs and the HIFα. Since taPHD has several disordered residues in the βIV/βV region (see Section A.4.), it could represent an evolutionary link between hsPHD2 and mbP4H. This is supported by the observation that the sequence in the βIV/βV region of taPHD differs from the PHD-sequences of other species shown in Figure B.4.7. According to secondary structure predictions based on taPHD and hsPHD2 structures, the βIV/βV insert of mbP4H is flexible and solvent-exposed. Potentially, the βIV/βV region could also represent a rapidly evolving region of mbP4H. In PHDs, this region has evolved into a more rigid, short turn.

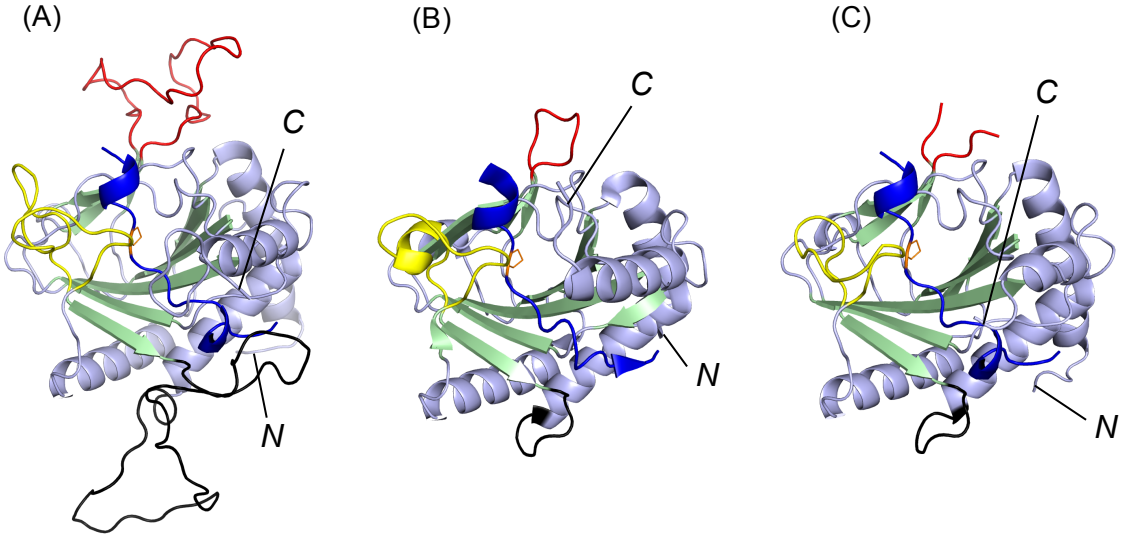


Figure C.1.4. Model of mbP4H with taODD (A) and crystal structures of hsPHD2 (B) and taPHD (C). The putative *N*- and *C*-terminal inserts in mbP4H are coloured black and red, respectively. The substrate-enclosing β2/β3 region is highlighted in yellow. The substrate, taODD in the case of mbP4H and taPHD and hsCODD in the case of hsPHD2, is highlighted in blue. The hydroxylated prolyl-residue is shown in orange.

C.1.5. Summary

Unlike poHPH and ddP4H, all mbP4H contains equivalents of all key catalytic residues identified in hsPHD2 and taPHD: 2OG- and iron-binding residues, a substrate-binding *C*-terminal arginine and a substrate-enclosing loop region. This suggests that mbP4H is more closely related to the PHDs that poHPH and ddPHD and may even accept ODD substrates.

C.2. 2OG Turnover Assays and Substrate Screening with mbP4H

C.2.1. His-Tag Purification

mbP4H was expressed in *E. coli* using a pET-28a vector and purified using a His-affinity column, followed by gel filtration (see Materials and Methods 2-3 for expression and purification protocols).

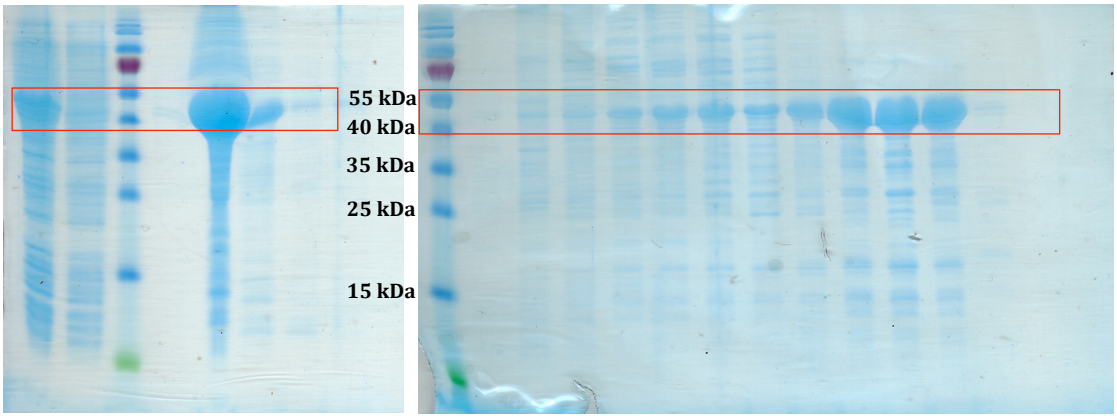


Figure C.2.1. SDS-PAGE gels of His-tag purification (A) and gel filtration (B) of mbP4H. (A) From left to right: Cell lysate; His-column flowthrough; protein ladder; His-column wash; elution fractions 9-11. Although the calculated mass of the 6xHis-tagged mbP4H protein is 41.001 kDa, ESI-MS experiments have yielded a mass of 40.078 kDa, suggesting that the *N*-terminal methionine is cleaved after expression in *E. coli*. A series of bands of matching mass, highlighted in red, is visible in the cell lysate and fractions. Fractions 9 and 10 were concentrated to 50 mg/mL and applied on a 300 mL S75 gel filtration column equilibrated in 50 mM Tris-HCl at pH=7.5. (B) From left to right: Protein ladder; fractions 1-14, collected upon peak detection by UV absorption of the eluate. The lower molecular mass bands in the gel filtration fractions are possibly degradation products of mbP4H. Fractions 8-11 were combined and concentrated to 20 mg/mL.

No aggregation or precipitation of the recombinant protein was observed at concentrations of up to 50 mg/mL in 50 mM Tris-buffer (pH=7.5). Expression at 19° C for 6 hours yielded 10 mg of protein per litre medium. The purity of the protein after a two-step purification was estimated at >95% by SDS-PAGE gel (see Figure C.2.1.). The predicted identity of the purified protein was supported by ESI-MS (see Figure C.2.2.).

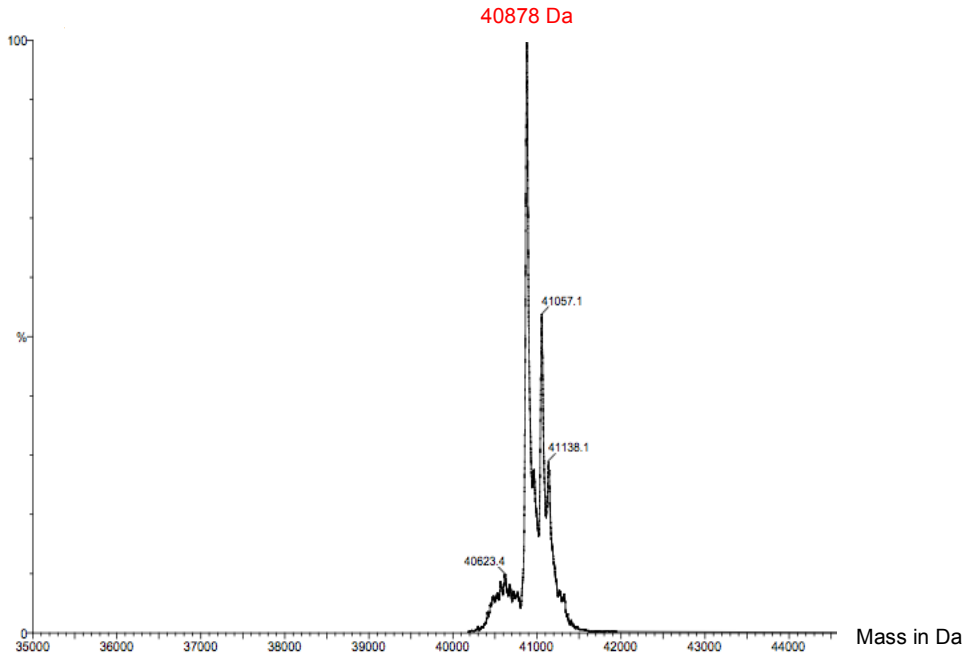


Figure C.2.2. ESI-MS spectrum of mbP4H. The calculated mass of mbP4H with the *N*-terminal methionine cleaved is 40878 Da.

C.2.2. Uncoupled 2OG-turnover by NMR

Since the native substrate of mbP4H has not been identified, the activity of the enzyme was initially assessed based on uncoupled turnover of 2OG to succinate. This was carried out in the same way as for taPHD (see Section A.2.). 2OG (final concentration of 300 µM) was added to a 3 mm NMR tube containing a 35 µM solution of enzyme, 4 mM ascorbate and 50 µM Fe(II). The tube was centrifuged briefly prior to insertion into the NMR-machine. The decrease in 2OG and production of succinate were monitored by taking a series of ^{1}H-NMR spectra at defined time intervals. While the intensity of the triplet caused by the β–protons of 2OG (2H) was decreasing, a new singlet, caused by the protons of succinate (4H) could clearly be seen appearing and gaining area with time. A plot of the integral of the 2OG peak and half the integral of the succinate peak over time shows that while the 2OG integrated signal decreases, the integrated signal of succinate increases (see Figure C.2.3.).

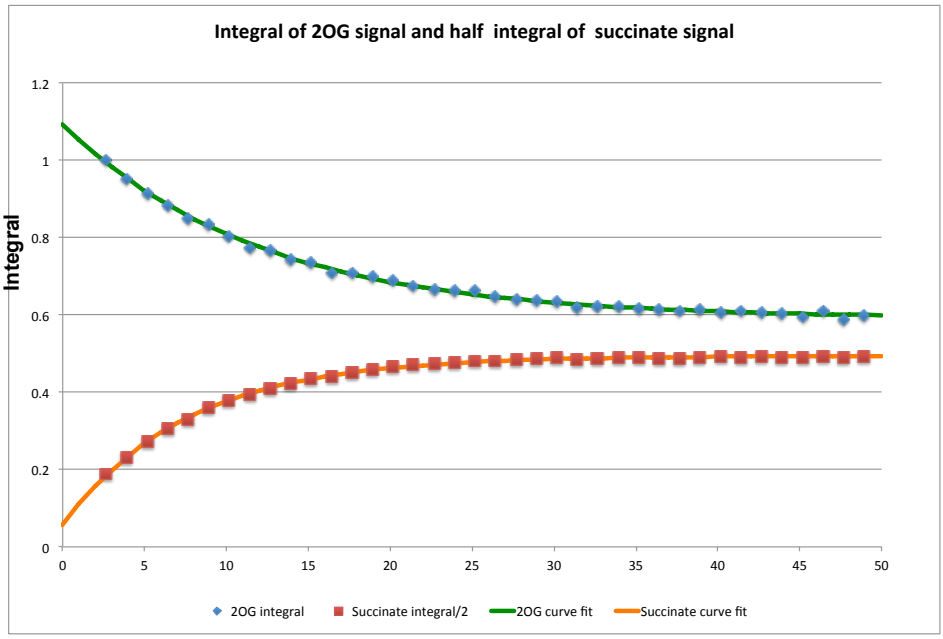


Figure C.2.3. A plot of the peak integrals of 2OG and succinate as a function of time. For this experiment, 300 µM 2OG was incubated with 35 µM mbP4H in the presence of 4 mM ascorbate and 50 µM Fe(II). ^{1}H-NMR spectra of the reaction mixture were taken at regular time intervals. The signals for the 2OG β–protons (2H) and four equivalent succinate protons (4H) were integrated. The succinate integral was divided by two for comparison with the 2OG integral. The data were fitted by the equation $y=(y_0-y_\infty)\times\exp(-K\times x)\pm y_\infty$ (see Table C.2.1.). Note that while an exponential fit can be used to describe first-order decay processes, it does not reflect on the real reaction order. Neither the reaction order nor the exact mechanism of this enzymatic conversion are known to this date.

	y_0	y_∞	K
2OG fitting curve	1.093±0.00512	0.5912±0.00198	0.08356±0.0016
Succinate fitting curve	0.5646±0.00378	0.4942±0.00066	0.13150±0.0015

Table C.2.1. Parameters of the fitting curves for the 2OG and succinate integrated peak data shown in Figure C.2.3.

Integration of the peaks and quantification of 2OG and succinate concentrations based on calibration experiments with solutions containing known 2OG and succinate concentrations (see Section A.2.) yielded data describing the variation of concentration of both species with time.



Figure C.2.4. The variation of 2OG and succinate concentration with time was calculated from data presented in Figure C.2.3. (see Section A.2. for a description of the calculation). After 40 minutes, the conversion stalls at around 135 µM succinate and 165 µM 2OG, which corresponds to a conversion of 45%.

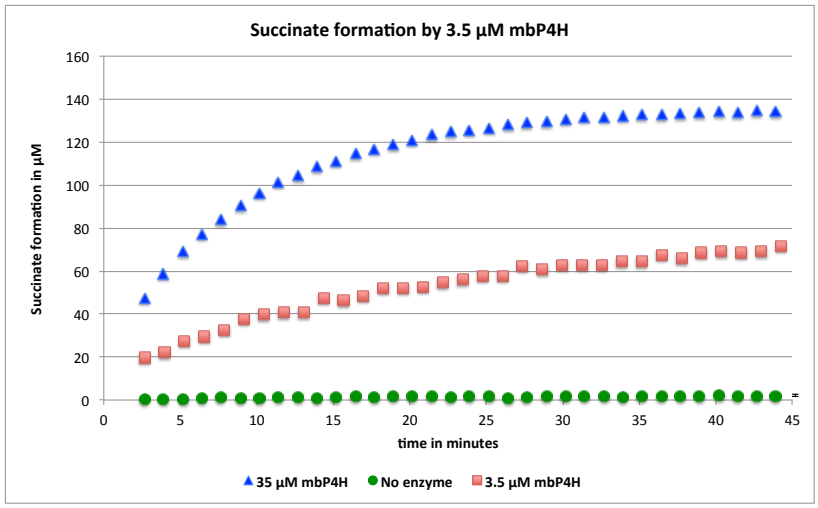


Figure C.2.5. Production of succinate in uncoupled turnover of 2OG by mbP4H, determined by ^{1}H-NMR at 35 µM and 3.5 µM enzyme and without added enzyme. The rate of succinate production in the absence of enzyme is negligible, suggesting that the formation of succinate is caused enzymatically.

To determine whether the 2OG turnover was enzymatically caused, time courses were carried out at 3.5 µM mbP4H. Comparison of initial rates of succinate production (see Table C.2.2.), determined as described in Section A.2, shows that the rate of succinate production depends on the concentration of mbP4H.

[mbP4H] / µM	r_i / µM/min
0	0
3.5	5.68
35	14.5

Table C.2.2. Initial rates of succinate production at different concentrations of mbP4H.

Furthermore, the uncoupled 2OG turnover experiments with 3.5 µM enzyme show that mbP4H turns over 2OG to succinate in the absence of a peptide at a rate ~4.7 times higher than the taPHD construct studied in Section A.2. This could either be an accurate reflection of the biochemical properties of mbP4H, or caused by varying amounts of active enzyme. These results show that the recombinant mbP4H enzyme prepared was active and could be used for substrate screening.

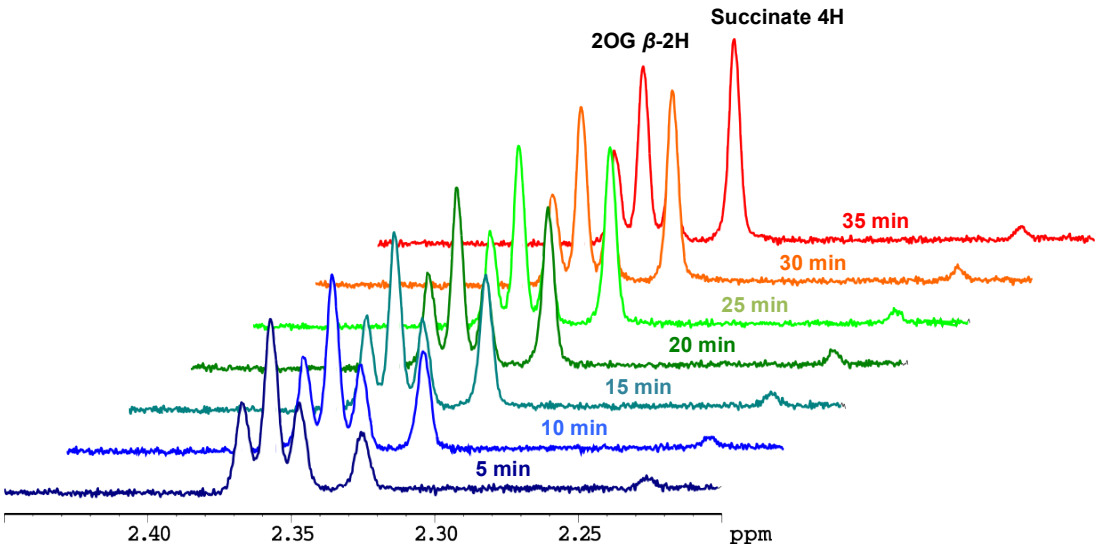


Figure C.2.6. Stacked ^{1}H-NMR spectra of the mbP4H 2OG turnover reaction mixture (3.5 µM mbP4H, 300 µM 2OG, 50 µM Fe(II), 4 mM ascorbate, 50 mM deuterated Tris-DCl pH-7.5), close-up of 2OG *β*-protons and succinate protons.

C.2.3. Substrate screening by MALDI-TOF MS

Once the activity of mbP4H had been confirmed, the set of ODD peptides of various species and their variants as used in Section B.3. (see Table C.2.3.) was tested for modifications by mbP4H using MALDI-TOF MS. The same assay conditions as described in Section B.3. were used (3.5 µM enzyme, 300 µM 2OG, 50 µM Fe(II), 4 mM ascorbate,

100 μM peptide, Tris-HCl at pH=7.5, 25°C). The reaction mix was incubated overnight prior to analysis.

Peptide	Turnover (6 hrs, 3.5 μM P4H)
taODD	75%
taODD Y481L	20%
CODD	none
DANRE	none
STRPU	none
NEMVE	none
PALPU	none
PALPU*	none
TRICA	none
TRICA*	none

Table C.2.3. ODD peptides (see Section B.3. for sequences) tested for hydroxylation by mbP4H using MALDI-TOF MS. The data were analysed taking into consideration 'no enzyme' controls as described in Section B.3.. Only two peptides, taODD and the taODD mutant Y4L were found to be modified relatively to the 'no enzyme' controls, both by +16 Da.

Of the 10 peptides tested, only taODD and the taODD variant Y481L were found to be modified, both by +16 Da. Subsequent experiments confirmed that this mass shift was caused by prolyl hydroxylation (Section C.3.2.).

It should, however, be kept in mind that the peptides used in the initial screening are relatively short. As the sequence and predicted structure of mbP4H suggests, the enzyme contains flexible loops possibly involved in *N*- and *C*-terminal interactions with the substrate. The peptides used may be simply too short to fulfill such interactions properly. To test for this possibility, a long human HIF-1α sequence covering the residues 530-698 was prepared by recombinant expression in *E. coli* (see Materials and Methods 2). Incubation with mbP4H under standard assay conditions (Materials and Methods 5) and analysis by ESI-MS (Materials and Methods 9), however, did not reveal any hydroxylations (data not shown).

C.2.4. 2OG turnover in the presence of ODD peptides

To gain further insight into the interactions between the ODD peptides and mbP4H, 2OG turnover in the presence of 100 µM peptide was quantified by ^{1}H-NMR (see Figures C.2.7.–C.2.8.).

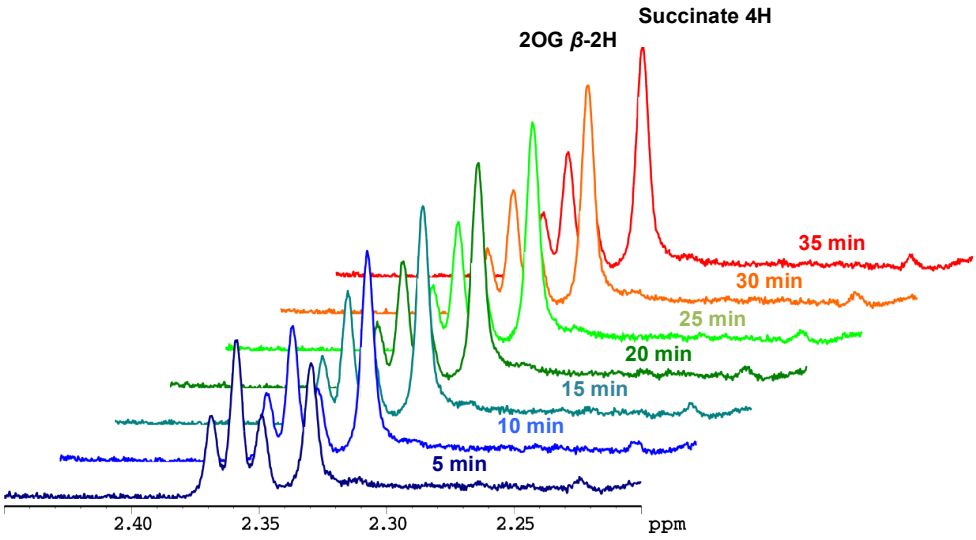


Figure C.2.7. Stacked ^{1}H-NMR spectra of a reaction mixture containing mbP4H, taODD and 2OG (3.5 µM mbP4H, 100 µM taODD, 300 µM 2OG, 50 µM Fe(II), 4 mM ascorbate, 50 mM deuterated Tris-DCl pH=7.5);close-up of 2OG β-protons and succinate.

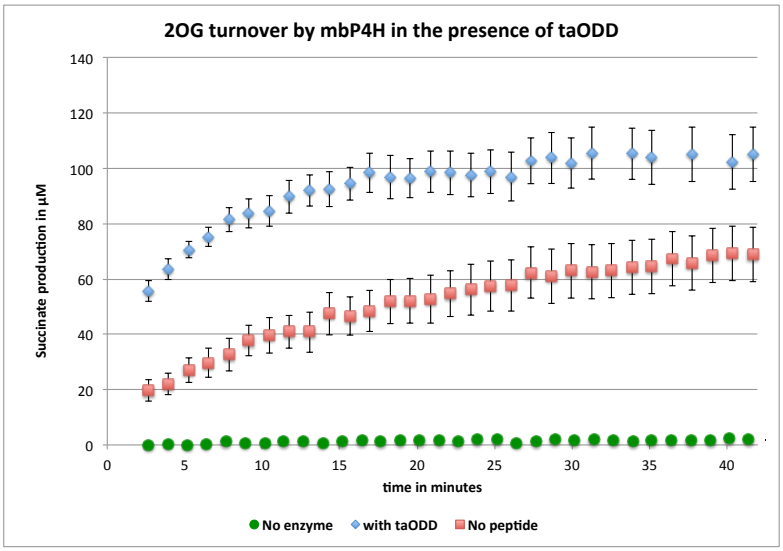


Figure C.2.8. Production of succinate by turnover of 2OG by mbP4H in presence of 100 µM taODD, without peptide and without enzyme, determined by ^{1}H-NMR at 3.5 µM enzyme concentration (n=3; error bars represent standard deviation).

In presence of taODD, the initial 2OG turnover rate by mbP4H was enhanced 2.7-fold (15.5 µM/min) compared to turnover detected in absence of a peptide. This suggests that the modification of taODD is indeed coupled with 2OG turnover to some extent. Repeating the same experiment with the taODD Y481L mutant and other CODD-homologs and determining the initial rates of succinate formation yielded interesting results (see Figure C.2.9. and Table C.2.4.).

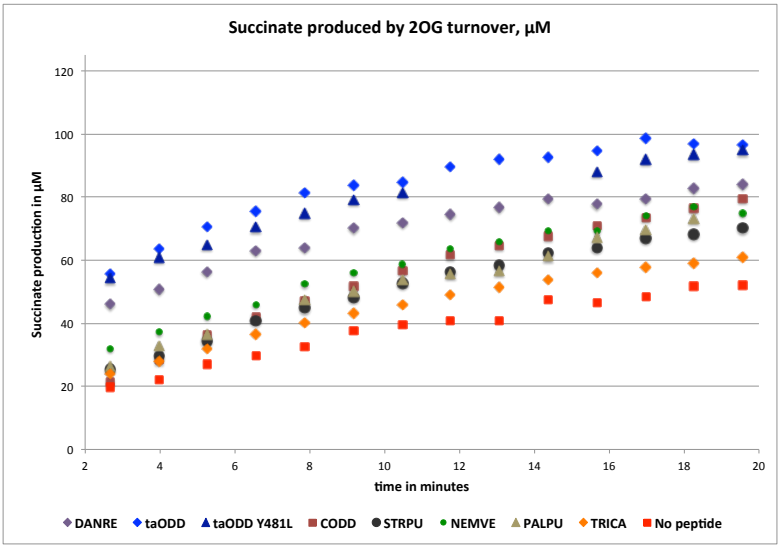


Figure C.2.9. Production of succinate in presence of 100 µM peptides, determined by ^{1}H-NMR at 3.5 µM enzyme concentration (see Materials and Methods 6 for protocol).

Peptide	r_i / µM/min
CODD	7.1
DANRE	12.3
STRPU	7.2
TRICA	6.8
PALPU	7.8
NEMVE	9.0
taODD	15.5
taODD Y481L	14.4
No peptide	5.7

Table C.2.4. Initial rates of succinate production by mbP4H in the presence of hsCODD, CODD-homolog peptides and the taODD variant Y481L.

Addition of the taODD variant Y481L led to comparable 2OG turnover levels as addition of the native taODD sequence, despite the lower levels of hydroxylation detected by MALDI-TOF MS (see Table C.2.4.). Apart from NEMVE and DANRE, other CODD-

homologs that were not hydroxylated by mbP4H also did not cause an enhancement in 2OG turnover initial rate compared to the data collected in the absence of a peptide. The stimulation of uncoupled 2OG turnover was similar with CODD, STRPU, PALPU, TRICA. While allosteric activation cannot be ruled out, it is also possible that the peptides NEMVE and DANRE enhance uncoupled 2OG turnover by binding to the active site, but lack key interactions required for hydroxylation by mbP4H. It is interesting that the peptide DANRE caused an enhancement in succinate formation comparable to taODD and the taODD Y481L variant, although no hydroxylation was detected by MALDI-TOF MS. Together, these results suggest that the enhancement of succinate formation in the presence of taODD, taODD Y481L, DANRE and NEMVE is caused to a large extent by the stimulation of uncoupled turnover by peptide binding. taODD, taODD Y481L and DANRE yield the highest enhancement in succinate formation independent of their hydroxylation rates, possibly because their sequences complement the active site of mbP4H. This is not a surprising observation for taODD and taODD Y481L, both of which are hydroxylated by mbP4H and vary in sequence only by one amino acid. DANRE, in contrast, is not hydroxylated by mbP4H to a detectable amount even after an overnight incubation. Subsequent experiments showed that DANRE is hydroxylated by the mbP4H variant C359V (see Section C.4.), supporting the hypothesis that the sequence of DANRE possesses sites of interactions with mbP4H that other non-substrate CODD-homologues do not. A comparison of the sequences of taODD, taODD Y481L, hsCODD and DANRE, followed by a discussion about possible substrate-enzyme interactions in mP4H is found in Section C.4.

C.2.5. Quantitative comparison of taODD turnover to succinate production

A peptide hydroxylation timecourse with mbP4H was carried out with simultaneous monitoring of 2OG turnover by NMR and peptide hydroxylation by MALDI-TOF MS (the assay was divided into two aliquots after starting, see Section B.1.2. for more detailed

description of experiment). This allowed for a direct comparison of the rates of succinate production and peptide hydroxylation.

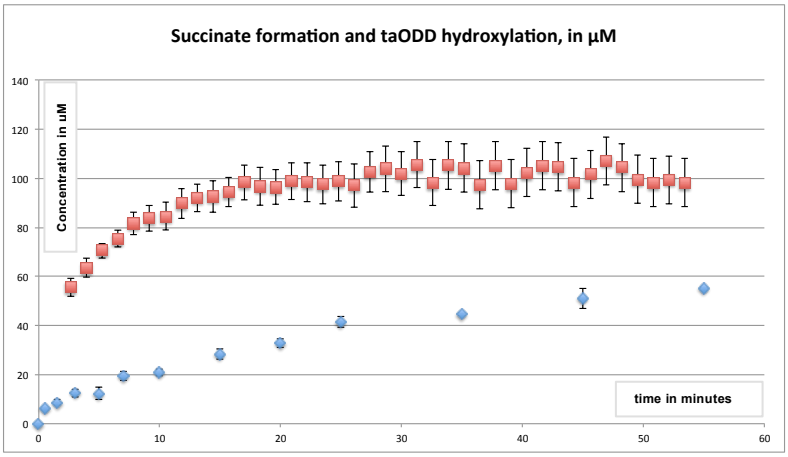


Figure C.2.10. taODD hydroxylation by mbP4H determined using MALDI-TOF MS (see Section A.2.4. for method of data analysis) and succinate production in µM, determined by ^{1}H-NMR. While the coupling between 2OG turnover and substrate hydroxylation was shown to be close to 1:1 for hsPHD2 and taPHD (see Section C.2.4.), the ratio for mbP4H was calculated to be 4.9:1. This could be caused by the fact that taODD is not the native substrate of mbP4H and therefore not be a true representation of biochemical properties of mbP4H.

With hsPHD2 and taPHD it was observed that the rates of peptide hydroxylation and succinate production during the same timecourse were very similar, indicating tight coupling between 2OG turnover and hsCODD/taODD hydroxylation (see Section A.2.4.). This was not the case with mbP4H and taODD. The coupling ratio was calculated from the initial rates of succinate production (see Table C.2.4.) and taODD hydroxylation (3.16 µM/min) and amounted to 4.9:1. It is possible that this ratio is caused by high uncoupled turnover because taODD is not a natural substrate of mbP4H. This is supported by the observation that succinate production in the presence of the taODD variant Y481L takes place at a very similar initial rate as in the presence of taODD, although the variant is being turned over more slowly.

C.2.6. Attempts to observe an mbP4H.Fe(II).2OG complex

Initial studies using non-denaturing mass spectrometry (see Materials and Methods 8) suggested that mbP4H forms a stable Fe(II)-2OG complex, similar to taPHD and hsPHD2 (data not shown). Only very low ion counts could be detected at a cone voltage of 50 V, which causes poor signal to noise ratios and makes deconvoluting and mass assignments ambiguous. A higher cone voltage should be attempted to get a more stable signal. Since the Fe(II)-2OG complexes with taPHD and hsPHD2 were observed at 50 V, reference experiments with taPHD and hsPHD2 at the new cone voltage should also be performed.

C.2.7. Summary and Discussion

The observations described in this Section are interesting for several reasons. On one hand, it was exciting to find that an ODD sequence could be modified by a mass of +16 Da by mbP4H, a PHD-like enzyme of protist origin and unknown function. These results established a link between the PHDs of the HIF-pathway and mbP4H. On the other hand, as none of the other ODD motives were modified, these results also suggest that the natural substrate of mbP4H interacts with mbP4H in a way rather different from the interaction of the ODDs with the PHDs. Of all ODD peptides tested, taPHD and hsPHD2 had essentially the same substrates, whereas only taODD, the substrate of the phylum evolutionary most closely related to *Monosiga brevicollis* within the library used, was modified by mbP4H. This could imply that the last common ancestor of mbP4H and the PHDs modified a sequence similar to taODD. With increasing distance from the basal placozoan phylum, this sequence could have changed too much to be accepted as a substrate by mbP4H. While the presence of a tyrosine instead of a leucine in the taODD sequence (Y481) could contribute to enzyme-substrate interactions, the lack of tyrosine

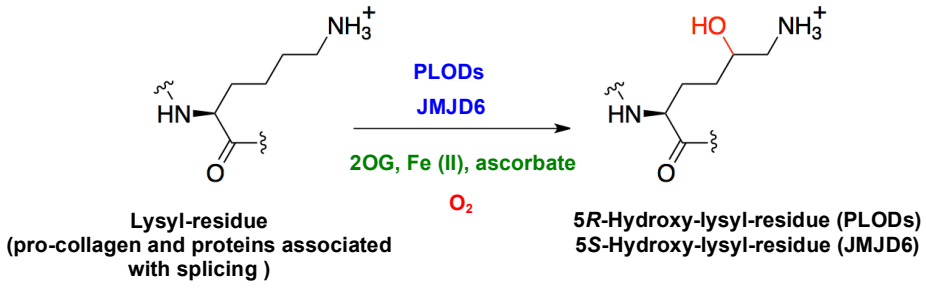
cannot be the only reason why the other ODD sequences are not hydroxylated. This is because the taODD mutant sequence lacking the tyrosine is also hydroxylated to a significant extent. It is therefore proposed that the natural substrate of mbP4H (or one of the natural substrates) is a protein containing a sequence similar to taODD, possibly with a YXXLAP motif.

C.3. Position and Stereochemistry of Hydroxylation by mbP4H

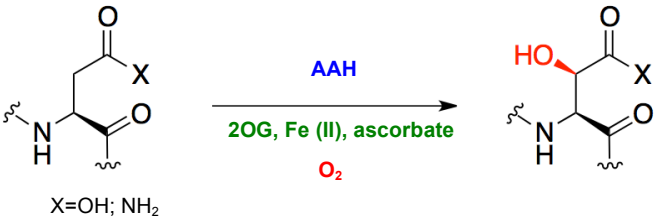
The +16 Da mass shift of taODD observed after incubation with mbP4H in the presence of 2OG, Fe(II) and ascorbate could theoretically have been caused by aspartyl-, asparaginyl-, lysyl- or prolyl-hydroxylation. All four modifications are characterised reactions of 2OG oxygenases. Tandem mass spectrometry was used to identify the modified residue, and subsequent amino acid analysis identified the position and stereochemistry of the hydroxylation. Although the nature of the taODD substrate and the biological context suggested that the +16 Da shift was caused by (4*R*)-prolyl-hydroxylation in homology to the modification carried out by taPHD, other possibilities needed to be considered.

C.3.1. Hydroxylations carried out by 2OG oxygenases

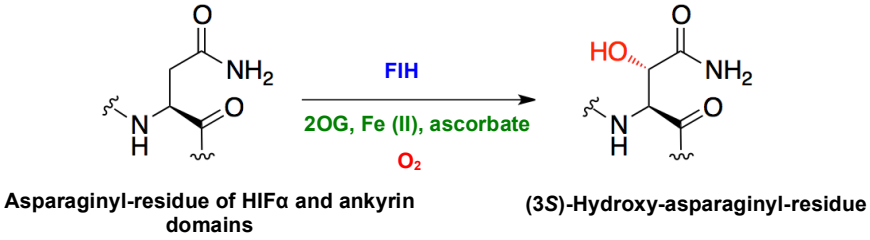
Lysyl-hydroxylation by 2OG oxygenases has been observed with PLODs (Procollagen-Lysine-2-Oxoglutarate-5-Dioxygenases) in the modification of pro-collagen and JMJD6 in the modification of U2AF65, a protein associated with RNA splicing (Webby et al. 2009).



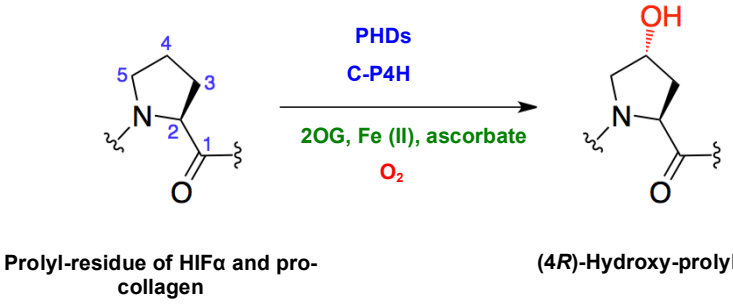
Aspartyl-asparaginyl-β-hydroxylase (AAH) has been shown to catalyse post-translational hydroxylation of aspartyl- and asparaginyl-residues in epidermal growth factor like domains of Notch and Jagged, among others (Joseph E Dinchuk et al. 2002; J E Dinchuk et al. 2000).

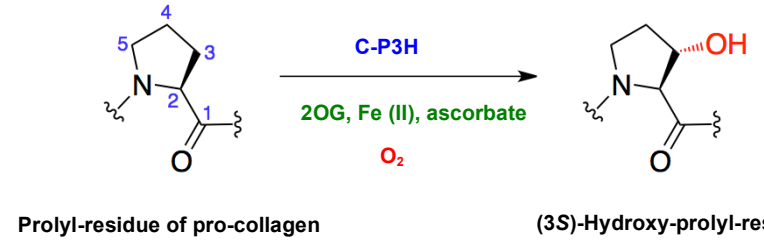


The asparaginyl-hydroxylase Factor Inhibiting HIF (FIH, see Introduction) hydroxylates HIFα as well as ankyrin repeat domain (ARD) containing proteins (Ferguson et al. 2007; Zheng et al. 2008).



Prolyl-hydroxylation is carried out by enzymes modifying pro-collagen and the PHDs.





While PHDs produce exclusively (4*R*)-hydroxy-prolyl products, the pro-collagen hydroxylases can catalyse (4*R*)-hydroxylation (prolyl-4-hydroxylase) as well as (3*S*)-hydroxylation (prolyl-3-hydroxylase) (Myllyharju 2004).

C.3.2. MS/MS Analysis of modified taODD peptide

In order to identify the modification giving rise to the +16 Da mass shift, an MS/MS analysis of the modified taODD peptide was carried out (see Figures C.3.1.-C.3.3.).

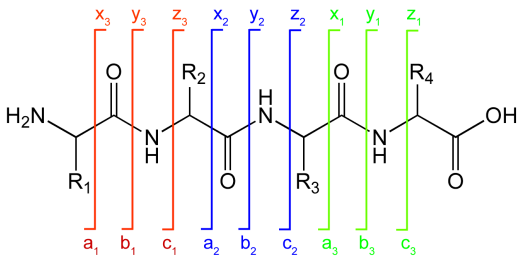


Figure C.3.1. Fragmentation of the selected peptide peak in MS/MS analyses can result in a variety of different fragments that are then assigned using predicted fragment masses. The standard notation (Roepstorff & Fohlman, 1984) uses letters to denote the position of the cleavage and the charge and numbers to denote the number of amino acid residues in the fragment. For peak identification in MS/MS, the observed peaks are assigned with masses of predicted fragments.

taODD was incubated overnight with mbP4H in the presence of Fe(II), 2OG and ascorbate (conditions as used in Section C.2.4.). Peptide turnover was determined to be at around 20% by MALDI-TOF MS. The enzyme was then precipitated with methanol and formic acid and removed from the solution. The resulting sample was analysed as described in Materials and Methods 13.

The peaks were assigned by comparison with predicted peak tables generated using Biolynx (WATERS). The presence of the b_{11}+O and b_{12}+O but no other +16 fragments clearly indicated that taODD is hydroxylated at either the prolyl- or the phenylalanyl-residue. Amino acid analysis (see next Section) confirmed that the +16 Da shift was caused by prolyl-hydroxylation. These results demonstrate that mbP4H acts on the same prolyl-residue that is hydroxylated by taPHD.

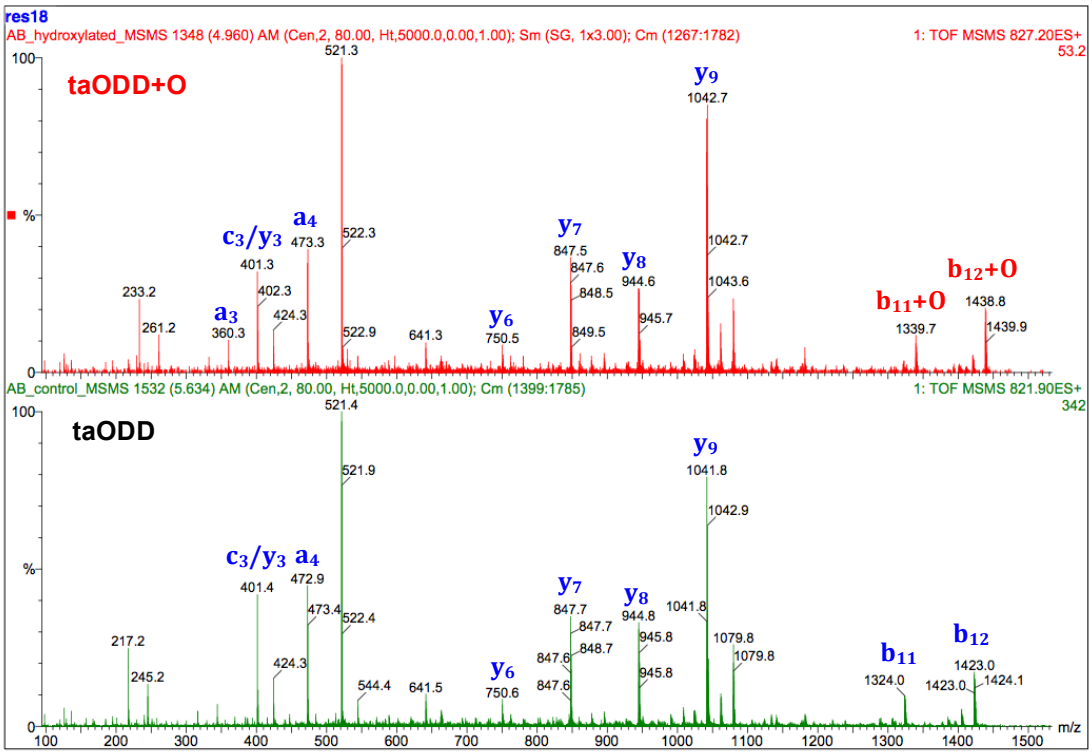


Figure C.3.2. MS/MS spectra of taODD peptide with and without treatment with mbP4H. Most peaks could be assigned to be fragments of the peptide. The presence of the b_{11}+O and the b_{12}+O peaks after treatment with mbP4H confirms the hydroxylation and narrows the position down to a prolyl- or a phenylalanyl-residue. Of these two possibilities, only the former is biologically relevant. This data rule out lysyl-, aspartyl- and asparaginyl-hydroxylation as well as prolyl-hydroxylation of the non-conserved triple-prolyl motif. The ~521 Da and ~233 Da peaks could not be assigned and have been assumed to be due to impurities. The comparably low ion count of the hydroxylated peptide is a result of the relatively low amount of hydroxylated peptide in the sample mixture.

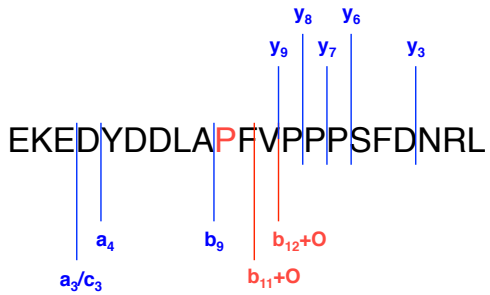


Figure C.3.3. Observed fragments of taODD.

taODD	Mass in Da		
Sequence	Predicted	Detected	Fragment
EKE/NRL	402.20	401.36	c_3/y_3
EKED	474.20	472.86	a_4
SFDNRL	751.40	750.59	y_6
PSFDNRL	848.40	847.68	y_7
PPSFDNRL	945.50	944.77	y_8
PPPSFDNRL	1042.50	1041.85	y_9
EKEDYDDLA	1079.50	1079.77	b_9
EKEDYDDLAPF	1323.60	1323.97	b_{11}
EKEDYDDLAPFV	1422.60	1423.07	b_{12}

taODD with mbP4H	Mass in Da		
Sequence	Predicted	Detected	Fragment
EKE	359.20	360.26	a_3
EKE/NRL	402.20	401.31	c_3/y_3
EKED	474.20	473.30	a_4
SFDNRL	751.40	750.00	y_6
PSFDNRL	848.40	847.53	y_7
PPSFDNRL	945.50	945.62	y_8
PPPSFDNRL	1042.50	1042.65	y_9
EKEDYDDLA	1079.50	1079.77	b_9
EKEDYDDLAHyPF	1339.60	1339.58	b_{11}+O
EKEDYDDLAHyPFV	1438.60	1438.76	b_{12}+O

Tables C.3.1. Assignment Tables for MS/MS analysis with predicted and detected fragment masses. HyP stands for hydroxy-proline.

C.3.3. Amino Acid Analysis

To determine the stereochemistry of the hydroxylation, amino acid analysis of the hydroxylated peptide was carried out. The protocol involves HCl-hydrolysis of the peptide to a mixture of amino acids, derivatisation of the amino acids followed by chromatographic separation and subsequent detection by mass spectrometry. taODD was incubated with mbP4H in the presence of Fe(II), ascorbate and 2OG overnight in MOPS buffer (Tris had previously been found to interfere with amino acid derivatisation by Nikita Loik, unpublished data). The amount of hydroxylated peptide was determined to be around 20% by MALDI-TOF MS. The enzyme was removed from the solution by methanol precipitation. After freeze-drying, the remaining sample was resuspended in 0.1% aqueous formic acid and purified by HPLC. The purified hydroxylated peptide was then hydrolysed with 6M HCl at 110°C for 24 hours. Derivatisation of all amino acids needed to be carried out prior to LC-MS analysis in order to block the polar amino group. Polar compounds are not retained by hydrophobic interaction with the stationary phase of the column used (see Materials and Methods 14). Blocking the polar amino group increases the hydrophobicity of the amino acids, thus allowing for separation.

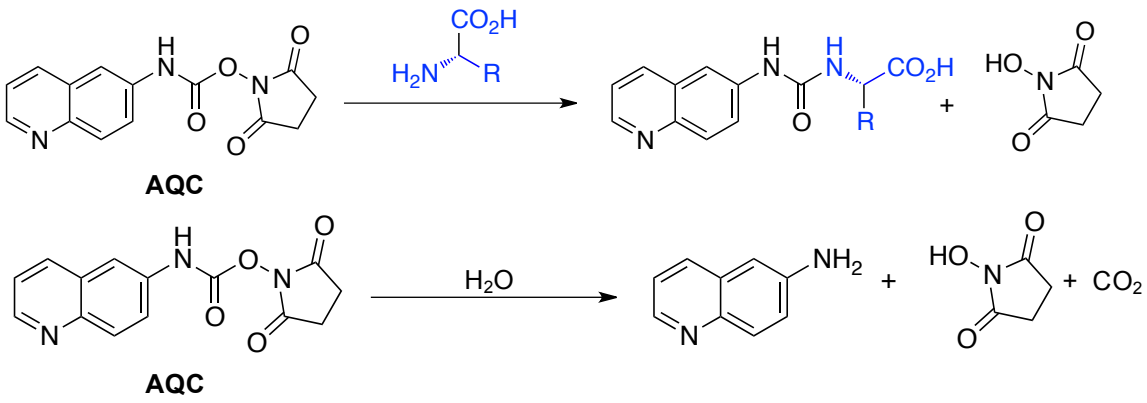


Figure C.3.4. Derivatisation of hydrolysate with AQC at 55°C. Excess AQC undergoes hydrolysis and decarboxylation.

For derivatisation, AccQTag solution protocol (Waters) was used (55°C, 1 hour). The derivatising agent AQC (6-Amino-Quinolyl-*N*-hydroxysuccinimidyl Carbamate) reacts with amino acids to form stable, asymmetric urea derivates. Surplus AQC undergoes hydrolysis followed by decarboxylation at 55°C.

The derivatised mixture was separated and analysed by LC-MS. Comparing the results with hydroxy-proline standards (4*R*)-hydroxy-proline, (4*S*)-hydroxy-proline, (3*R*)-hydroxy-proline, (3*S*)-hydroxy-proline) and controls (no added enzyme, no added peptide) as well as a taODD sample hydroxylated by taPHD supported the conclusion that both taPHD and mbP4H hydroxylate proline to yield (4*R*)-hydroxy-proline. Previous work on hsPHD2 has shown that it also catalyses the formation of (4*R*)-hydroxy-proline (Tarhonskaya, unpublished work).

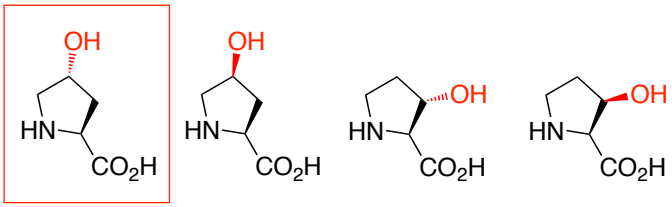


Figure C.3.5. Standards used for identification of position and stereochemistry of prolyl-hydroxylation of taODD by mbP4H. From left to right: (4*R*)-hydroxy-proline, (4*S*)-hydroxy-proline, (3*S*)-hydroxy-proline, (3*R*)-hydroxy-proline. Each isomer was used as a mixture of enantiomers, which cannot be separated using the method described. Since only the 2*S*-prolyl-derivates are of biological significance, the possibility of (2*R*)-enantiomers was ignored.

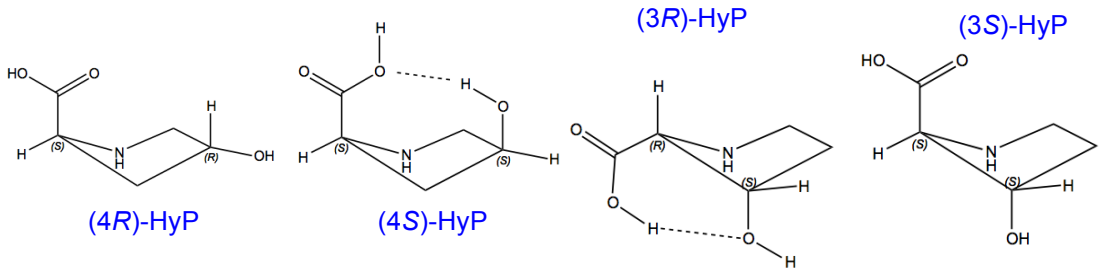


Figure C.3.6. The difference in elution times for the (4*R*)/(4*S*) and (3*R*)/(3*S*) diastereomeric urea derivates is unusually large and could be caused by the formation of intramolecular hydrogen bonds in the *cis*-isomers (4*S*)-hydroxy-proline and (3*R*)-hydroxy-proline as shown above (proposed by Nikita Loik; hydroxy-proline diastereomers rather than urea derivates shown above for clarity). Such interactions would decrease the polarity of the *cis*-hydroxyl-proline derivates and lead to longer elution times.

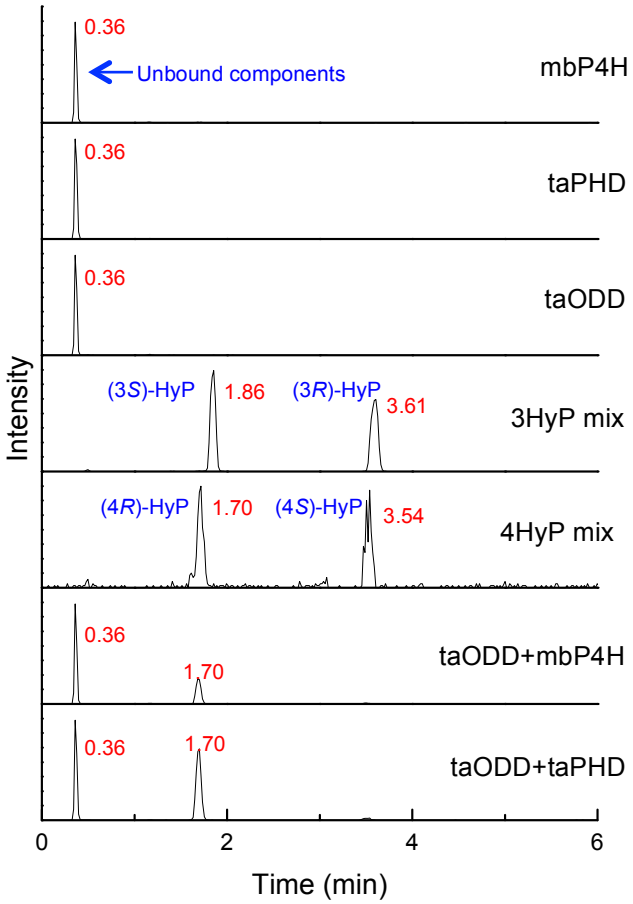


Figure C.3.7. Chromatograms representing results of amino acid analysis. The samples containing hydroxylated peptide were compared to 4-hydroxy-proline and 3-hydroxy-proline standards, also derivatised with AQC, and 'no enzyme'/'no peptide' controls. The two 4-hydroxy-proline *R/S* diastereomers and 3-hydroxy-proline *R/S* diastereomers were used as a 1:1 mix. This is possible because the elution volumes for the *cis*/*trans*-isomers are very different (see Figure C.3.6.) and there are no ambiguities in the resulting chromatograms. Comparison of the enzyme-treated peptide samples clearly indicates (4*R*)-prolyl-hydroxylation by both taPHD and mbP4H. The peak at 0.36 minutes is caused by unbound components contained in the sample mixture, including AQC and the byproducts of its hydrolysis.

C.3.4. Summary

Together with the MS/MS analyses in the previous chapter, these results demonstrate that mbP4H is capable of catalysing the hydroxylation of the same proline residue of taODD as do hsPHD2 and taPHD. The taODD hydroxylation by both mbP4H and taPHD yields (4*R*)-hydroxy-prolyl-taODD, while the hsCODD hydroxylation by hsPHD2 yields (4*R*) hydroxy-prolyl-CODD. These findings support the hypothesis that mbP4H is evolutionary related to taPHD and hsPHD2.

C.4. Modification of mbP4H Substrate Range Through Residue Substitution

Targeted substitutions of mbP4H residues were carried out with the aim of obtaining information about the interactions between mbP4H and taODD. Attempts to identify residue substitutions that might enhance the turnover rate of taODD or broaden the substrate range of mbP4H were of particular interest. To aid this analysis, the amino acid sequence of mbP4H was compared to the sequences of hsPHD2, taPHD and other PHDs (see Section C.1.). Three interesting differences were identified and individual variants of mbP4H were prepared to match the hsPHD2 sequence. The mbP4H variants C359V, P238D and G255R were expressed in *E. coli* and purified using a His-affinity column and subsequently screened for hydroxylase activity by MALDI-TOF MS with the ODD peptide library described in Section B.3. (see Materials and Methods 2-3 for expression and purification protocols).

C.4.1. mbP4H Variant C359V

The residue between the catalytic site histidine and aspartic acid (HXD motif) is conserved as hydrophobic valine or isoleucine (taPHD only) throughout the PHD sequences analysed in Section B.4.2. The crystal structures of taPHD.taODD and hsPHD.hsCODD suggest that this residue forms part of a hydrophobic pocket that can interact with $L559_{hsHIF\text{-}1\alpha}$ and $Y481_{taHIF\alpha}$ (see Section B.4.1. and Figure C.4.1.). In contrast, the equivalent residue in mbP4H is a cysteine (see alignment in Section C.). The fact that only taODD, a peptide not containing the complementary aliphatic leucine to fit the pocket was turned over was striking, especially when considering that the active site region is otherwise highly conserved between mbP4H and the PHDs. An mbP4H variant

C359V was therefore prepared and tested for hydroxylation of CODD-homologue peptides. The results are described in Section C.1.4.

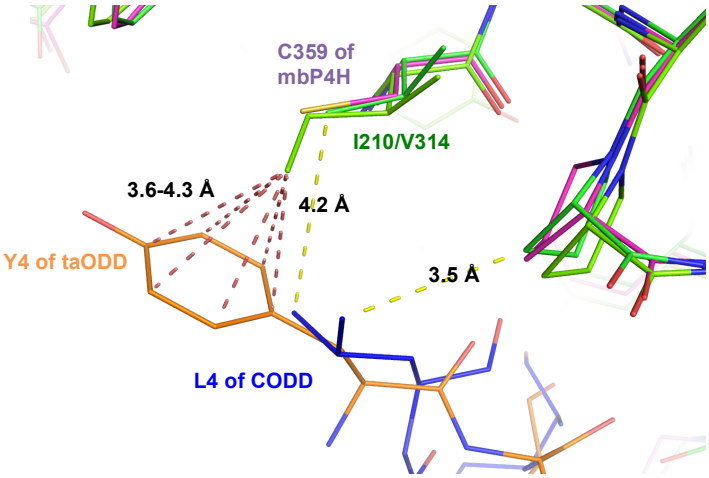


Figure C.4.1. Superimposed crystal structures of taPHD and hsPHD2 (both green) in complex with substrate (orange taODD and blue hsCODD) and a model of mbP4H (magenta). The distances between hsPHD2 and hsCODD residues are shown in yellow and the distances between taPHD and taODD are shown in salmon. The leucine residue of hsCODD can undergo hydrophobic interactions with $I210_{taPHD}/V314_{hsPHD2}$ and $P213_{taPHD}/P317_{hsPHD2}$. While the prolyl-residue is conserved in mbP4H, the residue between the active site histidine and aspartic acid (HXD motif) is a cysteine.

C.4.2. mbP4H Variants P239D and G256R

Residues 240-261 in the mbP4H sequence bear a close similarity to the substrate-enclosing β2/β3 finger-loop of the PHDs. Studies on hsPHD1-3 suggest that the loop sequence is important for hsCODD vs. hsNODD selectivity (Flashman et al. 2007). In addition, there is substantial variation of the predicted loop sequence between different metazoan species and the effect of this variation on substrate selectivity is not understood. While the β2/β3 finger-loop regions of hsPHD2 and taPHD differ in length and residue type, both enzymes had virtually the same range of substrates (see Section B.3.), with the exception that taPHD turns over hsNODD only very slowly (<10% hydroxylation of hsNODD after 3 hours but 70-80% hydroxylation of taODD and hsCODD at the same conditions; see Section B.5.). The overlapping substrate range of

hsPHD2 and taPHD may suggest that while the loop sequence could be key for hsNODD vs. hsCODD tuning, it may be of minor relative importance for selectivity between different CODD-like substrates.

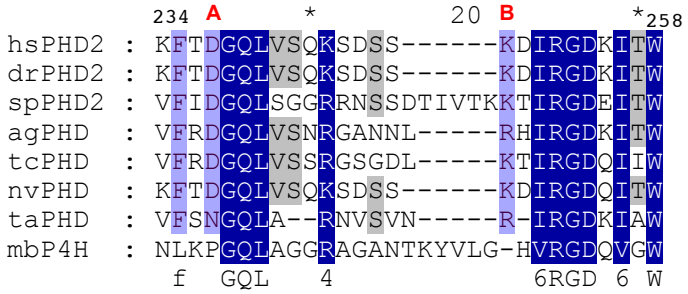


Figure C.4.2. Sequence alignment of the β2/β3 finger-loop region of hsPHD2 and PHD2-homologs (see Section B.4.2. for full alignment). While the *N*- and *C*-flanking regions are conserved, the mid-loop region varies substantially in length and character.

Additionally, since the β2/β3 finger-loop regions are only poorly resolved in the crystal structure taPHD.taODD, substitutions in this region would have had to be designed and interpreted on a rather speculative basis. For this reason, only two β2/β3 finger-loop mutants of mbP4H were prepared. Firstly, the mbP4H P239D variant was made based on the observation that a hydrophilic residue ($N135_{taPHD}$ and $D237_{hsPHD2}$; labelled 'A' in Figure C.4.2.) *N*-terminal to the loop motif is conserved. Inspection of the crystal structure taPHD.taODD suggested that $N135_{taPHD}$ might be involved in intramolecular polar interactions with $K151_{taPHD}$ (see Table A.4.4.) and possibly $N145_{taPHD}$, thereby stabilising the β2/β3 finger-loop conformation (see Figure C.4.4).

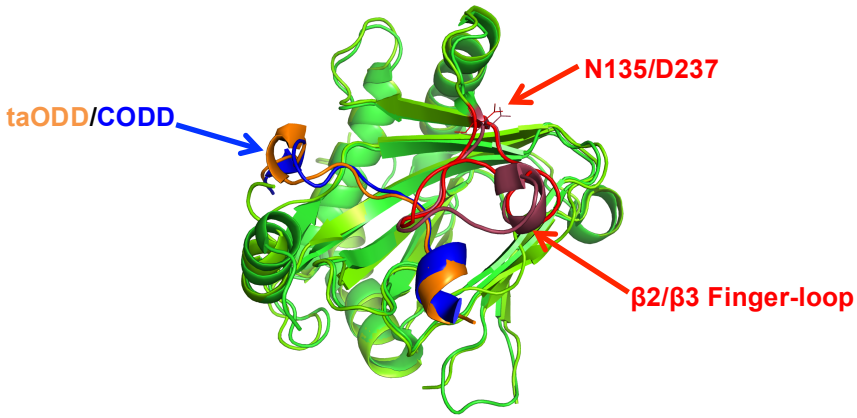


Figure C.4.3. Superimposition of crystal structures taPHD.taODD and hsPHD2.hsCODD (PDB ID=3HQR). The β2/β3 finger-loop is marked in red, while hsCODD is blue and taODD orange. The position of the conserved polar residues is highlighted in red. It is apparent that this position is not near the substrate.

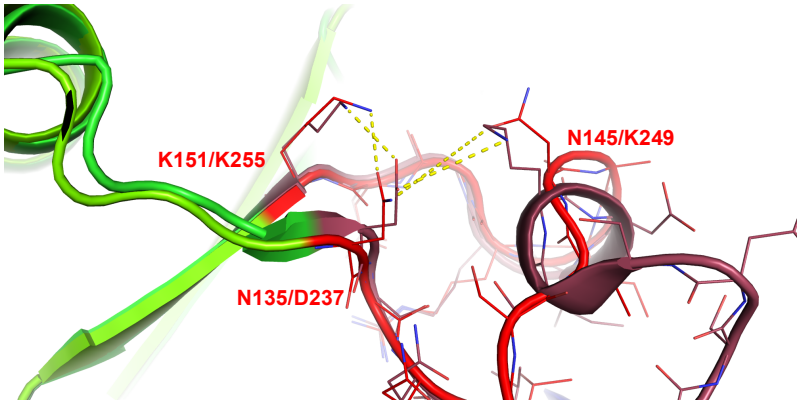


Figure C.4.4. Close-up view shows that $N135_{taPHD}/D237_{hsPHD2}$ could bind to $N145_{taPHD}/K249_{hsPHD2}$ and/or to $K151_{taPHD}/K255_{hsPHD2}$ and might be supporting the folded shape of the loop. The distances between the side chain heteroatoms range from 2.7 Å ($K151_{taPHD}/K255_{hsPHD2}$) to 6.6 Å ($N145_{taPHD}/K249_{hsPHD2}$) in the crystal structures, but the residues in question may be much closer to each other in solution at least at some stage throughout the catalytic cycle.

Secondly, $G226_{mbP4H}$ was substituted with an arginine because of conservation of a large polar residue towards the end of the β2/β3 finger-loop sequence throughout the PHD sequences examined in Section B.4.2. (labelled 'B' in Figure C.4.2.). A detailed structural analysis of this position is hindered by poor density in the crystal structures of taPHD and hsPHD2. The side chain of the equivalent residue R146 in the taPHD.taODD structure is disordered.

C.4.3. Preliminary screening results for mbP4H variants

The mbP4H G256R variant did not show any deviation in hydroxylation selectivity and no significant deviation in the amount of peptide turned over relative to the *wild-type* mbP4H (see Figure C.4.5.). Although more detailed experiments are required to be certain, this preliminary result implies that the lack of a conserved large polar residue towards the end of the loop region is not the reason why of all ODDs tested, only taODD and taODD Y481L were hydroxylated.

The mbP4H P239D variant turned over less taODD in the same timespan relative to the *wild-type* mbP4H. Further, hydroxylation of the taODD Y481L variant could not be detected. Hydroxylation of other substrates listed in Table C.2.4. could not be observed with P238D. The reason for this could be that region 240-261 of mbP4H has a different structure than the β2/β3 finger-loop region of hsPHD2 and taPHD. The prolyl-residue could be important in a structural feature such as a rigid turn. A substitution of the $P239_{mbP4H}$ would then result in the destabilization of the fold of region 240-261 and possibly result in poor substrate binding or release.

Interesting results were observed with the mbP4H variant C359V. While the amount of taODD turned over in an overnight incubation with mbP4H was lower compared to the *wild-type* mbP4H, the relative amount of taODD Y481L variant hydroxylation was increased by 40% compared to the *wild-type*. This result implies that introducing an aliphatic residue between H and D of the HXD motif might stabilise the enzyme-peptide interactions possibly by hydrophobic interactions with the leucine residue, provided by the taODD Y481L variant but lacking in taODD.

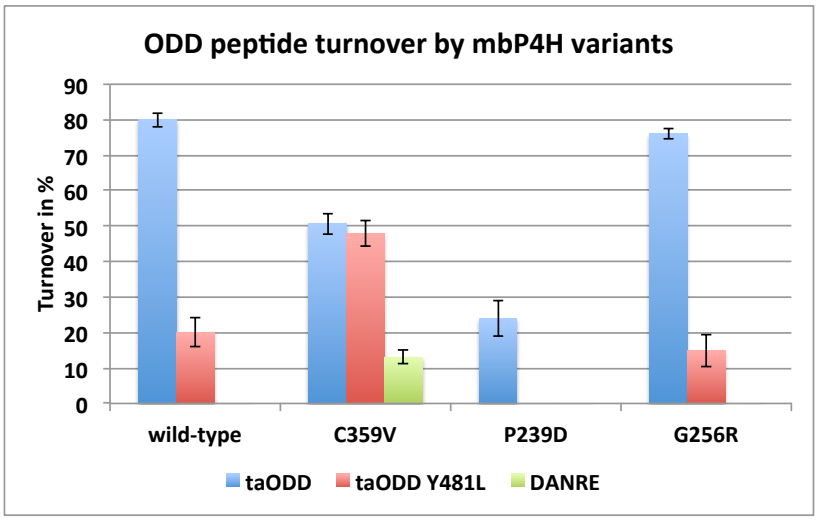


Figure C.4.5. Peptide hydroxylation by mbP4H variants in overnight incubation with 10 µM enzyme and 100 µM peptide, in % of peptide turnover relative to the 'no-enzyme' control (MALDI-TOF MS; data analysed as described in Materials and Methods 7). No difference in substrate selectivity and final peptide turnover was seen between the *wild-type* and the G256R variant. Over 65% less taODD hydroxylation compared to *wild-type* mbP4H were observed with the P239D variant. The C359V variant turned over 38% less taODD compared to the *wild-type* but 40% more taODD Y481L. DANRE, a peptide previously screened as negative was found to be hydroxylated by the C359V variant.

It was also found that the zebra fish ODD sequence DANRE was turned over by the mbP4H C359V variant (12% turnover after overnight incubation compared to 'no enzyme' sample). This result is significant because it confirms that the position between the active site histidine and aspartic acid of the HXD motif has the potential to be involved in determining substrate selectivity. The DANRE sequence is turned over very rapidly by both hsPHD2 and taPHD, both of which also rapidly turn over hsCODD. Since mbP4H, however, does not turn over hsCODD, comparison of the taODD, DANRE and hsCODD structures could yield clues about key residues involved in enzyme-peptide interactions with mbP4H and thereby assist in the search for the native *Monosiga brevicollis* substrate of mbP4H.

The initial rate of uncoupled 2OG turnover by mbP4H in the presence of DANRE was almost as high as the initial rates in the presence of taODD and taODD Y481L, both of

which were hydroxylated (Table C.2.4.). In the light of the C359V variant results, this observation implies that DANRE binds to wild-type mbP4H and enhances uncoupled turnover. The turnover of DANRE by the C359V variant suggest that the binding mode of DANRE at the active site may not be optimal with *wild-type* mbP4H for turnover but is improved by the presence of a valine in the HXD motif in the variant.

```
taODD        EKEDYDDLAPFVPPPSFDNRL   Trichoplax adhaerens
taODD Y481L  EKEDLDDLAPFVPPPSFDNRL   Trichoplax adhaerens variant
DANRE        ELDELDSLAPYIPMHGEDFLL   Danio Rerio
CODD          DLDLEMLAPYIPMDDDFQL    Human

NEMVE         SNELQNRAPYIPPPTGDAAL   Nematostella vectensis
PALPU         LDEFDMRAPFIPLSNELLM    Palaemonetes pugio
TRICA         ESDLVAKAPYITMNMGDDL    Tribolium castaneum
STRPU          DELAMRAPYIPMGEDFDL    Strongylocentrotus purpuratum
```

Figure C.4.6. Sequences of ODD peptides (see Section B.4.1. for sequences of related variants).

Like the taODD Y481L variant, DANRE also possesses an aliphatic leucine at the position equivalent to $481_{taHIF\alpha}$. Since all other peptides in Table C.2.4. that also possess this leucine tested negatively, including hsCODD, this can not be the distinguishing feature of the DANRE sequence. A comparison of the peptide sequences (Figure C.4.6.) shows that taODD and its variant and DANRE are the only 21mers, whereas the other peptides have 20 or 19 amino acids. It is possible that in mbP4H, interactions with the *N*-terminal acidic residue are key to peptide binding and turnover. Since the first three taODD amino acids are disordered in the taPHD.taODD structure, speculations as to which residues in mbP4H may be undergoing such an interactions can currently not be made reliably.

C.4.4. The significance of the HXD motif

The HXD motif is highly conserved in many 2OG-oxygenases, suggesting that it plays an important role in enzymatic functionality.

HXD is conserved as HVD in over 130 chordate PHD1-3 homologues and 6 PHD homologues of more simple species (5 aligned in Figure B.4.7 and EGL-9 of *Caenorhabditis elegans*). All of the 9 available FIH homologue sequences ranging from human to *Nematostella vectensis*, in contrast, contain an HYD motif (data not shown). DAOCS of *Streptomyces clavuligerus*, *Nocardia lactamdurans* and *Cephalosporium acremonium* contain a HYD motif. However, HXD motif conservation is not observed in all 2OG-oxygenases. IPNS of *Streptomyces jumonjinesis*, *Streptomyces clavuligerus* and *Streptomyces cattleya*, for example, contains an HLD motif, but other motifs were observed such as HED in *Cephalosporium acremonium*, HQD in *Lysobacter lactamgenus* and *Flavobacterium,* and HFD in *Nocardia lactamdurans.*

Interestingly, taPHD is the only known PHD-homologue sequence that contains an HID motif. This could be indicative of a larger evolutionary distance between taPHD and the other PHDs than among the other PHDs.

The presence of the HCD motif observed in mbP4H is supporting the proposal that it is only distantly related to the PHDs and has functionality distinct from the 2OG oxygenases described above. mbP4H is very different in sequence from the pro-collagen hydroxylases and an evolutionary relationship could not be established.

C.4.5. Summary

Overall, the results indicate that the substrate range of mbP4H can be broadened by substitution of key residues to match the sequence of hsPHD2 more closely, implying that mbP4H is potentially only several cumulative residue differences away from acting like an early metazoan PHD. This supports the hypothesis that mbP4H is the closest unicellular relative of the metazoan PHDs known to date. It is proposed that the active site HXD motif is a key site determining substrate selectivity in mbP4H. The PHD-

common HVD motif could have evolved from the taPHD HID motif, which in return could have the same evolutionary origin as the mbP4H HCD motif. Since this sequence occurs in the active site, it is possible that it has changed over time very slowly compared to other regions within the sequence. This proposal is in agreement with the observation that the HXD sequence differs between mbP4H and taPHD as well as taPHD and the other PHDs studied, but not within the other PHDs studied.

C.5. Crystallisation of mbP4H

It was of considerable interest to obtain a crystal structure of mbP4H as the structural information could reveal a substantial amount of information about the function and mode of action of this PHD-like enzyme and its similarity to metazoan PHDs.

C.5.1. Selection of screening conditions

The commercial screen JCSG-plus (HAMPTON RESEARCH) had generated many hits with 2OG oxygenases previously, including hsPHD2 and taPHD (see Section A.2.7.). This screen was therefore selected for initial screening with mbP4H. mbP4H was prepared using the method described in Section C.2.1. Initial tests with protein concentration varying between 40 mg/mL and 10 mg/mL have shown that the protein precipitates in the drops at concentrations higher than 20 mg/mL. Based on further initial tests, it was decided to set up all plates at 16 mg/mL. Several generic 2OG-oxygenase inhibitors acting as 2OG mimics were tested for inhibition of mbP4H (see Figure C.5.1.).

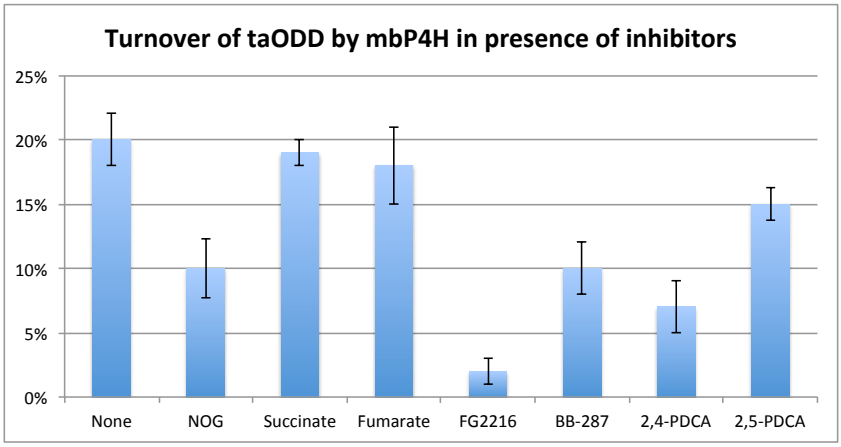


Figure C.5.1. Turnover of taODD (in % of starting peptide concentration) by mbP4H after 10 minutes in the presence of inhibitors.

Succinate and fumarate showed only very weak inhibition, suggesting poor binding to the active site. Of the remaining inhibitors, NOG and UN9 are known to have generated

hits with taPHD and hsPHD2 previously and were therefore chosen as ligands for initial screens. 12 sitting-drop round-well plates were set up automatically (PHOENIX), 6 at 20°C and 6 at 4°C.

C.5.2. Initial hits and optimization

A number of hits were detected in all of the plates at 20°C but none of the 4°C plates. Most of the crystals were too small for X-ray screening and required further optimisation. One crystal from well H.6. of plate A was big enough for an initial data collection (110x10 µm). The data were collected in-house (see Methods 11) and revealed that the crystal was protein rather than an inorganic salt. Data were collected to 3.7 Å. It was thought that the limiting factor for resolution in this case was crystal size and several methods of improving the crystal size were attempted.

Plate	Additives	Hit wells	Description
A	none	D.6, H.6	Haystacks/rectangular rods
B	NOG	D.6, E.11	Haystacks/rectangular rods
C	NOG; taODD	E.11	Rectangular rods
D	UN9	D.6	Haystacks

Tables C.5.1. Initial hits with the JCSG-plus commercial screen at 20°C

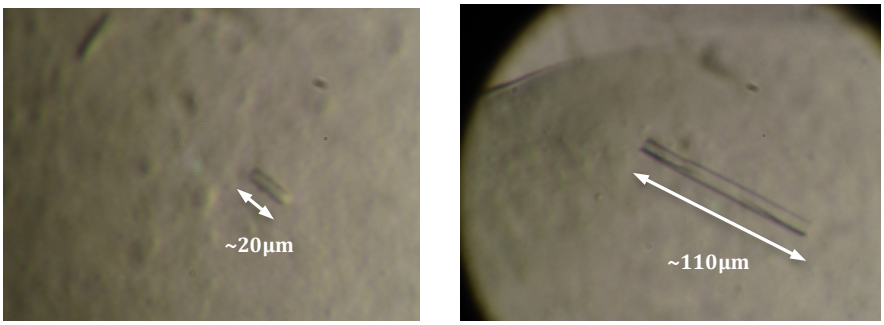


Figure C.5.2. mbP4H crystals obtained by automatised screening with the commercial screen JCSG-Plus. 2 mM NOG was added as a 2OG-mimic in the plate containing the crystals above (Plate B; see Table C.5.1.).

Streak seeding using the SEED-BEAD kit (HAMPTON RESEARCH) was attempted using small crystals from the wells D.6. and E.11. with NOG as the ligand. While crystals

appeared within hours after the introduction of seeds, single large crystals could not be generated. Instead, the seeding yielded either large amounts of small two-dimensional plates at concentrated seed solutions or no crystals at all at diluted seed solutions.



Figure C.5.3. Streak seeding with JCSG-Plus D.6. conditions and NOG ligand.

Additionally, three optimisation screens were designed based on the conditions in wells D.6., H.6. and E.11. Plates set up with NOG using the optimisation screens based on wells H.6. and D.6. yielded numerous crystals. All crystals obtained were of the same morphology: rectangular rods of a fixed width of ~20 µm and lengths varying form 100 µm to 350 µm. It appeared that the rods were hollow on the inside at the one end. It is possible that the crystals grow faster on the outside and that at some point not enough protein can diffuse into the tube-like opening of the growing crystal, resulting in the apparently hollow crystal shape.

The largest crystals obtained from the optimisation screens were harvested and irradiated at the Diamond synchrotron source (beam line I03). A data set to 1.8 Å was collected (see Figure C.5.4. for example image). Indexing suggested the space group P 21,21,21 and the presence of one molecule in the asymmetric unit, but the structure could not be solved by molecular replacement using hsPHD2 or taPHD. Attempts at re-indexing, varying the number of molecules in the asymmetric unit and trimming down the model to contain only the DSBH remained fruitless. It was concluded that the structure of mbP4H differs too strongly from hsPHD2 and taPHD for the use of the available molecular replacement techniques.

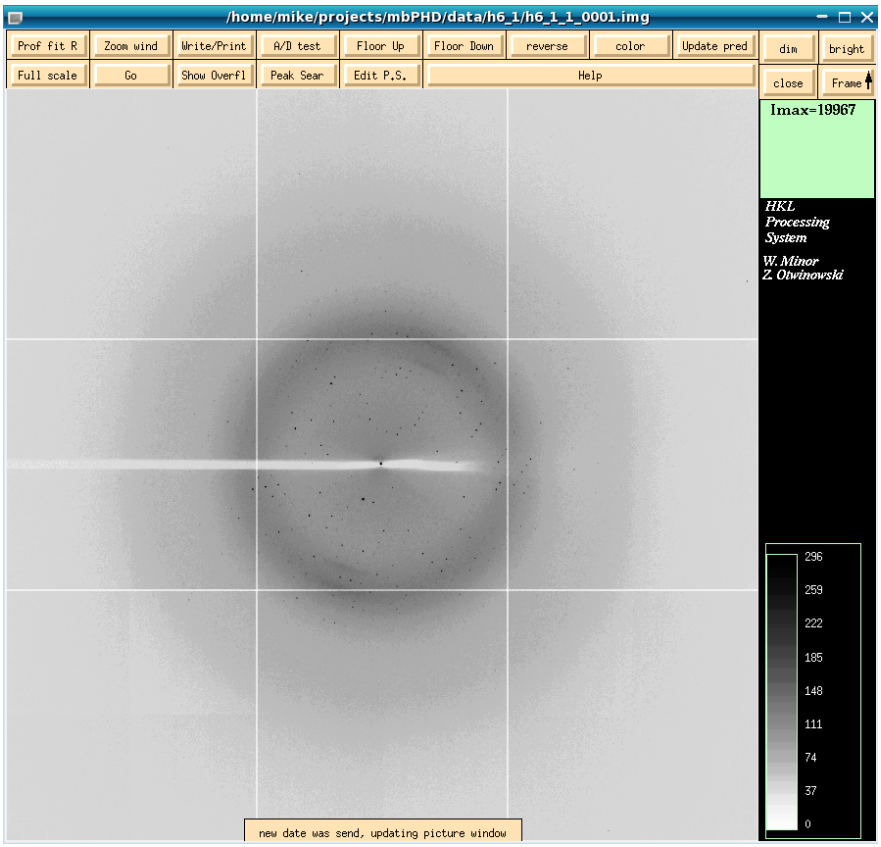


Figure C.5.4. Example of mbP4H crystal diffraction pattern (HKL-2000 screenshot).

C.5.3. Attempts at solving the structure: Halide and heavy metal soaks and preparation of seleno-methionine mbP4H

There are several ways to solve a crystal structure from diffraction data when molecular replacement is not available. *Ab initio* (direct) phasing methods based on the exploitation of known phase relationships between certain groups of reflections are only available for high-resolution data derived from small molecule crystals (Hauptman 1997). For macromolecules, including proteins, the introduction of electron-rich elements such as halides, for example iodine and bromine, or heavy metals, such as mercury or gold, can help solving the phase problem using multiple isomorphous replacement. This involves the use of direct phasing methods to determine the heavy atom location and to obtain the initial phases (Hauptman 1997). This technique is now increasingly being replaced by anomalous X-ray scattering methods using seleno-methionine containing proteins. This method is based on scanning the X-ray wavelength

past an absorption edge of an atom using a tunable X-ray source, which changes the scattering in a known way. If full sets of reflections far below the absorption edge, far above the absorption edge and in the middle of the absorption edge are recorded, the structure of the anomalously diffracting atoms can be solved, in theory. The obtained initial phases can then be used to solve the structure of the seleno-methionine protein (Ealick 2000).

mbP4H crystals generated under the same conditions as described in Section 5.2. were soaked for varying times in either 1 mM $HgCl_2$ or 0.1 mM K_2PtCl_4. Diffraction data were collected a the synchrotron source (Diamond, tunable beam line I04) for crystals soaked for 20 minutes in $HgCl_2$ (2.5 Å) and 25 minutes in K_2PtCl_4 (2.4 Å). The same space group as for the native data were suggested (orthorhombic; P 21,21,21). Attempts at multiple isomorphous replacement, however, were unsuccessful.

Seleno-methionine containing mbP4H (SeMet mbP4H) was prepared by growing the *E. coli* culture overexpressing the protein in a methionine-lacking medium (MOLECULAR DIMENSIONS) and adding 50 mg/L *L*-seleno-methionine (SIGMA-ALDRICH). Incorporation of seleno-methionine was confirmed by ESI-MS (described in Materials and Methods 9; see Figure C.5.5.).

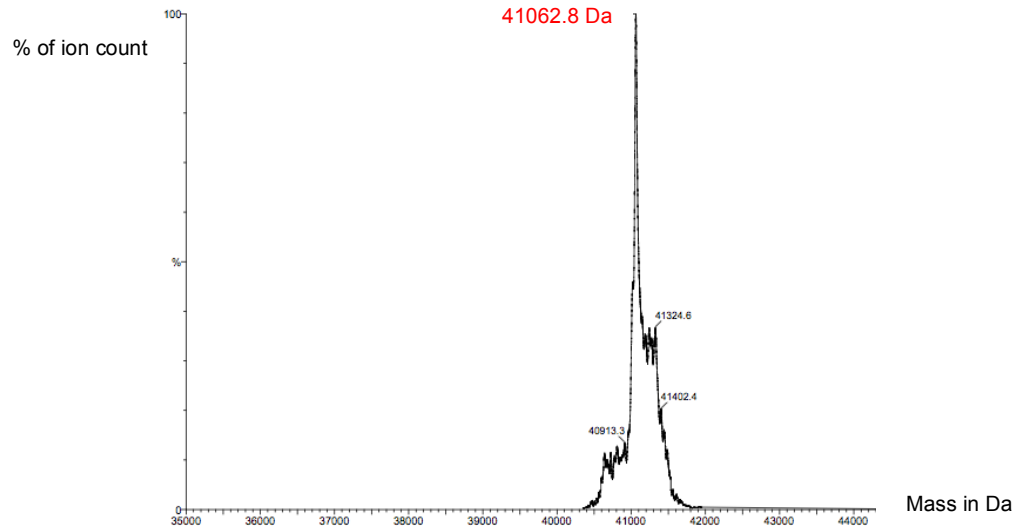


Figure C.5.5. ESI-MS spectrum of SeMet-mbP4H. The calculated mass of fully substituted mbP4H is 41065 Da. The slightly lower observed mass could be caused by incomplete substitution.

Plates set up using the H.6. optimisation screen that previously yielded crystals were set up. NOG was added to the protein solution to a final concentration of 2 mM. The first SeMet mbP4H crystals were observed after two days.

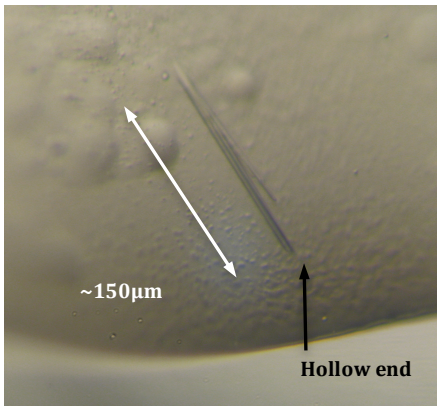


Figure C.5.6. Example of SeMet-mbP4H crystal, grown in 0.09 M ammonium acetate, 0.1 M Bis-Tris pH=5.5 and 18.5% w/v PEG 10k in the presence of 2 mM NOG. Note the hollow end of the crystal rod as pointed out.

The apparent morphology was the same as previously observed with native mbP4H. Several crystals were harvested. Two sets of SAD data (2.35 Å and 2.9 Å) and one set of MAD data (2.7 Å) were collected at the Diamond synchrotron source with a tunable beam line (I04). The same space group and similar unit cell parameters (±1 Å) as with

the native structure were detected (see Table C.5.2.). The structure, however, could not be solved using either of these data sets.

	Lengths in Å			Angles in degrees		
Parameter	**a**	**b**	**c**	**α**	**β**	**γ**
Native mbP4H	46.31	67.00	149.99	90	90	90
SeMet mbP4H	46.11	66.74	149.22	90	90	90

Tables C.5.2. Unit cell parameters determined with native mbP4H crystals and seleno-methionine substituted mbP4H (example). The unit cell parameters for native mbP4H crystals soaked in heavy atoms are similar (±1 Å for lengths, same angles). All mbP4H crystals irradiated were orthorhombic (P 21 21 21).

C.5.4. Summary

mbP4H was co-crystallised with NOG. A dataset at a resolution of 1.8 Å was collected at a synchrotron source, but the structure could not be solved by molecular replacement. SAD and MAD data were collected with seleno-methionine substituted mbP4H. Diffraction data were also collected with mbP4H crystals soaked in $HgCl_2$ and K_2PtCl_4. Attempts at structure solution, however, remained unsuccessful for reasons not fully known at this stage.

C.6. Strategies for Identifying the Native Substrate of mbP4H

The studies of mbP4H carried out so far have shown that this non-metazoan enzyme is similar to the PHDs in parts of its sequence and in some of its biochemical properties. It has been shown to catalyse prolyl-(4*R*)-hydroxylation in taODD, a reaction also carried out by taPHD and hsPHD2. At this stage, a question of interest is whether the natural substrate of mbP4H and the function of this enzyme in the biochemistry of *Monosiga brevicollis* are related to oxygen sensing.

C.6.1. HIF-pathway components in *Monosiga brevicollis*

Bioinformatic analyses involving the search of more than 60 eukaryote genomes for hypoxia-responsive elements (HREs; RCGTC) have suggested that the HIF-pathway is limited to metazoans (Loenarz et al. 2011). While it is possible that *Monosiga brevicollis* nevertheless possess a HIF-like pathway, no bHLH-PAS proteins have been identified in the genome of this organism (Loenarz 2010). Since *taHIFα* and *taPHD* are co-localised (within 18 kpb) in *Trichoplax adhaerens* (Loenarz et al. 2011), the proximate environment (± 50 kpb) of *mbP4H* was searched for predicted proteins containing a taODD-like sequence, but failed to yield any clues.

The genome of *Monosiga brevicollis* was searched for other components of the HIF-pathway using BLAST. No VHL homologues were found. Elongin B, which forms part of the VCB complex and is thought to be conserved throughout metazoans but not other eukaryotes as a key component of the HIF-pathway (Loenarz 2010), could not be identified in the genome of *Monosiga brevicollis* either. A FIH-like sequence was identified (Loenarz 2010), but has failed to generate any hits in screening with either FIH-substrates (ankyrins, CAD) or PHD-substrates (ODD sequences).

C.6.2. Possible strategies

C.6.2.1. Bioinformatic approach

In the face of the unsuccessful attempts at identifying a candidate for the natural substrate of mbP4H by BLAST-searches, two parallel strategies for further attempts are outlined. The first strategy involves hidden Markov model (HMM) searches of the *Monosiga brevicollis* genome using a search model based on the sequences of ODDs. A flaw of this approach is that only one of the ODDs tested was recorded to be a substrate, meaning that a model built from ODD-sequences may be based on wrong or incomplete assumptions.

C.6.2.2. Immunoprecipitation assay

The second strategy is to 'trap' the native substrate from a cell lysate of *Monosiga brevicollis* using mbP4H as the 'bait' and identifying it using MS/MS and protein fingerprinting techniques. Since no transfection techniques have been described with *Monosiga brevicollis* to this date, the experimental approach involves overexpression of affinity-tagged mbP4H in *E. coli*, followed by exposure of the mbP4H to *Monosiga brevicollis* cell lysate. The main drawbacks of this approach are that the interactions between the mbP4H and its substrate are likely to be of transient nature. Additionally, the substrate could be present at very low concentrations on the protein level, similar to HIFα in normoxia.

C.6.3. ODD-like sequences of mbP4H found by HMM searches

Bioinformatic searches of the *Monosiga brevicollis* genome using a Hidden Markov model as described above yielded several hits (performed by C. Loenarz). Six of the highest-scoring sequences were prepared using a commercial peptide synthesiser and screened for hydroxylation, however failed to yield any hits. This suggests that the

natural substrate bears less similarity to the ODD-sequences than the ODD-sequences to each other and is consistent with the hypothesis that *Monosiga brevicollis* does not possess a HIF-pathway.

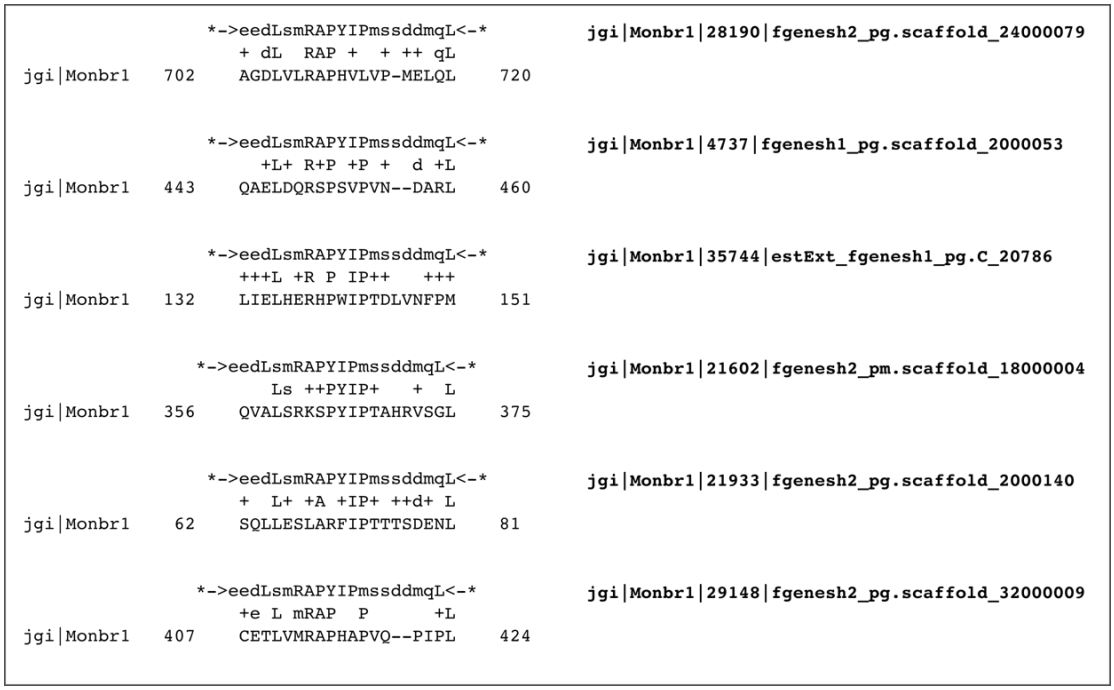


Figure C.6.1. Highest-scoring hits found using HMM searches based on metazoan ODD sequences. These six peptides were prepared but tested negative for hydroxylation by mbP4H.

C.6.4. Preparation of FLAG-tagged mbP4H construct for pull-down experiments

The 3xFLAG tag (DYKDHDGDYKDHDIDYKDDDDK) was chosen as the affinity tag, because it is relatively hydrophilic compared to other common tags and therefore less likely to denature or inactivate proteins to which it is appended. A vector for bacterial expression of 3xFLAG-tagged constructs was prepared by PCR, ligation and sub-cloning strategies (Inga Pfeffer). This vector is identical to pET-28a(+), with the 3xFLAG sequence encoded in place of the *N*-terminal 6xHis sequence. The mbP4H sequence was transferred into the 3xFLAG-pET28a(+) vector by sub-cloning. Expression trials

suggested that expression at 30°C for 12 hours is a suitable condition for protein expression.

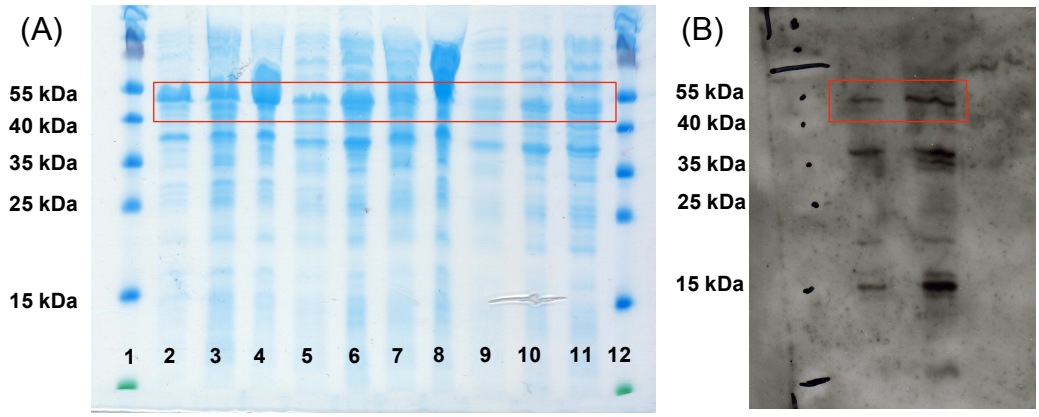


Figure C.6.2. SDS-PAGE gel of expression trials of 3xFLAG-tagged mbP4H (A) and Western blot of cell lysates (B). From left to right (A): 1. Marker; 2. 37°C, no induction, soluble fraction and 3. insoluble fraction; 4. 37°C, induction with 0.5 mM IPTG, soluble fraction and 5. insoluble fraction; 6. 30°C, no induction, soluble and 7. insoluble fractions; 8. 30°C, induction with 0.5 mM IPTG, soluble fraction and 9. insoluble fraction; 10. 18°C, induction with 0.5 mM IPTG, soluble fraction and 11. insoluble fraction; 12. marker. All samples were expressed for 12 hours. Similar to 6xHis-tagged mbP4H, the 41 kDa protein has an apparent mass of around 50 kDa on a 12% denaturing SDS-PAGE gel. (B) The Western blot was carried out using mouse anti-FLAG as the primary antibody and anti-mouse as the secondary antibody (both used at 1:10000, see Materials and Methods 18). From left to right: marker; soluble fraction of cell lysate from expression at 30° C for 4 hours; 8 hours. Apart from the band between 40 kDa and 55 kDa, which is assumed to be FLAG-mbP4H, several lower mass bands were stained. These bands are most likely degradation products of FLAG-mbP4H and could have arisen from overly harsh sonication of the lysed cells.

A high level of 'leaky', i.e. non-induced expression was also detected. It may be therefore of benefit to switch from BL21 (DE3) to a different cell line that is designed to minimise leaky expression, for example BL21 (DE3) pLysS/pLysE.

C.6.5. Possible future work for substrate identification

The 3xFLAG-tagged mbP4H construct can be used for immunoprecipitation experiments with *Monosiga brevicollis* cell lysate. This could involve culturing *Monosiga brevicollis*, lysing the protist cells and incubating the lysate with mbP4H immobilised on anti-FLAG beads. The incubation needs to be followed by washing steps, elution of immobilised mbP4H and any proteins bound to it by SDS-PAGE sample buffer and running the

sample on an SDS-PAGE gel. Any bands visible through coomassie staining will need to be excised, digested with trypsin in-gel and analysed by MS/MS techniques for identification.

To obtain the *Monosiga brevicollis* cell lysate, this protist needs to be cultured. Frozen *Monosiga brevicollis* culture can be obtained from ATCC (www.atcc.org), while detailed instructions of how to reconstitute and maintain such a culture can be found online in the choano wiki (www.choano.org). Prior to cell lysis, the bacterial load of the culture needs to be minimised in order to avoid unnecessary artefacts in the form of bacterial proteins interacting with mbP4H. Several strategies at reducing the bacterial load in *Monosiga brevicollis* cultures are also outlined in the choano wiki (www.choano.org). Prior to incubation of the lysate with immobilised mbP4H, a protease inhibitor should be added to prevent the degradation of proteins. Dimethyl-oxalyl-glycine (DMOG), an inhibiting 2OG analogue stable at physiological conditions, should also be used to prevent the turnover of the mbP4H substrate and stabilise the enzyme-substrate interactions. Any identified hits will need to be followed up thoroughly to exclude the possibility of artefacts.

While immunoprecipitation experiments with mbP4H are very interesting and could indeed lead to the identification of the native substrate, the drawbacks of this approach, described in Section 6.2.2, may lead to poor and/or inconclusive results. Another possible approach is to prepare variants of mbP4H and taODD to modulate the substrate range of mbP4H as well as the associated turnover rates. The information obtained through such an approach could be used to build a more suitable model for HMM searches.

C.7. Summary of Section C

Section C describes initial studies of mbP4H, a PHD-like enzyme from the choanoflagellate *Monosiga brevicollis*. Since choanoflagellates are thought to be the closest living unicellular relative of metazoans, mbP4H is likely to be an enzyme closely related to the metazoan PHDs. The study of mbP4H could therefore yield information about its last common ancestor with the metazoan PHDs and reveal clues about the origin of the function of the PHDs as oxygen sensors.

A sequence alignment of mbP4H with hsPHD2 and taPHD shows that the 2OG- and Fe(II)-binding residues and the substrate-binding *C*-terminal arginine are conserved in mbP4H and that it also contains a region resembling the β2/β3 finger-loop region. Key differences are that no MYND domain was found in mbP4H and that the protein sequence contains two long inserts, the first one *C*-terminal to the β2/β3 finger-loop region and the second one in the βIV/βV region. Since structural modelling suggested that the βIV/βV region is positioned along the substrate-binding groove, it was proposed that it could be involved in interactions with the substrate.

To study the biochemical properties of mbP4H, recombinant *E. coli* expression of a full-length mbP4H construct, obtained synthetically based on the predicted structure (Loenarz 2010), was carried out and followed by affinity tag purification. The activity of the purified protein was assessed by measuring turnover of 2OG to succinate in the absence of substrate. The CODD-homolog peptides described in Section B.3. and their amino acid substitution variants (see Section B.4.1.) were tested for hydroxylation by mbP4H. Interestingly, taODD and the taODD variant Y481L were found to be turned over by mbP4H. Tandem mass spectrometry studies and amino acid analysis of taODD incubated with mbP4H, Fe(II), 2OG and ascorbate have shown that mbP4H catalyses the same reaction as taPHD and hsPHD2, (4*R*)-prolyl-hydroxylation of P486. Comparing the

initial rates of 2OG turnover and taODD hydroxylation has shown that the coupling ratio between these two processes is 4.9:1. It was therefore proposed that the native substrate of mbP4H is a protein that contains a prolyl-hydroxylation site similar to taODD, possibly with a YXXLAP motif.

To obtain more information about the interactions between mbP4H and taODD, targeted substitutions of mbP4H residues were carried out. The residue between the catalytic site histidine and aspartic acid residues (HXD motif) is thought to be conserved as an aliphatic valine or isoleucine throughout the PHD sequences. mbP4H, however, has a cysteine at the equivalent position, C359. In Section B it was proposed that the aliphatic X in the HXD motif forms part of a hydrophobic pocket that binds taODD Y481/CODD L559. Therefore an mbP4H C359V variant was prepared. Peptide hydroxylation assays with this variant have shown that while the amount of taODD turned over was lower compared to the *wild-type* mbP4H, the relative amount of taODD Y481L variant hydroxylation was increased by 40%. The zebra fish hsCODD sequence DANRE was also found to be turned over by the mbP4H C359V variant, although slowly compared to taODD turnover. These results demonstrate that the HXD region is key to substrate selectivity and that the substrate range of mbP4H could be modified by substitutions of C359. It was also proposed that mbP4H might interact with the *N*-terminal acidic residues of taODD, taODD Y481L and DANRE. Overall, these results suggest that mbP4H is a non-metazoan PHD-like enzyme closely related to the PHDs.

Two β2/β3 finger-loop residue substitutions, P238D and G255R were also carried out. P239D was based on the observation that a polar residue *N*-terminal to the β2/β3 finger-loop is conserved. This variant turned over taODD more slowly than the *wild-type*. The variant G256R is based on the conservation of a large polar residue the end of the β2/β3 finger-loop sequence throughout the PHD sequences examined in Section B.4.1.2.

This variant has not shown any deviation in hydroxylation activity and selectivity compared to *wild-type* mbP4H.

It was attempted to obtain a crystal structure of mbP4H in order to carry out a structural comparisons between hsPHD2, taPHD and mbP4H. mbP4H was crystallised in the presence of NOG. A data set at 1.8 Å was collected at a synchrotron source, but the structure could not be solved by molecular replacement, possibly because of the long inserts present in mbP4H or because of slight variations in DSBH geometry. Seleno-methionine mbP4H was prepared and crystallised, but despite the collection of complete MAD and SAD datasets, the structure could not be solved.

The studies of mbP4H suggest that the precursor of PHDs may have had very similar properties to mbP4H. Finding out more about the structural and biochemical properties of mbP4H would therefore yield clues about the evolutionary origin of HIF-prolyl hydroxylases in oxygen sensing and probe the previously proposed connection between metazoan life and HIF oxygen sensing.

C.8. References

Bonner, J. & Lamont, D., 2005. Behavior of cellular slime molds in the soil. *Mycologia*, 97, pp.178–184.

Dinchuk, J. E. et al., 2000. Aspartyl beta -hydroxylase (Asph) and an evolutionarily conserved isoform of Asph missing the catalytic domain share exons with junctin. *The Journal of biological chemistry*, 275(50), pp.39543-54. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10956665 [Accessed April 30, 2012].

Dinchuk, J. E. et al., 2002. Absence of post-translational aspartyl beta-hydroxylation of epidermal growth factor domains in mice leads to developmental defects and an increased incidence of intestinal neoplasia. *The Journal of biological chemistry*, 277(15), pp.12970-7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11773073 [Accessed April 30, 2012].

Ealick, S. E., 2000. Advances in multiple wavelength anomalous diffraction crystallography. *Current Opinion in Chemical Biology*, 4, pp.495-499.

Ferguson, J. E. et al., 2007. ASB4 is a hydroxylation substrate of FIH and promotes vascular differentiation via an oxygen-dependent mechanism. *Molecular and cellular biology*, 27(18), pp.6407-19. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2099627&tool=pmcentrez&rendertype=abstract [Accessed April 30, 2012].

Flashman, E. et al., 2007. Kinetic Rationale for Selectivity toward N- and C-terminal Oxygen-dependent Degradation Domain Substrates Mediated by a Loop Region of Hypoxia-Inducible Factor Prolyl Hydroxylases. *Journal of Biological Chemistry*, 283(7), pp.3808-3815. Available at: http://www.jbc.org/cgi/doi/10.1074/jbc.M707411200 [Accessed May 5, 2012].

Hauptman, H., 1997. Phasing methods for protein crystallography. *Current Opinion in Structural Biology*, 7(5), pp.672-680. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0959440X97800772.

King, N. et al., 2008. The genome of the choanoflagellate Monosiga brevicollis and the origin of metazoans. *Nature*, 451(7180), pp.783-8. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2562698&tool=pmcentrez&rendertype=abstract [Accessed March 5, 2012].

Leite, R. B., Brito, A. B. & Cancela, M. L., 2008. An oxygen molecular sensor, the HIF prolyl 4-hydroxylase, in the marine protist Perkinsus olseni. *Protist*, 159(3), pp.355-68. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18539525 [Accessed March 22, 2012].

Loenarz, C., 2010. *Mechanistic and Evolutionary Studies on Oxygen Sensing in Animals*. Elsevier, Inc.

Loenarz, C. et al., 2011. The hypoxia-inducible transcription factor pathway regulates oxygen sensing in the simplest animal, Trichoplax adhaerens. *EMBO reports*, 12(1),

pp.63-70. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21109780 [Accessed April 19, 2012].

McDonough, M. A. et al., 2005. Structure of human phytanoyl-CoA 2-hydroxylase identifies molecular mechanisms of Refsum disease. *The Journal of biological chemistry*, 280(49), pp.41101-10. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16186124 [Accessed April 28, 2012].

Myllyharju, J., 2004. Collagens, modifying enzymes and their mutations in humans, flies and worms. *Trends in Genetics*, 20(1), pp.33-43. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0168952503003196 [Accessed March 6, 2012].

Nicholas, K., Nicholas, H. J. & Deerfield, D. I., 1997. GeneDoc: Analysis and Visualization of Genetic Variation. *EMBNEW. NEWS*, 4(14).

Raper, K., 1935. Dictyostelium discoideum, a new species of slime mold from decaying forest leaves. *J. Agr. Res.*, 50, pp.135-147.

Shao, Z., Zhang, Y. & Powell-Coffman, J. A., 2009. Two distinct roles for EGL-9 in the regulation of HIF-1-mediated gene expression in Caenorhabditis elegans. *Genetics*, 183(3), pp.821-9. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2778979&tool=pmcentrez&rendertype=abstract [Accessed March 19, 2012].

Singleton, C. K., Kirsten, J. H. & Dinsmore, C. J., 2006. Function of ammonium transporter A in the initiation of culmination of development in Dictyostelium discoideum. *Eukaryotic cell*, 5(7), pp.991-6. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1489295&tool=pmcentrez&rendertype=abstract [Accessed April 30, 2012].

Webby, C. J. et al., 2009. Jmjd6 catalyses lysyl-hydroxylation of U2AF65, a protein associated with RNA splicing. *Science (New York, N.Y.)*, 325(5936), pp.90-3. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19574390 [Accessed March 10, 2012].

van der Wel, H., Ercan, A. & West, C. M., 2005. The Skp1 prolyl hydroxylase from Dictyostelium is related to the hypoxia-inducible factor-alpha class of animal prolyl 4-hydroxylases. *The Journal of biological chemistry*, 280(15), pp.14645-55. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15705570 [Accessed April 28, 2012].

West, C., 2003. Comparative analysis of spore coat formation, structure, and function in Dictyostelium. *Int. Rev. Cytol.*, 222, pp.237–293.

West, C. M., Wang, Z. A. & van der Wel, H., 2010. A cytoplasmic prolyl hydroxylation and glycosylation pathway modifies Skp1 and regulates O2-dependent development in Dictyostelium. *Biochimica et biophysica acta*, 1800(2), pp.160-71. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2873859&tool=pmcentrez&rendertype=abstract [Accessed April 28, 2012].

West, C. M., van der Wel, H. & Wang, Z. A., 2007. Prolyl 4-hydroxylase-1 mediates O2 signaling during development of Dictyostelium. *Development (Cambridge, England)*,

134(18), pp.3349-58. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17699611 [Accessed April 28, 2012].

Willems, A. R., Schwab, M. & Tyers, M., 2004. A hitchhiker's guide to the cullin ubiquitin ligases: SCF and its kin. *Biochimica et biophysica acta*, 1695(1-3), pp.133-70. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15571813 [Accessed March 23, 2012].

Zheng, X. et al., 2008. Interaction with factor inhibiting HIF-1 defines an additional mode of cross-coupling between the Notch and hypoxia signaling pathways. *Proceedings of the National Academy of Sciences of the United States of America*, 105(9), pp.3368-73. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2265116&tool=pmcentrez&rendertype=abstract.

Materials and Methods

For all experiments, water from a MILLIPORE ELIX system sterifiltered through a 0.22 µm filter was used. All molecular biology techniques were carried out under sterile conditions, using pipette tips and tubes sterilised in a TOUCHCLAVE II autoclave (LTE SCIENTIFIC) at 121°C for 20 minutes. Bacterial plate cultures were grown at 37°C overnight in a HEREUS TYPB 6030 incubator (THERMO FISHER SCIENTIFIC). Cultures were incubated in a NEW BRUNSWICK SCIENTIFIC G25 environmental shaker at 200 rpm. Chemicals were supplied by SIGMA-ALDRICH unless otherwise stated.

1. Molecular Cloning Techniques

1.1. PCR for introduction of new restriction enzyme sites or for preparation of *N*-terminally truncated constructs

Oligonucleotide primers were synthesised to order by SIGMA-GENOSYS, obtained as lyophilised samples and resuspended in sterile water as 100 µM stocks. A typical PCR protocol is given below:

Component	**Volume in µL**
10x Taq buffer	2.5
25 mM $MgCl_2$	0.5
dNTP mix (10 mM each)	2
Primer 1	1
Primer 2	1
Taq DNA polymerase	1
Template DNA ~10 ng/µL	1
Autoclaved water	16
Total	**25 µL**

Typical cycling conditions are given below:

Step/temperature in °C	Duration
Initial denaturation/94	3 minutes
3-Step cycling (30x)	
Denaturation/94	1 minute
Annealing/50-68*	1 minute
Extension/72	1 minute
Final extension/72	**10 minutes**

*The annealing temperature depended on the primers and was chosen to be the melting temperature as estimated by the manufacturer minus 5°C.

All reagents were obtained from NEW ENGLAND BIOLABS.

1.2. Visualisation of PCR products on agarose gel

1% Agarose gels were prepared by boiling 0.4 g agarose (BIOLINE) in 40 mL 1x TAE buffer (see below for recipe), adding 4 µL SYBR SAFE DNA gel stain (INVITROGEN) and pouring the hot mixture in a gel cast (ANACHEM). Each gel was allowed to set at room temperature for 40 minutes before use. 2 µL of 6x DNA loading buffer (INVITROGEN) were added to 10 µL of sample taken directly from the PCR reaction tube. 10 µL of the resulting mixture were applied on the agarose gel immersed in TAE buffer. 4 µL of DNA ladder (1 kbp, INVITROGEN) were used for each gel. Gels were run using a BIO-RAD system at 90 V for a15 minutes. Bands were visualised by illumination of the gel with UV light using a GEL LOGIC 2000 imaging system (KODAK) and a 535 nm emission filter.

50x TAE buffer: 2 M Tris-HCl pH=8.0, 1 M acetic acid, 0.5 M EDTA

1.3. Extraction of PCR bands from agarose gel

The PCR products were purified by excision from agarose gel using a sterile scalpel and extraction using a commercial gel extraction kit (QIAGEN) according to the manufacturers' manual. DNA was eluted with 30 µL autoclaved water.

1.4. Quantification of purified DNA

The concentration of the purified PRC products was determined by measuring light absorbance at 260 nm using a UV/Vis Spectrophotometer (NANODROP).

1.5. DNA sequencing

All constructs were sequenced prior to use. DNA sequencing was performed by SOURCE BIOSCIENCE.

1.6. Ligation of PCR products into a plasmid vector

The insert and the target vector were digested with the same restriction enzymes (NEW ENGLAND BIOLABS) according to the manufacturers' protocol. The vector was dephosphorylated for 2 hours at 37°C using 1 µL alkaline phosphatase (NEW ENGLAND BIOLABS) to reduce self-ligation. The digestion products were purified by extraction from agarose gel as described in 1.2. and 1.3.

The following protocol was used for the ligation mix:

Component	**Volume in µL**
10x T4 ligase buffer	1
Digested vector ~5 ng/µL	1
Digested insert ~5 ng/µL	5
T4 DNA ligase	1
Autoclaved water	2
Total	**10 µL**

The ligation mixtures were incubated overnight at 4°C and transformed into XL10 Gold competent cells (STRATAGENE) according to the manufacturers' protocol. Before plating, the cells were pelleted by centrifugation and resuspended in 150 µL fresh 2TY medium (see below). If colonies were observed after overnight incubation of the plates, starter cultures in 100 mL 2TY medium containing the appropriate antibiotic (30 mg/L kanamycin for pET-28a(+) and 50 mg/L ampicillin for pGEX-6P-1) were set up. After 16 hours incubation at 37°C, 10 mL starter culture were used to extract plasmid DNA using a commercial plasmid extraction/miniprep kit (QIAGEN). The presence of the desired insert was confirmed by DNA sequencing as described in 1.5.

2TY medium: 16 g/L tryptone, 10 g/L yeast extract, 5 g/L sodium chloride

1.7. Sub-cloning of an insert into a different expression vector using the same restriction sites

The construct containing the desired insert and the new target vector were digested with appropriate restriction enzymes (NEW ENGLAND BIOLABS) according to the manufacturers' protocol. The insert and target vector were purified by extraction from agarose gel as described in 1.2. and 1.3. Ligation of the insert into the vector was carried out as described in 1.6. As in 1.6., DNA sequencing was used to verify the successful insertion of the desired DNA into the target vector.

1.8. Site-directed mutagenesis

taPHD and mbP4H amino acid variants were prepared by site-directed mutagenesis. Oligonucleotide primers were synthesised to order by SIGMA-GENOSYS as described in 1.1.

A typical PCR protocol is given below:

Component	**Volume in µL**
10x PFU buffer	2.5
dNTP mix (10 mM each)	2
Primer 1	1
Primer 2	1
PFU TURBO DNA polymerase	1
Template DNA ~10 ng/µL	1
Autoclaved water	16.5
Total	**25 µL**

Typical cycling conditions are given below:

Step/temperature in °C	**Duration**
Initial denaturation/94	5 minutes
3-Step cycling (18x)	
Denaturation/94	1 minute
Annealing/50-60	1 minute
Extension/68	8 minutes
Final extension/68	**10 minutes**

1 µL of the PCR product were transformed into XL10 Gold competent cells (STRATAGENE) according to manufacturers' protocol. Plating, preparation of starter cultures and verification of the constructs by sequencing were carried out as described in 1.6.

2. Protein Expression in *E. coli*

2.1. Expression trials

For expression trials, starter cultures were prepared as described in 1.6. and incubated overnight at 37°C. 1 mL of the starter cultures were used to start new cultures in fresh 2TY medium with the appropriate antibiotic (30 mg/L kanamycin for pET-type vectors and 50 mg/L ampicillin for pGEX-6P-1). The new cultures were incubated at 37°C until the OD_{600} reached 0.6 (NOVASPEC II spectrophotometer, PHARMACIA), at which point they were induced with IPTG and transferred to shakers equilibrated to the appropriate expression temperature. After overnight incubation, the cells were harvested by centrifugation (BECKMAN ALLEGRA). The cells were resuspended in 50 mM Tris-HCl (pH=7.5) and lysed by ultrasonication (MSE SONIPREP). The soluble and insoluble cell fractions were separated by centrifugation (BECKMANN) and analysed for the presence of the protein of interest by SDS-PAGE (see 3.6.).

2.2. Large-scale protein expression

To prepare large amounts of protein once a suitable expression condition was found, a 100 mL starter culture was prepared as described in 1.6. After overnight incubation at 37°C, the starter culture was distributed between 12 2000 mL PYREX narrow-mouth graduated Erlenmeyer flasks containing 600 mL 2TY medium and the appropriate antibiotic. The cultures were grown at 37°C in a shaker until the OD_{600} reached 0.6, at which point induction with IPTG was carried out and the temperature of the shaker was changed to the appropriate expression temperature. After the appropriate expression time had passed, the cells were harvested by centrifugation (BECKMANN ALLEGRA) and frozen at -80°C until purification.

3. Protein Purification

3.1. Preparation of cell lysate for FPLC

Prior to protein extraction, the cell pellets were thawed and resuspended in 50 mM Tris-HCl buffer containing DNAse I (INVITROGEN). The cells were lysed by ultrasonication (VIBRA CELL VCX 500). The soluble components were separated by centrifugation (BECKMAN ALLEGRA) and sterifiltered using 0.22 μm filters (MILLIPORE).

All proteins were purified by FPLC (ÄKTA, GE HEALTHCARE) at 4°C. Elution of proteins was monitored by UV absorbance at 280 nm (UPC-900 monitor). Elution fractions were collected using a FRAC-920 fraction collector. All columns were handled and cleaned according to the manufacturers' instructions.

3.2. His-affinity purification

The 6xHis tagged proteins expressed in the pET-28a(+) vector were purified by immobilised Ni(II) ion affinity chromatography using a 5 mL HISTRAP HP nickel affinity column (GE HEALTHCARE). Before use, the column was charged with 3 column volumes (CVs) of 400 mM $NiSO_4$, washed with 5 CVs water and equilibrated with 3 CVs loading buffer (see below for buffer recipes). The cell lysate (see 3.1.) was then loaded onto the column using the FPLC system, washed with 5 CVs wash buffer and eluted with elution buffer using a linear gradient of 0-100%. 5 mL elution fractions were collected and analysed by SDS-PAGE.

Loading buffer: 50 mM Tris-HCl pH=7.9, 0.5 M NaCl

Wash buffer: 50 mM Tris-HCl pH=7.9, 0.5 M NaCl, 4 mM imidazole

Elution buffer: 50 mM Tris-HCl pH=7.9, 0.5 M NaCl, 1 M imidazole

3.3. GST-tag purification

A glutathione-agarose column (GE HEALTHCARE) was used to purify GST-tagged taPHD. The cell pellet was resuspended and lysed as described in 3.1. but using buffer A (see below). The column was pre-equilibrated with buffer B. The cell lysate was applied to the column. The column was washed with 10 CVs of buffer C and then 10 CVs of buffer D. GST-tagged protein was eluted by applying a linear gradient (0-100%) of buffer E. 5 mL fractions were collected throughout the elution and analysed by SDS-PAGE.

Buffer A: 50 mM Tris-HCl pH=8.5, 0.1 M NaCl, 1 mM EDTA, 1 mM DTT
Buffer B: buffer A with 0.5% TRITON X-100
Buffer C: buffer A with 1% TRITON X-100
Buffer D: buffer A with 10 mM glutathione

3.4. Cation exchange

taPHD t64 has a theoretical isoelectric point pI of 9, suggesting that at pH=5.8, it would carry a net positive charge. Therefore, a cation exchange resin could be used as a crude purification step alternatively to the metal affinity purification outlined in 3.2. For cation exchange purification, the cell pellet was resuspended as described in 3.1. in buffer A (see below). The cell lysate was loaded on a 50 mL SP Sepharose column (AMERSHAP BIOSCIENCES) equilibrated in buffer A. Elution was carried out with a linear gradient (0-100%) of buffer B (see below). 5 mL fractions of the elution were collected and analysed by SDS-PAGE.

Buffer A: 100 mM MES pH=5.8
Buffer B: 100 mM MES pH=5.8, 1 M NaCl

3.5. Size exclusion chromatography

The elution fractions from the metal affinity, ion exchange or GST purification that contained the highest amount of protein were concentrated to a combined volume of 2 mL and injected into a 300 mL SUPERDEX 75 column (GE HEALTHCARE) that had been equilibrated with 50 mM Tris-HCl (1 CV, pH=7.5).

The chromatography was run in the same buffer as used for column equilibration. 5 mL fractions were collected upon detection of UV signal peaks and analysed by SDS-PAGE for purity. The fractions containing pure protein were concentrated to the desired concentration (5-20 mg/mL) using centrifuge filters of the appropriate size (MILLIPORE), divided into 5-50 µL aliquots, snap-frozen with liquid nitrogen and stored at -80°C.

3.6. SDS-PAGE analysis of purification fractions

12.5% SDS-PAGE gels were prepared by pouring mixture A (separating gel, see below) into a glass gel cast mounted in a BIO-RAD gel cast system, leaving about 1 cm space for the gel pockets. After the mixture had set, the cast was topped up with mixture B (stacking gel, see below) and the comb was inserted to form the pockets. Each gel was set for 40 minutes prior to use. 10 µL of elution fraction were mixed with 10 µL 2x SDS-PAGE loading buffer (see below) and heated at 110°C for 10 minutes to denature the proteins in the sample. The samples were loaded into the gel pockets and the gels were run in SDS running buffer at a constant potential of 200 V for 50 minutes. After the run, the gels were rinsed with water, stained in gel stain (see below) for 10 minutes and destained overnight.

12.5% Separating gel: 2 mL 30% polyacrylamide solution, 1.25 mL Tris-HCl pH=8.8, 50 µL 10% ammonium persulfate, 50 µL 10% SDS, 10 µL TEMED, 1.5 mL H_2O

5% Stacking gel: 0.8 mL 30% polyacrylamide solution, 0.6 mL Tris-HCl pH=6.8, 50 µL 10% ammonium persulfate, 50 µL 10% SDS, 10 µL TEMED, 3.4 mL H_2O

2x Sample loading buffer (50 mL): 0.379 g Tris-HCl pH=6.8, 5 g glycerol, 2.5 mL β-mercaptoethanol, 1.15 g SDS, 0.05 g bromophenol blue

SDS running buffer: 14.4 g/L glycine, 3.03 g/L Tris-HCl pH=6.8, 1 g/L SDS

Gel stain: 0.5% Coomassie Brilliant Blue, 50% methanol, 10% glacial acetic acid

Gel destain: 10% methanol, 10% glacial acetic acid.

3.7. Determination of protein concentration

The concentrations of protein in the elution fractions were determined by measuring absorption at 280 nm using an ND-1000 NANODROP spectrophotometer (THERMO SCIENTIFIC).

4. In-Gel Tryptic Digest

An in-gel tryptic digest of taPHD t64 was carried out prior to MS/MS analysis. The samples of interest were run on an SDS-PAGE gel as described in 3.6. After destaining, the gel was washed in water for 30 minutes. The bands of interest were excised using a scalpel, cut into ~1 mm^2 pieces and washed in 100 µL water, followed by 100 µL of solution A (see below), for 15 minutes each. After the washes, the liquid was removed and the gel pieces were dried by adding acetonitrile and shaking. The acetonitrile was subsequently removed and the gel pieces were rehydrated by incubation for 5 minutes in solution B. After incubation, an equal volume of acetonitrile was added and the mixture was incubated for another 15 minutes. The liquid was then removed and the gel pieces were dried by vacuum centrifugation at 50°C (BECKMAN COULTER MICROFUGE 22R).

Solution A: 50% water, 50% acetonitrile
Solution B: 0.1 M ammonium bicarbonate NH_4HCO_3
Solution C: 10 mM DTT in NH_4HCO_3
Solution D: 55 mM iodoacetamide in NH_4HCO_3
Solution E: 20 µg/mL trypsin in 50 mM NH_4HCO_3

After the gel pieces were dried, 50 µL of solution C was added to reduce cysteine residues. The mixture was incubated for 45 minutes at at 56°C. After incubation, the liquid was removed and 50 µL of solution D was added. The mixture was incubated in the dark at room temperature for 30 minutes. After incubation, all liquid was removed and the gel pieces were washed in 50 µL of 0.1 M NH_4HCO_3. The gel pieces were then washed by addition of acetonitrile as described above and rehydrated in 50 µL 0.1 M NH_4HCO_3. After 5 minutes incubation, an equal volume of acetonitrile was added and the mixture was incubated for 15 minutes. After incubation, the liquid was removed and the gel pieces were dried by vacuum centrifugation as described above. The gel pieces were then immersed in 50 µL solution E and incubated for 16 hours at 37 °C. After incubation, the liquid was transferred to a clean tube. The gel pieces were incubated for 5 minutes with shaking in 100 µL 25 mM NH_4HCO_3. After 5 minutes, an equal volume of acetonitrile was added and the mixture was incubated for further 60 minutes at room temperature. The wash was repeated twice with removal of the liquid in between and retaining it in clean tubes. Instead of NH_4HCO_3, however, 0.1% formic acid was used. All the wash liquid retained in clean tubes was combined and dried by vacuum centrifugation. The remaining powder contained the protein fragments. Prior mass spectrometric analysis, it was redissolved in 25 µL of 0.1% formic acid.

5. Activity Assay Conditions

The same conditions unless otherwise stated were used for all assays aimed at investigating enzymatic activity. These included the quantification of 2OG turnover by

^{1}H-NMR, screening for peptide substrates, determinations of initial rates of peptide hydroxylation, and determinations of K_m values. To start an assay, the appropriate volume of 35 µM enzyme stock solution was added to the reaction mixture (see below) to achieve a final enzyme concentration of 3.5 µM. Incubations were carried out at 25°C.

Reaction mixture: 300 µM 2OG, 50 µM Fe(II), 4 mM ascorbate, 50 mM Tris-HCl pH=7.5
Additional components: 100 µM peptide, 1 mM inhibitor

For each assay, fresh stock solutions were prepared. A 100 mM stock solution of 2-oxoglutaric acid disodium salt was prepared in Tris-HCl buffer and diluted to 3 mM prior to addition to the reaction mixture. A 100 mM stock solution of *L*-ascorbate sodium salt was used. A 250 mM solution of ammonium iron(II) sulfate was prepared in 20 mM HCl and then diluted to 5 mM in water prior to addition to the reaction mixture in order to prevent oxidation of Fe(II) to Fe(III). The peptide solutions were prepared once and then stored as 10 µL aliquots at -80°C. Some of the inhibitors were only soluble in the presence of DMSO (for example UN9). In such cases, the assay mixture contained a final amount of 1% DMSO. The same amount of DMSO was then added to all controls. Only pure protein (>95% purity) was used for assays. A new protein aliquot from the -80°C freezer was used for each assay. After 30 minutes on ice, a protein aliquot was discarded.

6. Quantification of 2OG Turnover by ^{1}H-NMR

All NMR experiments were conducted at 25°C using a BRUKER AVANCE AVIII 700 MHz spectrometer with an inverse TCI cryoprobe optimised for ^{1}H observation. The results were processed using TOPSPIN 2 software.

For monitoring of the 2OG and succinate concentrations as a function of time, an assay was prepared as described in 5. with the exception that 50 mM d11 Tris-DCl dissolved in D_2O was used. The assay was mixed by pipetting up and down and immediately transferred to a 3 mm NMR tube (HILGENBERG). The solvent deuterium signal was used as the internal lock signal. The spectrometer conditions were optimised before each experiment. After a fixed time of 2 minutes from the addition of enzyme to the assay, spectra were acquired continuously for the duration of the time course, separated by 1.5-minute time gaps. The 2OG and succinate peaks were integrated after background subtraction using an automated procedure. The integral ratios of the 2OG β–protons and the succinate protons were used to calculate the concentration ratios of 2OG and succinate, under consideration that the succinate signal is caused by 4H and the 2OG β-proton by 2H. The individual concentrations were calculated assuming that the sum of concentration of 2OG and succinate is constant at 300 μM (see Appendix 3.2).

7. Quantification of Peptide Hydroxylation by MALDI-TOF MS

For quantification of peptide turnover, the assay samples, carried out as described in 5. were snap-frozen in liquid nitrogen to stop the reaction. An equal amount of α-cyano-4-hydroxycinnamic acid (CHCA) was added to the frozen samples, thereby slowly defrosting them and quenching the reaction by change of pH. CHCA also acted as the crystallisation matrix for MALDI-TOF MS. The samples containing CHCA were applied onto a MALDI plate (WATERS) and left to dry for 10 minutes. The samples were then analysed using a MALDI MICRO MX mass spectrometer (WATERS) using appropriate settings for voltages and laser power, which varied from sample to sample dependent on the enzyme and peptide used. Data was processed using MASSLYNX 4.0 software.

For quantification of peptide turnover, the ion count (IC) of the peak with the mass of the peptide and the corresponding +16 Da peak was determined. Most peptides also contained an impurity appearing at -16 Da. Since this impurity also gets hydroxylated by the enzyme, the IC of the impurity was determined, if present. The fraction of turnover was calculated according to the following formula:

$$\text{Turnover fraction} = \frac{(\text{main})' / \text{TIC2}}{(\text{main})' / \text{TIC2} + (-16)' / \text{TIC2} - (-16) / \text{TIC1}}$$

Legend:

The ICs of the peaks before incubation are labelled **(main), (+16)** and **(-16)**.

The corresponding ICs after incubation are labelled **(main)'**, **(+16)'** and **(-16)'**.

The total IC before incubation, equal to **(main)+(+16)+(-16)** is labelled **TIC1**.

The total IC after incubation is labelled **TIC2** accordingly.

8. Non-Denaturing Mass Spectrometry

Protein mass spectra under non-denaturing conditions were recorded using a Q-TOF mass spectrometer (WATERS) and a NANOMATE sample injection system (ADVION BIOSCIENCES). A chip voltage of 1700 V was used. Protein samples were exchanged into 40 mM ammonium acetate buffer (pH=7.5) using 0.5 mL desalting columns (GE HEALTHCARE). The sample mixtures were prepared in 364-well plates (ABGENE) and automatically injected into the mass spectrometer by the NANOMATE. A delivery pressure of 0.25 psi was used. Different cone voltages were used (50 V – 120 V) and the data collected at different cone voltages was compared. The data was processed using MASSLYNX 4.0.

9. Electrospray Ionisation Mass Spectrometry

Electrospray Ionisation Mass Spectrometry (ESI-MS) was carried out by Dr. Adam Hardy. Protein masses were recorded by liquid chromatography/mass spectrometry (LC/MS) using a ZMD (WATERS) single quadrupole mass spectrometer, interfaced with a HEWLETT PACKARD Series 1050 liquid chromatography and sample handling system.

The protein sample (10 μM, 5 μL) was injected onto a DIONEX PROSWIFT RP-4H 1 x 50 mm Monolithic reverse phase HPLC column and eluted at 0.4 mL per minute using a gradient system from solvent A (see below) to solvent B according to the following conditions:

Time, minutes	%A	%B	Flow rate, mL/minute
0.0	100	0	0.4
1.0	100	0	
5.0	0	100	
7.0	0	100	
7.5	100	0	
10.0	100	0	

Solvent A: 95% water, 4.9% acetonitrile, 0.1% (v/v) formic acid
Solvent B: 95% acetonitrile, 4.9% water, 0.1% (v/v) formic acid

The eluent was injected directly into the mass spectrometer. The following MS parameters were used:

Capillary voltage: 3000 V
Sample cone voltage: 35 V
Extraction cone voltage: 1V;
Desolvation temperature: 250°C
Source temperature: 100°C
Cone gas flow: 100 L/hour
Desolvation gas flow (N_2): 830 L/hour

Sodium formate was used to calibrate the instrument. Spectra were processed using MASSLYNX 4.0 (WATERS) with the Maximum Entropy method (MaxEnt1).

10. Crystallisation Trials

Commercial crystallisation screens (HAMPTON RESEARCH and MOLECULAR DIMENSIONS) were used to screen for suitable crystallisation conditions. The protein samples were incubated with additives (metal ion, ligand, substrate) for 30 minutes at room temperature. Any precipitate formed during that time was separated by centrifugation for 15 minutes at 4°C at 10k rpm (BECKMANN COULTER MICROFUGE 22R). An automated dispensing robot (PHENIX) was used to prepare round-well sitting drop plates (HAMPTON RESEARCH). The plates were incubated at either 4°C or 20°C and inspected manually every 2 days for crystals. Hits were observed for a week. If no further crystal growth was detected, the crystals were harvested and irradiated in-house for initial data collection (see 11.).

11. Harvesting of Crystals, Data Collection and Data Processing

Crystal harvesting and data collection was carried out by Dr Michael McDonough. Harvesting loops (HAMPTON RESEARCH) of the appropriate size were used to retrieve the crystals, which were immersed in cryosolvent, typically the well solution of the hit conditions and 20-25% glycerol, and snap-frozen with liquid nitrogen.

For in-house data collection, the crystals in their loops were mounted in an OXFORD CRYOSYSTEMS COBRA cryostream at 100 K. Irradiation was carried out using a RIGAKU FR-E+ SUPERBRIGHT X-ray generator (CuKα rotating anode at 1.542 Å). Diffraction was detected by a RIGAKU SATURN 944+ detector.

Synchrotron data was collected at Diamond at the beamlines specified in the main text.

Crystallographic data was processed using HKL-2000. Structure solution was carried out by molecular replacement using PHASER and the search models with the PDB IDs 2G19 and 3HQR. Crystallographic structure refinement was carried out using PHENIX 1.7.3. COOT was used for model re-building between refinement cycles. Refinement was carried out until no further improvement in R_{free} could be achieved.

12. Peptide Synthesis and Purification

12.1. Synthesis

Peptides containing 19-21 amino acids were synthesised by solid phase peptide synthesis (SPPS) using automated peptide synthesisers (CS336S and CS336X synthesisers, CS BIO). Synthesis was carried out in a sequential manner, building the sequences from the *C*-termini. 4-(2',4'-Dimethoxyphenyl-Fmoc-aminomethyl)-phenoxyacetamido-norleucyl-MBHA resin (100-200 mesh) was reacted with N^{α}-Fmoc protected amino acids, which were activated with diisopropylcarbodiimide and hydroxybenzotriazole by incubation for 30 minutes before addition to the resin. After incubation with shaking for 3 hours, the solution phase reactants were drained from the reaction vessel and the resin was treated with a solution of 20% piperidine in DMF to remove the N^{α}-Fmoc group. The next activated amino acid was then added. These steps were repeated until all amino acids were coupled. After coupling of the last amino acid, the *N*-terminal Fmoc group was removed by piperidine treatment and the resin was washed with DCM and dried in a vacuum dessicator overnight.

The peptide-resin bond and peptide side chain protection group bonds were cleaved by 3-hour-long incubation with 4 mL of cleavage mix (97.5% TFA, 2.5% triisopropylsilane).

After evaporation of the cleavage mix with nitrogen, the peptides were precipitated with diethyl ether. The ether was subsequently decanted and the peptides were redissolved in 2 mL a 1:1 mix of water and acetonitrile. The solution was freeze-dried overnight, yielding a crude peptide solid. The peptide was further purified by HPLC (see 12.2.).

Amino acids used were purchased from CS BIO or NOVABIOCHEM. Piperidine was purchased from MERCK Chemicals. Diisopropylcarbodiimide and TFA were purchased from ALDRICH. Hydroxybenzotriazole was purchased from AGTC BIOPRODUCTS. Solvents (DMF, DCM) were purchased from RATHBURN Chemicals and SIGMA-ALDRICH, respectively.

12.2. HPLC purification of peptides

The peptides were purified by preparative reverse-phase HPLC using an ULTIMATE300 HPLC (DIONEX) and a VYDAC 218TP C18 10-15u column (GRACE DAVISON DISCOVERY SCIENCES). A typical HPLC protocol consisted of column equilibration in 20% HPLC-grade acetonitrile and 80% water (both containing 0.1% formic acid), followed by injection of 2 mg of peptide sample dissolved in 2 mL acetonitrile-water mix (1:1). A linear gradient from 20% to 80% acetonitrile was applied for 40 minutes for peptide elution. The UV absorption at 220 nm was monitored throughout the run and peaks were collected manually in 10 mL fractions. The fractions were freeze-dried, redissolved in water and analysed by MALDI-TOF MS for purity. Fractions containing pure peptide were combined.

13. MS/MS Analysis of Peptides

The peptide was incubated overnight with taPHD or mbP4H in the presence of Fe(II), 2OG and ascorbate (standard assay conditions, see 5.). After incubation, the enzyme was precipitated with methanol and formic acid and removed from the solution. The resulting sample was analysed using a nanoAcquity-Synapt-HDMS system with NANOLOCKSPRAY (WATERS). The peptide sample was injected and trapped on a 5 µm symmetry C18 column (180 µm x 20 mm) and washed for 1 minute at 15 µL/minute with mobile phase A (0.1% formic acid). The peptide was eluted for MS/MS analysis using a 30 minute reverse phase gradient at 400 nL/minute (0.1–60% acetonitrile over 30 minutes) on a BEH 130 C18 1.7 µm particle size 75 µm x 150 mm nanoAcquity UPLC column. The column temperature was set to 30°C. The reference for the NANOLOCKSPRAY was set to the ^{13}C peak of reserpine at a concentration of 3 mg/L flowing at 20 µL/minute with a splitter. This was constantly infused and sampled at 10-second intervals.

The eluted peptide MS/MS spectra were acquired by Synapt-HDMS (Q-TOF). The sample was analysed in positive V mode with a mass range of 140–2300 m/z and a scan time of 1 second. The on-line eluted peptides were analysed at low collision energy (6 eV) and high energy (35 V). Spectra were processed using BIOLYNX (WATERS).

14. Chromatographic Separation of the Aminoquinolyl Carbamate Derivatives of Amino Acids

Amino acid analysis was carried out by Nikita Loik. For the separation of derivatised

amino acids, an AccQTag Ultra reversed phase C18 column (2.1×100 mm) with 1.7µm particles (WATERS) was used according to the protocol outlined below:

Phase	Composition in (v/v)			
	H_2O	Solution A	Acetonitrile	Formic acid
A	95%	5%	-	-
B	-	-	98.7%	1.3%

A flow rate 0.7 mL/minute was used.

Separation was achieved by applying a linear one-step gradient was applied:

Time, minutes	%A	%B	Function
0.0	100	0	Separation
25.0	75	25	
25.5	40	60	Column cleaning
27.5	40	60	
28.0	100	0	Column reconditioning
30.0	100	0	

The column temperature was kept at 55°C. The sample injection volume was 1 µL. UV detection was carried out at 260 nm. The scan range of 250–800 Da was used. The amino acids were identified by comparison of their elution times to the elution times of standards.

15. Determination of K_m Values and Initial Rates

15.1. K_m values

For apparent K_m value determination, end-point peptide turnover assays at varying concentrations of the variable substrate (peptide or 2OG) and a fixed concentration of other substrates (peptide or 2OG) and co-factors (ascorbate, Fe(II)) were carried out as described in 5. The assay times were chosen so that 20-40% of peptide was turned over at the analysis time point. Peptide turnover was quantified by MALDI-TOF MS as

described in 7. The amount of peptide turned over was plotted against the concentration of the varying substrate. The resulting curve was fitted using the non-linear fitting function in GRAPH PAD PRISM, which also automatically determined the apparent K_m and V_{max} values.

15.2. Initial rates

To determine initial peptide turnover rates, time course assays were carried out and anaysed by MALDI-TOF MS as described in 7. The amount of peptide turned over was plotted against the assay time. By inspection of the shape of the resulting curve, a cut-off time was chosen, after which the curve shape deviated too strongly from the linear regime for initial rate determinations. The data points collected prior to this cut-off time were used for fitting of a straight line going through the origin. The slope of the straight line corresponded to an approximate initial rate of peptide turnover.

16. Preparation of Seleno-Methionine Substituted Protein

Protein containing seleno-methionine instead of methionine at all positions was prepared by expression in *E. coli* cultured in a medium containing seleno-methionine but no methionine. A commercial medium lacking methionine was purchased from MOLECULAR DIMENSIONS (SELENOMET MEDIUM BASE PLUS NUTRIENT MIX). *L*-Seleno-methionine was purchased from SIGMA-ALDRICH. The medium was prepared and autoclaved according to the manufacturers' instructions. *L*-seleno-methionine was sterifiltered through a 0.22 µm filter (MILLIPORE) into the sterile medium to yield a final concentration of 50 mg/L. Growth and expression of the protein was carried out as

described in 2. but using seleno-methionine containing medium for the starter culture and the large scale growth. Protein purification was carried out as described in 3.

17. Soaking Crystals in Heavy Atom Solutions

For heavy atom soaking, solutions equivalent to the well solution that the crystals were obtained in but containing 25% glycerol and 1 mM $HgCl_2$ or 0.1 mM K_2PtCl_4 were prepared. Crystals were harvested using loops (HAMPTON RESEARCH) and immersed in the soaking solution for varying amounts of time, typically 5, 10, 20 and 25 minutes. After soaking, some crystals were back-soaked in solution equivalent to the soaking solution but without the heavy atom. The crystals were then snap-frozen in liquid nitrogen.

18. Western Blotting

Protein samples for Western blots and a protein ladder (PAGERULER, FERMENTAS) were run on a denaturing SDS-PAGE gel as described in 3.6. but omitting the staining and destaining steps. After running, the gel was blotted in transfer buffer (see below) onto a HYDBOND ECL membrane (GE HEALTHCARE) using a MINI TRANS BLOT electrophoretic transfer cell system (BIO-RAD). The blotting was run 200 mA for 2-3 hours on ice. Protein transfer was verified by staining the membrane with Ponceau S dye. The membrane was subsequently destained in water. Blocking of the membrane was achieved by incubation with shaking in blocking buffer (see below) for 1 hour at room temperature. Primary and secondary antibodies were prepared to the appropriate dilution in 1 mL blocking buffer and applied to the membrane for 1 hour each at room temperature with rinsing. After each antibody incubation the membrane was washed

extensively in TBST buffer. After the last wash, ECL Detection Reagents (GE HEALTHCARE) were applied (1 mL of each reagents per membrane). The membranes were then incubated for 1 minute, immediately exposed to AMERSHAM HYPERFILM ECL film (GE HEALTHCARE) and developed using a COMPACT X4 (XOGRAPH) film developer.

Transfer buffer: 2.5 mM Tris, 192 mM glycine, 20 % MeOH
Blocking buffer: 5 % milk in 1 x Tris Buffered Saline 0.1 % Tween (TBST; 2
TBST: 5 mM Tris, 137 mM NaCl, 2.68 mM KCl, 0.1 % Tween 20; pH 8.8)

Appendix 1: Sequences, Vectors and Primers

1.1. hsPHD2 Full-Length Protein Sequence

```
  10         20         30         40         50         60
MANDSGGPGG PSPSERDRQY CELCGKMENL LRCSRCRSSF YCCKEHQRQD WKKHKLVCQG

        70         80         90        100        110        120
SEGALGHGVG PHQHSGPAPP AAVPPPRAGA REPRKAAARR DNASGDAAKG KVKAKPPADP

       130        140        150        160        170        180
AAAASPCRAA AGGQGSAVAA EAEPGKEEPP ARSSLFQEKA NLYPPSNTPG DALSPGGGLR

       190        200        210        220        230        240
PNGQTKPLPA LKLALEYIVP CMNKHGICVV DDFLGKETGQ QIGDEVRALH DTGKFTDGQL

       250        260        270        280        290        300
VSQKSDSSKD IRGDKITWIE GKEPGCETIG LLMSSMDDLI RHCNGKLGSY KINGRTKAMV

       310        320        330        340        350        360
ACYPGNGTGY VRHVDNPNGD GRCVTCIYYL NKDWDAKVSG GILRIFPEGK AQFADIEPKF

       370        380        390        400        410        420
DRLLFFWSDR RNPHEVQPAY ATRYAITVWY FDADERARAK VKYLTGEKGV RVELNKPSDS

VGKDVF
```

1.2. hsPHD2 Catalytic Domain Only (t181) Protein Sequence (Used for assays in Section B)

```
       190        200        210        220        230        240
PNGQTKPLPA LKLALEYIVP CMNKHGICVV DDFLGKETGQ QIGDEVRALH DTGKFTDGQL

       250        260        270        280        290        300
VSQKSDSSKD IRGDKITWIE GKEPGCETIG LLMSSMDDLI RHCNGKLGSY KINGRTKAMV

       310        320        330        340        350        360
ACYPGNGTGY VRHVDNPNGD GRCVTCIYYL NKDWDAKVSG GILRIFPEGK AQFADIEPKF

       370        380        390        400        410        420
DRLLFFWSDR RNPHEVQPAY ATRYAITVWY FDADERARAK VKYLTGEKGV RVELNKPSDS

VGKDVF
```

1.3. taPHD Predicted Sequence (TRIADDRAFT_50295)

```
        10         20         30         40         50         60
MVVNQLNTNG FCIINNFLGS SCSTEVLQQV LNLYQSGVFS NGQLARNVSV NRIRGDKIAW

        70         80         90        100        110        120
IGGDERGCEA IKYLSSCVDS LISRCNGRLG NYMITGRTKC MVACYPGSGL GYIRHIDNPN

       130        140        150        160        170        180
RDGRCVTVLY YLNPNWNSQD CGGQLWLYPN NENKVVKIDP IFDRLLLFWS DRRNPHEVKP

       190        200
AYAMRYAITL WYFDEKERAL SSQNGT
```

1.4. taPHD Full-Length Protein Sequence
(Suggested by C. Loenarz)

```
        10         20         30         40         50         60
MASCNEIPLC ESVTDGHLLV DQDAYCSLCF KKLISVKCSL CQSAYYCSKE HQQSHWKIHR

        70         80         90        100        110        120
ENCMDERRPK IGCDKTGLAL EKTWQLQKNE RLSNMVVNQL NTNGFCIINN FLGSSCSTEV

       130        140        150        160        170        180
LQQVLNLYQS GVFSNGQLAR NVSVNRIRGD KIAWIGGDER GCEAIKYLSS CVDSLISRCN

       190        200        210        220        230        240
GRLGNYMITG RTKCMVACYP GSGLGYIRHI DNPNRDGRCV TVLYYLNPNW NSQDCGGQLW

       250        260        270        280        290        300
LYPNNENKVV KIDPIFDRLL LFWSDRRNPH EVKPAYAMRY AITLWYFDEK ERALSSQNGT
```

The residues S177, P213 and R292 are highlighted in blue.

1.5. taPHD t64 6xHis-Tagged
(Used throughout Sections A and B)

```
MGSSHHHHHHSSGLVPRGSH (His-tag and thrombin cleavage site)

        70         80         90        100        110        120
   MDERRPK IGCDKTGLAL EKTWQLQKNE RLSNMVVNQL NTNGFCIINN FLGSSCSTEV

       130        140        150        160        170        180
LQQVLNLYQS GVFSNGQLAR NVSVNRIRGD KIAWIGGDER GCEAIKYLSS CVDSLISRCN

       190        200        210        220        230        240
GRLGNYMITG RTKCMVACYP GSGLGYIRHI DNPNRDGRCV TVLYYLNPNW NSQDCGGQLW

       250        260        270        280        290        300
LYPNNENKVV KIDPIFDRLL LFWSDRRNPH EVKPAYAMRY AITLWYFDEK ERALSSQNGT
```

1.6. taPHD t67 6xHis-Tagged

```
MGSSHHHHHHSSGLVPRGSH (His-tag and thrombin cleavage site)

        70         80         90        100        110        120
      RRPK IGCDKTGLAL EKTWQLQKNE RLSNMVVNQL NTNGFCIINN FLGSSCSTEV

       130        140        150        160        170        180
LQQVLNLYQS GVFSNGQLAR NVSVNRIRGD KIAWIGGDER GCEAIKYLSS CVDSLISRCN

       190        200        210        220        230        240
GRLGNYMITG RTKCMVACYP GSGLGYIRHI DNPNRDGRCV TVLYYLNPNW NSQDCGGQLW

       250        260        270        280        290        300
LYPNNENKVV KIDPIFDRLL LFWSDRRNPH EVKPAYAMRY AITLWYFDEK ERALSSQNGT
```

1.7. taPHD t72 6xHis-Tagged

```
MGSSHHHHHHSSGLVPRGSH (His-tag and thrombin cleavage site)

                   80         90         100        110        120
            GCDKTGLAL EKTWQLQKNE RLSNMVVNQL NTNGFCIINN FLGSSCSTEV

       130        140        150        160        170        180
LQQVLNLYQS GVFSNGQLAR NVSVNRIRGD KIAWIGGDER GCEAIKYLSS CVDSLISRCN

       190        200        210        220        230        240
GRLGNYMITG RTKCMVACYP GSGLGYIRHI DNPNRDGRCV TVLYYLNPNW NSQDCGGQLW

       250        260        270        280        290        300
LYPNNENKVV KIDPIFDRLL LFWSDRRNPH EVKPAYAMRY AITLWYFDEK ERALSSQNGT
```

1.8. taPHD t84 6xHis-Tagged

```
MGSSHHHHHHSSGLVPRGSH (His-tag and thrombin cleavage site)

                           90         100        110        120
                      WQLQKNE RLSNMVVNQL NTNGFCIINN FLGSSCSTEV

       130        140        150        160        170        180
LQQVLNLYQS GVFSNGQLAR NVSVNRIRGD KIAWIGGDER GCEAIKYLSS CVDSLISRCN

       190        200        210        220        230        240
GRLGNYMITG RTKCMVACYP GSGLGYIRHI DNPNRDGRCV TVLYYLNPNW NSQDCGGQLW

       250        260        270        280        290        300
LYPNNENKVV KIDPIFDRLL LFWSDRRNPH EVKPAYAMRY AITLWYFDEK ERALSSQNGT
```

1.9. taPHD Full-Length GST-Tagged

```
MPPYTVVYFPVRGRCAALRMLLADQGQSWKEEVVTVETWQEGSLKASCLYGQLPKFQDGD
LTLYQSNTILRHLGRTLGLYGKDQQEAALVDMVNDGVEDLRCKYISLIYTNYEAGKDDYV
KALPGQLKPFETLLSQNQGGKTFIVGDQISFADYNLLDLLLIHEVLAPGCLDAFPLLSAY
VGRLSARPKLKAFLASPEYVNLPINGNGKQ LEVFQGPLGS

(Glutathione S-transferase tag with cleavage site, 220 amino acids)

         10         20         30         40         50         60
MASCNEIPLC ESVTDGHLLV DQDAYCSLCF KKLISVKCSL CQSAYYCSKE HQQSHWKIHR

         70         80         90        100        110        120
ENCMDERRPK IGCDKTGLAL EKTWQLQKNE RLSNMVVNQL NTNGFCIINN FLGSSCSTEV

       130        140        150        160        170        180
LQQVLNLYQS GVFSNGQLAR NVSVNRIRGD KIAWIGGDER GCEAIKYLSS CVDSLISRCN

       190        200        210        220        230        240
GRLGNYMITG RTKCMVACYP GSGLGYIRHI DNPNRDGRCV TVLYYLNPNW NSQDCGGQLW

       250        260        270        280        290        300
LYPNNENKVV KIDPIFDRLL LFWSDRRNPH EVKPAYAMRY AITLWYFDEK ERALSSQNGT
```

1.10. mbP4H Predicted (JGI 27914)

```
        10         20         30         40         50         60
MIHQHRQQER ERESQREKKS HCMQAGMHCR YMLTIDLSDM YWSDIILFLL SRCLSLSLSL

        70         80         90        100        110        120
SLSLSLSLSL SLGCWQGEWA RGARHGRHRK ALKKALDGVE RALTQPKSST PDTALKRCVE

       130        140        150        160        170        180
LGQQLGCVEE MAAALTEIAQ AFQGTQPAPV ASSAAAAAAA AKQQPHWQHD LALDSTNPRY

       190        200        210        220        230        240
AHVSQLLTEA PDVAEYAHGL ALTLQRQHYV VVDRVLDDHV LRQFIHEVKS AHRDGNLKPG

       250        260        270        280        290        300
QLAGGRAGAN TKYVLGHVRG DQVGWFNQDP CHDSSSGRAE LAITEETASD PTPTTESAQA

       310        320        330        340        350        360
DRSQPRCWQG SGLRAFTSRM NTLVQLIAAF VPELQRIESR SEAMITCYPG NSTRYIRHCD

       370        380        390        400        410        420
NPHRNGRKLT VLFYLNEEWQ PGDGGEVRLF VTPDEHQHAA APADSDDETE ASASPTPSGS

       430        440        450        460        470        480
VEQAVTIPPR ANTMVIFFSD RRVPHEVLPT NVPRFAATHW FYDADERAEA EASAGTAFEL

       490        500        510        520        530        540
DQQIEEERIR REIEKFEATQ AQAQMVPAKN QVNSPRDEPM ASDVIAAAAT GGASHNKPAM

       550        560        570        580        590        600
TAATTAAAST TADLGMTSDS GRDEATPVSA PVGSPIESAS DSPGHTPTAW SHRVQKSDDA

       610        620        630        640        650        660
LIVRVDTDGL PVEVHVDNEQ YISITVKSQA PGTLSTRSPL VLDAKALGMD ATTLSVKHIA

       670
KSGVIKLKAQ ALRPGNT
```

1.11. mbP4H Residues 131-484 6xHis-Tagged

```
MGSSHHHHHHSSGLVPRGSH (His-tag and thrombin cleavage site)

               140        150        160        170        180
           MAAALTEIAQ AFQGTQPAPV ASSAAAAAAA AKQQPHWQHD LALDSTNPRY

       190        200        210        220        230        240
AHVSQLLTEA PDVAEYAHGL ALTLQRQHYV VVDRVLDDHV LRQFIHEVKS AHRDGNLKPG

       250        260        270        280        290        300
QLAGGRAGAN TKYVLGHVRG DQVGWFNQDP CHDSSSGRAE LAITEETASD PTPTTESAQA

       310        320        330        340        350        360
DRSQPRCWQG SGLRAFTSRM NTLVQLIAAF VPELQRIESR SEAMITCYPG NSTRYIRHCD

       370        380        390        400        410        420
NPHRNGRKLT VLFYLNEEWQ PGDGGEVRLF VTPDEHQHAA APADSDDETE ASASPTPSGS

       430        440        450        460        470        480
VEQAVTIPPR ANTMVIFFSD RRVPHEVLPT NVPRFAATHW FYDADERAEA EASAGTAFEL


DQQI
```

The residues P239, G256 and C359 are highlighted in red.

1.12. taPHD Full-Length DNA Sequence
(Predicted by C. Loenarz; codon optimised, synthesised by Geneart AG)

```
ATGGCGAGCTGCAACGAAATTCCGCTGTGCGAAAGCGTGACCGATGGCCATCTGCTGGTTGATCAGGATGCGTATTGCAGCCTGTG
CTTCAAAAAACTGATCAGCGTGAAATGCAGCCTGTGTCAGAGCGCGTATTATTGCAGCAAAGAACATCAGCAGAGCCATTGGAAAA
TTCATCGTGAAAACTGCATGGATGAACGTCGTCCGAAAATTGGCTGCGATAAAACCGGCCTGGCCCTGGAAAAAACCTGGCAGCTG
CAGAAAAACGAACGTCTGAGCAATATGGTGGTGAACCAGCTGAACACCAACGGCTTTTGCATCATCAACAACTTTCTGGGCAGCAG
CTGTAGCACCGAAGTTCTGCAGCAGGTGCTGAACCTGTATCAGAGCGGCGTGTTTAGCAACGGTCAGCTGGCCCGTAACGTGAGCG
TGAACCGTATTCGTGGCGATAAAATTGCGTGGATTGGCGGTGATGAACGTGGCTGCGAAGCGATTAAATATCTGAGCAGCTGCGTG
GATAGCCTGATTAGCCGCTGCAACGGCCGTCTGGGCAACTATATGATTACCGGCCGTACCAAATGCATGGTGGCGTGCTATCCGGG
TAGCGGCCTGGGCTATATTCGTCATATCGATAATCCGAACCGTGATGGCCGTTGCGTGACCGTGCTGTATTACCTGAATCCGAACT
GGAACAGCCAGGATTGCGGCGGTCAGCTGTGGCTGTATCCGAACAACGAAAACAAAGTGGTGAAAATCGATCCGATTTTTGATCGT
CTGCTGCTGTTTTGGAGCGATCGTCGTAATCCGCATGAAGTGAAACCGGCGTATGCCATGCGTTATGCGATTACCCTGTGGTACTT
CGATGAAAAAGAACGTGCCCTGAGCAGCCAGAACGGCACCTAATAA
```

1.13. Primer Design for taPHD *N*-terminal Truncations

For cloning into pET-28a(+) using NdeI and EcoRI

Primer for t67:
Forward (5'3'): ATG CAT ATG CGT CGT CCG AAA ATT G

Primer for t73
Forward (5'3'): CGT CAT ATG ATT GGC TGC GAT AAA ACC

Primer for t82
Forward (5'3'): GAA CAT ATG TGG CAG CTG CAG AAA AA

Reverse primer:
(5'3') GAA TTC TAT TAG GTG CCG TTC TGG C

1.14. Primers for Site-Directed Mutagenesis for taPHD Variants

S177R
Primer 1 (5'3'): GGA TAG CCT GAT TAG GCG CTG CAA CGG CCG
Primer 2 (5'3'): GCC GTT GCA GCG CCT AAT CAGGCT ATC CAC

P213R
Primer 1 (5'3'): CAT ATC GATAAT CGG AAC CGT GAT GGC CGT
Primer 2 (5'3'): GCC ATC ACG GTT CCG ATT ATC GAT ATG ACG

R292A
Primer 1 (5'3'): GAT GAA AAA GAT GCT GCC CTG AGC AGC CAG
Primer 2 (5'3'): TGG CTG CTC AGG GCA GCA TCT TTT TCA TCG

1.15. Primers for Cloning Full-Length taPHD into pGEX-6P-1

Forward primer (5'3'): GGT AAT CTG AAA GAC GGT CAG CTG GCT GGC
Reverse primer (5'3'): AGC CAG CTG ACC GTC TTT CAG ATT ACC ATC

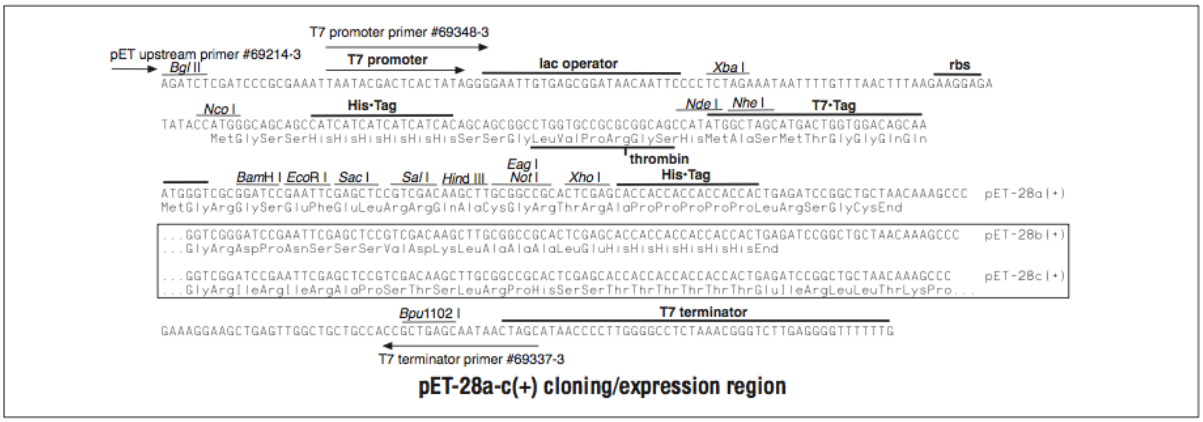
1.16. Primers for mbP4H Variants

P239D
Primer 1 (5'3'): GGT AAT CTG AAA GAC GGT CAG CTG GCT GGC
Primer 2 (5'3'): AGC CAG CTG ACC GTC TTT CAG ATT ACC ATC

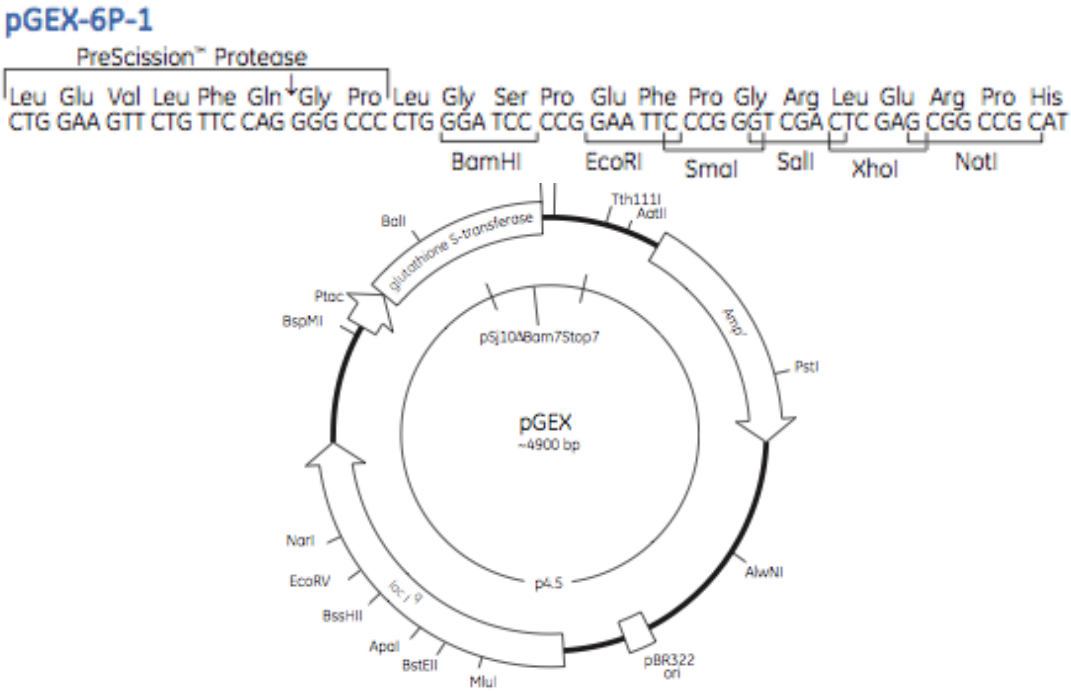
G256R
Primer 1 (5'3'): CC AAA TAT GTG CTG CGT CAT GTT CGT GGG G
Primer 2 (5'3'): C CCC ACG AAC ATG ACG CAG CAC ATA TTT GG

C359V
Primer 1 (5'3'): T TAT ATT CGC CAT GTT GAT AAT CCG CAT CG
Primer 2 (5'3'): ATG CGG ATT ATC AAC ATG GCG AAT ATA ACG

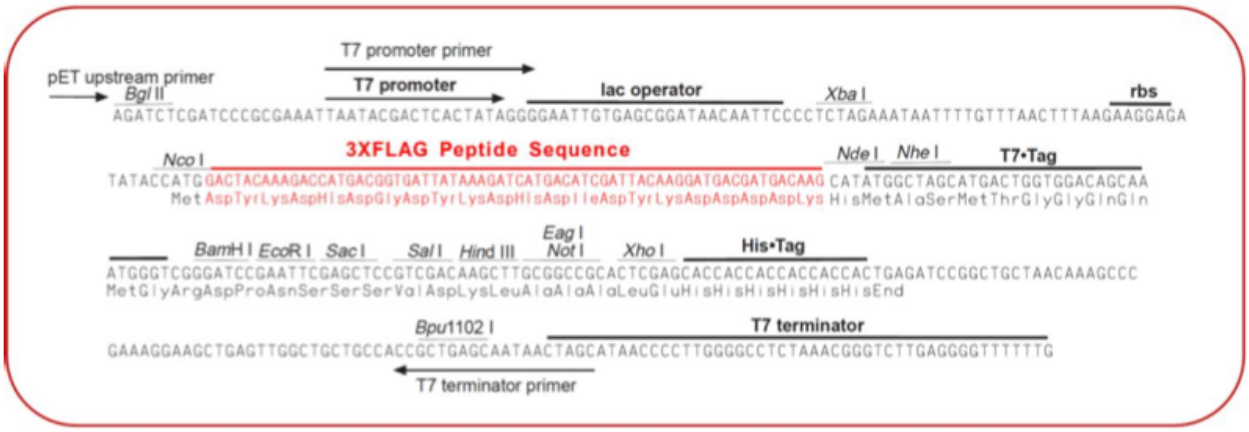
1.17. pET-28a(+) Expression Vector (From NOVAGEN Vector Manual)



1.19. pGEX-6P-1 Expression Vector (From NOVAGEN Vector Manual)



1.20. FLAG-pET-28a(+) Expression Vector (Prepared by Inga Pfeffer)



Appendix 2: Agarose and SDS-PAGE Gels

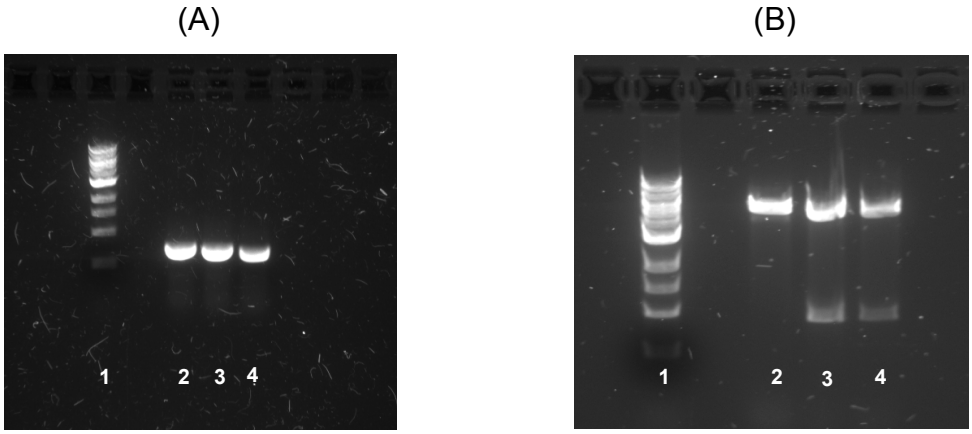


Figure 2.1. (A) 1% Agarose gel showing PCR products for preparing *N*-terminally truncated taPHD constructs for cloning into pET-28a(+). (B) 1% Agarose gel showing PCR products for cloning full-length taPHD into pGEX-6P-1. Lanes (A): 1=DNA ladder; 2=taPHD t67; 3=taPHD t74; 4=taPHD t82. Lanes in (B): 1=DNA ladder; 2=taPHD/pGEX primers PCR conditions a; 3=taPHD/pGEX primers PCR conditions b; 4=taPHD/pGEX primers PCR conditions c. The PCR protocol including conditions a-c can be found in Materials and Methods 1.

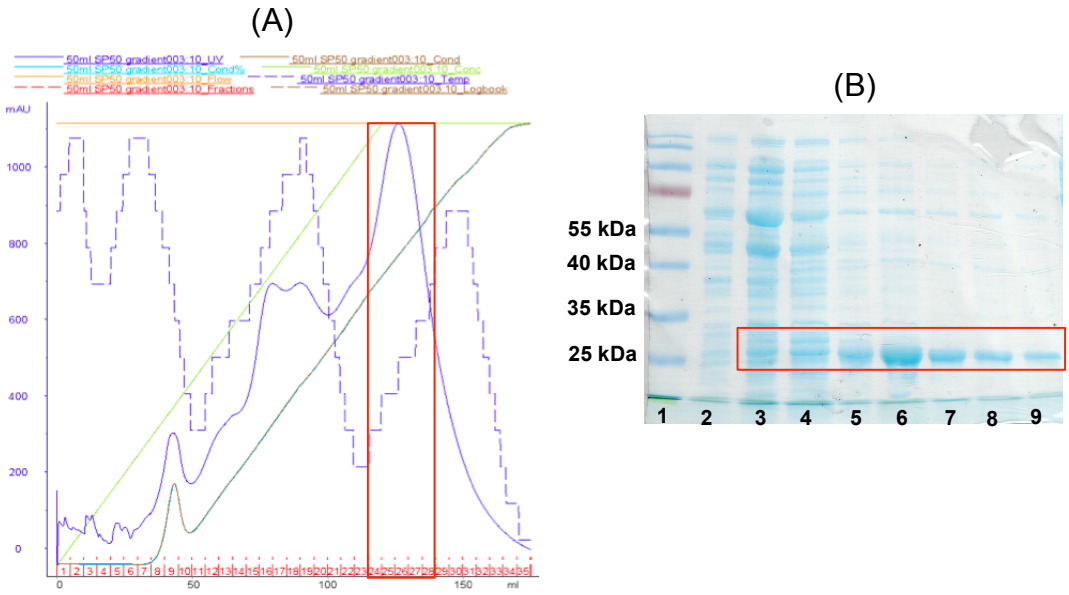


Figure 2.2. (A) FPLC trace of cation-exchange purification of tag-free t64 taPHD in pET-24a(+) and (B) SDS-PAGE gel of this purification run. Lanes: 1=Protein ladder; 2=Column flowthrough; 3=Fraction 16; 4=Fraction 18; 5-9Fractions 24-28. The taPHD peak and bands are highlighted in red. The calculated construct mass is 27.10 kDa.

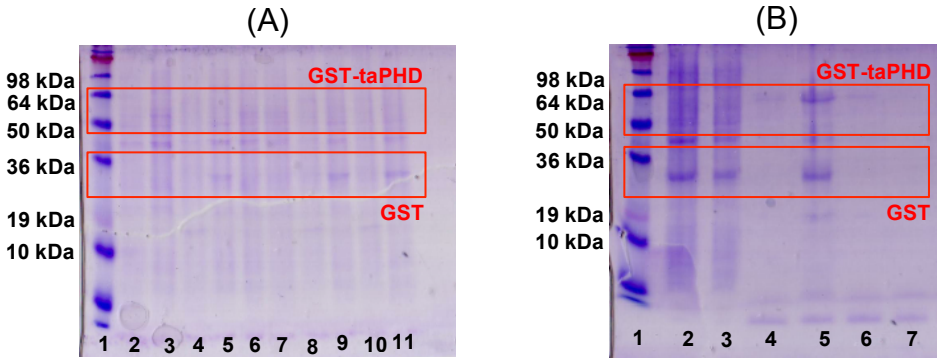


Figure 2.3. Expression of GST-tagged full-length taPHD in pGEX-6P-1; (A) Expression trials and (B) GST-column purification. Lanes in (A): 1=Protein ladder; 2=18°C, soluble fraction; 3=18°C, insoluble fraction; 4=25°C, soluble fraction; 5=25°C, insoluble fraction; 6=30°C, soluble fraction; 7=30°C, insoluble fraction; 8=37°C, soluble fraction; 9=37°C, insoluble fraction; 10=37°C, soluble fraction, no IPTG; 11=37°C, insoluble fraction, no IPTG. All conditions were induced with 0.5 mM IPTG unless otherwise stated. Lanes in (B): 1=Protein ladder; 2=Column flowthrough; 3=Column wash; 4-7=Fractions 1-3. The calculated mass of the construct is 58.58 kDa. The mass of the GST-tag is 24.38 kDa.

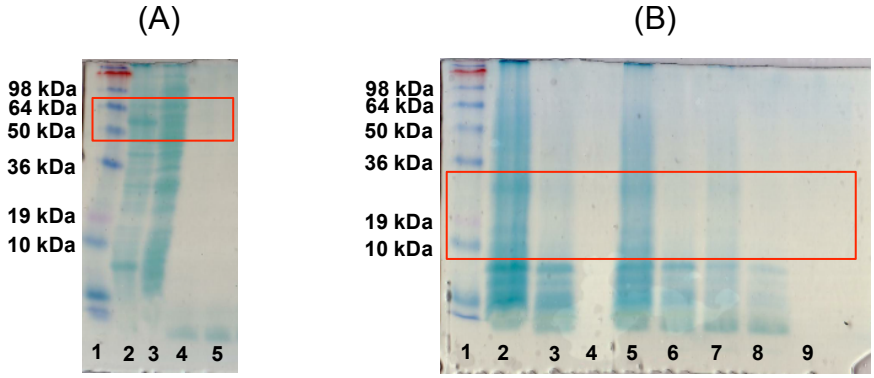


Figure 2.4. Expression of 6xHis-tagged taPHD in pET-28a(+); (A) Expression trials of full-length taPHD (B) Expression of *N*-terminally truncated taPHD constructs. Lanes in (A): 1=Protein ladder; 2=Insoluble fraction; 3=Soluble fraction; 4=0.5 M imidazole elution; 5=1 M imidazole elution. Lanes in (B): 1=Protein ladder; 2=taPHD t67, insoluble fraction; 3=taPHD t67, soluble fraction; 4=Empty lane; 5=taPHD t74, insoluble fraction; 6=taPHD t74, soluble fraction; 7=taPHD t82, insoluble fraction; 8=taPHD t82, soluble fraction; 9=Empty lane. All of the expressions above were carried out at 18°C for 12 hours after induction with 0.5 mM IPTG at OD_{600}=0.6. The calculated masses of the constructs are the following: full-length taPHD=58.58 kDa; taPHD t67=34.22 kDa; taPHD t74=28.80 kDa; taPHD t82=26.93 kDa.

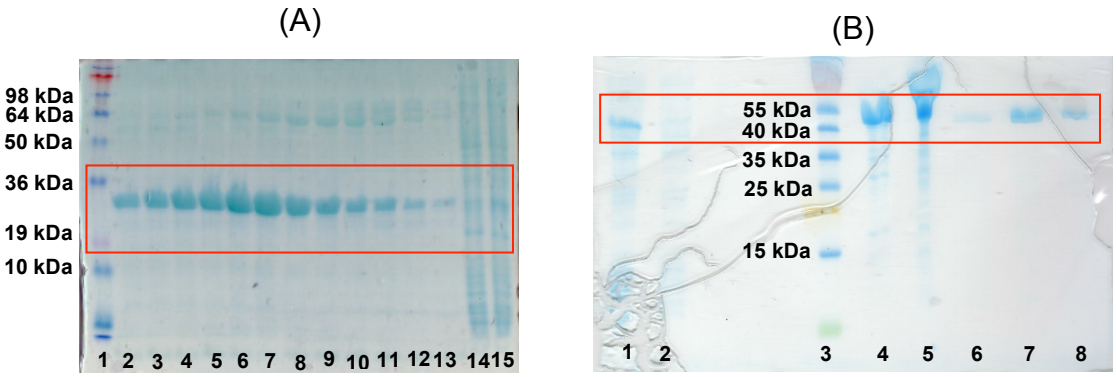


Figure 2.5. (A) His-affinity purification of taPHD S177R variant. (B) Selected fractions from His-affinity purification of mbP4H variants. Lanes in (A): 1=Protein ladder; 2-13=Fractions 12-23; 14=Column load; 15=Column flowthrough. The expression was carried out as described in methods X. Lanes in (B): 1=0.5 M imidazole elution of mbP4H P239D purification; 2=1 M imidazole elution of mbP4H P239D purification; 3=Protein ladder; 4=0.5 M imidazole elution of mbP4H G256R purification; 5=1 M imidazole elution of mbP4H G256R purification; 6=Empty lane; 7=0.5 M imidazole elution of mbP4H C359V purification; 8=1 M imidazole elution of mbP4H C359V purification. All of the expressions of mbP4H above were carried out as described in methods X. The calculated masses of the constructs are the following: taPHD S177R=29.24 kDa; mbP4H P239D=40.78 kDa; mbP4H G256R =40.86 kDa; mbP4H C359V =40.76 kDa.

Appendix 3: Complementary Data

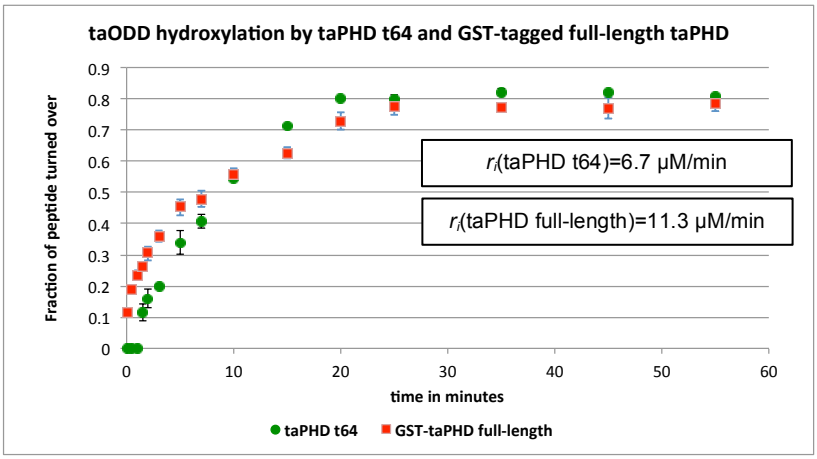


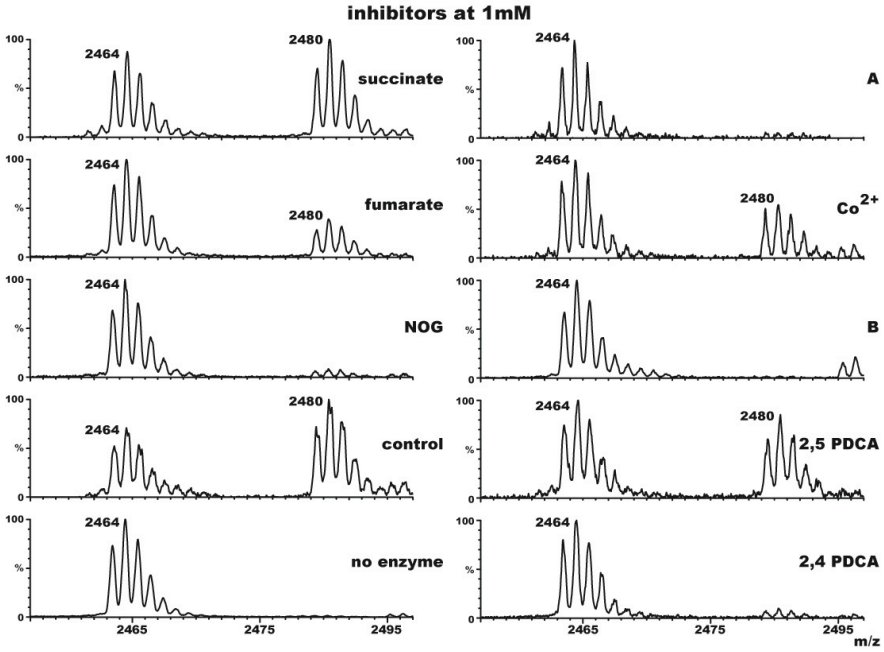
Figure 3.1. Comparison of taODD hydroxylation time course data with truncated taPHD t64 (catalytic domain only) and the GST-tagged full-length taPHD (MYND and catalytic domains). The initial rate of peptide turnover by the full-length construct is almost twice the initial rate with the truncated construct. The difference in initial rates could be of biological significance, suggesting that the MYND domain plays a role in substrate binding and turnover. Alternatively, they could be caused by varying amounts of active and/or properly folded enzyme. For different purification protocols, the amounts of active enzyme can vary because of different time lengths that the protein could be degrading for and different buffers used, some of which the enzyme could be sensitive to.

3.2. Considerations of factors affecting the ^{1}H-NMR signals of 2OG and succinate

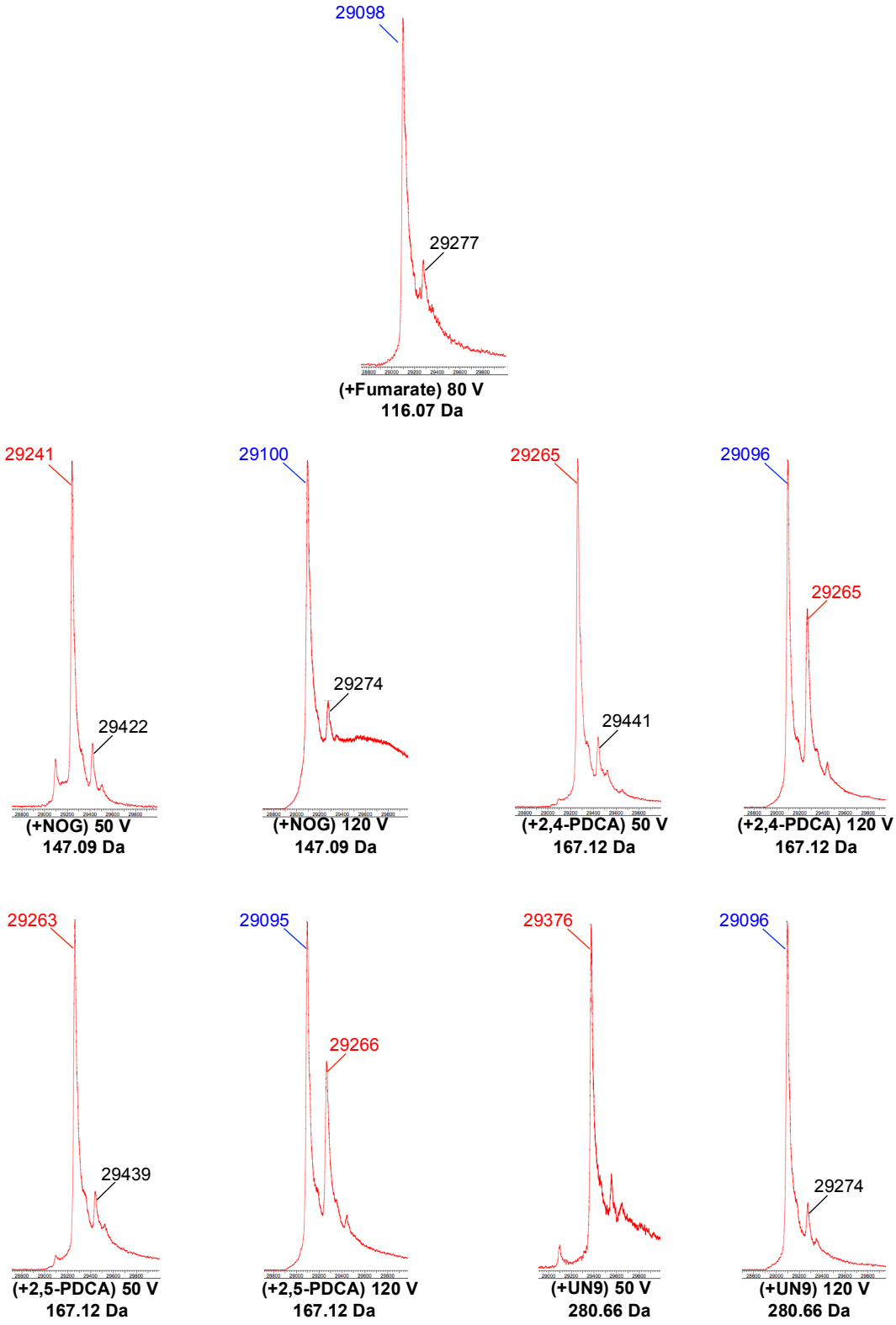
In the evaluation of ^{1}H-NMR experiments quantifying the enzymatic turnover of 2OG to succinate, it was assumed that the sum of the concentrations of 2OG and succinate is constant. To justify this assumption, competing processes affecting the signals of 2OG and succinate protons were considered.

The timescale of signal observation for the whole time course was around one hour, while the timescale for initial rate determination was 5 minutes. The processes affecting the β–proton 2OG signal in the reaction mixture are enzymatic turnover into succinate, exchange of the β–protons to deuterium from the solvent D_2O and non-enzymatic decarboxylation. The processes affecting the succinate proton signal are enzymatic formation of succinate, exchange of protons to deuterium and non-enzymatic formation

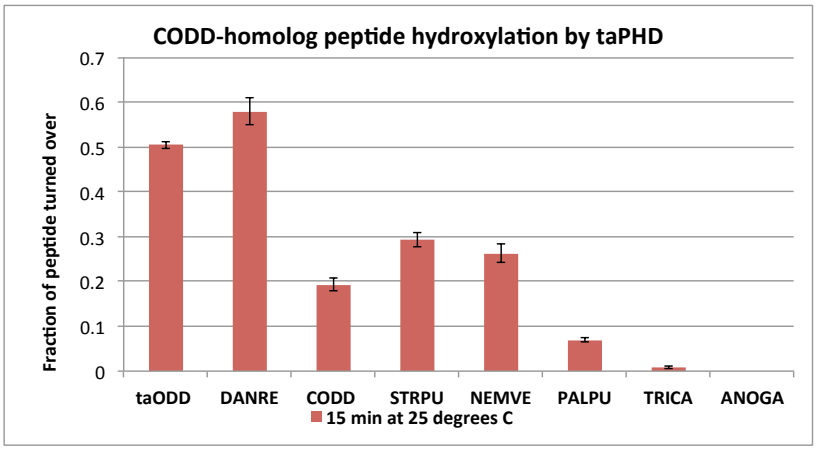
of succinate. Non-enzymatic turnover of 2OG to succinate is very slow compared to the experiment time scale and is estimated to have a half-life of more than a week. Proton-to-deuterium exchange for succinate is expected to be slow because the saturated succinate protons are not acidic (pK_a>40). The exchange of the β-protons of 2OG, however, is considerably faster than the processes above because of keto-enol tautomerism at the β-position (pK_a<20). Observations suggested that the half-life of the β–proton exchange of 2OG is about one day (observed by Andrea Szollossi, data not shown). Compared to enzymatic turnover of 2OG determined over five minutes, all of the processes above are negligibly slow, which justifies the assumption that all of the observed 2OG signal decrease is due to enzymatic turnover, giving rise to all of the succinate signal observed.



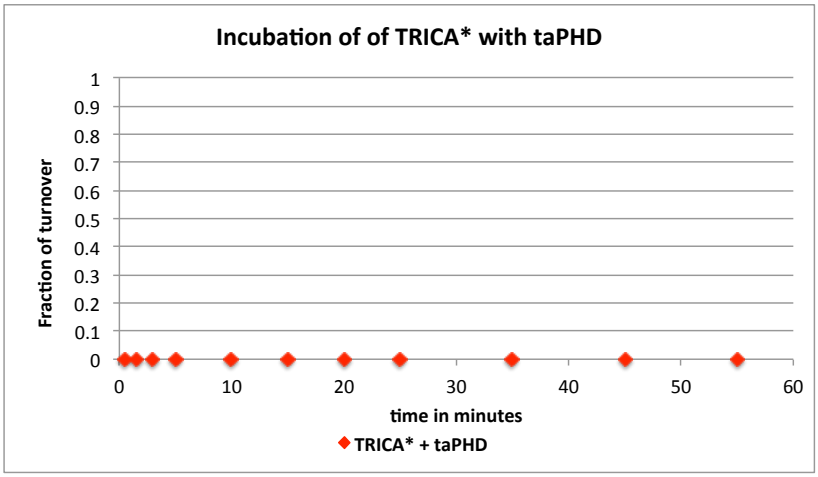
3.3. MALDI-TOF MS spectra of taODD after 10 minutes incubation with taPHD and an inhibitor. The labels above the peaks indicate the mass of the most intense peak of the isotopic progression. The fraction of turnover was calculated using peak intensities of the most intense peaks. The sample labelled 'control' does not contain any inhibitors. 'A' stands for UN9, 'B' is a control without added 2OG. The same assay conditions as described in Materials and Methods 5 were used.



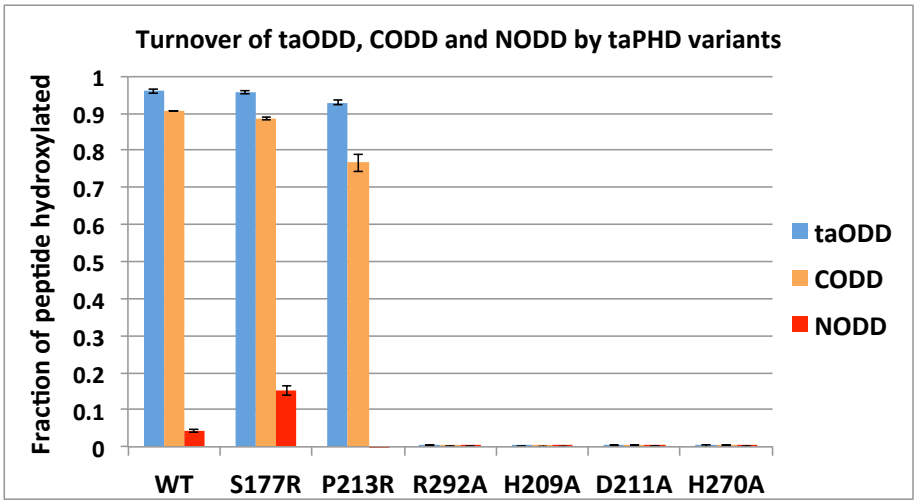
3.4. Non-denaturing mass spectra of taPHD with inhibitors taken at different cone voltages. The enzyme peak can be seen at 29095-29100 Da and is labelled with the observed mass in Da in blue. The calculated mass of taPHD is 29042 Da. The mass difference of 57-53 Da suggests that taPHD co-purifies with Fe(II), which has an average mass of 55.85 Da. The peaks labelled in red correspond to the enzyme-inhibitor complex. No fumarate binding was observed. The mass spectra of taPHD with inhibitors at 50 V and 120 V cone voltages show the cone voltage dependence of the enzyme-ligand complexes.



3.5. Hydroxylation of CODD-homologue peptides (100 µM) by taPHD, determined by MALDI-TOF MS after 15 minutes incubation with 3.5 µM taPHD, 300 µM 2OG, 50 µM Fe(II) and 4 mM ascorbate.



3.6. TRICA* shows no hydroxylation above background levels, neither with taPHD nor with hsPHD2. Assay conditions as in 3.5.



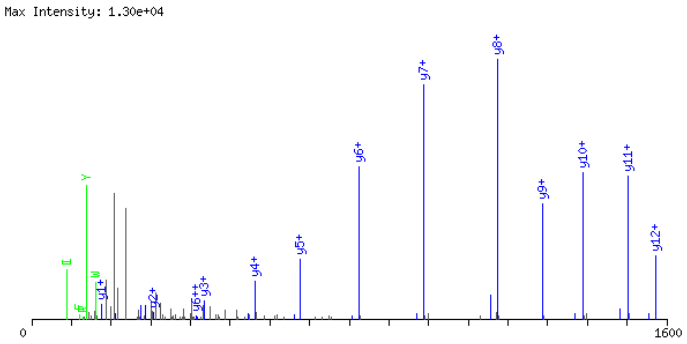
3.7. No turnover of taODD, CODD or NODD was detected with variants lacking residues thought to be key for substrate turnover. R292 forms a salt bridge to D494 of taODD. The hsPHD2 equivalent, R396, has been shown to be essential for CODD turnover through binding to D571 (unpublished data). H209, D211 and H270 are Fe(II)-coordinating residues. Assay conditions as in 3.5.

Appendix 4: Hydroxylated MS/MS Fragments

(A)

Y(+15.99)AITLWYFDEKER

Charge: 3, Exp. m/z: 583.966, Calc. m/z: 583.954
Data File: HKQ3095_CJS_AB_25K_220211_pkl, Scans: 573 - 573



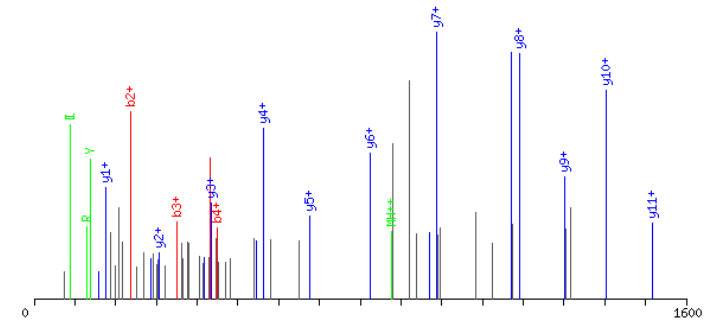
Showing fragment losses - Click to Hide

		b+	-H2O	-NH3	b++	-H2O	-NH3	y+	-H2O	-NH3	y++	-H2O	-NH3		
Y	1	180.07			90.54									13	Y
A	2	251.10			126.05			1570.79	1552.78	1553.76	785.90	776.89	777.39	12	A
I	3	364.19			182.60			1499.75	1481.74	1482.73	750.38	741.37	741.87	11	I
T	4	465.23	447.22		233.12	224.12		1386.67	1368.66	1369.64	693.84	684.83	685.32	10	T
L	5	578.32	560.31		289.66	280.66		1285.62	1267.61	1268.59	643.31	634.31	634.80	9	L
W	6	764.40	746.39		382.70	373.70		1172.54	1154.53	1155.51	586.77	577.77	578.26	8	W
Y	7	927.46	909.45		464.23	455.23		986.46	968.45	969.43	493.73	484.73	485.22	7	Y
F	8	1074.53	1056.52		537.77	528.76		823.39	805.38	806.37	412.20	403.20	403.69	6	F
D	9	1189.56	1171.55		595.28	586.28		676.33	658.32	659.30	338.67	329.66	330.15	5	D
E	10	1318.60	1300.59		659.80	650.80		561.30	543.29	544.27	281.15	272.15	272.64	4	E
K	11	1446.69	1428.68	1429.67	723.85	714.85	715.34	432.26	414.25	415.23	216.63	207.63	208.12	3	K
E	12	1575.74	1557.73	1558.71	788.37	779.37	779.86	304.16	286.15	287.14	152.58	143.58	144.07	2	E
R	13							175.12		158.09	88.06		79.55	1	R

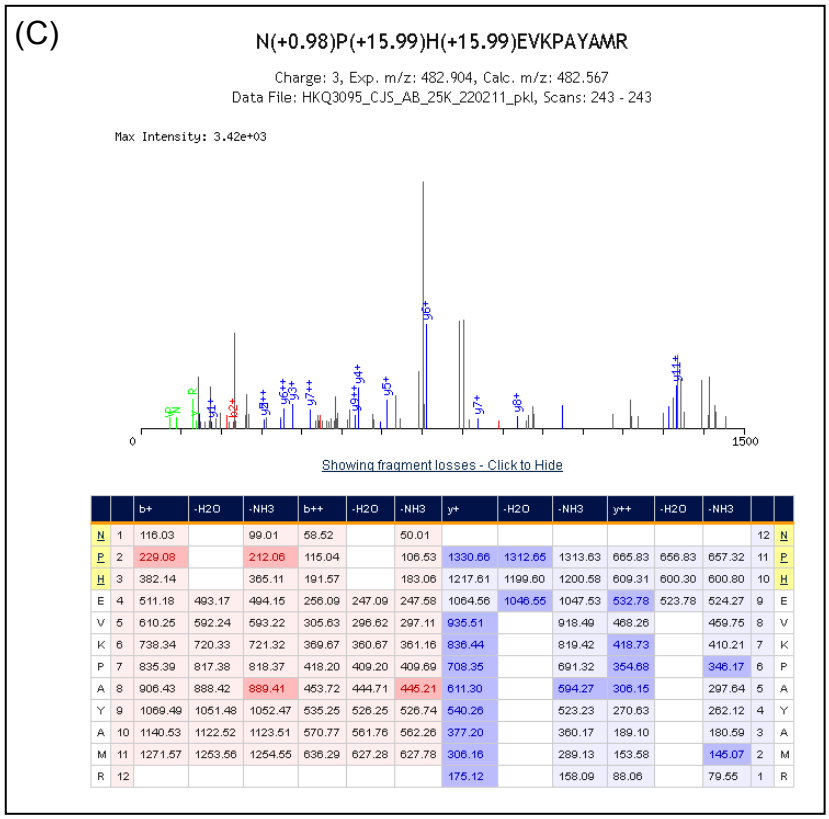
(B)

YAITLW(+15.99)YFDEKER

Charge: 2, Exp. m/z: 875.440, Calc. m/z: 875.427
Data File: HKQ3095_CJS_AB_25K_220211_pkl, Scans: 2299 - 2299



		b+	-H2O	-NH3	y+	-H2O	-NH3		
Y	1	164.07						13	Y
A	2	235.11			1586.78	1568.77	1569.76	12	A
I	3	348.19			1515.75	1497.74	1498.72	11	I
T	4	449.24	431.23		1402.66	1384.65	1385.64	10	T
L	5	562.32	544.31		1301.62	1283.61	1284.59	9	L
W	6	764.40	746.39		1188.53	1170.52	1171.51	8	W
Y	7	927.46	909.45		986.46	968.45	969.43	7	Y
F	8	1074.53	1056.52		823.39	805.38	806.37	6	F
D	9	1189.56	1171.55		676.33	658.32	659.30	5	D
E	10	1318.60	1300.59		561.30	543.29	544.27	4	E
K	11	1446.69	1428.68	1429.67	432.26	414.25	415.23	3	K
E	12	1575.74	1557.73	1558.71	304.16	286.15	287.14	2	E
R	13				175.12		158.09	1	R



		b+	-H2O	-NH3	b++	-H2O	-NH3	y+	-H2O	-NH3	y++	-H2O	-NH3		
N	1	116.03		99.01	58.52		50.01							12	N
P	2	229.08		212.06	115.04		106.53	1330.66	1312.65	1313.63	665.83	656.83	657.32	11	P
H	3	382.14		365.11	191.57		183.06	1217.61	1199.60	1200.58	609.31	600.30	600.80	10	H
E	4	511.18	493.17	494.15	256.09	247.09	247.58	1064.56	1046.55	1047.53	532.78	523.78	524.27	9	E
V	5	610.25	592.24	593.22	305.63	296.62	297.11	935.51		918.49	468.26		459.75	8	V
K	6	738.34	720.33	721.32	369.67	360.67	361.16	836.44		819.42	418.73		410.21	7	K
P	7	835.39	817.38	818.37	418.20	409.20	409.69	708.35		691.32	354.68		346.17	6	P
A	8	906.43	888.42	889.41	453.72	444.71	445.21	611.30		594.27	306.15		297.64	5	A
Y	9	1069.49	1051.48	1052.47	535.25	526.25	526.74	540.26		523.23	270.63		262.12	4	Y
A	10	1140.53	1122.52	1123.51	570.77	561.76	562.26	377.20		360.17	189.10		180.59	3	A
M	11	1271.57	1253.56	1254.55	636.29	627.28	627.78	306.16		289.13	153.58		145.07	2	M
R	12							175.12		158.09	88.06		79.55	1	R

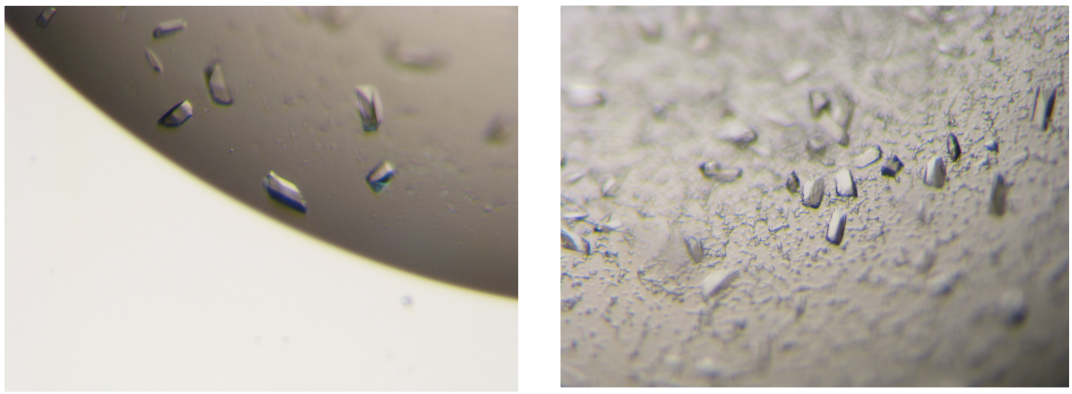
4.1. (A)-(C) MS/MS fragments of taPHD that were found to be hydroxylated. It was concluded that all of the observed hydroxylations were artefacts from the in-gel digest or the SDS-PAGE gel.

Appendix 5: Further Crystallisation Attempts

The high resolution of the taPHD structures were very interesting in the light of potential studies of reactive intermediates in the enzyme active site. Time-resolved crystallography at such a high resolution could yield great insight into the reaction mechanism and clear up ambiguities over the reactive iron species and the mode of interaction with the substrate during hydroxylation. It was therefore attempted to reproduce the crystals using NOG as the ligand, and to crystallise taPHD with 2OG and Fe(II) under anaerobic conditions.

5.1. Crystals with NOG and Mn(II)

The conditions based on well H.11. of the JCSG-Plus screen were reproduced manually in a hanging drop plate. Instead of compound UN9, NOG at a final concentration of 2 mM was used. Many crystals (average size 70 µm) were obtained with this condition using protein from the same purification batch that yielded the structures taPHD and taPHD.taODD. The largest crystals were harvested and irradiated in-house, however did not diffract X-rays. It was concluded that the crystals were possibly amorphous.

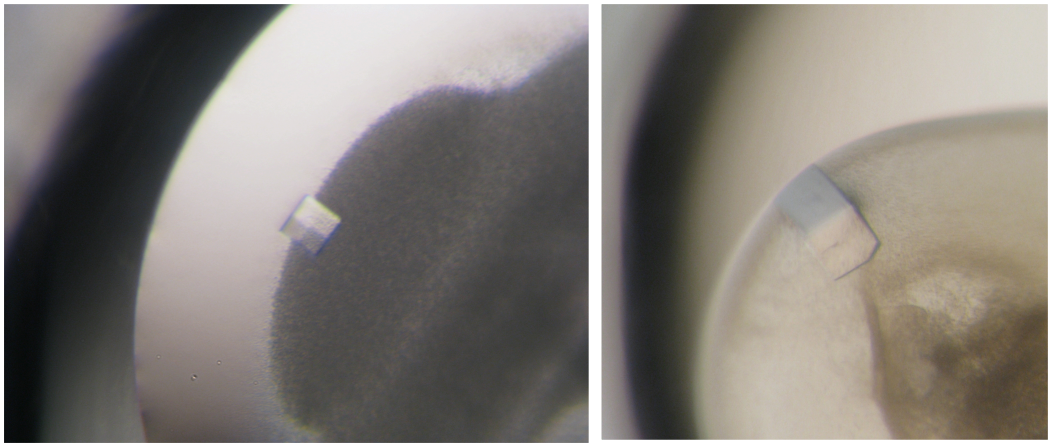


Crystals of taPHD with Mn(II) and NOG.

5.2. Reproducibility of taPHD and taPHD.taODD results

Small 30 µM two-dimensional plates were generated with 2OG and Fe(II) under anaerobic conditions (crystallization conditions: 0.2 M ammonium acetate, 25% w/v polyethylene glycol 3350, 0.1 M bis-tris pH=5.5). Neither these nor any of the previous results could, however, be reproduced with new enzyme purification batches. After several unsuccessful attempts, it was decided to change the strategy. A tag-free taPHD t64 construct was expressed in pET-24a(+). The enzyme was purified using anion exchange followed by gel filtration. The screening conditions were broadened again and the JCSG-Plus screen was used. Trays with Mn(II) and either NOG or UN9 were set up. The same enzyme concentration of 20 mg/mL as in the previously successful trays was used. The tray containing UN9 generated several hits:

JCSG-Plus at 20° C with UN9 (50-80 µm rectangles)
G.2.1 (0.02M $MgCl_2$, 22% polyacrylamide 5100 disodium salt, 0.1 M HEPES pH=7.5)
B.10.2 (0.2M $MgCl_2$, 50% PEG 200, 0.1 sodium cacodylate pH=6.5)
B.11.1: (1.6 M tri-sodium citrate pH=6.5)



Crystals of taPHD t64 purified by ion exchange and gel filtration.

All hits were harvested and screened in-house (RIGAKU). While the resolution was high at 1.3-1.5 Å, the structures could not be solved due to apparent twinning. An optimisation screen based on the conditions in well B.11. of JCSG-Plus generated further hits, but the twinning issue could not be resolved.

Further unsuccessful attempts at getting more crystal structures included:

- Using 1-5 mM fumarate or succinate as ligands
- Using a previously generated crystal to seed new enzyme batches
- Addition of a pre-incubated Zn(II)-2OG complex (NMR studies by Ivanhoe Leung suggested that hsPHD2 is very stable in presence of a Zn(II)-2OG complex)
- Trying 6 different purification batches and 3 different enzyme concentrations (10 mg/mL; 20 mg/mL; and 30 mg/mL)
- Use of different commercial screens (Structure screen, MIDAS, PACT premier, all from MOLECULAR DIMENSIONS)