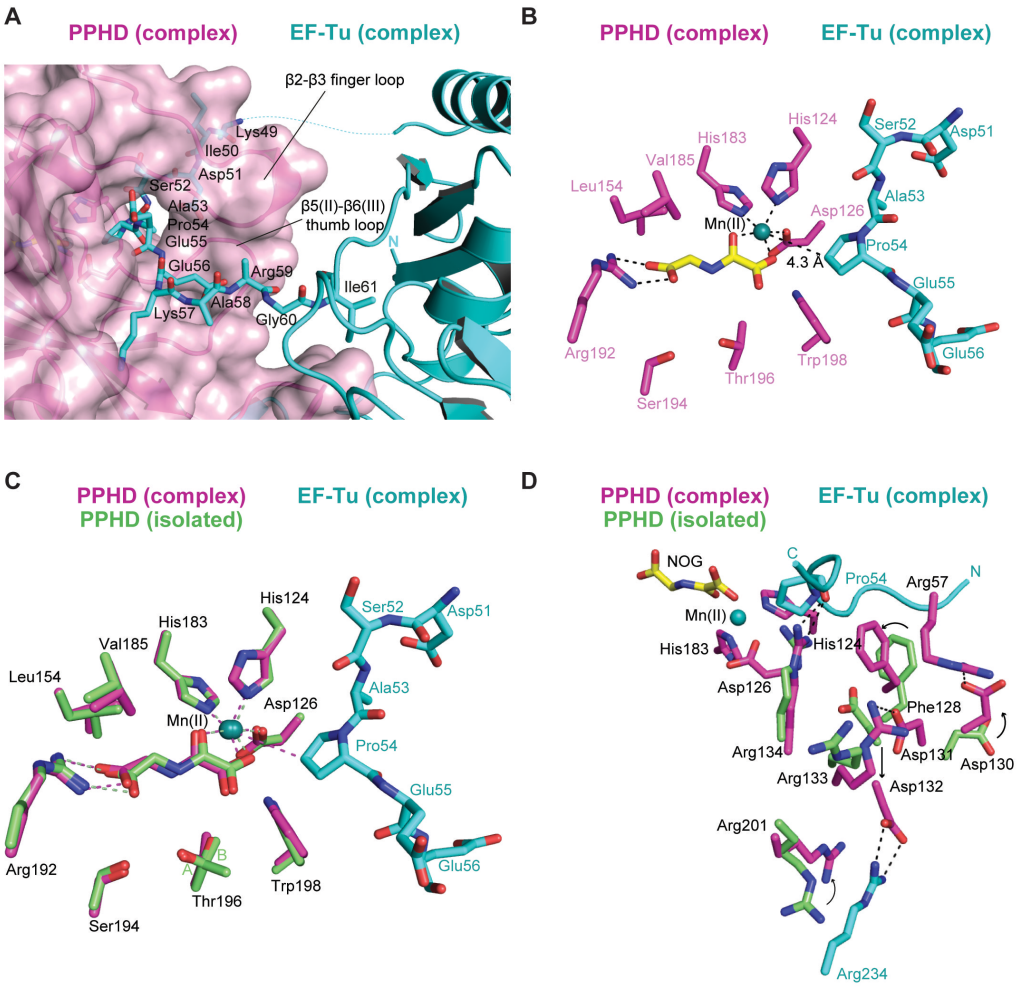


Figure 3.19. EF-Tu binding to PPHD is mediated by induced fit. (**A**) Surface representation of the PPHD(magenta):EF-Tu(cyan) complex shows that the PPHD β2-β3 finger loop and β5(II)-β6(III) thumb loop form a tunnel encompassing the EF-Tu switch I loop. (**B**) View of the PPHD:EF-Tu complex reveals Pro54 is poised for 4-*trans* hydroxylation with its C-4 4.3 Å from the Mn(II). (**C**) Superimposition of the active site residues of isolated PPHD (green) and the PPHD(magenta):EF-Tu(cyan) complex shows no significant conformational changes in the side chains of the residues in the core DSBH. (**D**) Superimposition of the β5(II)-β6(III) thumb loop of isolated PPHD (green) and PPHD (magenta) in complex with EF-Tu (cyan) PPHD(magenta):EF-Tu(cyan) reveals major conformational changes in the backbone and side chains of thumb residues upon EF-Tu binding. Figure and figure legend are adapted with permission from J.S. Scotti *et al*, Human oxygen sensing may have origins in prokaryotic elongation factor Tu prolyl-hydroxylation. *Proc Nat Acad Sci USA*. In press (2014).

3.4.4 Conformational changes in the PPHD β5(II)-β6(III) thumb loop

The β5(II)-β6(III) thumb loop of PPHD is highly polar and contains 3 arginines (Arg127, Arg129, and Arg133), 3 aspartates (Asp130, Asp131, and

Asp132), and one phenylalanine (Phe128), and makes multiple direct electrostatic interactions with EF-Tu (Table 3.20).

Table 3.20. Electrostatic interactions between the PPHD β2-β3 finger loop and EF-Tu.

β5(II)-β6(III) finger loop (residues 128-134)			
PPHD residue	**EF-Tu residue**	**Distance (Å)**	**Interaction type**
Arg129 backbone nitrogen	Asp51 carboxylate	3.0	Hydrogen bond
Asp132 carboxylate	Arg234 guanidinium group	2.9	Salt bridge
Arg133 guanidinium group	Glu56 carboxylate	3.2	Salt bridge
Arg134 guanidinium group	Pro54 backbone oxygen	2.7	Hydrogen bond
Arg134 guanidinium group	Glu55 backbone oxygen	3.1	Hydrogen bond

In addition to positioning Asp130 to form a salt bridge with Arg57, the backbone and side chains of the residues constituting the β5(II)-β6(III) thumb loop undergo distinct changes from their conformations in isolated PPHD (r.m.s.d. 1.4 Å)(Figure 3.19*D*). Phe128 rotates 90° to position its phenyl ring directly behind Pro54 locking the flexible switch I loop in the PPHD active site. Asp131, the side chain of which is disordered in isolated PPHD, is ordered in complexed PPHD and positioned to form a salt bridge with Arg133. Asp132 rotates 180° relative to its position in isolated PPHD where it forms a salt bridge with Arg134 in order to form a salt bridge with Arg234 of EF-Tu domain II (180 residues *C*-terminal to Pro54), an interaction further strengthened by π-stacking of EF-Tu Arg234 with Arg201 of the PPHD *C*-terminus, and thus freeing Arg134 to hydrogen bond to Pro54. In sum, the major conformational changes in the β5(II)-β6(III) thumb loop likely are driven by, and assist in, the induced fit binding of EF-Tu to PPHD.

3.4.5 PPHD and EF-Tu contacts outside the PPHD active site

The PPHD:EF-Tu protein-protein complex structure also reveals substantial contacts between PPHD and EF-Tu outside the immediate vicinity of

the PPHD active site. Complexation of the two proteins buries 1662 Å^2 of interaction surface (24 hydrogen bonds and 10 salt bridges; PISA analysis)(*36*).

However, only 965 Å^2 (13 hydrogen bonds and 3 salt bridges) constitute direct interactions between PPHD and the EF-Tu switch I loop, leaving 697 Å^2 of buried surface area attributed to PPHD:EF-Tu interactions between other regions of protein. Notable interactions include the positioning of Phe102 of the PPHD α3-helix (residues 81-101) in a hydrophobic pocket in the β-barrel fold of EF-Tu domain II formed by residues Phe222, Ile224, and Val230 (Figure 3.21*A*). In a similar manner, EF-Tu Phe265 is buried in a hydrophobic pocket in PPHD formed by Leu101, Leu103, and Leu177 of the PPHD α1- (residues 8-20) and α3-helices (Figure 3.21*A*). In addition to these hydrophobic contacts, Glu106 of PPHD β4(I) is positioned to form a salt bridge with Arg292 of EF-Tu domain II, Arg201 of PPHD β11(VIII) is positioned to hydrogen bond to Asn277, and Arg202 of the PPHD *C*-terminal loop is positioned to form a salt bridge to Glu236 (Figure 3.21*B*). Interactions between PPHD and EF-Tu domain II are of particular interest as they constitute a sizable fraction of PPHD:EF-Tu contacts in the complex and thus likely contribute to its overall stability, likely helping to order the PPHD β5(II)-β6(III) thumb loop (Figure 3.19*D* and Figure 3.21*B*).

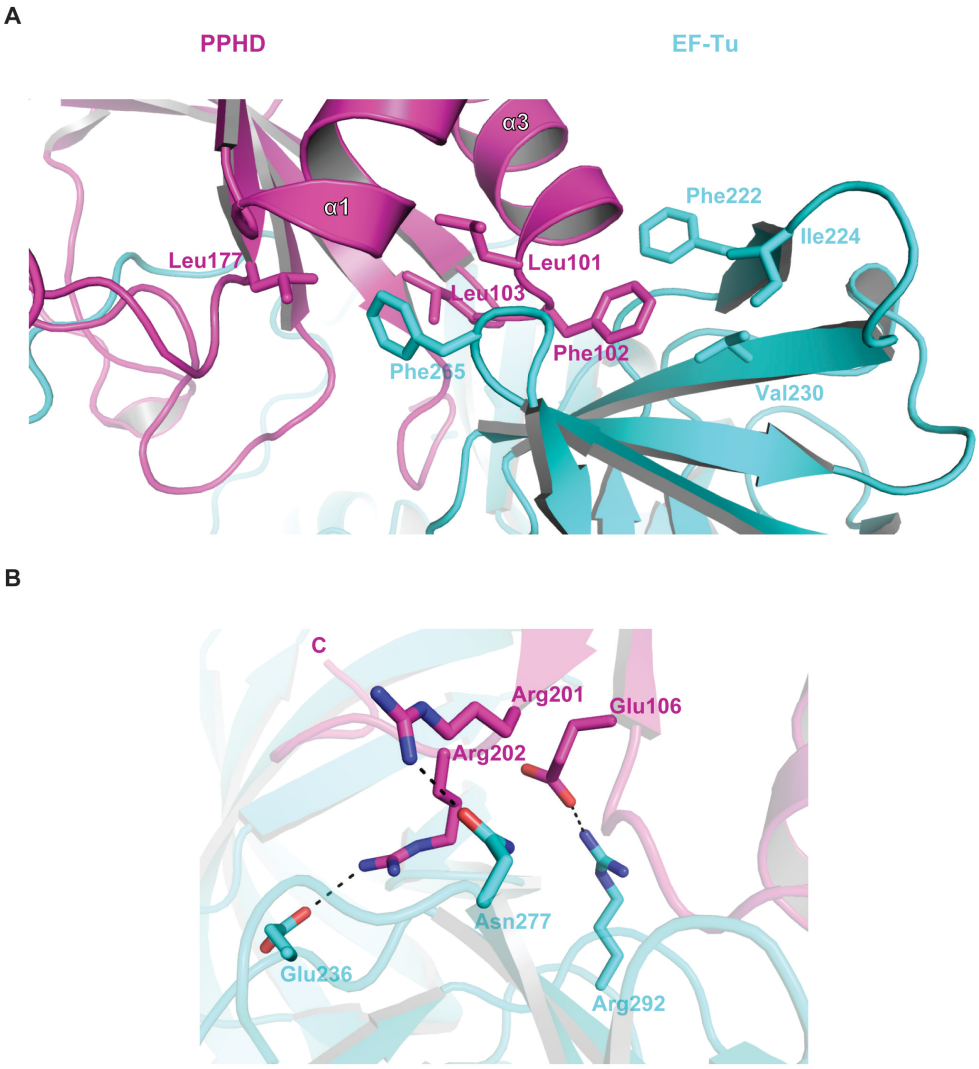


Figure 3.21. PPHD:EF-Tu contacts are present outside the vicinity of the catalytic site. (**A**) Residues from an α-helix (α3) of PPHD extend into the β-barrel fold of EF-Tu domain II. EF-Tu Phe265 is positioned in a hydrophobic pocket outlined by PPHD Leu101, Leu103, and Leu177; PPHD Phe102 is positioned in a hydrophobic pocket in the β-barrel fold of EF-Tu domain II formed by Phe222, Ile224, and Val230. (**B**) Polar residues in PPHD *C*-terminus and β4(I) interact with EF-Tu domain II. Figure and figure legend are adapted with permission from J.S. Scotti *et al*, Human oxygen sensing may have origins in prokaryotic elongation factor Tu prolyl-hydroxylation. *Proc Nat Acad Sci USA*. In press (2014).

3.5 Comparison with PHD2 and algal CP4H in complex with peptide substrate

The PPHD:EF-Tu complex structure has broader implications for understanding substrate recognition by prolyl-hydroxylases, including the PHDs and CP4Hs (and CP4H-related proteins). The structure solution of isolated human PHD2 (*1*) and of human PHD2 in complex with HIF-α CODD (residues 556-574)(PDB ID: 3HQR)(*23*) represented major breakthroughs in understanding the structural basis for HIF-α hydroxylation by the human PHDs, and allowed for more clarity in the design of PHD2 inhibitors through identification of catalytically important residues (*37*).

3.5.1 Comparison of the β2-β3 finger loop in substrate complex structures of PPHD, PHD2 and CrP4H

The positions of the β2-β3 finger loop in the isolated and substrate complex structures of PPHD are strikingly similar to that observed for PHD2 with and without HIF-α (PDB IDs: 2G1M and 3HQR) and to *Chlamydomonas reinhardtii* prolyl-4-hydroxylase in complex with a proline-rich peptide substrate (CrP4H; PDB ID: 3GZE; sequence PSPSPS)(Figure 3.22, *A* and *C*)(*38*). The finger and thumb loops of both PHD2 and CrP4H are ordered in the substrate-bound structure and function to enclose around and position substrate within the active site in a similar manner to PPHD (Figure 3.22, *A* and *C*).

Both HIF-α residue Pro564 and a proline residue are positioned in their C-4 *endo* conformations in the PHD2 and CrP4H active sites, respectively, consistent with the position of EF-Tu Pro54 in the PPHD active site (Figure 3.22, *B* and *D*). This conformation is in contrast to that of Hyp564 observed in

structures of HIF-α CODD fragment in complex with pVHL, which adopts a C-4 *exo* conformation (*39, 40*). Studies have shown that *trans*-4 hydroxylation induces a stereoelectronic conformational bias such that *trans*-4 hydroxyproline preferentially adopts the C-4 *exo* conformation (*41, 42*). The observation that the active site prolines in structures of PPHD, PHD2, and CrP4H share a C-4 *endo* conformation suggests that the role of *trans*-4 hydroxyproline in signalling (e.g. animal hypoxia sensing) may have emerged later than the evolutionary divergence of the PPHDs, PHDs, and CP4Hs from an ancestral prolyl-4-hydroxylase (for more discussion, see Chapter 6).

The PHD2 β2-β3 finger loop (23 residues; 236-258) is of a similar length to that of PPHD and contains an equal number (6) of flexible residues; 2 glycine (Gly238 and Gly253) and 4 serine (Ser242, Ser245, Ser247, and Ser248) residues. Superimpositions of PPHD:EF-Tu and PHD2:CODD reveals that the PHD2 β2-β3 finger loop and β5(6)(II)-β6(7)(III) thumb loop (6 residues; 316-321) function to enclose CODD substrate in a similar manner to those of PPHD (Figure 3.22*A*). In contrast to that of PPHD, the β5(6)(II)-β6(7)(III) thumb loop of PHD2 is ordered in the isolated PHD2 structure and all atoms move little from their positions in the isolated structure (r.m.s.d. 0.2 Å), suggesting that the PPHD thumb loop is more flexible in nature and that more extensive conformational changes comprise PPHD binding to EF-Tu than do PHD2 binding to HIF-α ODDs. Lastly, the PHD2 β5(6)(II)-β6(7)(III) thumb loop is significantly less polar (containing one aspartate residue and one arginine residue) and the salt bridge between Arg57 and Asp130 that contributes to closing the PPHD finger and thumb loops around the EF-Tu switch I loop is not observed in the PHD2:CODD structure.

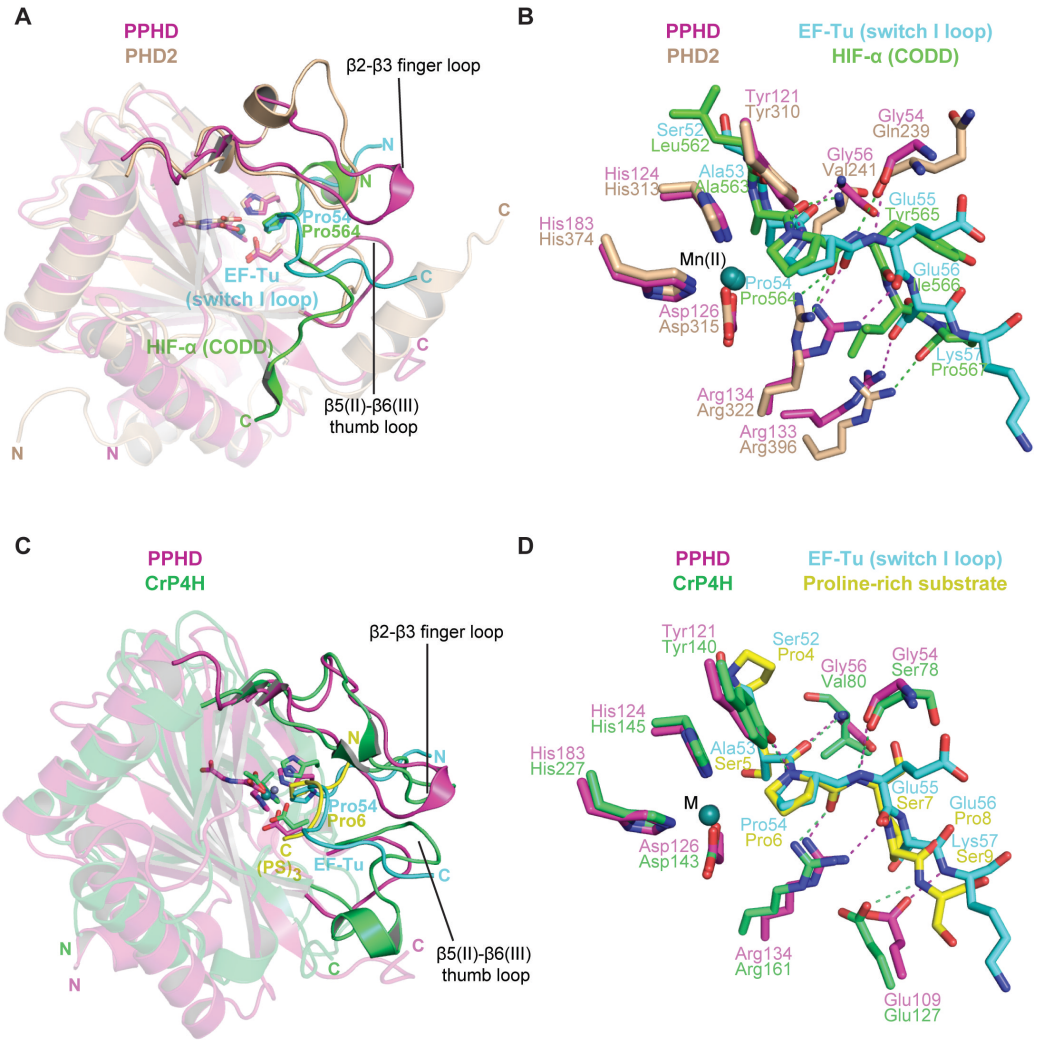


Figure 3.22. *P. putida* PPHD, *Homo sapiens* PHD2, and *Chlamydomonas reinhardtii* P4H share similar structural motifs underlying substrate recognition. (**A**) Overall and (**B**) active site superimpositions of the PPHD(magenta):EF-Tu(cyan) complex and PHD2(sand):CODD(green) reveals equivalent positions of the β2-β3 finger loop and the β5(II)-β6(III) thumb loop, location of substrate in the active site, and conservation of critical residues involved in substrate binding. (**C**) Overall and (**D**) active site superimpositions of the PPHD(magenta):EF-Tu(cyan) complex and CrP4H(green) bound to proline-rich collagen-like substrate (sequence: PSPSPS)(yellow). Figure and figure legend are adapted with permission from J.S. Scotti *et al*, Human oxygen sensing may have origins in prokaryotic elongation factor Tu prolyl-hydroxylation. *Proc Nat Acad Sci USA*. In press (2014).

3.5.2 Comparison of the β5(II)-β6(III) thumb loop in substrate complex structures of PPHD, PHD2 and CrP4H

The β2-β3 finger loop (76-99, 24 residues) of CrP4H is of similar length to those of PPHD and PHD2, yet the β5(II)-β6(III) thumb loop (146-160, 15

residues) of CrP4H is extended compared to those of PHD2 and PPHD and features an additional α-helix not present in the isolated CrP4H structure (Figure 3.22*C*), which suggests that the role of the β5(II)-β6(III) thumb loop in prolyl-4-hydroxylase substrate recognition may be conserved across subfamilies. Further, the elongated β5(II)-β6(III) thumb loop of CrP4H may play a role in its recognition of extended collagen-like like sequences.

3.5.3 Comparison of the active sites of PPHD, PHD2 and CrP4H bound to substrate

Superimposition of the substrate-bound PPHD active site residues with those of PHD2 and CrP4H reveals conservation of critical enzyme-substrate interactions (Figure 3.22, *B* and *D*). Notably, PPHD, PHD2, and CrP4H all share conserved interactions from backbone atoms of their respective β2-β3 finger loops to backbone atoms of substrate despite the lack of either enzyme or substrate sequence conservation: the backbone carbonyl of $Gly54_{PPHD}$/$Gln239_{PHD2}$/$Ser78_{CrP4H}$ is positioned to hydrogen bond to the backbone nitrogen of $Glu55_{EF\text{-}Tu}$/$Tyr565_{HIF\text{-}\alpha}$/$Ser7_{Collagen\text{-}like}$ and the backbone nitrogen of $Gly56_{PPHD}$/$Val241_{PHD2}$/$Val80_{CrP4H}$ is positioned to hydrogen bond to the backbone carbonyl of $Ala53_{EF\text{-}Tu}$/$Ala563_{HIF\text{-}\alpha}$/$Ser5_{Collagen\text{-}like}$ (Figure 3.22, *B* and *D*). These backbone interactions are of particular interest as their conservation in PPHD, PHD2, and CrP4H suggests an ancient importance in prolyl-hydroxylase substrate recognition, with subsequent divergent evolution of side chains resulting in different substrate-recognition profiles, and thus different families of prolyl-hydroxylases.

Interestingly, PPHD Arg133 and PHD2 Arg396 both interact with substrate ($Arg133_{PPHD}$ is positioned to form a salt bridge with $Glu56_{EF\text{-}Tu}$;

$Arg396_{PHD2}$ is positioned to hydrogen bond to the carbonyl oxygen of $Pro567_{HIF-\alpha}$), yet originate from different positions; Arg133 is positioned on the PPHD β5(II)-β6(III) thumb loop whereas Arg396 is positioned on the PHD2 *C*-terminal α-helix (Figure 3.22*B* and Figure 3.23). In contrast, CrP4H does not share a conserved $Arg133_{PPHD}$, yet it does share a conserved $Glu109_{PPHD}/Glu127_{CrP4H}$ that forms a hydrogen bond to backbone nitrogen of substrate.

3.5.4 PPHD and PHD2 share a role of their C-terminus in substrate recognition

Previous work has shown that the PHD2 *C*-terminal α-helix (residues 393-404) is important in substrate recognition and catalysis, as evidenced by mutagenesis studies showing that deletion of this helix abolishes PHD2 activity (*23, 43*). The PPHD *C*-terminus loop (residues 201-207, corresponding to PHD2 residues 392-408) is positioned in a gap between EF-Tu domains I and III and directly contacts the EF-Tu switch I loop and domain II, likely helping to stabilise the EF-Tu GDP-bound conformation. The side chain of $Arg201_{PPHD}$ is positioned to hydrogen bond to $Asn277_{EF-Tu}$; the side chain of $Arg202_{PPHD}$ is positioned to form a salt bridge with $Glu236_{EF-Tu}$)(Figure 3.23). In contrast, the PHD2 *C*-terminal α-helix interacts directly with CODD (the side chain of Arg396 is positioned to hydrogen bond to the carbonyl oxygen of Pro567; the Arg396 and Lys400 side chains are positioned to form a salt bridge with $Asp571_{HIF-\alpha}$)(Figure 3.23), suggesting that a role for the *C*-terminus in prolyl-hydroxylase substrate recognition is evolutionarily conserved.

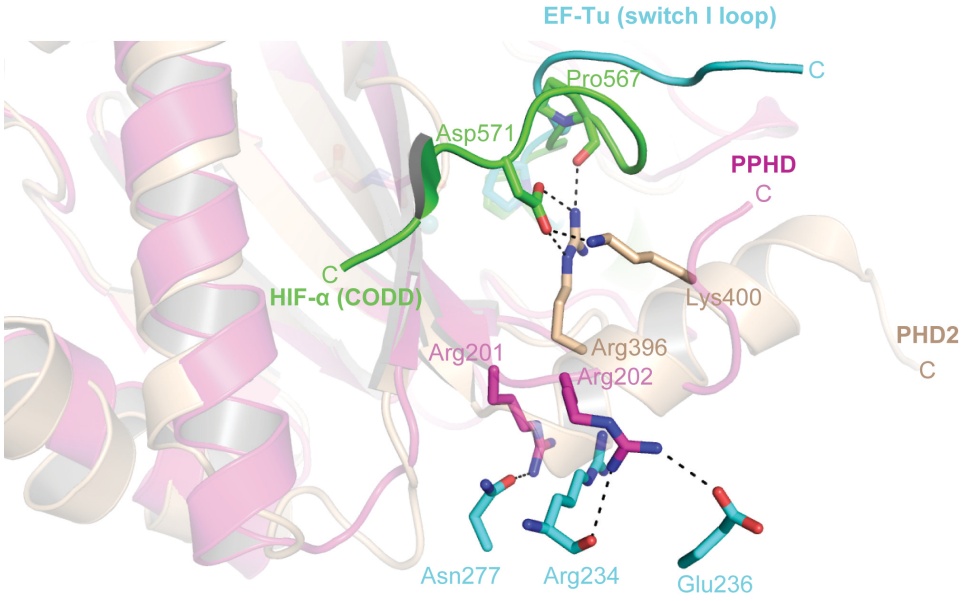


Figure 3.23. Interactions of the PPHD and PHD2 *C*-terminus with their respective substrates suggests a conserved role for the *C*-terminus in substrate recognition by prolyl-hydroxylases. Superimposition of PPHD(magenta):EF-Tu(cyan) and PHD2(sand):CODD(green) reveals that the PPHD *C*-terminal loop and the PHD2 *C*-terminal α-helix both interact with substrate (see text for details). Figure and figure legend are adapted with permission from J.S. Scotti *et al*, Human oxygen sensing may have origins in prokaryotic elongation factor Tu prolyl-hydroxylation. *Proc Nat Acad Sci USA*. In press (2014).

3.5.5 Structural conservation of EF-Tu, HIF-α and collagen-like substrates

Conservation of structural features extends beyond the prolyl-hydroxylases themselves. Remarkably, the seemingly evolutionarily distinct EF-Tu, HIF-α, and collagen-like substrates share similar conformations when complexed to their respective prolyl-hydroxylase partners; EF-Tu (residues 50-57) in complex with PPHD shares an analogous orientation and conformation to that of CODD (residues 560-567) in complex with PHD2 (backbone atoms r.m.s.d. 0.1 Å) and to that of proline-rich peptide substrate in complex with CrP4H (backbone atoms r.m.s.d. 0.1 Å)(Figure 3.22, *B* and *D*).

3.5.6 Comparison of prolyl-4- and prolyl-3-hydroxylases

Inspection of the PPHD active site not only reveals conservation of residues involved in substrate-recognition among the prolyl-hydroxylases, but also differences between the prolyl-4- (PPHD, PHD2, CrP4H) and prolyl-3-hydroxylases (OGFOD1/Tpa1)(Figure 3.24, *A*-*C* and Figure 3.25). These differences may be of interest in the design of selective inhibitors of prolyl-3- or prolyl-4-hydroxylases and may also reflect the evolutionary origins of these enzymes.

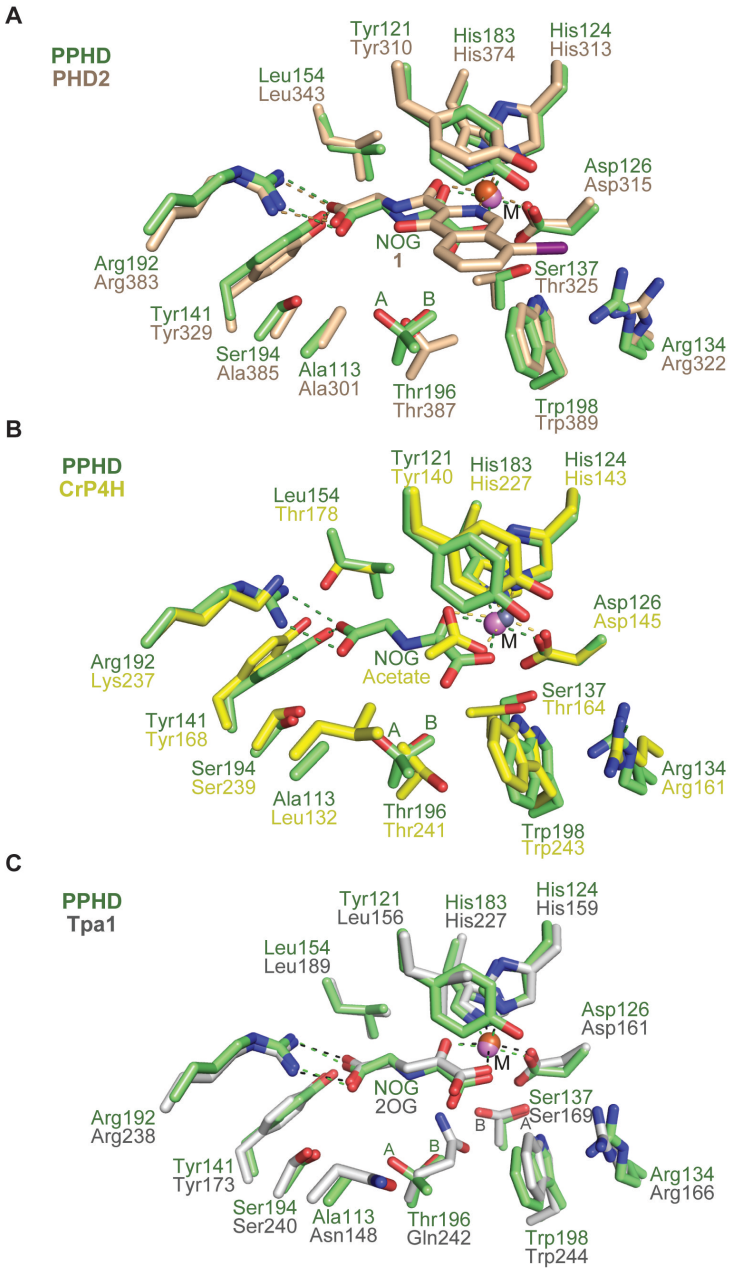


Figure 3.24. Comparison of prolyl-4-hydroxylase and prolyl-3-hydroxylase active site topologies reveals both conservation and differences in substrate-recognition residues. Superimpositions of the prolyl-4-hydroxylase PPHD and (**A**) prolyl-4-hydroxylase PHD2:**1** [**1**: *N*-(4-hydroxy-8-iodoisoquinolin-3-yl)carbonylglycine](*1*), (**B**) prolyl-4-hydroxylase CrP4H:collagen-like, and (**C**) prolyl-3-hydroxylase Tpa1 (PDB ID: 3KT7)(*44*) active site residues reveal conservation of $Arg^{134}{}_{PPHD}/Arg^{322}{}_{PHD2}/Arg^{161}{}_{CrP4H}/Arg^{166}{}_{Tpa1}$ but lack conservation of $Tyr^{121}{}_{PPHD}/Tyr^{310}{}_{PHD2}$ /$Tyr^{140}{}_{CrP4H}/Leu^{156}{}_{Tpa1}$ between prolyl-4- and prolyl-3-hydroxylases, suggesting that differences in active site residues may alter the position of the substrate proline in the active site.

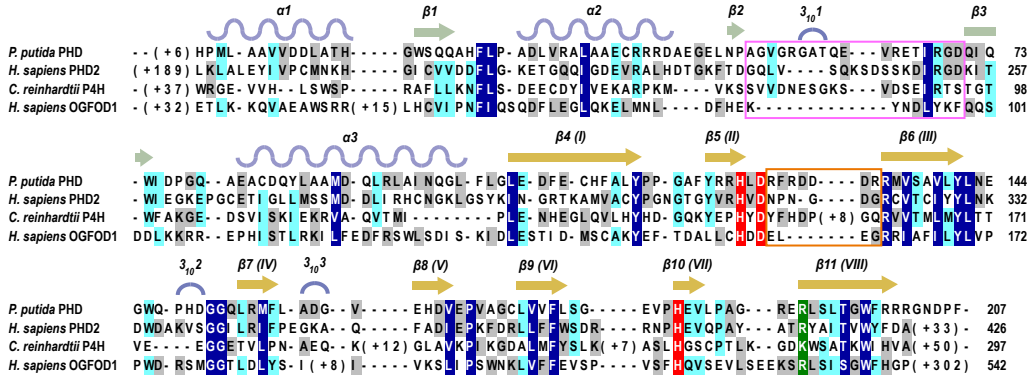


Figure 3.25. Structure-based sequence alignments of structurally characterised prolyl-hydroxylases. Alignments of *Pseudomonas putida* PPHD (PDB ID: 4IW3), *Homo sapiens* PHD2 (PDB ID: 3HQR)(*23*), *Chlamydomonas reinhardtii* collagen P4H (PDB ID: 3GZE)(*38*), and *Homo sapiens* OGFOD1 (PDB ID: 4NHX) reveal the presence of an extended β2-β3 finger loop (magenta box) and the β5(II)-β6(III) thumb loop (orange box) in prolyl-4- but not in prolyl-3-hydroxylases. Metal-binding residues (red), 2OG-binding residue (green), highly conserved residues (dark blue), mostly conserved residues (cyan), weakly conserved residues (grey), DBSH β-strands (yellow), non-DSBH β-strands (light green), α- and 3_{10}-helices (lavender). Secondary structure assignment is rooted in PPHD. Figure and figure legend are adapted with permission from J.S. Scotti *et al*, Human oxygen sensing may have origins in prokaryotic elongation factor Tu prolyl-hydroxylation. *Proc Nat Acad Sci USA*. In press (2014).

All structurally characterised prolyl-4-hydroxylases have a conserved arginine positioned at the catalytic site opening to hydrogen bond to the backbone carbonyl of the substrate proline ($Arg134_{PPHD}$, $Arg322_{PHD2}$, $Arg161_{CrP4H}$) (Figure 3.24, *A* and *B*). In addition, prolyl-4-hydroxylases possess a conserved tyrosine ($Tyr121_{PPHD}$, $Tyr310_{PHD2}$; $Tyr140_{CrP4H}$) of which the hydroxyl is positioned to hydrogen bond to the backbone nitrogen of the substrate proline (Figure 3.24, *A* and *B*). In contrast, prolyl-3-hydroxylases possess a hydrophobic leucine residue at the analogous position ($Leu152_{OGFOD1}$; $Leu156_{Tpa1}$), which may enable prolyl-3- over prolyl-4-hydroxylation by altering the position of the substrate proline-residue in the active site. The identification of these differences may be useful in developing inhibitors selective for the human PHDs over the CP4Hs or ribosomal prolyl-hydroxylases.

3.6 Conclusions

The identification and biochemical characterisation of PPHD in *Pseudomonas spp.* reveals that PPHD and PHD2 behaved similarly in solution and prompted structural studies on PPHD. A crystal structure of PPHD was determined and revealed a striking structural resemblance to PHD2, confirming that PPHD is the first known prokaryotic homologue of the human PHDs and supporting a common ancestry for the two proteins. Analyses of the PPHD structure and superimpositions with the structure of isolated PHD2 suggested that PPHD and PHD2 share a conserved mechanism of substrate-binding, enabled by substrate-induced conformational changes in their respective β2-β3 finger loops. Comparisons of PPHD with both prokaryotic and eukaryotic CP4Hs, *B. anthracis* P4H and *C. reinhardtii* P4H, respectively, revealed key differences in secondary structure and active site residues between the two prolyl-4-hydroxylase subfamilies.

The structure of *P. putida* EF-Tu was determined in its GDP-bound translationally inactive state. Importantly, no electron density was observed for the switch I loop, including Pro54, in the structure, hinting at a possible flexibility in solution conserved in substrates of PHDs and PHD homologues and suggesting that PPHD may order the loop upon binding.

To investigate the structural mechanisms underlying PPHD-catalysed hydroxylation of EF-Tu Pro54, a structure of a PPHD:EF-Tu protein-protein complex was successfully determined. Confirming the hypotheses derived from the solution studies, major conformational changes are observed in both PPHD and EF-Tu. The EF-Tu switch I loop is sequestered in the PPHD catalytic site, thereby positioning Pro54 for *trans* 4-hydroxylation. The observation that PPHD

interacts with EF-Tu outside the immediate vicinity of the active site suggests that interactions between PHD2 and other regions of HIF-α may be important for PHD2 catalysis.

Structural comparisons of PPHD:EF-Tu with PHD2 in complex with CODD peptide and CrP4H in complex with collagen-like peptide revealed a conserved movement of the β2-β3 finger loop upon substrate binding. Further analyses of PPHD/PHD2/CrP4H and OGFOD1/Tpa1 revealed insights into key structural elements differentiating prolyl-4- and prolyl-3-hydroxylases and raised questions regarding the evolutionary relationship of these enzymes (for a detailed discussion of prolyl-hydroxylase evolution, see Chapter 6). In sum, the crystal structures of *P. putida* PPHD, EF-Tu, and a PPHD:EF-Tu complex are vital for understanding induced fit during 2OG oxygenase catalysed protein-modification and provide a platform for the design of new types of 2OG oxygenase inhibitors based on various conformational states, rather than active site iron chelators, which comprise the majority of 2OG oxygenase inhibitors including those PHD inhibitors in clinical trials for the treatment of anemia (*37*).

3.7 Experimental Procedures

3.7.1 Protein purification

E. coli BL21(DE3) transformed with the pET28a_PPHD or pET28a_EF-Tu plasmid (encoding for protein with an *N*-terminal hexahistidine tag) were grown (37 °C; 180 rpm) to an OD_{600} of 0.6; recombinant protein production was induced with 0.5 mM β-D-1-thiogalactopyranoside (IPTG). The cells were then grown at 18°C overnight, harvested by centrifugation (10,000x*g*; 7 min), and stored at -80 °C. Cell pellets (for the PPHD:EF-Tu complex, 1:1 w/w cell pellets were used) were resuspended in 50 mM HEPES pH 7.5, 500 mM NaCl, 20 mM imidazole, 1 mM tris(2-carboxyethyl)phosphine (TCEP), one EDTA-free protease inhibitor tablet (Roche), and approximately 1 mg DNAseI (bovine pancreas, grade II, Roche) at room temperature with gentle stirring. Cells were lysed on ice by sonication and the lysate was cleared by centrifugation (50,000x*g*; 20 min). The supernatant was then loaded onto a 5 mL HisTrap FF column that had been preequilibrated with the resuspension buffer and purified using an AKTA FPLC system (GE Healthcare). The column was washed with 10 column volumes of 50 mM HEPES pH 7.5, 500 mM NaCl, 50 mM imidazole, 1 mM TCEP and protein was eluted with 50 mM HEPES pH 7.5, 500 mM NaCl, 250 mM imidazole, 1 mM TCEP. The purified sample was then exchanged to 50 mM HEPES pH 7.5, 200 mM NaCl using a PD-10 column (Millipore). PPHD, EF-Tu, or the PPHD:EF-Tu protein-protein complex were further purified using a Superdex 75 size exclusion column (GE Healthcare) preequilibrated in 50 mM HEPES pH 7.5, 200 mM NaCl (PPHD), 50 mM HEPES pH 7.5, 200 mM NaCl, 10 mM $MgCl_2$, 1 μM GDP (EF-Tu), or 50 mM HEPES pH 7.5, 10 mM $MgCl_2$, 1 μM GDP (PPHD:EF-Tu complex).

Proteins were eluted using the same buffer. PPHD-, EF-Tu-, or PPHD:EF-Tu-containing fractions were pooled, concentrated to 35 mg/mL (PPHD), 20 mg/mL (EF-Tu), or 15 mg/mL (PPHD:EF-Tu) by diafiltration, aliquoted (20 μL), flash frozen in liquid N_2, and stored at -80 °C.

3.7.2 Differential scanning fluorimetry (T_m shift) assays (*2*)

To determine the effect of various buffers, salts, metals, and small molecules on the stability of proteins, differential scanning fluorimetry was performed using a BioRad MiniOpticon real-time PCR detection system and MJ Mini thermal cycler. 5,000x SYPRO Orange dye (Invitrogen) was used for nonspecific binding to hydrophobic protein residues. For each buffer, metal, and inhibitor, samples were run in duplicate in a 48-well plate in accordance with the following general recipe.

For metal screen assays, PPHD was first incubated overnight in 50 mM EDTA (ethylenediaminetetraacetic acid) to generate the apo protein. The following metals were screen against apo PPHD and wild type, Fe(II) bound, PPHD, and buffer was used as a reference in both cases: $MgCl_2$, $MgSO_4$, $ZnCl_2$, $ZnSO_4$, $MnCl_2$, $MnSO_4$, $CoCl_2$, $CdCl_2$, $NiSO_4$, $FeCl_3$, $Fe_2(SO_4)_3$. Transition metals were chosen on their ability to exist in a stable +2 oxidation state, with the exception of Fe(III) as a negative control.

Table 3.26. Representative T_m shift assay conditions. The reagents were dissolved in 1x T_m shift buffer.

Reagent	Stock concentration (μM)	Final concentration (μM)
PPHD	500	2
SYPRO Orange	5,000x	1x
Metal	1,000*	50
Inhibitor	10	100

Table 3.27. Compositions of T_m shift buffers.

Buffer name	pH	[M]	Buffer name	pH	[M]	Buffer name	pH	[M]
phosphate-citrate	4.2	0.2	Tris	7.0	0.4	Tris	8.0	0.4
sodium acetate	4.5	0.4	bis-tris propane	7.0	0.4	tricine	8.0	0.4
sodium citrate	5.0	0.4	MOPSO	7.0	0.4	Tris	8.5	0.4
sodium citrate	5.5	0.2	PIPES	7.0	0.4	bicine	9.0	0.4
MES	6.0	0.2	bis-tris	7.0	0.4	CHES	9.5	0.4
Na/K phosphate	6.2	0.2	HEPES	7.5	0.4	CAPSO	9.5	0.4
bis-tris propane	6.5	0.4	TES	7.5	0.4	CAPS	10.5	0.4
HEPES	7.0	0.4	imidazole	8.0	0.4			

The fluorescence readings were taken every 1 °C from 25 – 95 °C with a linear temperature increase of 1 °C/min. In order to determine accurate T_m values, the data were exported to and analysed on GraphPad Prism software. The midpoint, or T_m, was calculated by fitting a Boltzmann curve between the minimum and maximum fluorescence intensities. The conditions that gave the highest ΔT_m compared to the reference indicated those buffer, metals, and/or inhibitors that best stabilised PPHD (*2*).

3.7.3 X-ray crystallography

Crystals of PPHD in complex with NOG and crystals of PPHD:EF-Tu complex were grown in sitting drops using the vapour diffusion method (drop size: 200-300 nL) at 293 K in 96-well Intelliplates (Art Robbins); crystals of EF-Tu were grown at 277 K. Crystals were cryo-protected by transfer to 25% (v/v) glycerol in well solution and then harvested in nylon loops (Hampton Research) and cryo-cooled by plunging in liquid nitrogen. Data were collected at 100 K using single crystals at Diamond Light Source beamline I04 (PPHD:Mn(II):NOG and PPHD:EF-Tu complex) with a ADSC Quantum 315r detector, and Diamond Light Source beamline I03 (EF-Tu:Mg(II):GDP) with a Pilatus 6M-F detector. Data were then indexed, integrated, and scaled using SCALA (PPHD:Mn(II):NOG)(*45*),

HKL3000 (EF-Tu:Mg(II):GDP)(*46*), and XSCALE (PPHD:EF-Tu complex)(*47, 48*). The structure of PPHD:Mn(II):NOG was determined by molecular replacement (MR) using the MR-PHASER (*18*) subroutine of PHENIX (*19*) using *H. sapiens* PHD2 (PDB ID: 2G1M) as the search model. The EF-Tu:Mg(II):GDP was determined by MR using *E. coli* EF-Tu (PDB ID: 1DG1) as the search model. The PPHD:EF-Tu protein-protein complex was determined by MR using *E. coli* EF-Tu (PDB ID: 1DG1) and the refined PPHD:Mn(II):NOG as search models. Model building and refinement were performed iteratively using COOT (*21*) and PHENIX until converging *R* and *R*free no longer decreased. Mn(II), Mg(II), NOG, and GDP and water molecules were modelled in the final stages of refinement based on the $F_{obs} - F_{calc}$ electron density map.

Table 3.28. Final crystallisation conditions.

	PPHD:NOG	EF-Tu:GDP	PPHD:EF-Tu complex
Protein solution	~10 mg/mL PPHD, 2 mM $MnCl_2$, 5 mM NOG, 5 mM EF-Tu_{44-63}	~18 mg/mL EF-Tu, 0.05 M HEPES pH 7.5, 10 m	~15 mg/mL PPHD:EF-Tu, 0.8 mM $MnCl_2$, 0.9 mM NOG
Reservoir solution	0.1 M MIB buffer pH 5.0, 25% (w/v) PEG 1500	0.05 M MES pH 5.6, 0.1 M MgOAc, 20% (v/v) MPD, 10 mM $SrCl_2$	0.1 M HEPES pH 7.5, 0.1 M $MgCl_2$, 22% (w/v) polyacrylate 5100, 1 mM GDP

Table 3.29. Crystallographic data and refinement statistics.

	PPHD:NOG	EF-Tu:GDP	PPHD:EF-Tu complex
X-ray source	Diamond Light Source beamline I04	Diamond Light Source beamline I03	Diamond Light Source beamline I04
Wavelength (Å)	0.97950	0.97630	0.97950
PDB Acquisition Code	4J25	4J0Q	4IW3
Resolution (Å)	1.97 (2.02–1.97)§	2.29 (2.38–2.29)§	2.70 (2.84-2.70)§
Space group	*P* 1	*C* 1 2 1	$P\,3_1\,2\,1$
Unit Cell Dimensions (*a* Å, *b* Å, *c* Å)	45.33 62.16 132.83	212.51 155.62 99.3	200.70 200.70 74.83
Molecules per a.u.	8	5	4 (2 PPHD; 2 EF-Tu)
Total Number of Reflections Observed	383204	856176	438242
Number of Unique Reflections	99718 (7268)§	129510 (12208)§	47866 (6913)§
Redundancy	3.8 (3.9)§	6.6 (5.6)§	9.1 (8.2)§
Completeness (%)	97.7 (96.7)§	99.3 (93.9)§	99.9 (99.2)§
Wilson B	27.8	64.1	74.7
I/σ(I)	10.6 (2.3)§	30.2 (1.7)§	13.4 (1.7)§
§§R_{merge}	0.069	0.056	0.086
*R_{cryst}	0.219	0.174	0.168
†R_{free}	0.257	0.213	0.218
¶RMS deviation	0.009 (1.0°)	0.01 (1.3°)	0.01 (1.2°)
Average *B* factors (Å²)	34.1	67.1	60.2
Number of Water Molecules	563	316	166

§ Parentheses indicate high resolution shell

§§ $R_{merge} = \sum_j \sum_h | I_{hj} - \langle I_h \rangle | / \sum_j \sum_h \langle I_h \rangle \times 100$

- $R_{cryst} = \sum ||Fobs| - |Fcalc|| / |Fobs| \times 100$

† R_{free}, based on 2-5% of the total reflections

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

3.8 References

1. M. A. McDonough, V. Li, E. Flashman, R. Chowdhury, C. Mohr, B. M. Lienard, J. Zondlo, N. J. Oldham, I. J. Clifton, J. Lewis, L. A. McNeill, R. J. Kurzeja, K. S. Hewitson, E. Yang, S. Jordan, R. S. Syed, C. J. Schofield, Cellular oxygen sensing: Crystal structure of hypoxia-inducible factor prolyl hydroxylase (PHD2). *Proc Natl Acad Sci U S A.* **103**, 9814-9819 (2006).
2. F. H. Niesen, H. Berglund, M. Vedadi, The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nature Protocols.* **2**, 2212-2221 (2007).
3. E. C. Y. Woon, A. Tumber, A. Kawamura, L. Hillringhaus, W. Ge, N. R. Rose, J. H. Y. Ma, M. C. Chan, L. J. Walport, K. H. Che, S. S. Ng, B. D. Marsden, U. Oppermann, M. A. McDonough, C. J. Schofield, Linking of 2-Oxoglutarate and Substrate Binding Sites Enables Potent and Highly Selective Inhibition of JmjC Histone Demethylases. *Angew Chem Int Edit.* **51**, 1631-1634 (2012).
4. N. Poklar, J. Lah, M. Salobir, P. Macek, G. Vesnaver, pH and temperature-induced molten globule-like denatured states of equinatoxin II: a study by UV-melting, DSC, far- and near-UV CD spectroscopy, and ANS fluorescence. *Biochemistry.* **36**, 14345-14352 (1997).
5. M. W. Pantoliano, E. C. Petrella, J. D. Kwasnoski, V. S. Lobanov, J. Myslik, E. Graf, T. Carver, E. Asel, B. A. Springer, P. Lane, F. R. Salemme, High-density miniaturized thermal shift assays as a general strategy for drug discovery. *J Biomol Screen.* **6**, 429-440 (2001).
6. E. C. Woon, A. Tumber, A. Kawamura, L. Hillringhaus, W. Ge, N. R. Rose, J. H. Ma, M. C. Chan, L. J. Walport, K. H. Che, S. S. Ng, B. D. Marsden, U. Oppermann, M. A. McDonough, C. J. Schofield, Linking of 2-oxoglutarate and substrate binding sites enables potent and highly selective inhibition of JmjC histone demethylases. *Angew Chem Int Ed Engl.* **51**, 1631-1634 (2012).
7. N. E. Good, G. D. Winget, W. Winter, T. N. Connolly, S. Izawa, R. M. M. Singh, Hydrogen Ion Buffers for Biological Research. *Biochemistry.* **5**, 467-& (1966).
8. R. Page, S. K. Grzechnik, J. M. Canaves, G. Spraggon, A. Kreusch, P. Kuhn, R. C. Stevens, S. A. Lesley, Shotgun crystallization strategy for structural genomics: an optimized two-tiered crystallization screen against the Thermotoga maritima proteome. *Acta Crystallogr D Biol Crystallogr.* **59**, 1028-1037 (2003).
9. J. Newman, D. Egan, T. S. Walter, R. Meged, I. Berry, M. Ben Jelloul, J. L. Sussman, D. I. Stuart, A. Perrakis, Towards rationalization of crystallization screening for small- to medium-sized academic laboratories: the PACT/JCSG+ strategy. *Acta Crystallogr D Biol Crystallogr.* **61**, 1426-1431 (2005).
10. A. McPherson, A comparison of salts for the crystallization of macromolecules. *Protein Sci.* **10**, 418-422 (2001).

11. C. Grimm, A. Chari, K. Reuter, U. Fischer, A crystallization screen based on alternative polymeric precipitants. *Acta Crystallogr D Biol Crystallogr.* **66**, 685-697 (2010).
12. A. D'arcy, A. Mac Sweeney, M. Stihle, A. Haber, The advantages of using a modified microbatch method for rapid screening of protein crystallization conditions. *Acta Crystallogr D.* **59**, 396-399 (2003).
13. A. McPherson, B. Cudney, Searching for silver bullets: An alternative strategy for crystallizing macromolecules. *J Struct Biol.* **156**, 387-406 (2006).
14. B. Cudney, S. Patel, K. Weisgraber, Y. Newhouse, Screening and Optimization Strategies for Macromolecular Crystal-Growth. *Acta Crystallogr D.* **50**, 414-4123 (1994).
15. S. Trakhanov, F. A. Quiocho, Influence of divalent cations in protein crystallization. *Protein Sci.* **4**, 1914-1919 (1995).
16. R. Sousa, Use of Glycerol, Polyols and Other Protein-Structure Stabilizing Agents in Protein Crystallization. *Acta Crystallogr D.* **51**, 271-277 (1995).
17. A. Mcpherson, S. Koszelak, H. Axelrod, J. Day, L. Robinson, M. Mcgrath, R. Williams, D. Cascio, The Effects of Neutral Detergents on the Crystallization of Soluble-Proteins. *J Cryst Growth.* **76**, 547-553 (1986).
18. A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, R. J. Read, Phaser crystallographic software. *J Appl Crystallogr.* **40**, 658-674 (2007).
19. P. D. Adams, P. V. Afonine, G. Bunkoczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, P. H. Zwart, PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr.* **66**, 213-221 (2010).
20. P. H. Zwart, R. W. Grosse-Kunstleve, A. A. Lebedev, G. N. Murshudov, P. D. Adams, Surprises and pitfalls arising from (pseudo)symmetry. *Acta Crystallogr D Biol Crystallogr.* **64**, 99-107 (2008).
21. P. Emsley, K. Cowtan, Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr.* **60**, 2126-2132 (2004).
22. W. Aik, M. A. McDonough, A. Thalhammer, R. Chowdhury, C. J. Schofield, Role of the jelly-roll fold in substrate binding by 2-oxoglutarate oxygenases. *Curr Opin Struct Biol.* **22**, 691-700 (2012).
23. R. Chowdhury, M. A. McDonough, J. Mecinovic, C. Loenarz, E. Flashman, K. S. Hewitson, C. Domene, C. J. Schofield, Structural basis for binding of hypoxia-inducible factor to the oxygen-sensing prolyl hydroxylases. *Structure.* **17**, 981-989 (2009).
24. M. A. Culpepper, E. E. Scott, J. Limburg, Crystal structure of prolyl 4-hydroxylase from Bacillus anthracis. *Biochemistry.* **49**, 124-133 (2010).
25. M. K. Koski, R. Hieta, C. Bollner, K. I. Kivirikko, J. Myllyharju, R. K. Wierenga, The active site of an algal prolyl 4-hydroxylase has a large structural plasticity. *Journal of Biological Chemistry.* **282**, 37112-37123 (2007).
26. H. Hasegawa, L. Holm, Advances and pitfalls of protein structural alignment. *Curr Opin Struct Biol.* **19**, 341-348 (2009).

27. K. Abel, M. D. Yoder, R. Hilgenfeld, F. Jurnak, An alpha to beta conformational switch in EF-Tu. *Structure.* **4**, 1153-1159 (1996).
28. H. Wolf, G. Chinali, A. Parmeggiani, Mechanism of the inhibition of protein synthesis by kirromycin. Role of elongation factor Tu and ribosomes. *Eur J Biochem.* **75**, 67-75 (1977).
29. A. Parmeggiani, I. M. Krab, T. Watanabe, R. C. Nielsen, C. Dahlberg, J. Nyborg, P. Nissen, Enacyloxin IIa pinpoints a binding pocket of elongation factor Tu for development of novel antibiotics. *J Biol Chem.* **281**, 2893-2900 (2006).
30. M. Valle, A. Zavialov, W. Li, S. M. Stagg, J. Sengupta, R. C. Nielsen, P. Nissen, S. C. Harvey, M. Ehrenberg, J. Frank, Incorporation of aminoacyl-tRNA into the ribosome as seen by cryo-electron microscopy. *Nature Structural Biology.* **10**, 899-906 (2003).
31. A. Parmeggiani, I. M. Krab, S. Okamura, R. C. Nielsen, J. Nyborg, P. Nissen, Structural basis of the action of pulvomycin and GE2270 A on elongation factor Tu. *Biochemistry.* **45**, 6846-6857 (2006).
32. S. E. Heffron, F. Jurnak, Structure of an EF-Tu complex with a thiazolyl peptide antibiotic determined at 2.35 angstrom resolution: Atomic basis for GE2270A inhibition of EF-Tu. *Biochemistry.* **39**, 37-45 (2000).
33. H. Wolf, G. Chinali, A. Parmeggiani, Kirromycin, an inhibitor of protein biosynthesis that acts on elongation factor Tu. *Proc Natl Acad Sci U S A.* **71**, 4910-4914 (1974).
34. X. Agirrezabala, J. Frank, Elongation in translation as a dynamic interaction among the ribosome, tRNA, and elongation factors EF-G and EF-Tu. *Q Rev Biophys.* **42**, 159-200 (2009).
35. T. M. Schmeing, V. Ramakrishnan, What recent ribosome structures have revealed about the mechanism of translation. *Nature.* **461**, 1234-1242 (2009).
36. E. Krissinel, K. Henrick, Inference of macromolecular assemblies from crystalline state. *J Mol Biol.* **372**, 774-797 (2007).
37. N. R. Rose, M. A. McDonough, O. N. King, A. Kawamura, C. J. Schofield, Inhibition of 2-oxoglutarate dependent oxygenases. *Chem Soc Rev.* **40**, 4364-4397 (2011).
38. M. K. Koski, R. Hieta, M. Hirsila, A. Ronka, J. Myllyharju, R. K. Wierenga, The crystal structure of an algal prolyl 4-hydroxylase complexed with a proline-rich peptide reveals a novel buried tripeptide binding motif. *J Biol Chem.* **284**, 25290-25301 (2009).
39. J. H. Min, H. Yang, M. Ivan, F. Gertler, W. G. Kaelin, Jr., N. P. Pavletich, Structure of an HIF-1alpha -pVHL complex: hydroxyproline recognition in signaling. *Science.* **296**, 1886-1889 (2002).
40. W. C. Hon, M. I. Wilson, K. Harlos, T. D. Claridge, C. J. Schofield, C. W. Pugh, P. H. Maxwell, P. J. Ratcliffe, D. I. Stuart, E. Y. Jones, Structural basis for the recognition of hydroxyproline in HIF-1 alpha by pVHL. *Nature.* **417**, 975-978 (2002).
41. C. Loenarz, J. Mecinovic, R. Chowdhury, L. A. McNeill, E. Flashman, C. J. Schofield, Evidence for a stereoelectronic effect in human oxygen sensing. *Angew Chem Int Ed Engl.* **48**, 1784-1787 (2009).

42. C. J. R. Illingworth, C. Loenarz, C. J. Schofield, C. Domene, Chemical Basis for the Selectivity of the von Hippel Lindau Tumor Suppressor pVHL for Prolyl-Hydroxylated HIF-1 alpha. *Biochemistry.* **49**, 6936-6944 (2010).
43. E. Flashman, E. A. Bagg, R. Chowdhury, J. Mecinovic, C. Loenarz, M. A. McDonough, K. S. Hewitson, C. J. Schofield, Kinetic rationale for selectivity toward N- and C-terminal oxygen-dependent degradation domain substrates mediated by a loop region of hypoxia-inducible factor prolyl hydroxylases. *J Biol Chem.* **283**, 3808-3815 (2008).
44. H. S. Kim, H. L. Kim, K. H. Kim, J. Kim do, S. J. Lee, J. Y. Yoon, H. J. Yoon, H. Y. Lee, S. B. Park, S. J. Kim, J. Y. Lee, S. W. Suh, Crystal structure of Tpa1 from Saccharomyces cerevisiae, a component of the messenger ribonucleoprotein complex. *Nucleic Acids Res.* **38**, 2099-2110 (2010).
45. M. D. Winn, C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin, K. S. Wilson, Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr.* **67**, 235-242 (2011).
46. Z. Otwinowski, W. Minor, Processing of X-ray diffraction data collected in oscillation mode. *Method Enzymol.* **276**, 307-326 (1997).
47. W. Kabsch, Xds. *Acta Crystallogr D Biol Crystallogr.* **66**, 125-132 (2010).
48. W. Kabsch, Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr D Biol Crystallogr.* **66**, 133-144 (2010).

4 Crystal structures of the ribosomal prolyl-hydroxylases OGFOD1 and Tpa1

4.1 OGFOD1 and Tpa1 are members of the ribosomal hydroxylase family

The discovery that 2OG oxygenases catalysed the post-translational modification of ribosomal proteins expanded their roles to include potential regulators of both transcription as well as of translation (*1-5*). The ribosomal oxygenases (ROX) include MYC-induced nuclear antigen 53 (MINA53) and nucleolar protein 66 (NO66), which catalyse histidyl-hydroxylation of ribosomal protein L27A (RPL27A) and ribosomal protein L8 (RPL8), respectively, in humans; ycfD, which catalyses arginyl-hydroxylation of ribosomal protein L16 (RPL16) in *E. coli*; and OGFOD1 and Tpa1, which catalyse prolyl-hydroxylation of RPS23 in humans and yeast, respectively (Figure 4.1) (*1-4*). OGFOD1 and Tpa1 are reported to catalyse C-3 *trans* (2*S*,3*S*) prolyl-hydroxylation of the highly conserved Pro62 and Pro64 of RPS23 in humans and in yeast, respectively, and therefore differ from the PHDs, PPHDs, and CP4Hs, which catalyse C-4 *trans* (2*S*,4*R*) prolyl-hydroxylation. The further observation that prolyl-hydroxylation of RPS23 affects translational accuracy in yeast in a codon dependent manner (*2*) expands the roles of prolyl-hydroxylases in eukaryotes to extend beyond oxygen sensing and collagen stabilisation to potential regulators of ribosomal processivity, translation rate, and translational accuracy (*6, 7*).

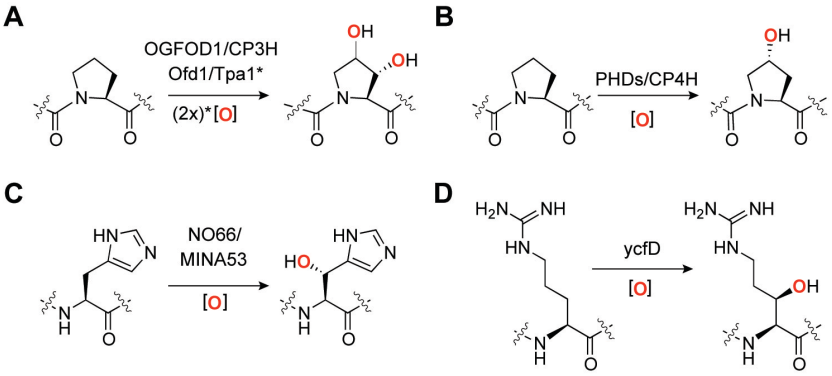


Figure 4.1. Post-translational hydroxylations. (**A**) OGFOD1 catalyses *trans* C-3 prolyl-hydroxylation, where its yeast homologues Tpa1 and Ofd1 catalyse *trans* C-3 as well as C-4 hydroxylations. (**B**) The HIF prolyl-hydroxylases (PHDs) and collagen prolyl-hydroxylases (CP4Hs) catalyse *trans* C-4 prolyl-hydroxylation. (**C**) MYC-induced nuclear antigen 53 (Mina53) and nucleolar protein 66 (NO66) are human ribosomal hydroxylases that catalyse the C-3 histidyl-hydroxylations. (**D**) ycfD is a bacterial ribosomal hydroxylase that catalyses C-3 arginyl-hydroxylation. All hydroxylations are coupled to the oxidation of 2-oxoglutarate to give succinate and carbon dioxide.

Crystal structures of Tpa1 reported in 2010, prior to its assignment as a RPS23 hydroxylase (*8, 9*), revealed that Tpa1 contained two DSBH domains in tandem, only one of which contains the catalytic machinery. There are currently no reported structures for human OGFOD1. Crystal structures of the RPS23 prolyl-hydroxylases would provide insight into the catalytic mechanisms of these enzymes and provide a basis for understanding the structural requirements of prolyl-3- vs prolyl-4-hydroxylation, including important active site differences with the clinically relevant PHDs. The determination of crystal structures of OGFOD1 and Tpa1 in complex with 2OG-mimetic inhibitors is, therefore, of intense interest from both evolutionary and immediate therapeutic perspectives.

4.2 Expression and purification of recombinant OGFOD1 and Tpa1

pET28b plasmid containing the *H. sapiens OGFOD1* gene (residues 1-542) with an *N*-terminal hexahistidine tag was obtained from Dr. Alexander Wolf (Chemistry, University of Oxford). pNIC28-Bsa4 plasmid containing the *Tpa1*

gene with a 20-residue *N*-terminal truncation (residues 21-644) and an *N*-terminal hexahistidine tag was obtained from Dr. Armin Thalhammer (Chemistry, University of Oxford). For heterologous protein expression, the plasmid pET28a_OGFOD1 or pNIC28_Tpa1 was transformed into *E. coli* BL21 (DE3), cells were allowed to grow to OD_{600} ~0.6, and OGFOD1 or Tpa1 overexpression was induced through addition of 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG).

Prior purification strategies for OGFOD1 and Tpa1 relied on a two-step protocol reminiscent of PPHD/EF-Tu purification, consisting of initial Ni-affinity and subsequent size exclusion chromatography. This purification was found to produce protein of sufficient quality for biochemical assays; yet, in an attempt to produce highly pure protein for crystallography, a tertiary chromatographic step was deemed necessary. Anion exchange chromatography using a MonoQ column was found to produce protein of >95% purity as determine by SDS-PAGE (Figure 4.2). Both OGFOD1 and Tpa1 were noted to be very stable in buffer conditions without added NaCl.

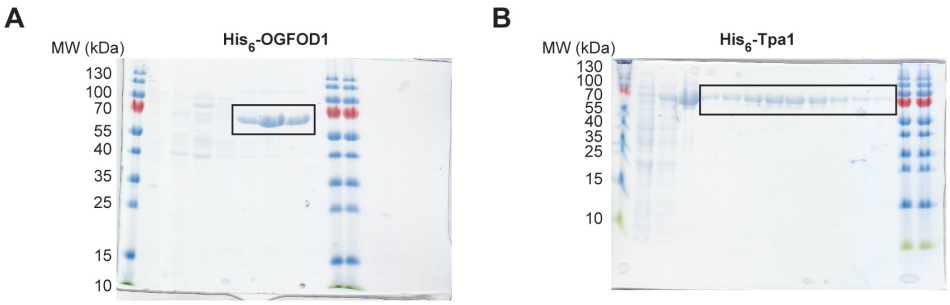


Figure 4.2. Purification of His_6-OGFOD1 and His_6-Tpa1. SDS-PAGE reveals that (**A**) OGFOD1 and (**B**) Tpa1 are >95% pure after anion exchange chromatography using a MonoQ column.

4.3 Kinetic characterisation of OGFOD1

Experiments in this section were performed by Cyrille Thinnes.

4.3.1 Michaelis-Menten kinetics of OGFOD1 and comparison with PHD2

It was of interest to characterise the biochemical properties of OGFOD1 and Tpa1 for use in the development of selective inhibitors and for comparison with human oxygen sensing prolyl-4-hydroxylases such as PHD2. Unfortunately, previous work (Dr. Armin Thalhammer and Dr. Christoph Loenarz) has demonstrated that recombinant Tpa1 is not active against peptide substrates. Therefore, the Michaelis-Menten kinetic parameters of OGFOD1 were then investigated using the 20-residue RPS23 peptide fragment, $RPS23_{51-70}$, and MALDI-TOF mass spectrometry (Figure 4.3).

Compared to PHD2, OGFOD1 was observed to have a higher K_m for oligopeptide substrate fragment [195.3 ± 59.0 μM for $RPS23_{51-70}$, vs 36.7 ± 9.0 μM for 19mer HIF-1α CODD (residues 556-574)](*10*). Studies on intact human ribosomes have shown that RPS23 is constitutively hydroxylated *in vivo* (*2*), suggesting that the higher observed K_m of OGFOD1 for RPS23 compared to that of PHD2 for HIF-1α may have evolved to mimic the relative difference in intracellular concentration of RPS23 (higher) compared to HIF-α (lower).

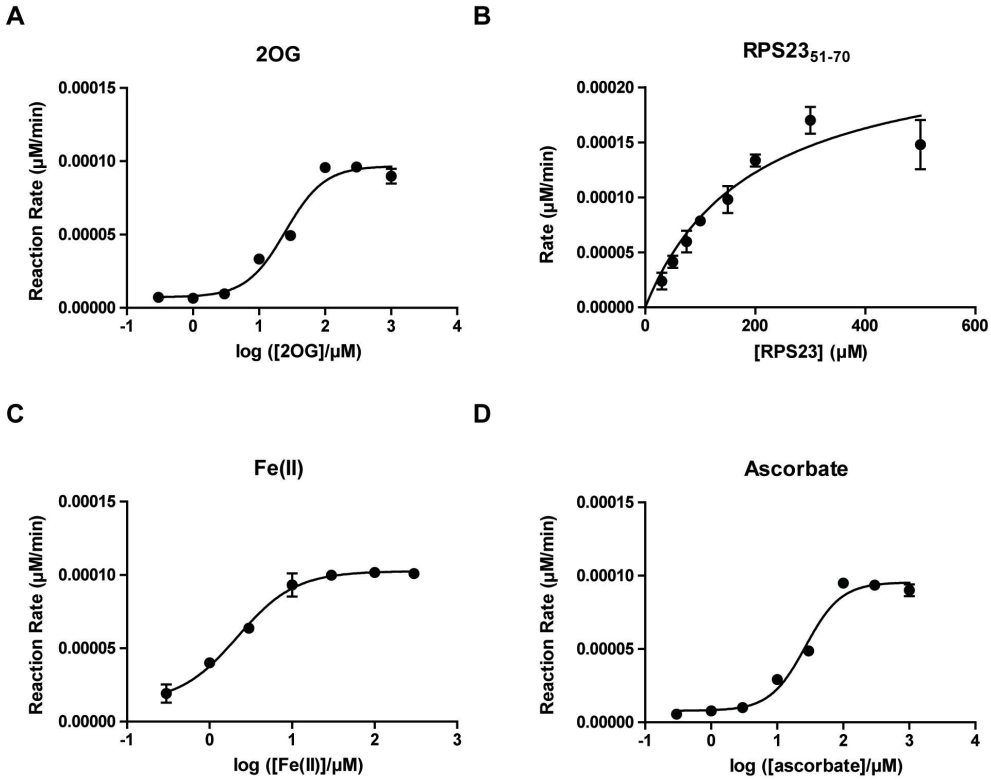


Figure 4.3. Kinetic characterisation of OGFOD1. (**A**) The apparent K_m of 2OG for OGFOD1 is 25.0 ± 1.2 μM. (**B**) The apparent K_m of RPS23$_{51-70}$ peptide substrate for OGFOD1 is 195.3 ± 59.0 μM. (**C**) The apparent K_m of Fe(II) for OGFOD1 is 2.8 ± 1.2 μM. (**D**) The apparent K_m of sodium ascorbate for OGFOD1 is 26.8 ± 1.1 μM. Experiment performed by Cyrille Thinnes.

In contrast, OGFOD1 was observed to have a lower K_m value for 2OG (25.0 ± 1.2 μM for OGFOD1; 55 ± 11 μM for PHD2)(*11*) and ascorbate (26.8 ± 1.1 μM for OGFOD1; 54 ± 10 μM for PHD2)(*12*). The observation that the apparent K_m for 2OG and ascorbate are of the same order of magnitude for each enzyme, yet those for OGFOD1 (~25 μM) are lower than those of PHD2 (~55 μM), suggests that RPS23 hydroxylation is not subject to the same level of intracellular regulation and may reflect the role of OGFOD1 in a constitutive process, rather than a switch-like mechanism that is sensitive many stimuli. Finally, OGFOD1 was observed to have a similar K_m Fe(II) (2.8 ± 1.2 μM for PPHD; <1 μM for PHD2)(*13*).

4.3.2 Initial inhibitor screening against OGFOD1

The biochemical characterisation of OGFOD1 facilitated initial inhibitor screening of 2OG-mimetic inhibitors, which inhibit via competition for the 2OG binding site and chelation of active site iron. The IC_{50} for NOG to OGFOD1 (0.82 ± 0.18 μM) was found to be similar to that observed for NOG to PHD2 (0.8 μM)(*14, 15*), whereas the IC_{50} for 2,4-PDCA and IOX3 to OGFOD1 (0.32 ± 0.05 μM and 0.52 ± 0.08 μM, respectively) were found to be lower than those to PHD2 (6 μM and 1.4 μM, respectively)(*15-17*) (Figure 4.4). These observations are of interest as IOX3 and related lead compounds are currently in phase III clinical trials for anemia (*18*), and suggest that off-target inhibition of OGFOD1 may produce unwanted clinical side effects, some of which may include increased stop-codon readthrough.

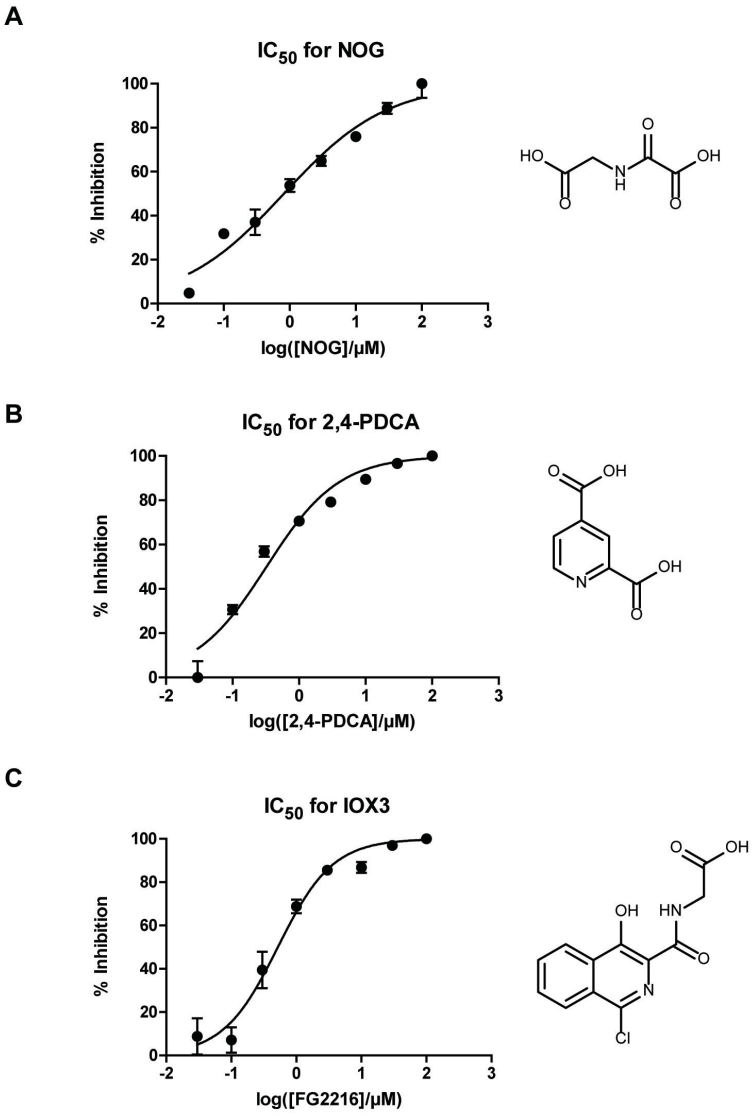


Figure 4.4. OGFOD1 is inhibited by broad spectrum 2OG oxygenase inhibitors. (**A**) OGFOD1 IC_{50} for *N*-oxalylglycine (NOG) is 0.82 ± 0.18 μM. (**B**) OGFOD1 IC_{50} for 2,4-pyridine dicarboxylic acid (2,4-PDCA) is 0.32 ± 0.05 μM. (**C**) OGFOD1 IC_{50} for *N*-[(1-chloro-4-hydroxyisoquinolin-3-yl)carbonyl]glycine (IOX3) is 0.52 ± 0.08 μM. Experiment performed by Cyrille Thinnes.

4.4 Crystallisation of OGFOD1 and Tpa1 in complex with inhibitors

Crystallisation of OGFOD1 (not Tpa1) was performed in collaboration with Dr. Shoichiro Horita.

4.4.1 Initial crystallisation experiments on OGFOD1 and Tpa1

It was of interest to crystallise OGFOD1 and Tpa1 in complex with 2OG mimetic inhibitors in order to both aid in the design of selective inhibitors to affect translational accuracy *in vivo* (and to test those proposed to be selective for PHD2), and to probe the structural mechanisms behind substrate recognition and differentiation of prolyl-3- and prolyl-4-hydroxylases. Whereas Tpa1 structures have been reported in complex with iron and 2OG (*8, 9*), there are currently no reported structures for OGFOD1. A structure of OGFOD1 would thus represent the first known visualisation of a human prolyl-3-hydroxylase.

Crystallisation of OGFOD1 and Tpa1 was approached in the same systematic manner as for PPHD. Protein solution for OGFOD1 crystallisation trials typically contained 11 mg/mL, 0.7 mM $MnCl_2$, 1.0 mM inhibitor (NOG or 2,4-PDCA) or substrate peptide C036 in 10 mM Tris-HCl pH 7.5. Protein solution for Tpa1 crystallisation trials typically contained 6 mg/mL Tpa1, 0.8 mM $MnCl_2$, 1.1 mM inhibitor and/or 1.1 mM substrate peptide in 50 mM Tris-HCl pH 7.5.

Initial broad commercial screening on OGFOD1 and Tpa1 (using the JCSG+, PACT Premier, PEGION, Structure, Natrix, Index, SaltRx, and MIDAS screens)(*19-22*) afforded crystals in numerous different conditions (Table 4.6 and Table 4.7). For crystallisation experiments on OGFOD1 and Tpa1 in the presence of NOG, cyclic RPS23 substrate peptide (C036; *S. cerevisiae* RPS23 residues 54-73), synthesised by Dr. Martin Munzel, was added to protein solution (1.0 mM) in the hope of attaining a substrate complex structure.

Previous work in the Schofield group had demonstrated that OGFOD1 hydroxylates the cyclic peptide *in vitro.* Despite the inclusion of the peptide in the production solution, no density for peptide substrate was observable in any crystal structures. Therefore, no peptide was added to the protein solution for broad screening of OGFOD1 with 2,4-PDCA, or Tpa1 with 2,4-PDCA, IOX3, and BB287 (Figure 4.5).

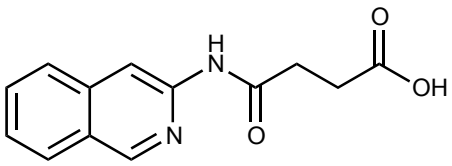


Figure 4.5. Chemical structure of BB287.

Table 4.6. OGFOD1 crystal hits. Red shading indicates a crystal that yielded a complete data set and structure solution. Crystallisation and refinment of OGFOD1 was performed in collaboration with Dr. Shoichiro Horita.

Condition	Observed after	Temp (°C)	Additive	Inhibitor	Best Diffraction (Å)
JCSG D10	2 days	20	-	NOG	
JCSG A5	5 days	20	-	NOG	-
PACT C10	6 days	20	-	NOG	-
PACT B3	13 days	20	-	NOG	-
PACT B4	13 days	20	-	NOG	-
Structure G4	13 days	20	-	NOG	-
PACT C9	13 days	20	-	NOG	-
PACT D3	20 days	20	-	NOG	-
PACT B3/4	**2 months**	**20**	**0.1 M glycine**	**NOG**	**2.1**
PACT B3/4	1 month	20	3% (w/v) sucrose	NOG	-
PACT B3/4	1 month	20	3% (w/v) D-(+)-Trehalose dihydrate	NOG	-
PACT C10	**4 days**	**20**	**-**	**2,4-PDCA**	**2.6**
PACT E5	11 days	20	-	2,4-PDCA	-
PACT B10	1 month	20	-	2,4-PDCA	-
JCSG H11	7 weeks	20	-	2,4-PDCA	-
Index G1	1 month	20	-	2,4-PDCA	-

Table 4.7. Tpa1 crystal hits. Red shading indicates a crystal that yielded a complete data set and structure solution.

Condition	Observed after	Temp (°C)	Additive	Inhibitor	Best Diffraction (Å)
Natrix A10	7 days	20	-	NOG	-
Natrix D8	7 days	20	-	NOG	>10
Natrix E1	7 days	20	-	NOG	-
Natrix F3	7 days	20	-	NOG	-
Natrix F5	7 days	20	-	NOG	-
Natrix F6	7 days	20	-	NOG	-
Natrix F7	7 days	20	-	NOG	-
Natrix F8	7 days	20	-	NOG	-
Natrix F9	7 days	20	-	NOG	-
Natrix F10	7 days	20	-	NOG	-
Natrix G12	7 days	20	-	NOG	-
Natrix H1	7 days	20	-	NOG	-
Natrix H2	7 days	20	-	NOG	>10
Natrix H3	7 days	20	-	NOG	-
Natrix H4	7 days	20	-	NOG	-
Structure H7	7 days	20	-	NOG	-
Index B7	7 days	20	-	NOG	-
Index F9	7 days	20	-	NOG	-
Index H7	7 days	20	-	NOG	-
PACT G9	7 days	20	-	NOG	-
PACT G11	7 days	20	-	NOG	>10
JCSG C4	7 days	20	-	NOG	-
JCSG B8	7 days	20	-	NOG	>10
PEGION D4	7 days	20	-	NOG	-
PEGION E5	7 days	20	-	NOG	-
PEGION E8	7 days	20	-	NOG	-
PEGION F7	**7 days**	**20**	**-**	**NOG**	**2.9**
PEGION G3	7 days	20	-	NOG	-
PEGION G11	7 days	20	-	NOG	-
PACT E4	9 days	20	-	IOX3	-
PACT E11	**9 days**	**20**	**-**	**IOX3**	**1.9**
PACT A5	**7 days**	**20**	**-**	**2,4-PDCA**	**1.9**
JCSG G7	7 days	20	-	2,4-PDCA	-
PACT G12	7 days	20	-	BB287	-
PACT A4	**1 month**	**20**	**-**	**BB287**	**2.0**
PACT A6	1 month	20	-	BB287	2.2
PACT E10	1 month	20	-	BB287	2.2

4.4.2 Attempts at crystal optimisation

Optimisation of Tpa1:NOG hits was carried out by varying the precipitant and salt concentrations and by increasing drop/reservoir size 10-fold (0.1 µL to 1 µL). Unfortunately, increasing drop size generally yielded poorly formed crystals, or no crystals at all. Optimisation of OGFOD1:NOG crystals from the PACT B3 and B4 conditions, identical except for a 1.0 difference in pH [B3 pH 6.0,

B4 pH 7.0; both contain 25% (w/v) PEG 1500] was performed using the Additive screen (Hampton Research)(*23, 24*) on a B3/B4 hybrid condition (pH 6.5). After 2 months, diffraction-quality crystals of OGFOD1:NOG appeared in the presence of 3% (w/v) glycine. Broad crystallisation screening experiments were performed for OGFOD1 in complex with IOX3, yet no diffraction quality crystals were obtained. In general, with the exception of OGFOD1:NOG, the highest quality crystals of OGFOD1 and Tpa1 in complex with inhibitors were obtained directly from broad commercial screens. The optimised crystallisation conditions are shown in Table 4.8.

Table 4.8. Details of final crystallisation conditions.

	OGFOD1:NOG	OGFOD1:2,4-PDCA	Tpa1:NOG	Tpa1:2,4-PDCA	Tpa1:IOX3	TPA1:BB287
Protein solution	~10 mg/mL OGFOD1, 0.7 mM $MnCl_2$, 1.0 mM NOG, 1.0 mM C036 peptide	~10 mg/mL OGFOD1, 2.0 mM $MnCl_2$, 2.0 mM 2,4-PDCA	~9.5 mg/mL TPA1, 0.8 mM $MnCl_2$, 1.1 mM NOG, 1.1 mM C036 peptide	~6 mg/mL TPA1, 0.8 mM $MnCl_2$, 1.1 mM 2,4-PDCA	~6 mg/mL TPA1, 0.8 mM $MnCl_2$, 1.1 mM IOX3	~6 mg/mL TPA1, 0.8 mM $MnCl_2$, 1.1 mM BB287
Reservoir solution	0.1 M MIB buffer pH 6.5, 0.1 M glycine, 25% PEG 1500	0.1 M HEPES pH 7.0, 0.2 M $MgCl_2$, 20% PEG 6000	0.1 M succinic acid pH 7.0, 12% PEG 3350	0.1 M SPG buffer pH 8.0, 25% PEG 1500	0.2 M trisodium citrate, 20% PEG 3350	0.1 M SPG buffer pH 7.0, 25% PEG 1500

Abbreviations: PEG, polyethylene glycol; **SPG buffer (Molecular Dimensions)**, succinic acid, phosphate, glycine; **MIB buffer (Molecular Dimensions)**, malonic acid, imidazole, boric acid. C036 peptide; LEKLGIESKQPNSAIRKAVR(d-Cys).

4.4.3 Data collection and structure solution

Crystallisation experiments resulted in 6 diffraction quality crystals: i) OGFOD1:NOG, ii) OGFOD1:2,4-PDCA, iii) Tpa1:NOG, iv) Tpa1:2,4-PDCA, v) Tpa1:IOX3, vi) Tpa1:BB287. Crystals i-vi were subjected to X-ray radiation and complete data sets were collected to:

i) 2.1 Å resolution in the hexagonal space group $P3_221$
ii) 2.6 Å resolution in the orthorhombic space group $P2_12_12$
iii) 2.9 Å resolution in the monoclinic space group C 1 2 1
iv) 1.9 Å resolution in the monoclinic space group C 1 2 1
v) 1.9 Å resolution in the monoclinic space group C 1 2 1
vi) 2.0 Å resolution in the monoclinic space group C 1 2 1.

The fact that OGFOD1:NOG and OGFOD1:2,4-PDCA crystallised in two different crystal forms with one or four molecules in the asymmetric unit, respectively, was of interest, as it may reflect differences in a final OGFOD1 models. All Tpa1 structures were processed in the monoclinic space group *C* 1 2 1 and contained one molecule in the asymmetric unit. Crystals of Tpa1 in complex with 2,4-PDCA, IOX3, and BB287 seemed to have a diffraction limit around 2.0 Å resolution. All structures were solved by molecular replacement using the MR-PHASER (*25*) subroutine of the PHENIX (*26*) software package using Tpa1 (PDB ID: 3KT4)(*8*) as a search model. In the case of OGFOD1 structures, Tpa1 *C*-terminal non-DSBH associated surface loops were manually removed prior to molecular replacement. Crystallographic data and refinement statistics are shown in Table 4.9.

Table 4.9. Crystallographic data collection and refinement statistics.

	OGFOD1:NOG	OGFOD1:2,4-PDCA	Tpa1:NOG	Tpa1:2,4-PDCA	Tpa1:IOX3	Tpa1:BB287
X-ray source	Diamond Light Source beamline I04	Diamond Light Source beamline I04-1	Diamond Light Source beamline I03	Diamond Light Source beamline I04	Diamond Light Source beamline I04	Diamond Light Source beamline I04
Wavelength (Å)	0.83440	0.97949	1.5418	0.83440	1.2716	0.83440
PDB Acquisition Code	4NHX	4NHY	4NHL	4NHK	4NHM	4NHN
Resolution (Å)	2.10 (2.18–2.10)§	2.60 (2.69–2.60)	2.84 (2.94–2.84)	1.90 (1.97–1.90)	1.90 (1.97–1.90)	2.00 (2.07–2.00)
Space group	$P3_221$	$P2_12_12$	*C* 1 2 1	*C* 1 2 1	*C* 1 2 1	*C* 1 2 1
Unit Cell Dimensions (*a* Å, *b* Å, *c* Å)	64.401 64.401 232.041	108.68 130.473 175.784	168.216 67.251 70.976	168.044 67.716 70.871	169.419 67.564 71.495	166.739 67.392 70.735
Molecules per a.u.	1	4	1	1	1	1
Total Number of Reflections Observed	536556	419736	68118	396484	404304	307714
Number of Unique Reflections	33097 (2981)	76983 (7587)	18332 (1806)	59886 (5937)	61321 (6037)	51127 (4724)
Redundancy	16.2 (6.1)	5.5 (5.5)	3.7 (3.7)	6.6 (5.9)	6.6 (6.3)	6.0 (3.9)
Completeness (%)	99.1 (91.7)	100.0 (100.0)	100.0 (100.0)	99.0 (98.2)	99.4 (98.5)	98.9 (92.0)
I/σ(I)	17.4 (2.5)	12.7 (1.9)	7.4 (2.1)	24.7 (2.5)	26.5 (2.3)	20.9 (1.9)
$^{*}R_{cryst}$	0.1887	0.1854	0.1810	0.1546	0.1449	0.1620
$^{\dagger}R_{free}$	0.2154	0.2278	0.2425	0.1758	0.1704	0.1954
¶RMS deviation	0.006 (1.0°)	0.007 (0.9°)	0.011 (1.3°)	0.010 (1.2°)	0.010 (1.3°)	0.010 (1.2°)
Average *B* factors (Å^2)	55.5	59.9	42.7	44.8	42.5	40.7
Number of Water Molecules	175	83	74	359	535	398

§Parentheses indicate high resolution shell. *R_{cryst} =∑||Fobs| – |Fcalc||/|Fobs|×100. †R_{free}, based on 10% of the total reflections. ¶RMS deviation from ideality for bonds (followed by the value for angles).

4.5 Crystal structure of OGFOD1

4.5.1 Overall structure of OGFOD1

The structure of OGFOD1 in complex with NOG represents the first visualisation of a human prolyl-3-hydroxylase. The overall crystal structure of OGFOD1 consists of nine α-helices, six 3_{10}-helices, and 24 β-strands, and is comprised of two distinct tandem double-stranded β-helix (DSBH) domains (Figure 4.10 and Figure 4.11). Of the two tandem DSBH domains, only the *N*-terminal domain (NTD, 24-238) contains the catalytic machinery. The *C*-terminal domain (CTD, 270-542) is connected to the NTD by a linker region (239-269). The NTD and CTD domains pack against each other via their minor β-sheets and display low sequence and structural similarity to each other (sequence identity 17%; r.m.s.d. of 2.6 Å over 186 Cα atoms). The final models of OGFOD1 consist of residues 24–371, 430–451, and 456–542 for the NOG complex; residues 24–371, 430–542 in chains A, B, and D and 24–371, 430–451 and 455–542 in chain C for the 2,4-PDCA complex. The disordered region (452-455) of OGFOD1:NOG was observed as ordered in OGFOD1:2,4-PDCA in chains A, B, and D. Both crystal forms share two disordered regions: residues 1-23 at the *N*-terminus and residues 371-430 between α9 and β17 in the CTD (for discussion, see section 4.9.3).

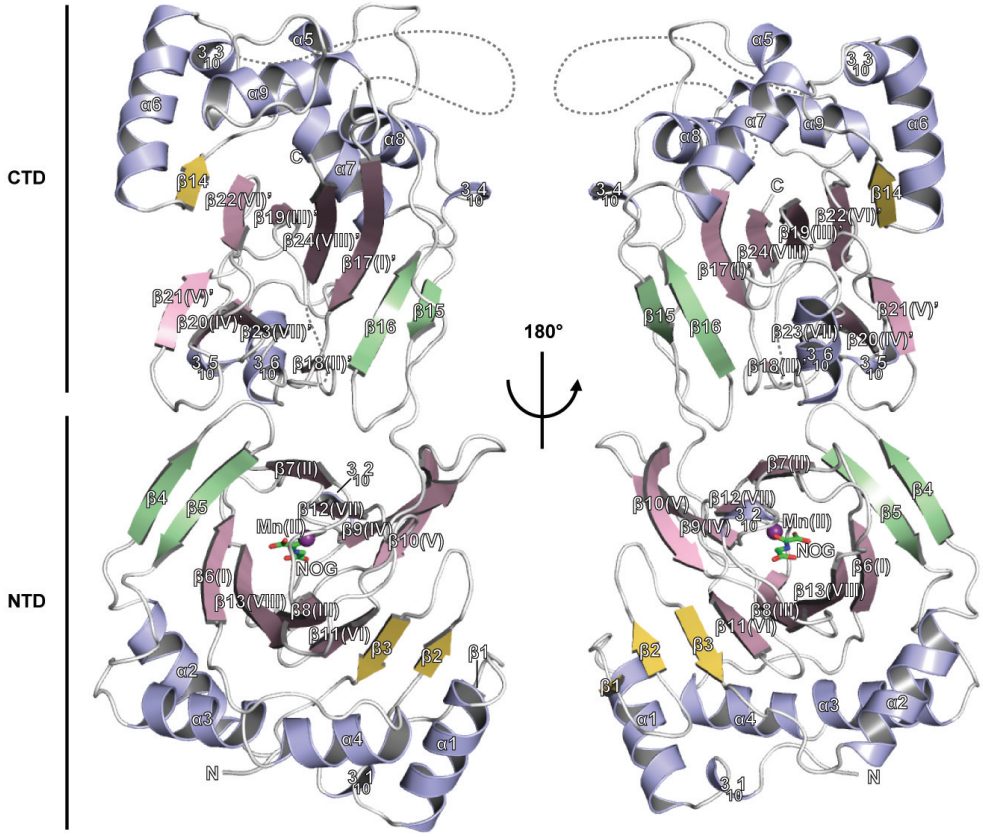


Figure 4.10. Overall view of the crystal structure of OGFOD1. Secondary structure elements are shown as ribbons [α-helix and 3_{10}-helix (blue), β-strand (pink/green/yellow)]. The eight β-strands forming the core double-stranded beta-helix (DSBH) fold are in pink, the β-strands forming the β4-β5 hairpin are in green, and all other strands are in yellow. The DSBH β-strands are additionally labelled with Roman numerals as in Clifton et al. (*27*). The long disordered loop linking α9 and β17 in the CTD is represented by a dashed line.

The NTD contains six helices ($3_{10}1$, $3_{10}2$, α1-α4) and 13 β-strands (β1-β13), 8 of which (I-VIII) comprise the core DSBH fold and form two b-sheets (major and minor) that serve to enclose the 2OG binding pocket and metal-binding site. Four antiparallel β-strands [β7(II)-β12(VII)-β9(IV)-β10(V)] form the minor β-sheet. Nine antiparallel β-strands [β1-β2-β3-β11(VI)-β8(III)-β13(VIII)-β6(I)-β5-β4] form the major β-sheet. A cluster of helices ($3_{10}1$, α1-α4) was observed to pack against the major β-sheet.

The CTD contains a similar number of secondary structure elements as the NTD. The CTD was observed to contain 9 helices ($3_{10}3$-$3_{10}6$, α5-α9) and 11 β-

strands (β14-β24). In order to differentiate the tandem DSBH folds of the NTD and CTD, β-strands that comprise the DSBH in the CTD are noted with by ', e.g. βII'. Seven antiparallel β-strands [β14-β22(βVI')-β19(βIII')-β24(βVIII')-β17(βI')-β16-β15] form the major β-sheet, and four antiparallel β-strands [β18(βII')-β23(βVII')-β20(βIV')-β21(βV')] form the minor β-sheet. As in the NTD, a cluster of helices ($3_{10}3$, $3_{10}4$, a6-a9) was observed to pack against the major β-sheet.

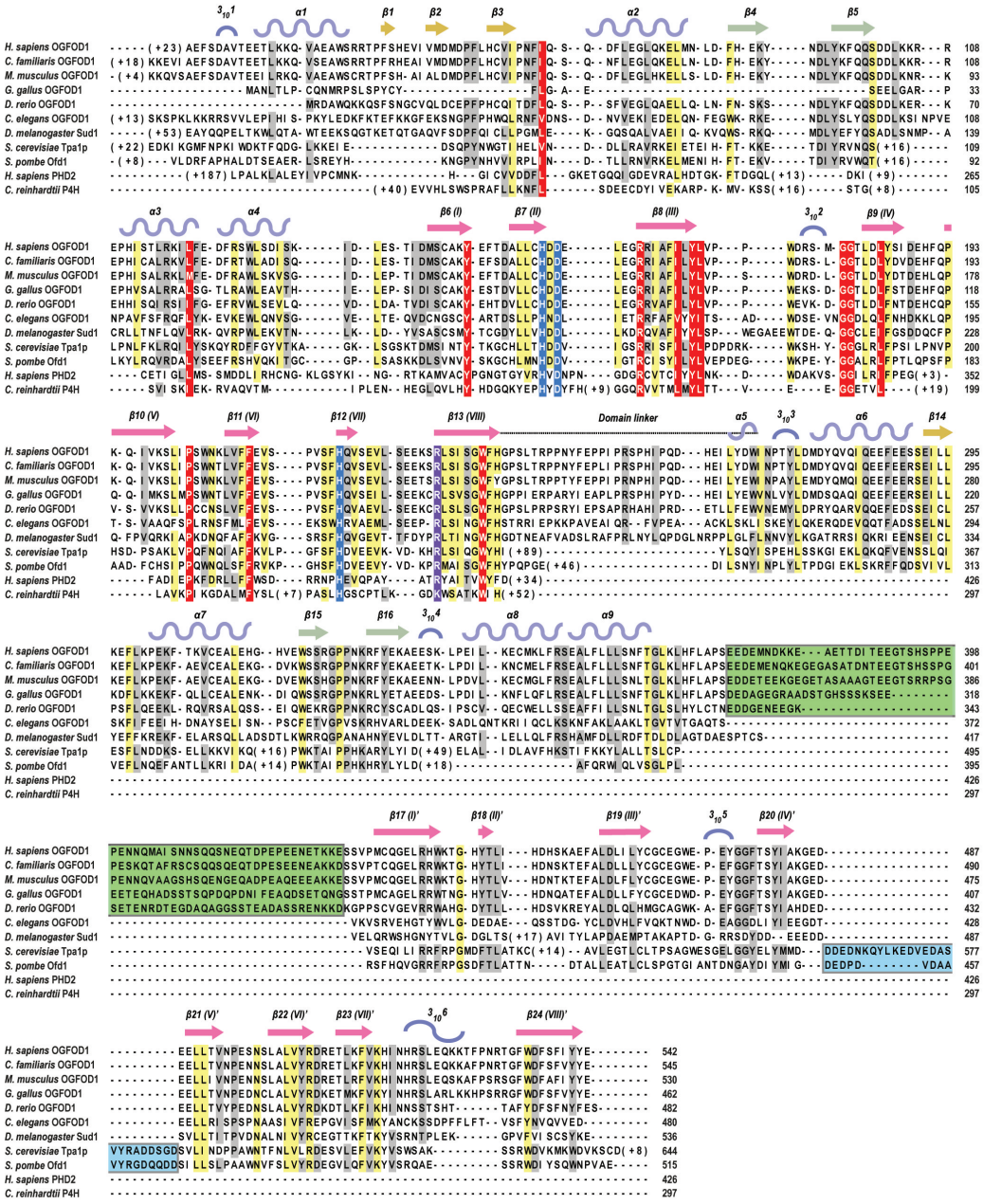


Figure 4.11. Structure-based sequence alignment of OGFOD1 and homologs. Structure-based sequence alignment of *Homo sapiens* OGFOD1 (gi:94536836), *Saccharomyces cerevisiae* Tpa1 (gi:731462), *Homo sapiens* PHD2 (gi:32129514), and *Chlamydomonas reinhardtii* P4H (gi:159478673) (STRAP) (34). ClustalW(35)-generated sequence alignment of OGFOD1 and homologues from higher eukaryotes, *Canis familiaris* (gi:73949826), *Mus musculus* (gi:34850072), *Gallus gallus* (gi:118096214), *Danio rerio* (gi:41054417), *Caenorhabditis elegans* (gi:17531931), *Drosophila melanogaster* (gi:74942745), *Schizosaccharomyces pombe* (gi:2894283). Strictly conserved residues (red), highly conserved residues (yellow), semi-conserved residues (grey), the conserved metal-binding triad (blue) residue that binds the 2OG C-5 carboxylate (purple). Boxed regions represent the disordered 'acidic' loops in OGFOD1 (light green; residues 371-430) and Tpa1 (light blue; residues 561-586).

4.5.2 OGFOD1 active site and inhibitor binding

The OGFOD1 active site is located in the NTD, where an HXD...H metal-binding motif comprising residues His155, Asp157, and His218, is observed to octahedrally coordinate Mn(II) along with bidentate coordination from the bound inhibitor (oxalate group in NOG; pyridine N and C-2 carboxylate in 2,4-PDCA) and a water molecule [note: the non-catalytically active Mn(II) was used as a surrogate for Fe(II) to promote stable complex formation as Mn(II) is less prone to oxidation]. His155 and Asp157 are located on βII(β7) and the loop between strands βII(β7) and βIII(β8). The 'distal' metal-binding histidine in the HxD...H triad, His218, is positioned on strand βVII(β12) (Figure 4.12).

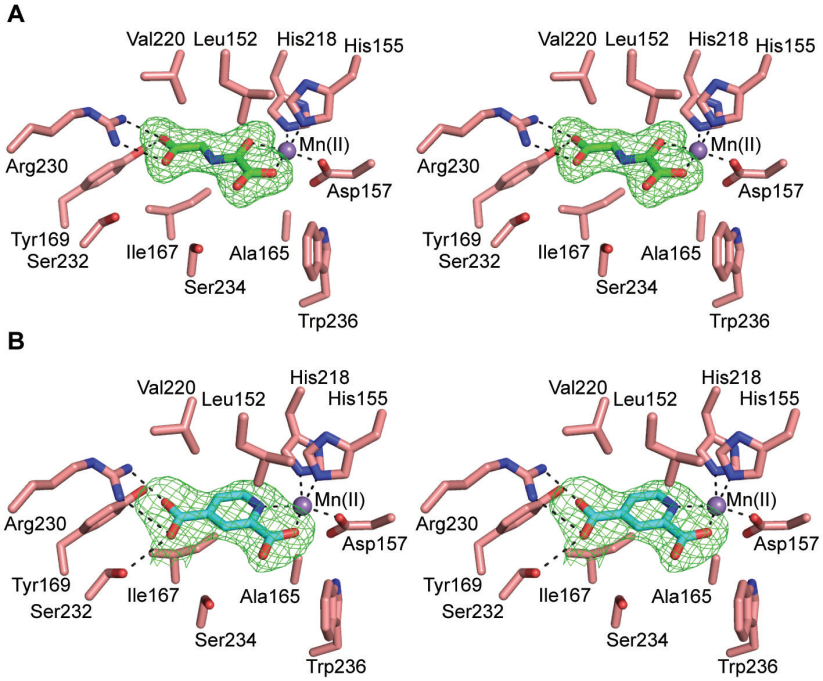


Figure 4.12. The OGFOD1 active site in complex with 2OG oxygenase inhibitors. Wall-eyed stereoviews of the active sites of the (**A**) OGFOD1:Mn:NOG and (**B**) OGFOD1:Mn:2,4-PDCA complexes. The electron density maps OMIT |mF_o - DF_c| (shown in green mesh) are contoured to 3.0 σ.

4.6 Crystal structure of Tpa1

4.6.1 Overall structure of Tpa1

Similar to OGFOD1, the overall structure of Tpa1 consists of tandem DSBH containing domains, only one of which, the NTD, contains the catalytic machinery. The Tpa1 overall fold consists of 12 α-helices, 8 3_{10}-helices, and 29 β-strands. As in the OGFOD1 structures, the catalytic *N*-terminal domain (NTD, 23-246) and the *C*-terminal domain (CTD, 342-644) are connected by a linker region (247-341) and pack tightly against each other via their minor β-sheets. The Tpa1 NTD and CTD domains display low sequence and structural similarity to each other (sequence identity 19%; r.m.s.d. of 2.5 Å over 169 Cα atoms). The final models of Tpa1 consist of residues 23–259, 279-305, 328-560, and 587–635 for the NOG, 2,4-PDCA, and IOX3 complexes; residues 23–95, 100-259, 278-305, 328-560, and 587–635 for the BB287 complex. The disordered region (residues 96-99) of Tpa1:BB287 was observed as ordered in Tpa1:NOG, Tpa1:2,4-PDCA, and Tpa1:IOX3. All crystal structures of Tpa1 in complex with inhibitors share three disordered regions: residues 260-277 (278 in the Tpa1:BB287 structure), residues 306-327, 561-586, and *C*-terminal residues 636-644.

4.6.2 Tpa1 active site and inhibitor binding

The NTD active site HXD…H metal-binding motif comprises residues His159, Asp161, and His227, which octahedrally coordinate Mn(II) along with bidentate coordination from the bound inhibitor (oxalate group in NOG; pyridine N and C-2 carboxylate in 2,4-PDCA; isoquinoline N and C-6 carbonyl in IOX3; isoquinoline N and C-12 carbonyl in BB287) and a water molecule. His159 and Asp161 are positioned on βII(β5) and the loop between strands βII(β5) and

βIII(β6). The distal metal-binding histidine, His227, is positioned on strand βVII(β12) (Figure 4.13).

The binding of IOX3 to Tpa1 was observed to be similar to that for NOG with its glycyl side chain forming a salt bridge with Arg238. In comparison to the NOG and 2,4-PDCA complexes, the side chain of Gln242 was observed in two alternative conformations, likely due to π-stacking with the aromatic system of the bicyclic inhibitor (Figure 4.13*C*). The hydroxyl group of Tyr150 is positioned to form a hydrogen bond to the carbonyl oxygen of the bicyclic ring of IOX3.

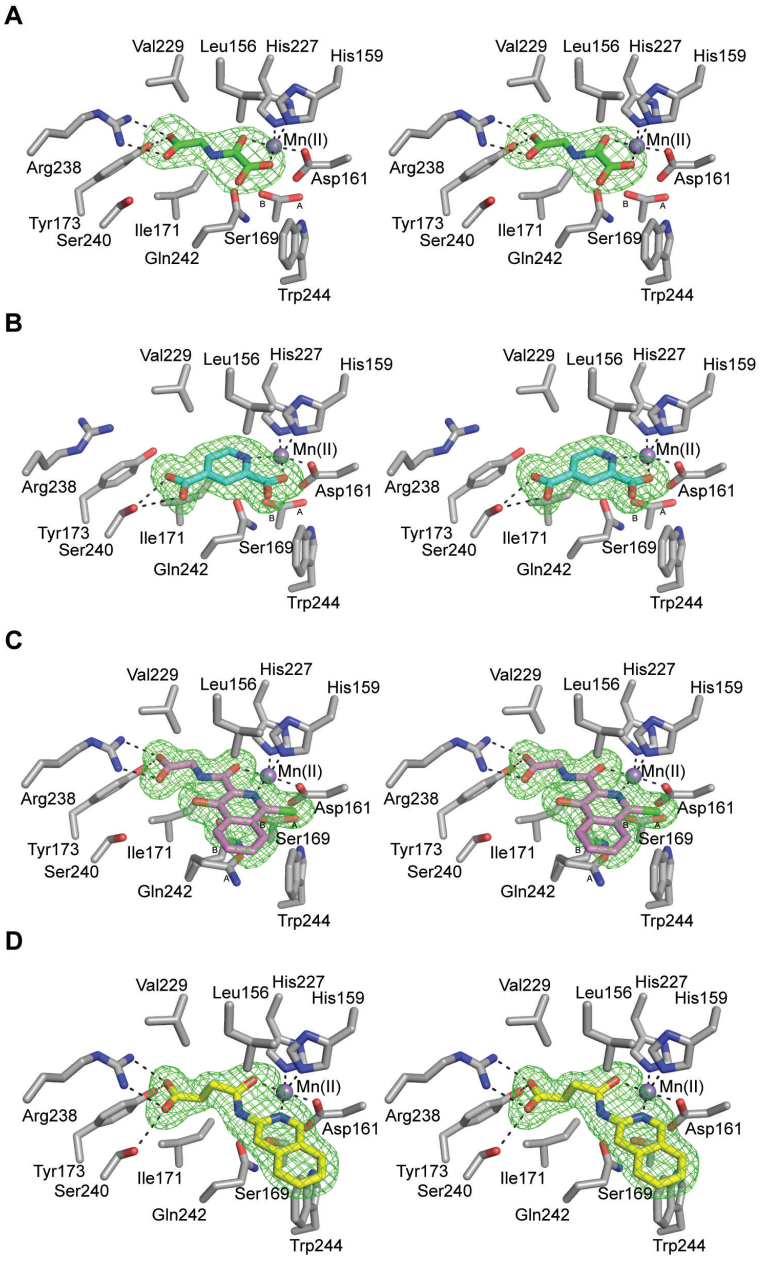


Figure 4.13. The active site of Tpa1 bound to 2OG oxygenase inhibitors. Wall-eyed stereoview of (**A**) Tpa1:Mn:NOG, (**B**) Tpa1:Mn:2,4-PDCA, (**C**) Tpa1:Mn:IOX3, and (**D**) Tpa1:Mn:BB287 active sites. The electron density map OMIT $|mF_o - DF_c|$ is shown in green mesh contoured to 3.0 σ.

4.7 Structural comparison of OGFOD1 and Tpa1

4.7.1 Overall structural comparison

Superimposition of the OGFOD1 and Tpa1 structures (r.m.s.d. 2.2 Å over 387 Cα atoms) reveals analogous positions of both the NTD and CTD core DBSH,

yet differences in non-DBSH associated secondary structure elements. Superposition of the NTDs of OGFOD1 and Tpa1 reveals that the DBSH and metal-binding sites align well (r.m.s.d. 2.0 Å over 200 Cα atoms). Similar to Tpa1, nine antiparallel β-strands [β1-β2-β3-β11(VI)-β8(III)-β13(VIII)-β6(I)-β5-β4] form the major β-sheet. In contrast to Tpa1, OGFOD1 contains two additional β-strands in the NTD (β1 and β2). The NTD-surrounding helices ($3_{10}1$, $3_{10}2$, α1-α4) are also similar to Tpa1 and are positioned relative to the core DSBH.

Whereas superimposition of the OGFOD1 and Tpa1 NTDs reveals structural similarity, superimposition of their CTDs reveals major differences in non-DSBH associated secondary structure (r.m.s.d. value of 2.0 Å over 186 Ca atoms)(Figure 4.14). OGFOD1 contains abbreviated helices and a shorter loop between α7 and the β15-β16 hairpin (residues 313-318, corresponding Tpa1 residues 383-402), no helices or loops between $3_{10}4$ and α8 (residues 332-339, corresponding Tpa1 residues 417-470), and no extension between the βII'(β18) and βIII'(β19) strands in the CTD DSBH (Tpa1 residues 518-531)(Figure 4.14). In contrast, the CTD of OGFOD1 contains an additional 3_{10}-helix [($3_{10}6$) OGFOD1 residues 518-531, corresponding Tpa1 residues 618-620) between β23 and β24. The remaining helices in the OGFOD1 CTD ($3_{10}3$-$3_{10}5$, α6-α9) are structurally conserved in Tpa1. In general, the catalytic NTDs of OGFOD1 and Tpa1 are very similar and the CTDs show more pronounced differences in agreement with more discrepancies in sequence in the CTD, possibly reflecting functional diversity of this domain among homologues.

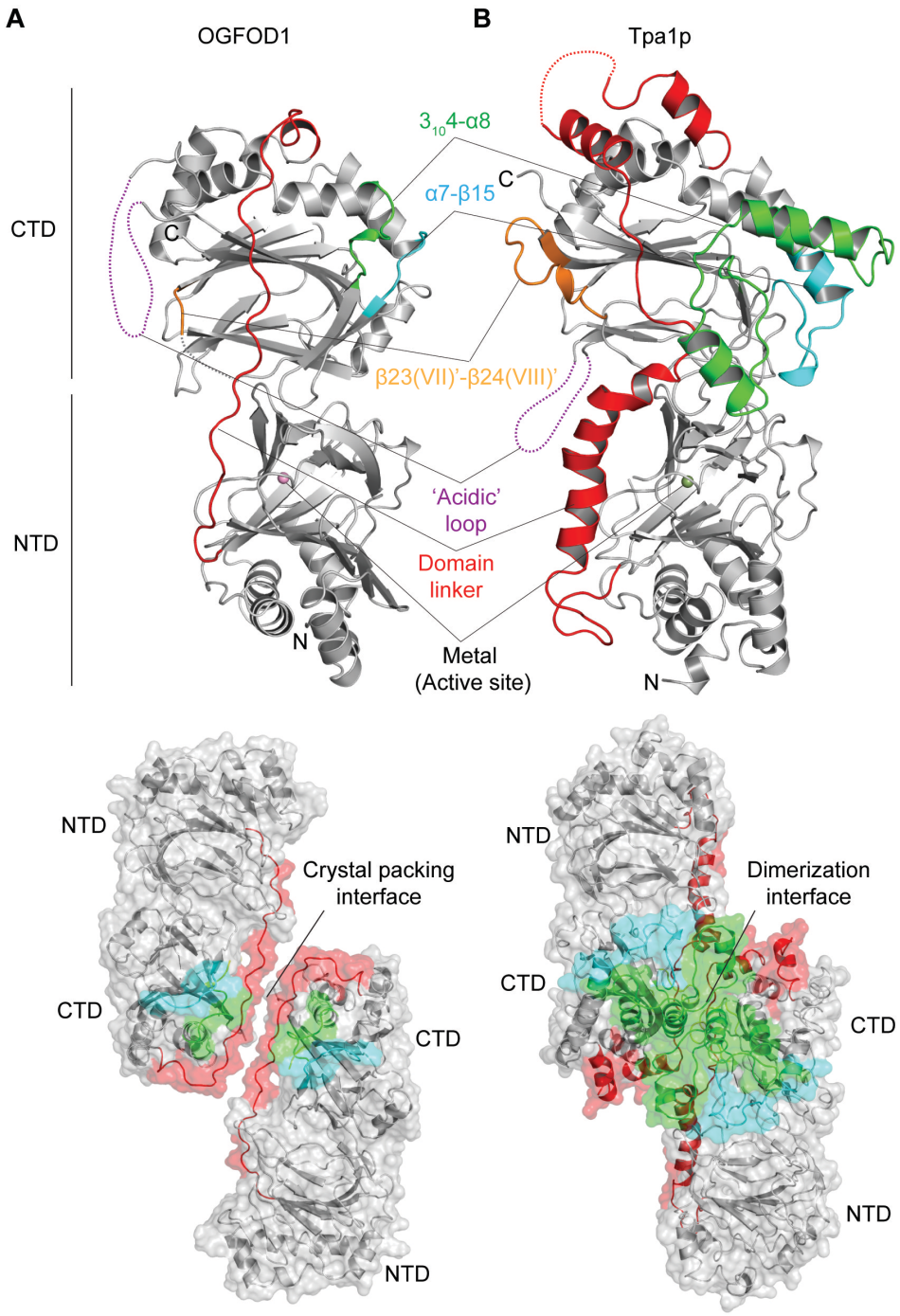


Figure 4.14. Structures of OGFOD1 and Tpa1 reveal differences in their CTD and resulting presence/absence of a CTD:CTD dimerisation interface. Structural comparison of (**A**) OGFOD1 and (**B**) Tpa1 (PDB ID: 3KT4)(*8*) highlighting the linker regions in OGFOD1 (residues: 239-269) and Tpa1 (residues: 247-341) (red), and additional structural elements in Tpa1 at the proposed dimer interface within residue range of 383-402 (OGFOD1 residues 313-318)(cyan), 417-470 (OGFOD1 residues 332-339)(green), and 518-531 (OGFOD1 residues 456-458)(yellow).

Linker regions that serve to connect the NTD and CTD are present in both OGFOD1 and Tpa1, albeit of different lengths and secondary structures. In the case of OGFOD1, the linker comprises 31 residues (239-269), has loop secondary structure, and contains eight proline residues, likely serving to rigidify the connection between the two domains (Figure 4.14). Relative to that of OGFOD1, the Tpa1 linker is substantially longer, totalling 95 residues (247-341) and is rigidified by four α-helices (residues 259-266, 269-277, 294-304 and 332-339) and one 3_{10}-helix (279-282)(Figure 4.14).

4.7.2 Presence/absence of oligomerisation interface in Tpa1 and OGFOD1

In all known Tpa1 crystal structures (*8, 9*), including those solved in this chapter, an oligomerisation interface is observed between the CTD of two molecules (buried surface area ~1900 Å^2) that is independent of crystal packing (Figure 4.14). This observation is further validated by previous studies that demonstrate Tpa1 dimerises in solution (*8, 9*). In contrast, both crystal forms of OGFOD1 (NOG and 2,4-PDCA bound structures) reveal no calculable buried interface between neighbouring CTDs, despite crystal packing arrangements comparable to those observed for Tpa1 structures (PISA analysis).

The complete CTD dimerisation interface in Tpa1 is comprised of structural elements that are absent from OGFOD1; including the elongated NTD to CTD linker (residues 247-341), the region between 3_{10}4 and α8 (residues 417-470), and the region between βII'(β18) and βIII'(β19) (residues 518-531)(Figure 4.14). Importantly, the position of the NTD relative to the CTD is identical in both OGFOD1 and Tpa1 and the buried surface area between the domains is similar (OGFOD1 buried ~700 Å^2, four hydrogen bonds and two salt bridges; TPA1

buried ~1000 Å^2, 17 hydrogen bonds and four salt bridges, excluding the NTD to CTD linker region). Thus, the Tpa1 CTD may have evolved to support its oligomerisation, the precise biological function of which remains unknown. It is of interest that the structural dimerisation elements observed in Tpa1 are not present in the *S. pombe* Ofd1 sequence (Figure 4.11)(*8*), suggesting that Ofd1 is more similar to OGFOD1.

4.7.3 Active site and inhibitor binding

Consistent with overall similarity observed in the OGFOD1 and Tpa1 NTDs, the metal-binding site of OGFOD1 superimposes well on that of Tpa1 and the 2OG oxygenase inhibitors, NOG and 2,4-PDCA, are bound in the 2OG binding site in similar modes between the major and minor β-sheets of the DSBH fold and have bidentate metal coordination (oxalate group in NOG; pyridine N and C-2 carboxylate in 2,4-PDCA) and salt bridges (C-5 carboxylate in NOG; C4 carboxylate in 2,4-PDCA) to $Arg230_{OGFOD1}$/$Arg238_{Tpa1}$.

Interestingly, the binding mode of 2,4-PDCA varies slightly between OGFOD1 and Tpa1 (Figure 4.15). In Tpa1, the 2,4-PDCA C-4 carboxylate is positioned to form a hydrogen bond to $Ser240_{Tpa1}$, yet is not positioned to form a salt bridge to $Arg238_{Tpa1}$ ($Arg230_{OGFOD1}$) as does the NOG C-5 carboxylate. Further, $Arg238_{Tpa1}$ is in a different conformation to that observed in other Tpa1 inhibitor (NOG, IOX3, and BB287) structures, and is positioned too far from the 2,4-PDCA C-4 carboxylate to form a hydrogen bond, although electrostatics are still possible (4.8 Å). Notably, the $Arg238_{Tpa1}$ rotamer (61% allowed) in the TPA1:2,4-PDCA complex is sterically allowed due to a proximal $Pro175_{Tpa1}$; however, in OGFOD1, the equivalent position to $Pro175_{Tpa1}$ is $Val171_{OGFOD1}$, of

which the isopropyl side chain sterically precludes $Arg230_{OGFOD1}$ from adapting a conformation similar to that of $Arg238_{Tpa1}$.

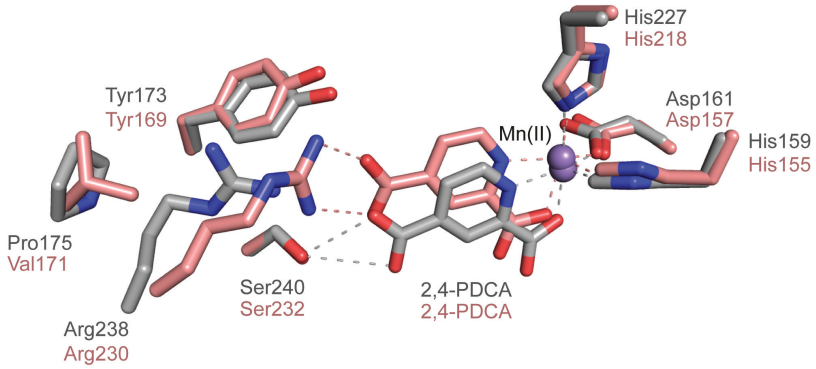


Figure 4.15. OGFOD1 and Tpa1 structures bound to 2,4-PDCA reveal differences in its binding mode. Superimposition of OGFOD1:2,4-PDCA (chain A) (pink) and Tpa1p:2,4-PDCA (silver) active sites revealing different binding mode of 2,4-PDCA and conformation of Arg230/Arg238. OGFOD1 Arg230 is prevented from accessing the conformation of Tpa1 Arg238 due to the side chain of a proximal Val171.

Superimposition of the individual four OGFOD1 molecules in the asymmetric unit of the OGFOD1:2,4-PDCA structure reveal that 2,4-PDCA has some flexibility with regard to its binding to OGFOD1 (Figure 4.16), such that its C-4 carboxylate occupies a somewhat different spacial position than the C-5 carboxylate of NOG, and is positioned to form a hydrogen bonding to the hydroxyl group of $Ser232_{OGFOD1}$ and, in the case of chain D, the hydroxyl group of $Tyr169_{OGFOD1}$. These results are of interest given the low IC_{50} of 2,4-PDCA for OGFOD1 and suggest that the flexibility of 2,4-PDCA more accurately mimics its behaviour in solution and may not adversely affect (raise) its IC_{50} value, an observation that may extend more broadly to inhibitor binding to 2OG oxygenase active sites.

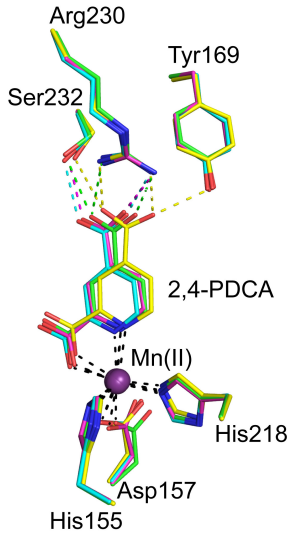


Figure 4.16. The binding mode of 2,4-PDCA to OGFOD1 is flexible. Superimposition of OGFOD1 chains A (green), B (cyan), C (magenta), and D (yellow) in complex with 2,4-PDCA. The active site residues are shown as stick, Mn(II) as sphere (purple).

4.8 Comparison of Tpa1 and PHD2 bound to IOX3

Superimposition of the catalytic sites of Tpa1 and PHD2 bound to IOX3 (*28*) reveals that the bicyclic 2OG-mimetic inhibitor binds in a similar mode to both proteins and shows high conservation in the vicinity of the 2OG C-4 carboxylate (Figure 4.17) (conserved: $Tyr150_{Tpa1}/Tyr303_{PHD2}$, $Ile171_{Tpa1}/Ile327_{PHD2}$, $Tyr173_{Tpa1}/Tyr329_{PHD2}$, $Phe218_{Tpa1}/Phe366_{PHD2}$, $Val229_{Tpa1}/Val376_{PHD2}$, $Arg238_{Tpa1}/Arg383_{PHD2}$, $Trp244_{Tpa1}/Trp389_{PHD2}$; semi-conserved, $Ser169_{Tpa1}/Thr325_{PHD2}$, $Ser240_{Tpa1}/Ala385_{PHD2}$) and differences near the opening of the active site (not conserved: $Val90_{Tpa1}/Asp254_{PHD2}$, $Ser146_{Tpa1}/Met299_{PHD2}$, $Asn148_{Tpa1}/Ala301_{PHD2}$, $Leu156_{Tpa1}/Tyr310_{PHD2}$, $Gln242_{Tpa1}/Thr387_{PHD2}$), likely responsible for differences in primary substrate-recognition. Given the similarity of the OGFOD1 and Tpa1 active sites, it can be inferred that binding of IOX3 to OGFOD1 would likely mimic its binding to Tpa1. Therefore, the striking structural similarity of IOX3 bound to PHD2 and Tpa1 suggests that inhibitors targeting PHD2 currently in phase III clinical trials for

anemia may also target OGFOD1 *in vivo* (*18*). However, it may be possible to generate inhibitors for PHD2 with selectivity over OGFOD1 by elaborating the second heterocycle and capitalising on the lack of conservation at the opening to the active site.

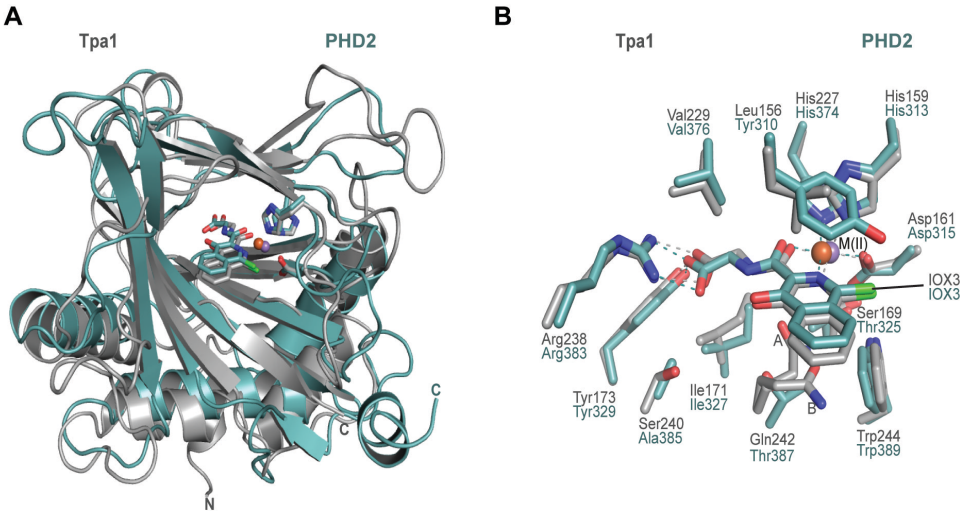


Figure 4.17. Structures of Tpa1 and PHD2 bound to IOX3 reveal similar binding modes. (**A**) Overall and (**B**) active site superimposition of TPA1:IOX3 (grey) and PHD2:IOX3 (turquoise) (PDB ID: 3HQU)(*28*) active sites.

4.9 Putative substrate-binding groove in OGFOD1 may provide insight into RPS23 binding

4.9.1 Identification of a putative substrate-binding groove in OGFOD1

Superimposition of the OGFOD1 NTD and either human PHD2 in complex with HIF-α CODD (*28*) or *C. reinhardtii* prolyl-4-hydroxylase (CrP4H) in complex with collagen-like peptide substrate (r.m.s.d value of 2.9 Å over 184 Ca atoms)(*29*) reveals the presence of an analogous substrate-binding groove in OGFOD1 as observed for PHD2 and CrP4H substrate complexes (Figure 4.18).

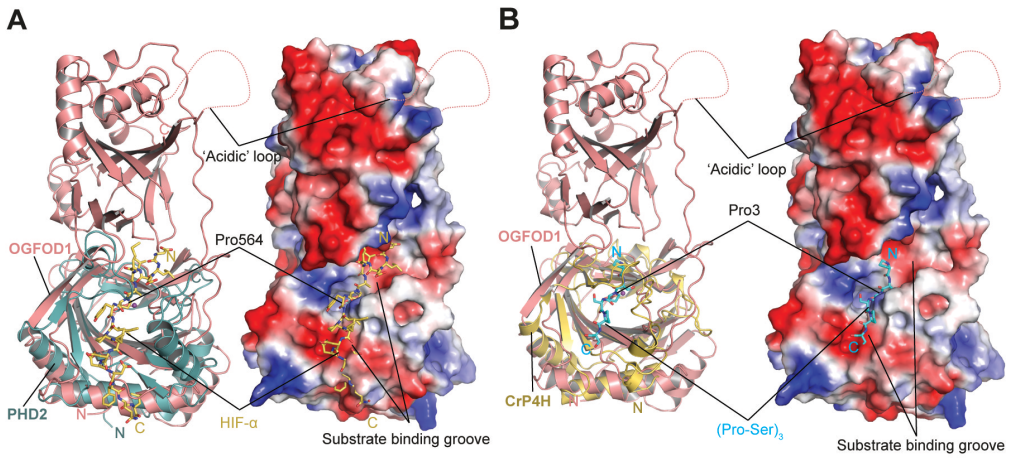


Figure 4.18. The putative OGFOD1 substrate-binding groove. Cartoon and surface representations of the OGFOD1:NOG structure (salmon/surface) superimposed on a (**A**) PHD2 structure (PDB ID: 3HQR)(*28*) (green) in complex with its CODD peptide fragment (yellow; HIF-α residues 556-574) and a (**B**) CrP4H structure (PDB ID: 3GZE)(*29*) in complex with peptide substrate (sequence: PSPSPS) highlighting the putative substrate-binding groove of OGFOD1.

4.9.2 Comparison of the PHD2 β2-β3 finger loop and the OGFOD1/Tpa1 β4-β5 hairpin

The β4-β5 hairpin of OGFOD1/Tpa1 (OGFOD1: 14 residues, 88-101; Tpa1: 14 residues, 80-93) is equivalent to the flexible β2-β3 finger loop of PHD2 and CrP4H (corresponding residues of OGFOD1/Tpa1 β4-β5 hairpins in PHD2: 23 residues, 235-257; in CrP4H: 25 residues, 74-98), which have been shown to secure and position primary substrate in the catalytic site for *trans* C-4 prolyl-hydroxylation (*28, 29*). Structures of PHD2 and CrP4H reveal that the flexible β2-β3 finger loop is closed in its substrate-bound conformation, and open in its *apo* (no primary substrate bound) conformation. In contrast, the β4-β5 hairpin of OGFOD1/Tpa1 is well defined in *apo* structures and its conformation mimics that of closed β2-β3 finger loop in substrate-bound structures of PHD2/CrP4H (Figure 4.19). These observations suggest that the shorter β4-β5 hairpin of OGFOD1/Tpa1 is less flexible in nature than the longer β2-β3 finger loop of

PHD2/CrP4H and likely does not play an analogous role in OGFOD1/Tpa1 substrate-binding (yet is still likely interacts with substrate in some way).

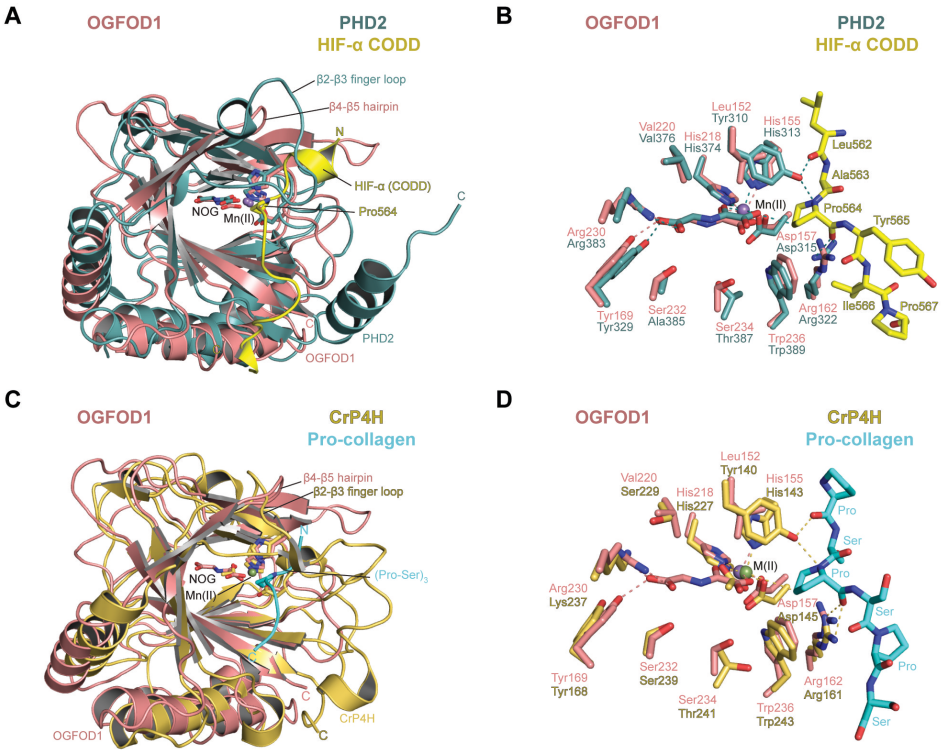


Figure 4.19. Superimpositions of OGFOD1 with prolyl-4-hydroxylases reveal differences in substrate binding residues. (**A**) Overall and (**B**) active site superimposition of OGFOD1 (salmon) and PHD2 (PDB ID: 3HQR)(*28*) (green) in complex with CODD peptide fragment (yellow). The colouring scheme is the same as with other figures. The active site residues are shown as stick, Mn(II) as sphere. (**C**) Overall and (**D**) active site superimposition of OGFOD1:NOG (pink) and *Chlamydomonas reinhardtii* P4H (PDB ID: 3GZE)(*29*) (yellow) in complex with substrate peptide (cyan). The active site residues are shown as stick, Mn(II) as a sphere.

4.9.3 Observation and potential role for an 'acidic loop' in OGFOD1/Tpa1

Instead, a disordered 'acidic' loop between α9 and β17 in the CTD of OGFOD1 (60 residues, 371-430; calculated net charge of -14 at neutral pH) could possibly play an analogous role of a lid in closing over the substrate in the active site (Figure 4.18 and Figure 4.20). Tpa1 also contains an 'acidic' loop (36 residues, 561-586; calculated net charge -8 at neutral pH), albeit originating from a different position in the CTD that that of OGFOD1, and is equivalent to the slightly 'acidic' βIV'(β20)-βV'(β21) loop of OGFOD1 (residues 484-488;

calculated net charge -2 at neutral pH)(Figure 4.20). Although the OGFOD1 and Tpa1 'acidic' loops differ in their length and position, both are predicted to have the length necessary to reach the active site. The 60 residue 'acidic' loop of OGFOD1 could extend 50 Å from the far end of the CTD to the active site in the NTD; the 'acidic' loop of Tpa1 is positioned adjacent to the NTD/CTD interface and thus may still play a role in substrate recognition.

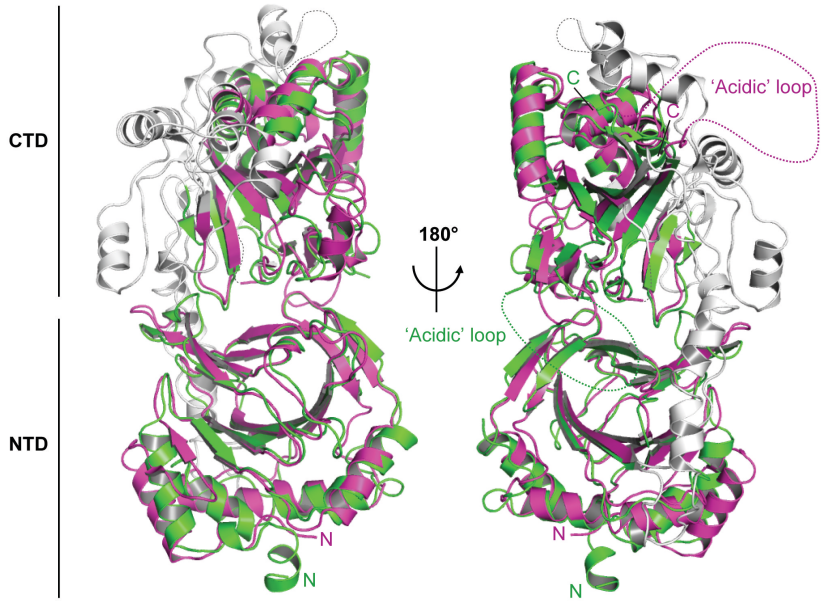


Figure 4.20. Structure of OGFOD1 and Tpa1 reveal the presence of 'acidic' loops in different positions, despite a similarity in their overall tandem DSBH architecture. Superimposition of OGFOD1:NOG (magenta) and TPA1 (PDB ID: 3KT4) (green). Regions not common between OGFOD1 and TPA1 (white). The location of the 'acidic' loops of OGFOD1 (371-430) and TPA1 (561-586) shown in dashed line.

Although the role of the 'acidic' loop in both enzymes remains unknown, it is of interest that OGFOD1 and Tpa1 each have 'acidic' loops, albeit in different regions, and whether their location may play a role in substrate recognition and/or catalysis. Previous work has demonstrated that the CTD of Tpa1 interacts with nucleic acid (*8*). The 'acidic' loop could possibly be used for this interaction via bridging metal ions.

4.9.4 Conservation of residues in the putative substrate-binding groove

Further analysis of the putative OGFOD1 substrate-binding groove reveals conservation of key groove-lining residues with Tpa1 including $Lys91_{OGFOD1}/Lys83_{Tpa1}$, $Asp94_{OGFOD1}/Asp86_{Tpa1}$, $Tyr96_{OGFOD1}/Tyr88_{Tpa1}$, $Gln100_{OGFOD1}/Gln92_{Tpa1}$, $Ser101_{OGFOD1}/Ser93_{Tpa1}$, $Asp140_{OGFOD1}/Asp144_{Tpa1}$, $Ser142_{OGFOD1}/Ser146_{Tpa1}$, $Leu152_{OGFOD1}/Leu156_{Tpa1}$, $Asp156_{OGFOD1}/Asp160_{Tpa1}$, $Arg162_{OGFOD1}/Arg166_{Tpa1}$ and $Trp236_{OGFOD1}/Trp244_{Tpa1}$. The residues $Leu95_{OGFOD1}/Ile87_{Tpa1}$, $Phe98_{OGFOD1}/Val90_{Tpa1}$, $Leu159_{OGFOD1}/Ile163_{Tpa1}$ are semi-conserved; whereas residues $Asp103_{OGFOD1}$, $Leu104_{OGFOD1}$, $Lys105_{OGFOD1}$, $Glu158_{OGFOD1}$, $Ile450_{OGFOD1}$, and $Asp452_{OGFOD1}$ are not conserved. A more global analysis of OGFOD1 conserved residues using ConSurf (*30-33*), which automatically generates a multiple-sequence alignment of OGFOD1 homologues using a structural input and then scores conserved residues according to a graded visual output, reveals a high level of conservation in and around the immediate vicinity of the active site, within the core DSBH of both the NTD and CTD, and in the vicinity of the putative OGFOD1 substrate-binding groove (Figure 4.21).

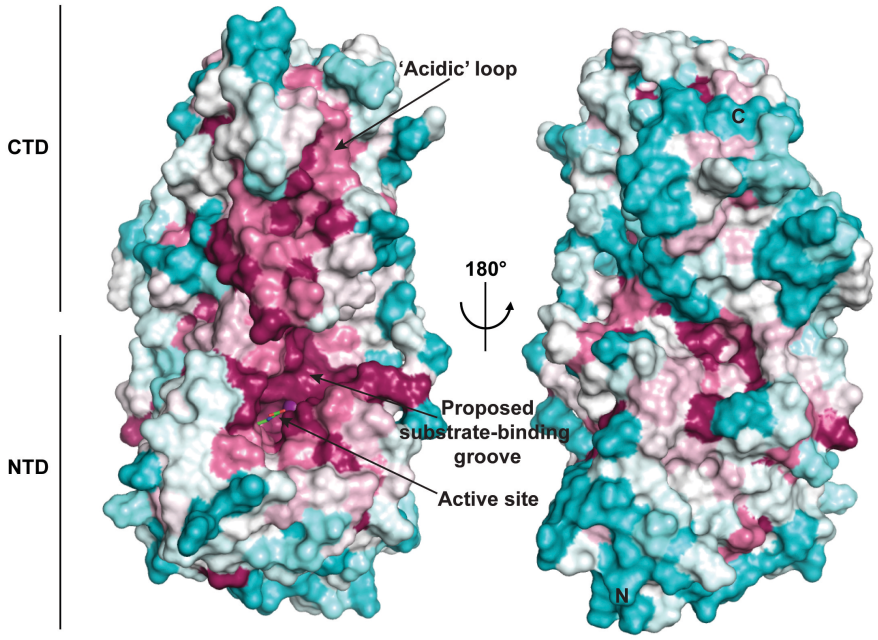


Figure 4.21. Consurf analysis of OGFOD1 reveals conservation of residues in the putative substrate-binding groove. OGFOD1 ConSurf conservation not conserved (cyan) to strictly conserved (magenta). The most conserved region is located around the groove near the catalytic site in the NTD. Mn(II) (purple) and NOG (green) are shown. The active site is located in the middle of a groove formed by OGFOD1 residues Lys91, Asp94, Leu95, Tyr96, Phe98, Gln100, Ser101, Asp103, Leu104, Lys105, Asp140, Ser142, Leu152, Asp156, Glu158, Leu159, Arg162, Trp236, Ile450 and Asp452. Lys91, Asp94, Leu95, Tyr96, Phe98, Gln100 and Ser101 are located on the β4-β5 hairpin region of the NTD; Asp103, Leu104 and Lys105 are located at the loop region between β5 and α3; Asp140 at the loop region between α4 and β6(βI); Ser142 and Leu152 on the βI(β6) and βII(β7) strands, respectively; Asp156, Glu158, Leu159 at the loop region between βII(β7) and βIII(β8); Arg162 and Trp236 on the βIII(β8) and βVIII(β13) strands, respectively; Ile450 and Asp452 at the loop region between βII'(β18) and βIII'(β19).

4.9.5 Concluding insights into an OGFOD1 substrate-binding groove

Overall, the observations that i) OGFOD1 contains a substrate-binding groove analogous to those of PHD2 and CrP4H, ii) the OGFOD1/Tpa1 β4-β5 hairpin likely does not share a conserved role in substrate-binding with the PHD2/CrP4H β2-β3 finger loop, iii) OGFOD1/Tpa1 possess additional secondary structure in their CTDs that may implicate in substrate-recognition, and iv) residues lining the OGFOD1 substrate-binding groove are highly conserved combine to suggest that OGFOD1/Tpa1, CrP4H, and PHD2 target different

protein substrates thus have evolved to possess different substrate-binding groove topologies, yet also that, in addition to the β4-β5 hairpin in the NTD, other regions such as the βII'(β18)-βIII'(β19) loop in the CTD of OGFOD1, which forms part of the putative extended substrate-binding groove leading from the NTD, could be involved in substrate recognition.

4.10 Prolyl-4-hydroxylation vs prolyl-3-hydroxylation

The observation that OGFOD1 contains a putative substrate-binding groove analogous to those of PHD2 and CrP4H suggests that, broadly, prolyl-3- and prolyl-4-hydroxylases may have evolved from a common ancestor, and that divergent evolution of residues lining the groove resulted in differential regioselectivity at the prolyl C-3 and C-4 positions (for more evolutionary discussion, see Chapter 6). As discussed in Chapter 3 in reference to PPHD, all prolyl-hydroxylase substrate complex structures reveal the presence of a conserved arginine ($Arg322_{PHD2}$, $Arg161_{CrP4H}$, $Arg134_{PPHD}$) at the active site opening that is conserved in OGFOD1 and Tpa1 ($Arg162_{OGFOD1}$/$Arg166_{Tpa1}$), which is positioned to form a hydrogen bond to the carbonyl oxygen of the substrate proline. Prolyl-4-hydroxylases also share a conserved tyrosine ($Tyr310_{PHD2}$; $Tyr140_{CrP4H}$; $Tyr121_{PPHD}$)(Figure 4.11) that is positioned to form a hydrogen bond to the amide of the substrate proline, likely further stabilising the substrate proline in the active site and aligning its stereochemistry for C-4 hydroxylation. Instead, prolyl-3-hydroxylases contain a leucine residue ($Leu152_{OGFOD1}$; $Leu156_{Tpa1}$)(Figure 4.11) in the equivalent position, which, combined with their shorter, more rigid β4-β5 hairpin, may facilitate prolyl-3-hydroxylation by altering the position of the proline within the active site.

Possibly important topological differences between prolyl-3- and prolyl-4-hydroxylase active site imply that a more sterically open catalytic pocket may be necessary for correct positioning of the 'linear' peptide backbone for prolyl-3-hydroxylation, whereas prolyl-4-hydroxylation may depend more on individual enzyme-substrate interactions.

4.11 Towards the purification and crystallisation of a Tpa1:RPS23 protein-protein complex

4.11.1 Expression and purification of a Tpa1:RPS23 protein-protein complex

The PPHD:EF-Tu protein-protein complex structure revealed the presence of substantial PPHD:EF-Tu contacts beyond the immediate vicinity of the active site (Chapter 3). A similar structure of OGFOD1/Tpa1 in complex with RPS23 protein would be of intense interest as it would represent the first structure of a ribosomal oxygenase in complex with protein substrate, and possibly reveal the roles, if any, of the OGFOD1/Tpa1 acidic loop and other regions in the CTD in substrate binding. Initial work on OGFOD1 has demonstrated that OGFOD1 and RPS23 form a tight complex (*2*), yet complex purification and crystallisation studies were performed using the more readily crystallisable Tpa1, which had also been shown to form a complex with human RPS23 (Dr. Christoph Loenarz, Chemistry, University of Oxford).

Previous work by Dr. Armin Thalhammer (D.Phil. thesis, University of Oxford) had demonstrated that successful production of soluble RPS23 protein can be achieved using a pGEX vector encoding for an *N*-terminally truncated variant of *H. sapiens* RPS23 (residues 44-143) with an *N*-terminal GST tag, and that expression of RPS23 with a hexahistidine proved insoluble (*3*). It is likely

that the large GST tag (relative to hexahistidine) increases the soluble of RPS23. Production of soluble RPS23-GST in *E. coli* BL21 cells was achieved by growing the cells at 30 °C overnight without addition of IPTG.

Purification of an intact Tpa1:RPS23 protein-protein complex was initially attempted using the same co-lysis method that proved successful for the purification of PPHD:EF-Tu. 10g of *E. coli* BL21 (DE3) cell pellet containing overproduced Tpa1 was mixed with 10g of *E. coli* BL21 (DE3) cell pellet containing overproduced RPS23-GST. To the cell pellets was added $MnCl_2$ to a final concentration of 2 mM in order to inhibit Tpa1 catalysed RPS23 hydroxylation, which was thought to inhibit complex formation in the same manner as for PPHD and EF-Tu. Ni-affinity chromatography was performed on the mixture, but, unfortunately, no RPS23 protein was observed to copurify with Tpa1, suggesting that Tpa1:RPS23 binding is weaker than Tpa1 binding to Ni-affinity resin.

Further attempts at Tpa1:RPS23 complex purification were pursued using glutathione affinity chromatography on isolated RPS23-GST as an initial step, followed by the addition of pure recombinant Tpa1 protein and size exclusion chromatography (Figure 4.22). Purification of isolated RPS23 is challenging given its low stability and tendency to precipitate in buffer at low salt concentrations (<100 mM NaCl); therefore, GST-affinity chromatography was performed using a binding buffer of 50 mM HEPES pH 7.2, 500 mM NaCl and an elution buffer of 50 mM HEPES pH 8.0, 500 mM NaCl, 10 mM reduced glutathione, and 1 mM DTT. Tpa1 (7 mg/mL) pre-equilibrated with 0.7 mM $MnCl_2$ and 1.0 mM NOG was then added to purified RPS23-GST and the mixture

was buffer exchanged to 50 mM HEPES, 250 mM NaCl pH 7.5 and allowed to incubate for 30 mins at 4 °C.

Subsequent size exclusion chromatography (Superdex 200) using 50 mM HEPES pH 7.5 (to promote complex formation) revealed 3 significant peaks and a void volume peak (Figure 4.22). SDS-PAGE analyses revealed that the first peak contained a Tpa1:RPS3-GST protein-protein-complex, the second peak contained isolated Tpa1, and the third peak contained GST, likely due to RPS23-GST degradation. It is of interest that RPS23-GST was only identified when bound to Tpa1; no peak corresponding to isolated RPS23-GST was identified, likely due to the fact that isolated RPS23-GST has been shown to be unstable in low salt concentrations. Finally, it is important to note that this procedure was demonstrated to be time-sensitive; using the same cell pellets, GST-affinity and size exclusion chromatography performed on the same day resulted in the formation of Tpa1:RPS23 complex, whereas purification over two days yielded no complex formation, possibly due to rapid degradation of RPS23-GST.

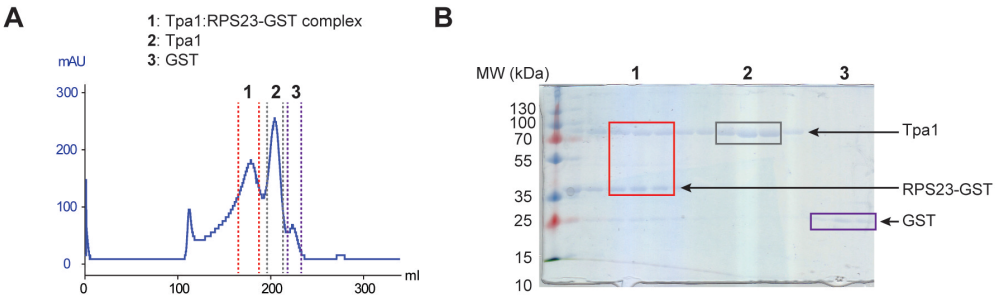


Figure 4.22. Purification of a Tpa1:RPS23-GST protein-protein complex. (**A**) Size exclusion chromatogram revealed 3 peaks (1, 2 and 3). (**B**) SDS-PAGE analysis revealed that peak 1 corresponds to the pure Tpa1:RPS23-GST complex, peak 2 to isolated Tpa1, and peak 3 to GST.

4.11.2 Towards the crystallisation of a Tpa1:RPS23 complex

Crystallisation trials on the Tpa1:RPS23-GST protein-protein complex were performed immediately using broad commercially available screens (JCSG+, PACT Premier, Index). Within one week crystals were observed in similar conditions to those observed for isolated Tpa1 crystallisation trials. Subsequent in-house screening of the crystals revealed that the crystals had identical unit cell constants to isolated Tpa1 crystals and appeared to be the same crystal form, suggesting that RPS23-GST degrades during incubation in the crystallisation drop; a view substantiated by the fact that Tpa1:RPS23 complex purification is time-sensitive and dependent on RPS23 stability. Therefore, Tpa1 binding to RPS23-GST is not sufficient to completely prevent its degradation in solution. Future attempts at purification and crystallisation of a stable Tpa1:RPS23 should involve prior optimisation of recombinant RPS23 production.

4.12 Conclusions

This chapter presents the first crystal structures of a human prolyl-3-hydroxylase, OGFOD1, as well as those of Tpa1 after its assignment as an RPS23 hydroxylase. The structures reveal that OGFOD1 and Tpa1 have similar overall tandem DSBH folds, yet contain more pronounced differences in their CTDs. Other than oligomerisation in Tpa1, the role of the OGFOD1 and Tpa1 CTDs is still unknown, yet the observation that both contain an 'acidic' loop is of interest and suggests that the CTD may be involved in substrate-binding or recognition.

The structures of OGFOD1 and Tpa1 in complex with 2OG mimetic inhibitors, combined with kinetic characterisation and initial inhibitor screening

of OGFOD1, (NOG, 2,4-PDCA, IOX3, and BB287) will be useful in the development of inhibitors selective for OGFOD1/Tpa1 over prolyl-4-hydroxylases, and open the door to targeting prolyl-hydroxylases not only for the treatment of anemia and other ischemia related diseases, but also for nonsense mutation disorders such as Duchenne muscular dystrophy.

4.13 Experimental Procedures

4.13.1 Protein purification

E. coli BL21(DE3) transformed with the pET28b_OGFOD1 or pNIC28_TPA1 plasmid (encoding for protein with an *N*-terminal hexahistidine tag) were grown (37 °C; 180 rpm) to an OD_{600} of 0.6, after which recombinant protein production was induced with 0.5 mM β-D-1-thiogalactopyranoside (IPTG) and the cells were allowed to continue to grow at 18°C overnight, harvested by centrifugation, and stored at -80 °C. Cell pellets, one complete EDTA-free protease inhibitor tablet (Roche), and approximately 1 mg DNAseI (bovine pancreas, grade II, Roche) were resuspended in 50 mM HEPES pH 7.5, 500 mM NaCl, 20 mM imidazole at room temperature with gentle stirring. Cells were lysed on ice by sonication and the lysate was cleared by centrifugation. The supernatant was then loaded onto a 5 mL HisTrap FF column preequilibrated with the resuspended buffer and purified using an AKTA FPLC system (GE Healthcare). The column was washed with 50 mM HEPES pH 7.5, 500 mM NaCl, 50 mM imidazole and the protein was eluted with 50 mM HEPES pH 7.5, 500 mM NaCl, 250 mM imidazole. The purified sample was then exchanged to 25 mM HEPES pH 7.5 and 100 mM NaCl using a PD-10 column (Millipore). Following buffer exchange, 1 mM EDTA was added and the sample incubated for 3 h at 4°C. OGFOD1 or TPA1 were further purified using a Superdex 200 size exclusion column (GE Healthcare) preequilibrated in 25 mM HEPES pH 7.5 and 100 mM NaCl. Subsequently, the proteins were further purified by anion exchange using a 20 mL MonoQ column. Proteins were eluted using a gradient of 25 mM HEPES pH 7.5 and 1 M NaCl. OGFOD1- or TPA1-containing fractions were pooled,

concentrated to 20 mg/mL by diafiltration and buffer exchanged into 10 mM Tris-HCl pH 7.5, then aliquoted, flash frozen in liquid N_2, and stored at -80 °C.

For purification of the RPS23-GST protein used in the formation of a Tpa1:RPS23-GST complex, *E. coli* BL21 transformed with a pGEX vector encoding for an *N*-terminally truncated variant of *H. sapiens* RPS23 (residues 44-143) with an *N*-terminal GST tag were grown at 30 °C overnight without addition of IPTG. Cell pellet, one complete EDTA-free protease inhibitor tablet (Roche), and approximately 1 mg DNAseI (bovine pancreas, grade II, Roche), were resuspended in binding buffer (50 mM HEPES pH 7.2, 500 mM NaCl) and loaded onto a 5 mL GST-Trap HP column (GE Healthcare). Protein elution was achieved using elution buffer (50 mM HEPES pH 8.0, 500 mM NaCl, 10 mM reduced glutathione, 1 mM DTT). To the purified sample was added a solution of pure Tpa1 (6 mg/mL), 0.7 mM $MnCl_2$, and 1.0 mM NOG, and the mixture was exchanged into 50 mM HEPES pH 7.5, 250 mM NaCl using a PD-10 column (Millipore), concentrated by diafiltration, and incubated for 30 mins at 4 °C. The protein complex sample was then further purified by Superdex 200 size exclusion column (GE Healthcare) preequilibrated in 50 mM HEPES pH 7.5, then concentrated by diafiltration, aliquoted, and either used directly for crystallographic experiments or flash frozen in liquid N_2 and stored at -80 °C.

4.13.2 OGFOD1 activity assays and apparent K_m determination

Concentrated OGFOD1 used for biochemical assays was stored at -80 °C in 50 mM HEPES (pH 7.5), 150 mM NaCl, 1 mM DTT and 5% (*w*/*v*) glycerol after HisTrap HP (5 mL) column purification. Standard activity assays were carried out by preparing a reaction mix of 20 μM OGFOD1, 50 μM $(NH4)_2Fe(SO_4)_2$, 200

μM 2OG, 1 mM sodium L-ascorbate, and 100 μM $RPS23_{51-70}$ (H_2N-VLEKVGVEAKQPNSAIRKCV-$CONH_2$). For determinations of apparent K_m of Fe(II), 2OG, and ascorbate, each cofactor concentration was varied and reactions were incubated at 37 °C for 15 minutes. For determination of apparent K_m of $RPS23_{51-70}$ oligopeptide time courses at various concentrations of $RPS23_{51-70}$ were carried out and steady-state reaction rate was determined. All reactions were quenched with an equal volume of 20% (v/v) formic acid. The extent of substrate hydroxylation was analysed by MALDI-TOF-MS (matrix-assisted laser-desorption ionisation–time-of-flight mass spectrometry). Recrystallised CHCA (α-cyano-4-hydroxycinnamic acid) MALDI matrix (1 μL) and quenched assay solution (1 μL) were spotted onto a 96-well MALDI sample plate and analysed using a Waters Micromass™ MALDI micro MX™ mass spectrometer in the positive ion reflectron mode. Raw data was further analysed with MassLynx™ version 4.1. Data were fitted to the Michaelis-Menten equation using GraphPad Prism and kinetic parameters were determined from the fit. All assays were performed in triplicate and plotted as mean ± standard deviation.

4.13.3 Inhibitor IC_{50} determination

For IC_{50} determination, representative assay conditions contained 1 μM OGFOD1, inhibitor (1% *v*/*v* in DMSO), 50 μM Fe(II), 25 μM 2OG, 100 μM sodium ascorbate, and 25 μM $RPS23_{51-70}$ in 50 mM HEPES pH 7.5 and were allowed to incubate at 37 °C for 15 min. All reactions were quenched with an equal volume of 20% (v/v) formic acid. Recrystallised CHCA (α-cyano-4-hydroxycinnamic acid) MALDI matrix (1 μL) and quenched assay solution (1 μL) were spotted onto a 96-well MALDI sample plate and analysed using a Waters Micromass™ MALDI

micro MX™ mass spectrometer in the positive ion reflectron mode. Dose-response studies employed 8-point assays in triplicate and plotted as mean ± standard deviation; data were analysed with GraphPad Prism 5.04.

4.13.4 Protein crystallisation and X-ray crystallography

Crystals of OGFOD1 in complex with NOG and 2,4-PDCA and crystals of TPA1 in complex with NOG, 2,4-PDCA, and IOX3 were grown in sitting drops by the vapour diffusion method (drop size: 200-300 nL) at 293 K in 96-well Intelliplates (Art Robbins). Crystals were cryo-protected in 25% glycerol in well solution and then flash cooled in liquid nitrogen. Data were collected at 100 K using single crystals at Diamond Light Source beamline I04 [OGFOD1:Mn(II):NOG, TPA1:Mn(II):2,4-PDCA, TPA1:Mn(II):IOX3] and Diamond Light Source beamline I04-1 [OGFOD1:Mn(II):2,4-PDCA], with a Pilatus 6M-F detector. Data were collected at 100 K for single crystals of TPA1:Mn(II):NOG using a Rigaku FR-E+ Superbright copper rotating anode diffractometer equipped with Osmic HF optics and a Saturn 944+ CCD detector. Data for all structures were indexed, integrated, and scaled using HKL3000 (*34*). All structures were solved by molecular replacement with MR-PHASER (*25*) subroutine of PHENIX (*26*). Model building and refinement were performed iteratively using COOT (*35*) and PHENIX until *R* and *R*free no longer converged.

4.14 References

1. W. Ge, A. Wolf, T. Feng, C. H. Ho, R. Sekirnik, A. Zayer, N. Granatino, M. E. Cockman, C. Loenarz, N. D. Loik, A. P. Hardy, T. D. Claridge, R. B. Hamed, R. Chowdhury, L. Gong, C. V. Robinson, D. C. Trudgian, M. Jiang, M. M. Mackeen, J. S. McCullagh, Y. Gordiyenko, A. Thalhammer, A. Yamamoto, M. Yang, P. Liu-Yi, Z. Zhang, M. Schmidt-Zachmann, B. M. Kessler, P. J. Ratcliffe, G. M. Preston, M. L. Coleman, C. J. Schofield, Oxygenase-catalyzed ribosome hydroxylation occurs in prokaryotes and humans. *Nat Chem Biol.* **8**, 960-962 (2012).
2. R. S. Singleton, P. Liu-Yi, F. Formenti, W. Ge, R. Sekirnik, R. Fischer, J. Adam, P. J. Pollard, A. Wolf, A. Thalhammer, C. Loenarz, E. Flashman, A. Yamamoto, M. L. Coleman, B. M. Kessler, P. Wappner, C. J. Schofield, P. J. Ratcliffe, M. E. Cockman, OGFOD1 catalyzes prolyl hydroxylation of RPS23 and is involved in translation control and stress granule formation. *Proc Natl Acad Sci U S A.* **111**, 4031-4036 (2014).
3. C. Loenarz, R. Sekirnik, A. Thalhammer, W. Ge, E. Spivakovsky, M. M. Mackeen, M. A. McDonough, M. E. Cockman, B. M. Kessler, P. J. Ratcliffe, A. Wolf, C. J. Schofield, Hydroxylation of the eukaryotic ribosomal decoding center affects translational accuracy. *Proc Natl Acad Sci U S A.* **111**, 4019-4024 (2014).
4. M. J. Katz, J. M. Acevedo, C. Loenarz, D. Galagovsky, P. Liu-Yi, M. Perez-Pepe, A. Thalhammer, R. Sekirnik, W. Ge, M. Melani, M. G. Thomas, S. Simonetta, G. L. Boccaccio, C. J. Schofield, M. E. Cockman, P. J. Ratcliffe, P. Wappner, Sudestada1, a Drosophila ribosomal prolyl-hydroxylase required for mRNA translation, cell homeostasis, and organ growth. *Proc Natl Acad Sci U S A.* **111**, 4025-4030 (2014).
5. R. Chowdhury, R. Sekirnik, N. C. Brissett, T. Krojer, C. H. Ho, S. S. Ng, I. J. Clifton, W. Ge, N. J. Kershaw, G. C. Fox, J. R. Muniz, M. Vollmar, C. Phillips, E. S. Pilka, K. L. Kavanagh, F. von Delft, U. Oppermann, M. A. McDonough, A. J. Doherty, C. J. Schofield, Ribosomal oxygenases are structurally conserved from prokaryotes to humans. *Nature.* **510**, 422-426 (2014).
6. K. M. Keeling, J. Salas-Marco, L. Z. Osherovich, D. M. Bedwell, Tpa1p is part of an mRNP complex that influences translation termination, mRNA deadenylation, and mRNA turnover in Saccharomyces cerevisiae. *Mol Cell Biol.* **26**, 5237-5248 (2006).
7. K. A. Wehner, S. Schutz, P. Sarnow, OGFOD1, a novel modulator of eukaryotic translation initiation factor 2alpha phosphorylation and the cellular response to stress. *Mol Cell Biol.* **30**, 2006-2016 (2010).
8. H. S. Kim, H. L. Kim, K. H. Kim, J. Kim do, S. J. Lee, J. Y. Yoon, H. J. Yoon, H. Y. Lee, S. B. Park, S. J. Kim, J. Y. Lee, S. W. Suh, Crystal structure of Tpa1 from Saccharomyces cerevisiae, a component of the messenger ribonucleoprotein complex. *Nucleic Acids Res.* **38**, 2099-2110 (2010).
9. J. Henri, D. Rispal, E. Bayart, H. van Tilbeurgh, B. Seraphin, M. Graille, Structural and Functional Insights into Saccharomyces cerevisiae Tpa1, a Putative Prolylhydroxylase Influencing Translation Termination and Transcription. *Journal of Biological Chemistry.* **285**, 30767-30778 (2010).
10. E. Flashman, E. A. Bagg, R. Chowdhury, J. Mecinovic, C. Loenarz, M. A. McDonough, K. S. Hewitson, C. J. Schofield, Kinetic rationale for selectivity

toward N- and C-terminal oxygen-dependent degradation domain substrates mediated by a loop region of hypoxia-inducible factor prolyl hydroxylases. *J Biol Chem.* **283**, 3808-3815 (2008).
11. D. Ehrismann, E. Flashman, D. N. Genn, N. Mathioudakis, K. S. Hewitson, P. J. Ratcliffe, C. J. Schofield, Studies on the activity of the hypoxia-inducible-factor hydroxylases using an oxygen consumption assay. *Biochem J.* **401**, 227-234 (2007).
12. E. Flashman, S. L. Davies, K. K. Yeoh, C. J. Schofield, Investigating the dependence of the hypoxia-inducible factor hydroxylases (factor inhibiting HIF and prolyl hydroxylase domain 2) on ascorbate and other reducing agents. *Biochem J.* **427**, 135-142 (2010).
13. L. A. McNeill, E. Flashman, M. R. Buck, K. S. Hewitson, I. J. Clifton, G. Jeschke, T. D. Claridge, D. Ehrismann, N. J. Oldham, C. J. Schofield, Hypoxia-inducible factor prolyl hydroxylase 2 has a high affinity for ferrous iron and 2-oxoglutarate. *Mol Biosyst.* **1**, 321-324 (2005).
14. R. Chowdhury, K. K. Yeoh, Y. M. Tian, L. Hillringhaus, E. A. Bagg, N. R. Rose, I. K. Leung, X. S. Li, E. C. Woon, M. Yang, M. A. McDonough, O. N. King, I. J. Clifton, R. J. Klose, T. D. Claridge, P. J. Ratcliffe, C. J. Schofield, A. Kawamura, The oncometabolite 2-hydroxyglutarate inhibits histone lysine demethylases. *EMBO Rep.* **12**, 463-469 (2011).
15. R. J. Hopkinson, A. Tumber, C. Yapp, R. Chowdhury, W. Aik, K. H. Che, X. S. Li, J. B. L. Kristensen, O. N. F. King, M. C. Chan, K. K. Yeoh, H. Choi, L. J. Walport, C. C. Thinnes, J. T. Bush, C. Lejeune, A. M. Rydzik, N. R. Rose, E. A. Bagg, M. A. McDonough, T. J. Krojer, W. W. Yue, S. S. Ng, L. Olsen, P. E. Brennan, U. Oppermann, S. Muller, R. J. Klose, P. J. Ratcliffe, C. J. Schofield, A. Kawamura, 5-Carboxy-8-hydroxyquinoline is a broad spectrum 2-oxoglutarate oxygenase inhibitor which causes iron translocation. *Chem Sci.* **4**, 3110-3117 (2013).
16. O. N. F. King, X. S. Li, M. Sakurai, A. Kawamura, N. R. Rose, S. S. Ng, A. M. Quinn, G. Rai, B. T. Mott, P. Beswick, R. J. Klose, U. Oppermann, A. Jadhav, T. D. Heightman, D. J. Maloney, C. J. Schofield, A. Simeonov, Quantitative High-Throughput Screening Identifies 8-Hydroxyquinolines as Cell-Active Histone Demethylase Inhibitors. *PLoS One.* **5**, (2010).
17. R. Chowdhury, J. I. Candela-Lena, M. C. Chan, D. J. Greenald, K. K. Yeoh, Y. M. Tian, M. A. McDonough, A. Tumber, N. R. Rose, A. Conejo-Garcia, M. Demetriades, S. Mathavan, A. Kawamura, M. K. Lee, F. van Eeden, C. W. Pugh, P. J. Ratcliffe, C. J. Schofield, Selective small molecule probes for the hypoxia inducible factor (HIF) prolyl hydroxylases. *ACS Chem Biol.* **8**, 1488-1496 (2013).
18. N. R. Rose, M. A. McDonough, O. N. King, A. Kawamura, C. J. Schofield, Inhibition of 2-oxoglutarate dependent oxygenases. *Chem Soc Rev.* **40**, 4364-4397 (2011).
19. J. Newman, D. Egan, T. S. Walter, R. Meged, I. Berry, M. Ben Jelloul, J. L. Sussman, D. I. Stuart, A. Perrakis, Towards rationalization of crystallization screening for small- to medium-sized academic laboratories: the PACT/JCSG+ strategy. *Acta Crystallogr D Biol Crystallogr.* **61**, 1426-1431 (2005).
20. A. McPherson, A comparison of salts for the crystallization of macromolecules. *Protein Sci.* **10**, 418-422 (2001).

21. C. Grimm, A. Chari, K. Reuter, U. Fischer, A crystallization screen based on alternative polymeric precipitants. *Acta Crystallogr D Biol Crystallogr.* **66**, 685-697 (2010).
22. A. D'arcy, A. Mac Sweeney, M. Stihle, A. Haber, The advantages of using a modified microbatch method for rapid screening of protein crystallization conditions. *Acta Crystallogr D.* **59**, 396-399 (2003).
23. A. McPherson, B. Cudney, Searching for silver bullets: An alternative strategy for crystallizing macromolecules. *J Struct Biol.* **156**, 387-406 (2006).
24. R. Sousa, Use of Glycerol, Polyols and Other Protein-Structure Stabilizing Agents in Protein Crystallization. *Acta Crystallogr D.* **51**, 271-277 (1995).
25. A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, R. J. Read, Phaser crystallographic software. *J Appl Crystallogr.* **40**, 658-674 (2007).
26. P. D. Adams, P. V. Afonine, G. Bunkoczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, P. H. Zwart, PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr.* **66**, 213-221 (2010).
27. I. J. Clifton, M. A. McDonough, D. Ehrismann, N. J. Kershaw, N. Granatino, C. J. Schofield, Structural studies on 2-oxoglutarate oxygenases and related double-stranded beta-helix fold proteins. *J Inorg Biochem.* **100**, 644-669 (2006).
28. R. Chowdhury, M. A. McDonough, J. Mecinovic, C. Loenarz, E. Flashman, K. S. Hewitson, C. Domene, C. J. Schofield, Structural basis for binding of hypoxia-inducible factor to the oxygen-sensing prolyl hydroxylases. *Structure.* **17**, 981-989 (2009).
29. M. K. Koski, R. Hieta, M. Hirsila, A. Ronka, J. Myllyharju, R. K. Wierenga, The crystal structure of an algal prolyl 4-hydroxylase complexed with a proline-rich peptide reveals a novel buried tripeptide binding motif. *J Biol Chem.* **284**, 25290-25301 (2009).
30. G. Celniker, G. Nimrod, H. Ashkenazy, F. Glaser, E. Martz, I. Mayrose, T. Pupko, N. Ben-Tal, ConSurf: Using Evolutionary Data to Raise Testable Hypotheses about Protein Function. *Isr J Chem.* **53**, 199-206 (2013).
31. H. Ashkenazy, E. Erez, E. Martz, T. Pupko, N. Ben-Tal, ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Research.* **38**, W529-W533 (2010).
32. M. Landau, I. Mayrose, Y. Rosenberg, F. Glaser, E. Martz, T. Pupko, N. Ben-Tal, ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Research.* **33**, W299-W302 (2005).
33. F. Glaser, T. Pupko, I. Paz, R. E. Bell, D. Bechor-Shental, E. Martz, N. Ben-Tal, ConSurf: Identification of Functional Regions in Proteins by Surface-Mapping of Phylogenetic Information. *Bioinformatics.* **19**, 163-164 (2003).
34. W. Minor, M. Cymborowski, Z. Otwinowski, M. Chruszcz, HKL-3000: the integration of data reduction and structure solution - from diffraction

images to an initial model in minutes. *Acta Crystallogr D.* **62**, 859-866 (2006).
35. P. Emsley, K. Cowtan, Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr.* **60**, 2126-2132 (2004).

5 Biochemical and crystallographic studies on human collagen prolyl-4-hydroxylase

All experiments in this chapter were performed as part of a strictly collaborative project with Dr. Wei Shen Aik, who has described some of the work in his D.Phil. thesis.

5.1 Challenges of performing research on human CP4H

Although it was discovered nearly 50 years ago (*1*), human collagen prolyl-hydroxylase (CP4H) has proven difficult to study for two key reasons. The first lies in its purification. Human collagen prolyl-4-hydroxylase is a 240 kDa tetramer consisting of two α subunits, which contain the catalytic machinery, and two β subunits, which are identical to protein disulfide isomerase (PDI)(*2*). As a result, the CP4H $\alpha_2\beta_2$ likely contains significant conformational heterogeneity. Early techniques for producing mammalian CP4H have involved chick embryos and baculovirus (*3, 4*). It has only been within the last decade that *E. coli* expression techniques became prevalent (*5*). In addition to increased ease of production, expression in *E. coli* results in the lack of glycosylation of the CP4H:PDI tetramer α subunits, which may be favourable for crystallisation. A second important difficulty in working with CP4H is its promiscuous substrate profile (it hydroxylates numerous prolyl residues in various sequence contexts in human collagen).

Therefore, despite nearly 50 years of research, both a crystal structure of CP4H and a reliable assay for inhibition studies on CP4H remain unreported. Such data are of immediate therapeutic importance as some of the PHD inhibitors currently in phase III clinical trials may also inhibit CP4H (*6*).

5.2 Creation of a novel CP4H expression protocol

5.2.1 Strategy for vector design and construction

Synthetic genes encoding for human CP4H (Uniprot ID: P13674, residues 1-517, including a *N*-terminal hexahistidine tag and 5′ NdeI and 3′ KpnI restriction sites) and human PDI (Uniprot ID: P07237, residues 1-491 with 5′ EcoRI and 3′ NotI restriction sites) lacking their respective *N*-terminal transmembrane localisation domains were designed and obtained from GeneArt (Invitrogen)(for DNA fasta sequences, see section 5.10.1). Previous work has demonstrated that the yield of CP4H:PDI tetramer from *E. coli* expression can be increased ~50-fold though gene optimisation (*7*), and so the DNA sequences for both genes were codon optimised for protein production in *E. coli*.

In an effort to deviate from previous approaches that involved coexpression of CP4H and PDI each subcloned in different vectors containing different promoters (*5*) and, likely, different intracellular copy numbers, a strategy was devised in which both CP4H and PDI genes were subcloned into a single pETduet-1 vector (Novagen). In this way, expression of both genes in *E. coli* is controlled by the same T7 polymerase promoter, likely minimising promoter bias and consequent variation in expression levels. Further, the use of a single plasmid simplifies the expression protocol, in that a single addition of IPTG induces expression of both proteins, rather than necessitating separate additions of two different compounds, e.g. arabinose and IPTG. The pETduet-1 vector contains two multiple cloning sites, in which expression of inserts is modulated by T7 polymerase and compatible with the standard pET vector and *E. coli* BL21 (DE3) cell expression system. However, only the first (5′) multiple

cloning site contains a *N*-terminal hexahistidine tag, whereas the second (3′) contains a *C*-terminal S tag.

Plasmid design was approached in light of the successful purification of the PPHD:EF-Tu protein-protein complex in which both proteins possessed an *N*-terminal hexahistidine tag (Chapter 3, section 3.3). Therefore, the PDI gene was subcloned into the first (5′) multiple cloning site in pETduet-1, which contains a *N*-terminal hexahistidine tag, and the CP4H gene was subcloned into the second (3′) multiple cloning site (Figure 5.1). A *N*-terminal hexahistidine tag was added to the CP4H gene during synthesis such that both PDI and CP4H would be expressed with *N*-terminal hexahistidine tags in *E. coli*. Finally, design of the CP4H gene incorporated a stop codon such that the S tag would not be translated.

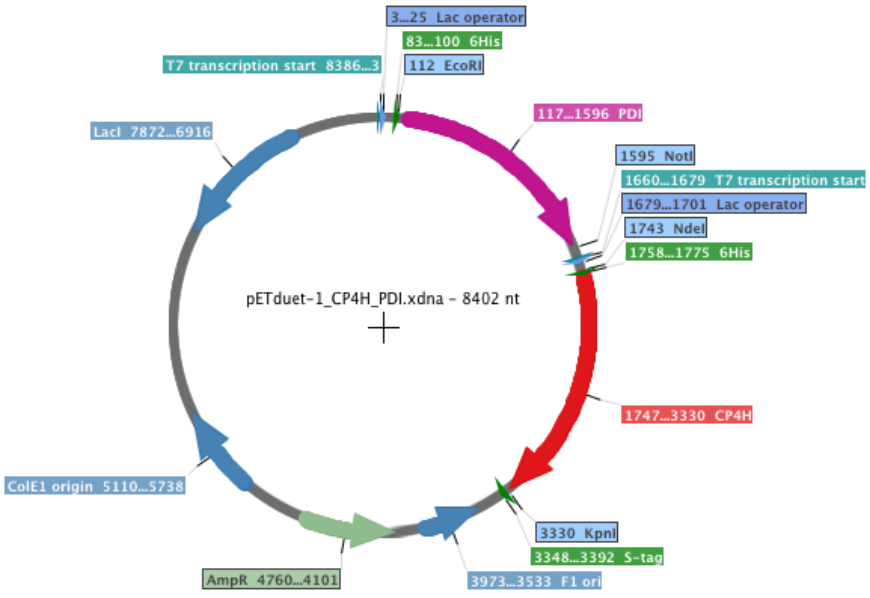


Figure 5.1. Design of the pETduet-1 CP4H:PDI expression vector used in this Chapter. Figure was prepared using the Serial Cloner 2.6 software package.

5.3 Creation of a novel method for CP4H purification

5.3.1 CP4H requires an oxidising environment for effective heterologous expression

Studies have shown that the choice of *E. coli* expression strain can affect the expression levels of CP4H/PDI significantly (*8*). Previous research has shown that attempts to express the CP4H:PDI tetramer in *E. coli* BL21 (DE3) cells resulted in no observed tetramer formation and no prolyl-4-hydroxylase activity (*8*). Mutation studies have shown that CP4H probably contains two intramolecular disulfide bonds (Cys276:Cys293 and Cys486:Cys511) that are critical for successful assembly of active CP4H:PDI tetramer (*9, 10*). Further, treatment of CP4H:PDI tetramer with reducing agents such as dithiothreitol have been shown to result in tetramer disassembly (*3, 11*). Successful production of CP4H:PDI tetramer has been achieved by using the *E. coli* Origami 2 (DE3) strain (*8*), which is a derivative of the BL21 (DE3) strain with substitutions in both the thioredoxin reductase (*trxB*) and glutathione reductase (*gor*) genes, thus creating an oxidising environment in the cytoplasm that greatly enhances disulfide bond formation. In this way, formation of the intramolecular disulfide bonds in CP4H that are critical to CP4H:PDI tetramer assembly is promoted. Therefore, *E. coli* Origami 2 (DE3) was chosen over BL21 (DE3) as the expression strain.

5.3.2 CP4H expression conditions

The pETduet-1 vector containing the synthetic CP4H and PDI genes was transformed into *E. coli* Origami 2 (DE3) for expression. *E. coli* Origami 2 (DE3) cells were grown at 37 °C until OD_{600}=0.6 [requires 12 hours due to the lower doubling time of *E. coli* Origami 2 (DE3) cells], upon which expression of CP4H

and PDI were induced simultaneously with 0.5 mM IPTG and the cells were allowed to continue to grow at 18 °C overnight. This method was found to result in soluble protein production and deemed satisfactory for large-scale purification.

5.3.3 An alternative strategy for CP4H purification

Purifications of CP4H:PDI tetramer typically utilise polyproline chromatography as a first chromatographic step as polyproline has been shown to bind to full-length CP4H (*5, 8*). Interestingly, expression of only the CP4H 2OG oxygenase domain (residues 235-534), and therefore lacking the *N*-terminal dimerisation and PSB domains, does not bind to polyproline resin (*8*), suggesting that polyproline may interact with the peptide binding motif in the PSB domain. In this case, the use of polyproline resin during purification was avoided due to residual polyproline that leaches from the resin being a potent inhibitor of the enzyme, which may lock the CP4H:PDI tetramer in a conformational state not suitable for crystallisation. Therefore, there may be a link between the widespread use of polyproline chromatography for CP4H:PDI tetramer purification and the absence of a crystal structure of the CP4H:PDI tetramer.

Instead, an alternative purification protocol was developed (Figure 5.2). Similar to the purification of the PPHD:EF-Tu complex (Chapter 3, section 3.3), Ni-affinity chromatography was chosen as the first chromatographic step, as both CP4H and PDI subunits contain a *N*-terminal hexahistidine tag. However, in contrast to purification of the PPHD:EF-Tu complex, subsequent size exclusion chromatography revealed a 'mountain range' of peaks, all of which contained pure CP4H and PDI as shown by SDS-PAGE analyses, indicating that CP4H:PDI

exists in multiple different oligomeric states or that CP4H/PDI subunits are also present in their isolated (i.e. not complexed) forms.

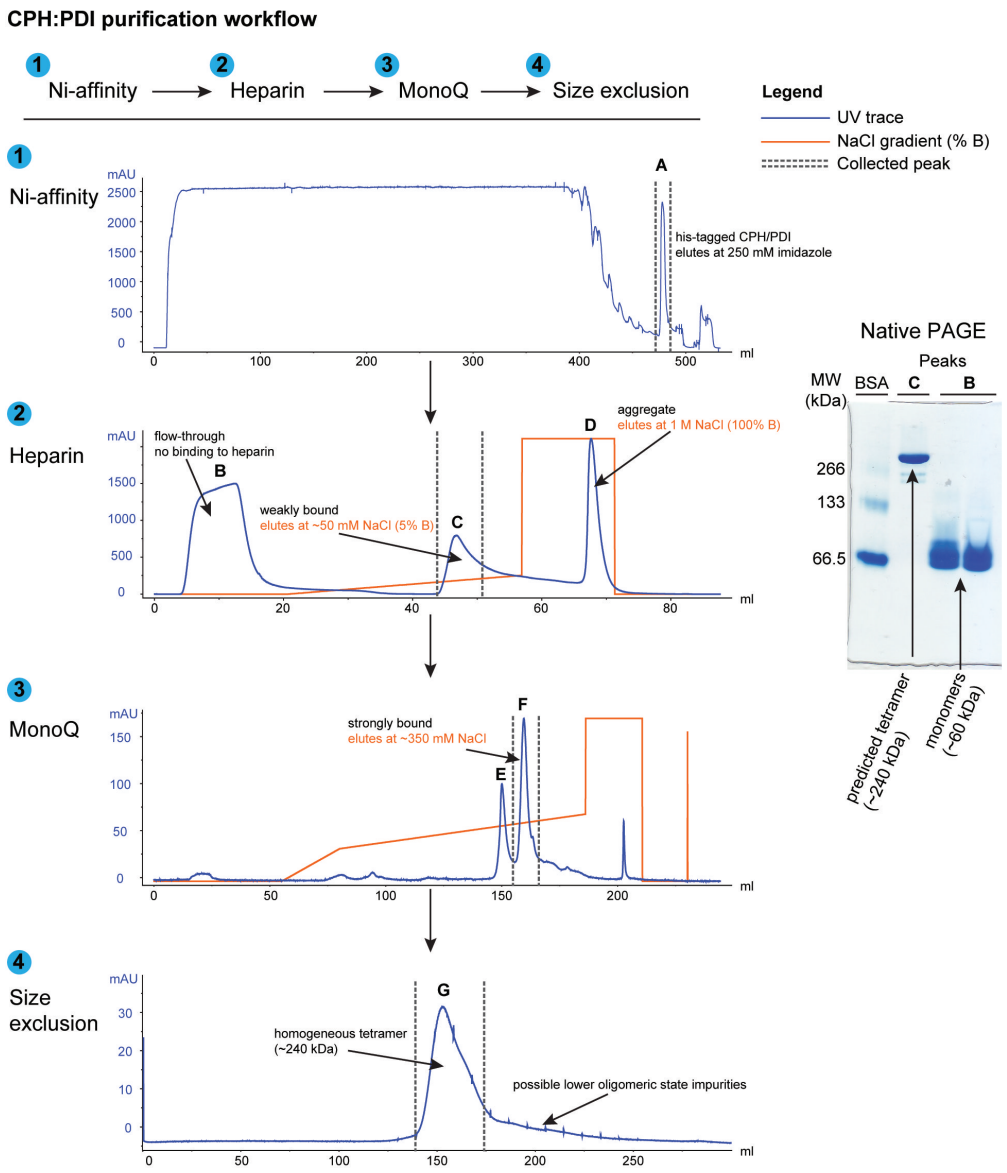


Figure 5.2. Workflow for the purification of an active and crystallisable CP4H:PDI $\alpha_2\beta_2$ tetramer. A four-step chromatographic procedure was pursued in which homogeneous CP4H:PDI tetramer was able to be purified from both bacterial proteins and CP4H:PDI monomer contaminants.

Size exclusion chromatography was thus side-lined, and a heparin cation exchange chromatography procedure was chosen as a second chromatographic step. Heparin is a polysaccharide of the glycosamino family that contains many branched sulfate groups (*12*). Its highly negative charge makes it a suitable resin for cation exchange chromatography. Previous work in the Schofield group has shown that heparin chromatography is critical for the purification of crystallisable nucleic acid binding 2OG oxygenases, such as fat mass and obesity-associated protein (FTO) (*13, 14*), where the negative charge and tubular structure of heparin is thought to mimic that of nucleic acid and bind to FTO in the vicinity of its active site (in a locally basic environment, despite the overall low pI of FTO 5.1). The pI in the local environment of the active site (where nucleic acid and heparin are predicted to bind) is likely higher to due an abundance of basic residues. It was hypothesised that the tubular structure of heparin may mimic that of polyproline and interact with the CP4H:PDI tetramer, possibly by burying itself within a (substrate-binding) groove in the tertiary structure, despite the overall negatively charged CP4H (pI 5.7) and PDI (pI 4.7).

In order to maximise binding of CP4H to the heparin column, which was predicted to be weak due to the negatively charged heparin and low pI of CP4H/PDI, it was important to use a solution of low ionic strength. Therefore, CP4H/PDI containing fractions after Ni-affinity chromatography were buffer exchanged into 25 mM Tris-HCl pH 7.5 that contained no additional salt. Notably, the protein mixture was found to be stable in 25 mM Tris-HCl pH 7.5 at both 4 °C and room temperature and no precipitation was observed. The protein mixture was then loaded onto the heparin column. Some immediate elution (Figure 5.2, peak B), indicative of no binding to the resin, was observed. The eluent that did

not bind to the heparin resin was allowed to fully exit the column by continuous elution with buffer A and the UV absorbance return to baseline, after which a gentle gradient (0-10% B over 10 mL; buffer A corresponds to 25 mM Tris-HCl pH 7.5, buffer B corresponds to 25 mM Tris-HCl pH 7.5, 1 M NaCl) was initiated. A subsequent sharp peak was observed (Figure 5.2, peak C) at approximately 50 mM NaCl (5% B). The column was then washed with 100% B, during which a large, sharp peak eluted (Figure 5.2, peak D). The three peaks were then buffer exchanged back into buffer A, after which precipitant was immediately visible in fractions corresponding to peak D, confirming the instability of peak D in solution and its tendency toward aggregation, whereas peaks B and C remained stable.

Native PAGE was performed on peaks B and C in order to analyse their oligomeric state using BSA as a standard, which oligomerises in a concentration dependent manner and possesses a molecular weight roughly equivalent to those of CP4H and PDI (BSA monomer: ~66.5 kDa). Strikingly, the native PAGE revealed that peak B consisted of CP4H/PDI monomers (they both migrated a comparable distance to the 66.5 kDa BSA monomer), whereas peak C contained a CP4H/PDI mixture of a higher oligomeric state (migrated a comparable distance to ~266 kDa BSA tetramer) (Figure 5.2). Native PAGE can be useful for assigning the relative molecular weights of native proteins, but is unable to assign precise molecular weights to protein complexes because the way in which a given protein migrates through a PAGE gel is determined by its 3-dimensional structure (*15*). Nonetheless, the differences in migration of peaks B and C were clear: peak B appeared to contain two diffuse bands at low molecular weight, hypothesised to be both CP4H and PDI monomers, whereas peak C contained a

single, sharp band at a much higher molecular weight. Therefore, it was proposed that peak C predominately contained the CP4H:PDI $\alpha_2\beta_2$ tetramer, free from contamination with its uncomplexed forms.

Although the purification and study of the CP4H monomer was of interest (especially if it was able to function without PDI), the original goal of the project was to purify and crystallise the CP4H:PDI $\alpha_2\beta_2$ tetramer, and therefore purification was continued on peak C. Mono Q, a strong anion exchange column, was chosen as a third chromatographic step to follow the weak cation exchanger heparin column. To maintain consistency, the Mono Q column was pursued in the same way as the heparin column, with buffer A and buffer B unchanged. Two peaks (Figure 5.2, peaks E and F) were eluted at high salt concentrations (~350 mM NaCl, 35% B), which indicated that the CPI:PDI complex as purified from the heparin column existed in two distinct forms, and, likely, conformations. At this stage in the purification, since the yield was low, the second, and more abundant peak, was chosen for further purification (Figure 5.2, peak F).

After having successfully performed three distinct chromatographic steps (affinity, cation, and anion exchange chromatography), size exclusion chromatography was chosen as a final chromatographic step in order to verify the oligomeric homogeneity of the CP4H:PDI tetramer. A single peak (Figure 5.2, peak G) was eluted from the size exclusion column (S200 resin, exclusion limit of 600,000 Da) at approximately 150 mL. According to experience within the group, smaller proteins on the order of <100 kDa, including PPHD, EF-Tu, OGFOD1, and Tpa1, elute closer or after the 200 mL mark, whereas a 600 kDa protein would elute immediately after the 100 mL mark (100 mL corresponds to the void column of the 300 mL S200 column). The early elution of peak G (only 50 mL

after the void volume) provided further empirical evidence that indeed a CP4H:PDI $\alpha_2\beta_2$ tetramer had been successfully purified.

5.4 Design of a novel CP4H activity assay platform

5.4.1 Challenges for development of a rigorous CP4H activity assay

It was then of interest to determine the *in vitro* activity of the purified CP4H:PDI $\alpha_2\beta_2$ tetramer. One of the difficulties in monitoring the activity of CP4H:PDI $\alpha_2\beta_2$ tetramer *in vitro* stems from the fact that it hydroxylates prolyl-residues at numerous sites in human collagen and so directly monitoring the oxidation of a single proline residue is challenging. Previous kinetic studies on CP4H:PDI $\alpha_2\beta_2$ tetramer have relied on the turnover of 2OG to succinate as an indicator for hydroxylation and have used $(PPG)_n$ motif containing peptides (typically, n=5,10) as substrates (*16, 17*). However, previous studies with PPHD and EF-$Tu_{44\text{-}63}$ peptide substrate revealed that prime substrate uncoupled 2OG turnover may be present to a high degree in prolyl-hydroxylase dependent oxidation of peptides *in vitro*. Such uncoupled turnover would likely result in systematic errors within kinetic parameters derived from monitoring reaction by 2OG turnover. Therefore, it was desirable to development an assay in which substrate oxidation could be directly measured.

5.4.2 Rational design of synthetic CP4H peptide substrates

A 30mer peptide consisting of 10 Pro-Pro-Gly repeats [termed HCOL1, sequence: $(PPG)_{10}$] is commonly used in the literature to probe the activity of CP4H (*16*) and so was chosen as a substrate for biochemical studies. However, the observation that HCOL1 has multiple potential hydroxylation sites and an

absence from natural human collagen, suggested that CP4H kinetic parameters derived using HCOL1 as a substrate may be not be indicative of actual values.

Therefore, in addition to HCOL1, two synthetic peptides were designed to probe the reactivity of recombinant CP4H: i) HCOL3: a 31mer encompassing residues 186-216 of human collagen α-1(I) chain (Uniprot code P02452; sequence: PRGLPGPPGAPGPQGFQGPPGEPGEPGASGP) and possessing 6 potential hydroxylation sites ii) HCOL5: a 30mer encompassing residues 606-635 of human collagen α-1(I) chain (Uniprot code P02452; sequence: AVGPAGKDGEAGAQGPPGPAGPAGERGEQG) and possessing only one hydroxylation site. Peptides were designed as 30mers (or 31mer for HCOL3) for two reasons: i) previous studies using $(PPG)_{5,10}$ have shown that the K_m of the 30mer repeat $[(PPG)_{10}]$ is >20-fold lower than that of the 15mer repeat $[(PPG)_5]$(*16*), and ii) it was hypothesised that a peptide substrate of longer length would be necessary in order to span the distance from the peptide binding motif in the PSB domain to the 2OG oxygenase domain catalytic site. The HCOL3 and HCOL5 peptides were designed to rectify the issues besetting HCOL1, in that both HCOL3 and HCOL5 naturally occur in human collagen and have fewer XPG hydroxylation sites.

The simplicity of HCOL5, which contains only one potential XPG hydroxylation site, was of intense interest as its CP4H dependent prolyl-hydroxylation would likely result in only one +16 Da shift, which could directly be monitored by MALDI MS in the same way as for *in vitro* kinetic assays performed with other prolyl-hydroxylases [PPHD (Chapter 2, section 2.5) and OGFOD1 (Chapter 4, section 4.3)] and their respective peptide substrates. In this way, the use of HCOL5 eliminates the need to monitor CP4H reaction by 2OG

turnover (*16*) and the kinetic parameters derived from CP4H reaction with HCOL5 may potentially be accurately compared with those of other prolyl-hydroxylases. Further, the viability of HCOL5 as a CP4H substrate would open the door to more high-throughput activity assays (such as AlphaScreen®) or binding assays (e.g. using fluorescence polarisation) for the development of selective CP4H inhibitors, which is of immediate utility as PHD inhibitors currently in clinical trials are likely to also inhibit human CP4H.

5.4.3 CP4H hydroxylates prolyl-residues in human collagen

Both the purified CP4H:PDI $\alpha_2\beta_2$ tetramer (derived from heparin peak C) and the CP4H/PDI monomer mixture (derived from heparin peak B) were tested against all three peptide substrates (HCOL1, HCOL3, and HCOL5)(Figure 5.3). Assay conditions were consistent throughout and closely mimicked those for PPHD and PHD2 [50 μM Fe(II), 400 μM 2OG, 1 mM ascorbate, 100 μM substrate in 25 mM Tris-HCl pH 7.5 at room temperature for 20 mins]. Importantly, when tested at similar concentrations, the CP4H/PDI monomer mixture was also found to be active against all three substrates, albeit slightly less so than the CP4H:PDI $\alpha_2\beta_2$ tetramer. As active, non-aggregated CP4H is known to be produced only in the presence of PDI (*2, 18*), this result is of interest as it suggests either that i) PDI may play a role in proper CP4H folding, rather than directly in catalysis or ii) that the monomeric PDI and CP4H can interact in solution under the incubation conditions.

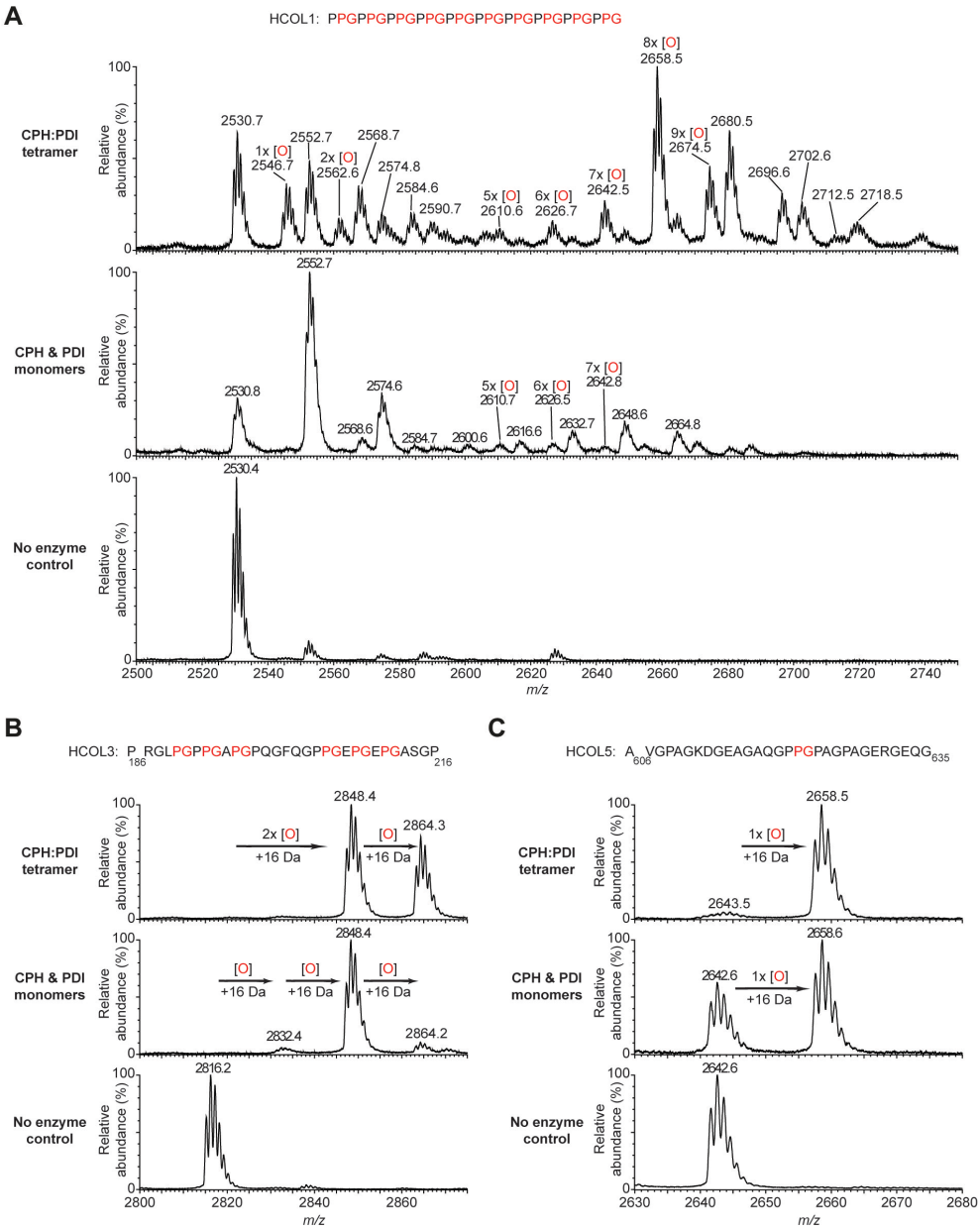


Figure 5.3. CP4H catalyses the prolyl-hydroxylation of synthetic and natural pro-collagen substrates. (**A**) CP4H hydroxylates HCOL1, an unnatural peptide substrate that contains 10 XPG hydroxylation sites. For simplicity peaks without hydroxylation notation correspond to an adduct of a mono- or polyhydroxylated product plus one or two sodium ions. (**B**) CP4H hydroxylates HCOL3, a natural peptide substrate derived from the sequence of human collagen that contains 6 hydroxylation sites. (**C**) CP4H hydroxylates HCOL5, a natural peptide substrate that contains only 1 hydroxylation site. For all peptides, XPG sites are shown in red.

CP4H catalysed hydroxylation of HCOL1 produced, as expected, multiple compounded +16 Da shifts, many of which ionised in the presence of mono- or

disodium adducts (Figure 5.3). Interestingly, the CP4H:PDI $\alpha_2\beta_2$ tetramer was observed to catalyse the hydroxylation of all 10 possible XPG hydroxylation sites (2712.5 Da corresponds to 10 hydroxylations plus one sodium ion). The results represent the first evidence of the direct CP4H-catalysed oxidation of substrate by MS and suggest that the active site of CP4H is flexible enough to accommodate the differences in polarity and, likely, secondary structure, that polyprolyl-hydroxylation would confer on the HCOL1 peptide.

Despite the presence of 6 potential XPG hydroxylation sites in HCOL3, it was found to undergo only a maximum of 3 hydroxylations, over the stated assay conditions. A possible explanation for the difference between HCOL3 and HCOL1 includes the fact that 3 of HCOL3 XPG hydroxylation sites reside towards the *C*-terminus of HCOL3 (sequence: PPGEPGEPG), with two (sequence: EPGEPG) containing a glutamate residue in the X position, which may preclude hydroxylation of all three of those prolyl residues by preventing binding to the CP4H catalytic site. Incubation of HCOL3 with both CP4H:PDI $\alpha_2\beta_2$ tetramer and CP4H/PDI monomers resulted in complete conversion to the dihydroxylated product (Figure 5.3), whereas the third hydroxylation occurred appreciably only in the case of the CP4H:PDI $\alpha_2\beta_2$ tetramer (Figure 5.3). Future studies will help to probe the substrate specificity of human CP4H.

As predicted, HCOL5 was found to undergo a single hydroxylation (Figure 5.3), catalysed to complete conversion by the CP4H:PDI $\alpha_2\beta_2$ tetramer and to a lesser extent by the CP4H/PDI monomers (Figure 5.3). The conversion of HCOL5 to the monohydroxylated product, which can be routinely monitored due to the presence of a single reactant and product species, represents the foundation of a

hitherto non-existent, robust assay platform that should open the door to kinetic characterisation of CP4H as well as inhibitor design.

5.5 Preliminary determination of the Michaelis-Menten kinetic parameters of a CP4H:PDI tetramer

As early as 1966, it had been shown that Fe(II), 2OG, and sodium ascorbate were required for CP4H dependent conversion of prolyl residues to hydroxyprolyl residues (*1, 19, 20*), providing a putative link between CP4H activity and scurvy (*21*). Fe(II) and 2OG were found to be indispensable for CP4H catalysis. On the other hand, reducing agents other than sodium ascorbate, such as β-mercaptoethanol, dihydroxymaleate, were also found to increase CP4H activity, although to a lesser extent than was achieved by sodium ascorbate (*20*).

Early assays for CP4H activity relied on tritium release from 3,4-^{3}H-proline (*22*), which, despite radioactivity and handling concerns, had the advantage that direct conversion of substrate to product could be monitored; many subsequent CP4H activity assays employed the quantification of 2OG turnover (*5, 16*). However, both assays suffer from the fact that the substrates used contained multiple XPG hydroxylation sites, and so direct analysis of kinetic parameters was difficult.

Using HCOL5 as a primary substrate, the apparent K_m values for Fe(II), 2OG, sodium ascorbate and HCOL5 peptide were determined (Figure 5.4). Compared to PHD2, CP4H was found to have a similar K_m for Fe(II) (0.19 ± 0.04 μM for CP4H:PDI tetramer/HCOL5; <1 μM for PHD2/HIF-1α CODD), 2OG (38.2 ± 6.0 μM for CP4H:PDI tetramer/HCOL5; 55 ± 11 μM for PHD2/HIF-1α CODD), and ascorbate (39.1 ± 14.9 μM for CP4H:PDI tetramer/HCOL5; 54 ± 10 μM for

PHD2/HIF-1α CODD) (*23-25*). While the results likely do not reflect cellular activities of CP4H, the similar *in vitro* kinetic parameters of the human collagen and HIF prolyl-4-hydroxylases is of interest from an evolutionary perspective and potentially suggests a common ancestor for the two types of prolyl-hydroxylases, possibly a PPHD-like prokaryotic prolyl-hydroxylase.

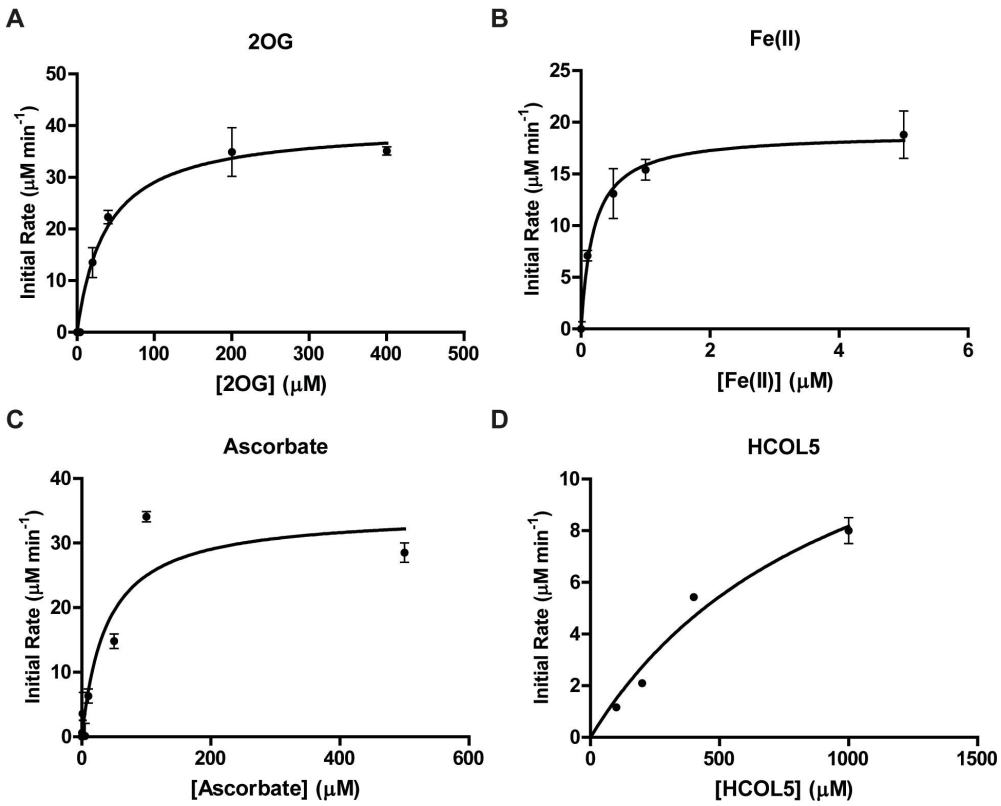


Figure 5.4. CP4H catalysis is dependent of Fe(II), 2OG, and sodium ascorbate. (**A**) 2OG (**B**) Fe(II) (**C**) ascorbate and (**D**) HCOL5.

It is likely that CP4H was not saturated by HCOL5 at a concentation of 1 mM (Figure 5.4*D*). Therefore, the CP4H:PDI tetramer K_m for HCOL5 was determined to be beyond the limits of detection in the assay conditions and is predicted to be >1 mM. Studies have shown that increasing the length of peptide or that of recombinant human collagen substrates can dramatically decrease

their K_m (*16*). However, it is important to note that CP4H hydroxylates XPG motifs in a plethora of different sequence contexts (*26*), and so direct comparison of K_m values for different CP4H substrates, even of the same length, may be difficult. Nonetheless, CP4H was shown to be very active against HCOL5, with substantial activity within manageable time scales (~1-2 hours) even at 1000:1 substrate:enzyme concentration, suggesting that lower enzyme concentrations (and thus a higher assay throughput) are amenable to these assay conditions.

In sum, the basis of a medium-throughput activity assay for human CP4H was developed that is of notable future utility for the screening and generation of CP4H selective inhibitors. Having shown that recombinant CP4H:PDI can be produced in *E. coli* in pure form and is active against a variety of synthetic and natural substrates, it was of interest to return to the initial goal of the project, the crystallisation of the CP4H:PDI $\alpha_2\beta_2$ tetramer.

5.6 Crystallisation and preliminary diffraction of a CP4H:PDI tetramer

5.6.1 Using DSF to screen for optimal CP4H stability

Prior to crystallisation trials with CP4H:PDI, it was of interest to determine the effect (if any) of common cofactors of planned use in crystallisation experiments on the stability of the CP4H:PDI $\alpha_2\beta_2$ tetramer. DSF (T_m shift) was employed to analyse the relative affect of Mn(II), NOG, and peptide substrates (HCOL1, HCOL3, and HCOL5) on the stability of CP4H:PDI tetramer (Figure 5.5). The results revealed that Mn(II) slightly stabilises the tetramer (1.05 ± 0.1 °C shift), yet the further addition peptide substrates did not appreciably affect the stability over that of isolated Mn(II). Mn(II) and NOG were

observed to have a cooperative effect in enhancing CP4H stability [NOG 5.86 ± 0.19 °C shift; Mn(II) and NOG 6.52 ± 0.01 °C shift]. As in the absence of NOG, a peptide-dependent increase in stability was not observed in the presence of both Mn(II) and NOG [Mn(II), NOG, HCOL1 6.72 ± 0.09 °C shift; Mn(II), NOG, HCOL3 6.69 ± 0.15 °C shift; Mn(II), NOG, HCOL5 6.55 ± 0.02 °C shift], suggesting that prior binding of Mn(II) and NOG, which likely serves to order the CP4H active site, is the major contributor to CP4H stabilisation. Although all three peptide substrates tested were observed to stabilise CP4H to the same extent (within error), HCOL5 was chosen as the preferred substrate for crystallisation trials as it contains only one XPG hydroxylation site and so was predicted to bind to CP4H in only one conformation, important for maintaining the conformational homogeneity of the CP4H:PDI tetramer for crystallisation.

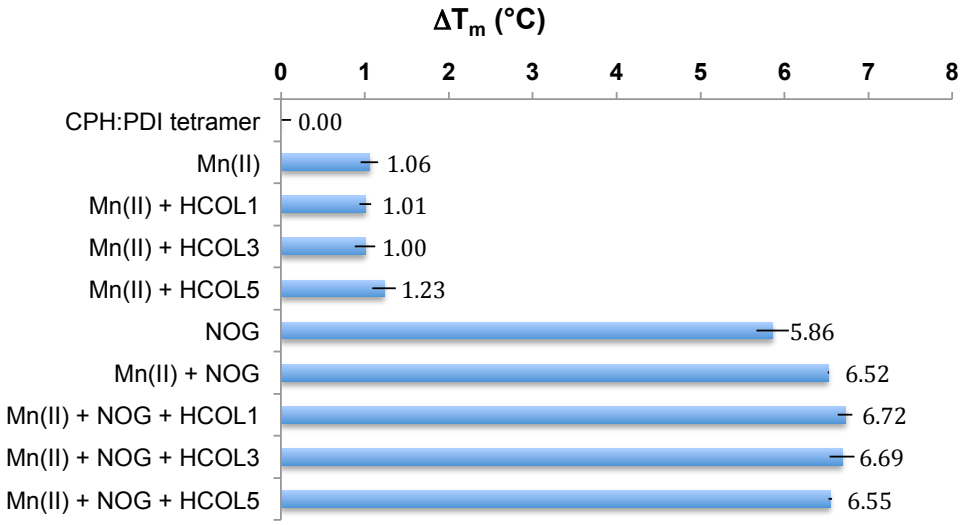


Figure 5.5. Differential scanning fluorimetry on CP4H:PDI tetramer. Results shown are the mean ± s.d. of triplicates with *apo* (no metal bound) CP4H:PDI tetramer as a reference.

5.6.2 Initial crystallisation trials on CP4H

Crystallisation trials on pure, recombinant CP4H:PDI tetramer were then pursued as for other prolyl-hydroxylases, such as PPHD, OGFOD1 and Tpa1 (see

Chapters 3 and 4). The protein solution consisted of 5-7 mg/mL fresh (typically directly off the column, not frozen) CP4H:PDI, 0.5 mM $MnCl_2$, 1.0 mM NOG, and 1.1 mM HCOL5 in 25 mM Tris-HCl pH 7.5. Crystallisation screening experiments on CP4H:PDI began with multiple commercial crystallisation screens at both 20 and 4 °C – i.e. the JCSG+, PACT Premier, Proplex, Index, MIDAS screns (Molecular Dimensions and Hampton Research) (*27-30*).

Several hits were observed in the commercial screens, appearing over a period of two months (for an abbreviated summary of conditions that afforded crystals, see Table 5.6). After 6 days of incubation at 20 °C, the first crystal hits for CP4H:PDI were observed in JCSG well A3 [0.2 M di-ammonium hydrogen citrate, 20% (w/v) PEG 3350]. Shortly thereafter (9 days total incubation), JCSG well A2 [0.1 M sodium citrate pH 5.5, 20% (w/v) PEG 3000] and JCSG well B9 [0.1 M sodium citrate pH 5.0, 20% (w/v) PEG 6000] also developed crystals. Further, after 7 weeks, crystals were observed in JCSG well H7 [0.2 M $(NH_4)_2SO_4$, 0.1 M bis-tris pH 5.5, 25% (w/v) PEG 3350]. In general, conditions in the JCSG broad screen that produced crystals tended to contain citrate as a buffer/salt (exception of well H7), a <7 pH, and a middle molecular weight PEG as a precipitant. The Proplex screen afforded one hits in the well C4 [0.2 $LiSO_4$, 0.1 M MES pH 6.0, 20% (w/v) PEG 4000], which appeared after 3 weeks. In all conditions, crystals were of a hexagonal morphology, with the largest crystals obtained in JCSG well B9 and Proplex well C4. JCSG H7 and Proplex C4 differ from other conditions both in the timescale required for crystal growth (weeks as opposed to days) and in their components (they do not contain citrate, yet they both contain 0.2 M sulfate anion).

Table 5.6. Initial CP4H crystal hits derived from broad screening.

Condition	Observed after	Temp (°C)	Additive	Size (μm)	Morphology	Best Diffraction (Å)
JCSG A2	9 days	20	-	75	Hexagonal	7.0
JCSG A3	6 days	20	-	75	Hexagonal	-
JCSG B9	9 days	20	-	150	Hexagonal	4.0
JCSG B9	4 days	20	2% (w/v) benzamidine	50	Hexagonal	-
JCSG H7	7 weeks	20	-	150	Hexagonal	-
Proplex C4	3 weeks	20	-	150	Hexagonal	4.0
JCSG B9	5 days	20	Silver bullet G1*	175	Hexagonal	3.8

5.6.3 Attempts at optimisation of CP4H crystal hits

Optimisation of hits was immediately undertaken to investigate the following parameters: buffer pH, salt concentration, PEG concentration, protein:reservoir ratio within the drop, CP4H:PDI concentration within the protein solution, addition of additives and type of cryo-protectant solution. Hits in JCSG A2 and B9 were manually optimised by varying the buffer pH and PEG concentrations [0.1 M sodium citrate pH 5.0/5.5/5.6 and PEG 3000/6000 at 12-22% (w/v) in steps of 2%] as well as the protein:reservoir solution ratio (1:1 and 1:2). For manual trials, the total drop size was scaled up 10-fold (0.1 μL to 1 μL). An additional optimisation of JCSG B9 was performed in the same manner, except that the original drop sizes were maintained and the plates were set up in nanoscale using the high-throughput robot, as for broad commercial screening. The concentration of the citrate buffer was varied instead of the pH [25-200 mM citrate pH 5.0 in steps of 25 mM; 13-24% (w/v) PEG 6000 in steps of 1%]. Manual optimisation of Proplex C4 by varying the $LiSO_4$ and PEG (w/v) concentrations [0.1 M MES pH 6.0 throughout, 150-237.5 mM $LiSO_4$ in steps of 12.5 mM, 18-20.75% (w/v) PEG 4000 in steps of 0.25%] was also performed.

Unfortunately, no crystals were observed in any of the aforementioned optimisation screens.

In a further attempt to optimise the JCSG B9 condition, Additive, Silver Bullets HT, and Silver Bullets Bio screens were performed (Hampton Research) (*31*). Out of all the conditions in the three screens (288 total), only two hits were observed in the presence of the following additives: i) 2% (w/v) benzamidine hydrochloride from well F9 of the Additive screen and ii) a cocktail of organic acids from well G1 of the Silver Bullets HT screen including 0.16% (w/v) 5-sulfosalicylic acid dihydrate, 0.16% (w/v) dodecaneoic acid, 0.16% (w/v) hippuric acid, 0.16% (w/v) mellitic acid, 0.16% (w/v) oxalacetic acid, 0.16% (w/v) suberic acid, and 20 mM HEPES pH 6.8. Benzamidine hydrochloride was purchased and manual optimisation of JCSG B9 in the presence of 2% (w/v) benzamidine hydrochloride was performed in the same manner. For the first time outside of commercially available broad screens, small (50 μm) crystals were observed [125 mM citrate pH 5.0, 20% (w/v) PEG 6000, 2% (w/v) benzamidine hydrochloride]. However, despite their reproducibility, these crystals were of an irregular morphology and were found to diffract poorly at the synchrotron, to a maximum of 7 Å, and so were not pursued further.

5.6.4 CP4H crystal formation is highly sensitive to external stimuli

In general, CP4H:PDI crystals (Figure 5.7) were found to be very fragile and sensitive to any changes in condition that deviated from the original hit in the broad screen. The ratio of protein solution to reservoir solution in the crystallisation drop proved to be important, with the largest and most well-formed crystals observed in crystallisation drops of a 1:1 ratio. Slightly more

nucleation and smaller crystals were observed in drops containing a 1:2 protein:reservoir ratio. No crystals were ever observed in identical conditions at 4 °C. All hits were screened in-house before being brought to the synchrotron, yet no in-house diffraction was observed, suggesting that the high energy synchrotron beamline would be required for diffraction.

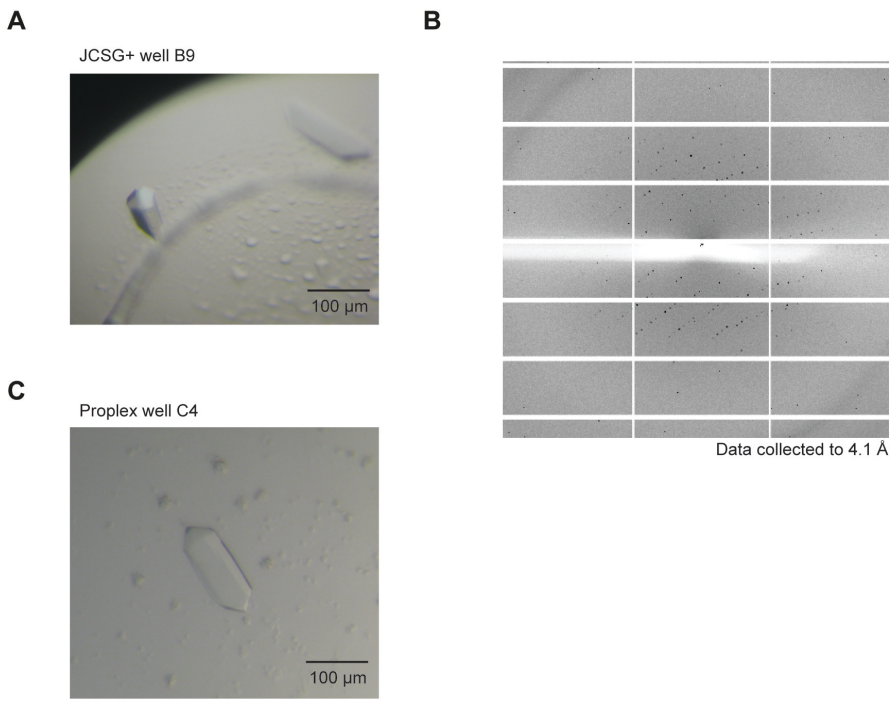


Figure 5.7. Crystallisation and preliminary diffraction of a CP4H:PDI tetramer. (**A**) Representative crystals from the JCSG+ screen well B9. (**B**) Synchrotron diffraction pattern of crystals from the JCSG+ well B9; the particular dataset was complete to 4.1 Å. (**C**) Representative crystals from the Proplex screen well C4.

5.6.5 CP4H crystals are sensitive to choice of the cryo-protectant

Independent of the condition, CP4H:PDI crystals were found to be highly sensitive to cryo-protectant solution. The addition of 25% (v/v) glycerol or 25% (v/v) ethylene glycol, diluted in well solution, quickly resulted in crystal disintegration. To overcome this issue, rather than diluting the cryo-protectant

(glycerol, ethylene glycol, etc.) in well solution, a separate cryo-protectant solution was prepared in which the concentrations of the precipitants in the well solution were not diluted. Interestingly, only crystallisation experiments performed using fresh protein (never frozen) afforded diffraction quality crystals. Finally, allowing crystals to grow undisturbed for at least 1 month, even from conditions in which they appeared after one week, was found to improve both their size and morphology.

5.6.6 Data collection

Collecting datasets complete to high resolution also proved problematic as the CP4H:PDI crystals degraded quickly upon exposure to synchrotron radiation, resulting in increasingly lower observed diffraction even after the first few images.

Despite these challenges, enough crystals of sufficient quality were obtained such that complete datasets could be collected at the synchrotron. High-quality crystals typically diffracted no further than a limit of ~4 Å. The highest resolution complete dataset diffracted to 3.8 Å and was collected using a single crystal obtained from JCSG well B9 in the presence of the Silver Bullet well G1 additive (for data collection statistics, see Table 5.8). The data was processed in the $P6_3$ hexagonal space group, in empirical agreement with the observed hexagonal crystal morphology. Calculation of the Matthew's coefficient revealed that the asymmetric unit likely contained one intact CP4H:PDI tetramer and ~45% solvent.

Table 5.8. CP4H:PDI data collection statistics. Statistics correspond to the highest quality dataset to date for CP4H:PDI using a single crystal from JCSG well B9 and 25% glycerol in well solution as a cryo-protectant. § indicates statistics from the highest resolution bin.

	CP4H:PDI
X-ray source X	Diamond Light Source beamline I04-1
Wavelength (Å)	0.92000
Resolution (Å)	3.84 (3.93–3.84)§
Space group	$P6_3$
Unit Cell Dimensions (*a* Å, *b* Å, *c* Å)	177.71, 177.71, 111.44
Total Number of Reflections Observed	307010
Number of Unique Reflections	19281 (1266)§
Redundancy	16.1 (15.7)§
Completeness (%)	100.0 (100.0)§
I/σ(I)	11.8 (2.0)§
Average *B* factor (Å^2)	124.34

5.7 Attempts to solve the structure of CP4H:PDI using various methods

5.7.1 Molecular replacement

Initial attempts to solve the structure of CP4H:PDI involved molecular replacement using the MR-PHASER (*32*) subroutine of the PHENIX (*33*) crystallographic software package. For the human CP4H 2OG oxygenase domain, for which there is no reported structure, the structure of *C. reinhardtii* P4H (CrP4H) (PDB ID: 3GZE) was used as a search model (*34*). The reported structure of the human *N*-terminal dimerisation and PSB domains of CP4H (PDB ID: 4BTA) (*35*) were used to search for the same domains in the asymmetric unit of CP4H:PDI. There are two reported structures of human PDI, either in its oxidised (PDB ID: 4EL1)(*36*) or reduced form (PDB ID: 4EKZ)(*36*), which show large conformational changes between the two states. It was predicted that PDI, the β subunit of the tetramer, would be present in its oxidised form due to the oxidising environment of the cytosol of *E. coli* Origami 2 (DE3) cells and also the

demonstrated requirement of two disulfide bonds within the CP4H α subunit for tetramer formation. It was thought to be highly unlikely that the CP4H α subunit would be oxidised and the PDI β subunit would be reduced. Unfortunately, these initial attempts were not successful.

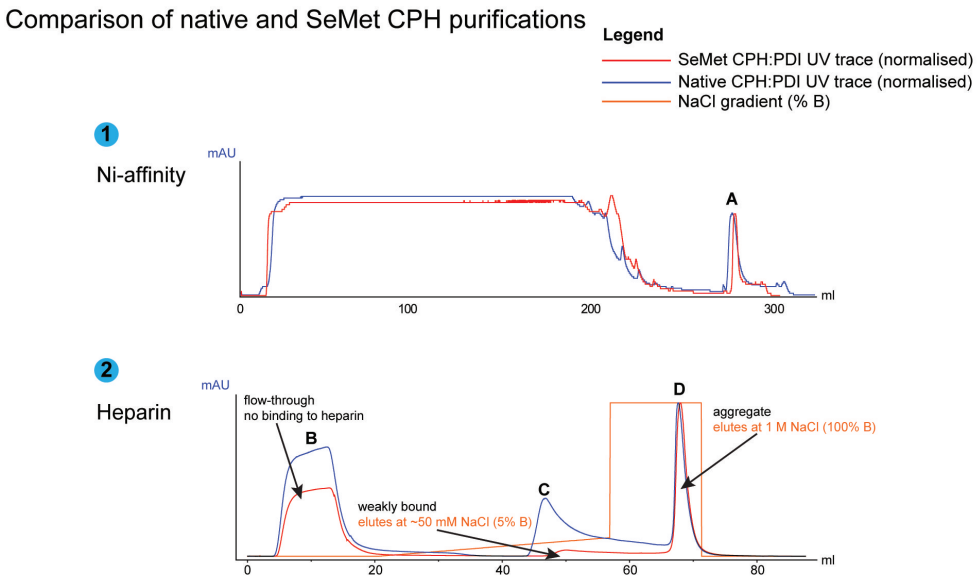
Subsequent attempts to solve the CP4H:PDI structure by molecular replacement involved fine-tuning various parameters including i) manually trimming surface loops on the CrP4H structure such that only the core DSBH remained, ii) individually searching for the PSB domain and *N*-terminal dimerisation domain, iii) individually searching for the a, a', b, and b' PDI subdomains given that it is likely that large conformational changes occur in the structure of PDI in the CP4H:PDI tetramer relative to isolated PDI, iv) changing the search order of components within the asymmetric unit, and v) manipulating the allowable clashes between chains within the molecular replacement algorithm (from 5-10% of atoms/chain) in an attempt to accommodate major domain movements. The aforementioned strategies (i-v) were performed in isolation and in combination, yet still no molecular replacement solutions were identified.

The continued inability to solve the structure by molecular replacement could be due to the fact that the quality of the 3.8 Å resolution data set is not sufficient or that there are substantial conformational changes in the CP4H and/or PDI subunits from their previously crystallised forms that render molecular replacement highly unlikely, even after adjusting clash constraints.

5.7.2 Single-wavelength anomalous dispersion (SAD) using selenomethionine (SeMet)-labelled protein

Given that efforts to solve the structure of CP4H:PDI $\alpha_2\beta_2$ tetramer by molecular replacement continued to be unsuccessful, alternative phasing strategies using single-wavelength anomalous dispersion (SAD) were attempted.

Expression and purification of SeMet-labelled CP4H:PDI was pursued as the first alternative method of phasing. In addition, SeMet-labelled protein has been shown to crystallise in different space groups, resulting in crystals with diffraction limits altered relative to those of native protein crystals. *E. coli* Origami 2 (DE3) cells transformed with the same pETduet-1 plasmid encoding for hexahistidine tagged CP4H:PDI used for native protein production were grown in minimal media supplemented with selenomethionine (Molecular Dimensions). The cells were harvested and lysed as for *E. coli* Origami 2 (DE3) grown in 2TY medium and purification of SeMet-labelled CP4H:PDI was attempted using the same purification protocol as was successful for native protein. Comparison of Ni-affinity chromatograms of SeMet-labelled and native CP4H:PDI revealed similar levels of Ni-affinity eluent (Figure 5.9). However, subsequent heparin cation exchange chromatography revealed a striking difference in the relative amount of SeMet-labelled tetramer compared to native tetramer, despite a similar level of aggregate (Figure 5.9). Indeed, in practice there was little to no recoverable yield of SeMet-labelled CP4H:PDI tetramer.



Peak	Corresponding species	Relative yield of SeMet preparation
A	Heterogeneous mixture of CPH/PDI	Same
B	CPH/PDI monomers	Lower yield (rougly half)
C	Predicted CPH:PDI tetramer	No recoverable yield
D	Aggregate	Same

Figure 5.9. Comparison of native and SeMet-labelled CP4H:PDI tetramer purifications. UV traces for native and SeMet-labelled CP4H:PDI were normalised to the most abundant peak for easier relative comparison of peak height.

The notable absence of tetramer production in the case of the SeMet-labelled protein, despite efficient protein production, could possibly be due to propensity of selenomethionine itself to undergo oxidation (selenium is more prone to oxidation than sulfur), which likely would be increased by the oxidising environment in the *E. coli* Origami 2 (DE3) cytosol. Typical protocols for SeMet-labelled protein production involve the addition of a reducing agent, such as DTT, to the growth medium to prevent protein aggregation caused by SeMet oxidation. However, given that previous studies had shown that the presence of reducing agent precludes tetramer formation by preventing the formation of the two critical intramolecular disulfide bonds within the CP4H 2OG oxygenase

domain (*3, 11*), the addition of reducing agent to the growth medium was avoided.

To test the inverse of this hypothesis, *E. coli* B834 (DE3), a methionine auxotroph without the oxidising cytosol of *E. coli* Origami 2 (DE3), was used as an expression strain. Purification of SeMet-labelled CP4H:PDI was pursued in the same way. Similar levels of hexahistidine tagged protein were obtained after Ni-affinity chromatography, yet, reminiscent of SeMet-labelled CP4H:PDI purification from *E. coli* Origami 2 (DE3), no recoverable yield of CP4H:PDI tetramer was observed during heparin cation exchange chromatography. In this case, the lack of an oxidising cytosol likely precluded tetramer formation, as has been shown using *E. coli* BL21 (DE3) cells (*8*), rather than aggregation resulting from SeMet selenium oxidation.

Unfortunately, the paradoxical requirements for an oxidising environment for CP4H:PDI tetramer formation and a reducing environment for SeMet stability resulted in an inability to produce SeMet-labelled tetramer. Therefore, other SAD phasing methods using heavy halides were attempted.

5.7.3 Single-wavelength anomalous dispersion (SAD) using halide ions

Supplementation of cryo-protectant solution with sodium bromide (NaBr) followed by quick soaking (usually on the order of seconds) has been shown to result in the presence of an arbitrary number of ordered Br ions in the asymmetric unit, having been coordinated by electrostatic interactions to the protein surface (*37, 38*). These Br ions can then be detected by collected data at their absorption edge (0.92 Å) and used as anchors for phase determination (*38*).

Other methods for halide phasing have involved I atoms. 5-amino-2,4,6-triiodoisophthalic acid (I3C), otherwise known as the 'magic triangle', provides a more specific method of phasing because the three I atoms form three vertices of an equilateral triangle, easily recognisable within electron density maps as empirical evidence of correct structure determination (*39*). I3C can be added to cryo-protectant solution similar to NaBr, or it can be co-crystallised as a ligand, similar to NOG or HCOL5 (*40*).

Initial experiments consisted of soaking CP4H:PDI crystals with NaBr (0.4-2 M) or I3C (2-10 mM) supplemented cryo-protectant solution. However, both NaBr and I3C were found to cause significant damage to the fragile CP4H:PDI crystals. Consequently, only minimal (>8 Å) diffraction was observed at the synchrotron, which deteriorated almost completely after the collection of a few images (typically under 20 images, indicative of a ~4 secs lifetime). Crystals from JCSG well A2 and JCSG well B9 grown in the presence of with 2% (w/v) benzamidine hydrochloride were screened with both NaBr and I3C. Given that I3C can be co-crystallised with proteins as a ligand, 5 mM I3C was added to protein solution and then screened manually in the JCSG B9 condition that included 2% (w/v) benzamidine hydrochloride, but no crystals were obtained.

5.7.4 Summary of challenges to solving the CP4H structure

In sum, multiple strategies were attempted in order to solve the structure of CP4H:PDI tetramer, which include molecular replacement and SAD phasing via SeMet-labelled protein production or NaBr/I3C soaking, yet none was successful. Further, the propensity of CP4H:PDI crystals to rapidly degrade upon exposure to synchrotron radiation would likely preclude the collection of a

highly redundant dataset optimal for SAD. Therefore, future attempts to solve the structure using higher resolution data may find the most success should they revisit molecular replacement.

5.8 Small-angle X-ray scattering (SAXS) provides a working model of CP4H

5.8.1 SAXS can provide a molecular envelope

The difficulty of solving a crystal structure of CP4H:PDI tetramer prompted the search for other methods to determine a 3-dimensional model of CP4H:PDI that did not require a crystalline sample. Small-angle X-ray scattering (SAXS), which focuses on the low-resolution scattering about a small range of angles (<10°) of an in-solution sample, can be used to determine a molecular envelope of a protein (*41, 42*). Since the sample need not be crystalline, an advantage of SAXS lies in its ability to map the general location of flexible regions of protein that would be otherwise disordered and thus undetectable in a crystal structure. SAXS was predicted to be especially useful for CP4H:PDI tetramer due to its predicted irregular oblong structure.

5.8.2 Ab initio model of CP4H

SAXS experiments on CP4H:PDI tetramer were conducted by Dr. Ed Lowe (Biochemistry, University of Oxford) at the European Synchrotron Radiation Facility BioSAXS beamline BM29 equipped with a Pilatus 1M detector. The protein solution was identical to that used for crystallisation experiments. Data were processed under the supervision of Dr. Ed Lowe at the ESRF BioSAXS BM29 (for details, see Materials and Methods).

An *ab initio* molecular envelope for the CP4H:PDI tetramer was calculated by the DAMMIN software package at the beamline under the supervision of Dr. Ed Lowe (Figure 5.10)(*43*). The shape of the SAXS molecular envelope for CP4H:PDI tetramer was oblong and had dimensions of 210 Å and 90 Å at its widest and longest points, respectively (Figure 5.10), and was in agreement with a SAXS envelope previously reported for the *N*-terminal dimerisation domain and PSB domain of the α subunit (*35*), which possessed a similar oblong Z-shape shape and a distance of 180 Å at its longest point (the 30 Å discrepancy can be attributed to the presence of the full-length α and β subunits in the tetramer).

The molecular envelope was of sufficient resolution such that individual domains could be placed in order to determine a possible working model of the CP4H:PDI tetramer, and a clear narrowing of density halfway along its x-axis (based on Figure 5.10) likely corresponded to the interlocking *N*-terminal dimerisation domains of two CP4H α subunits. Interestingly, whereas two complete CP4H α subunits could be placed into the molecular envelope, the *ab initio* model did not appear to be completely symmetric about its y-axis (based on Figure 5.10), and only one complete PDI β subunit could be modelled.

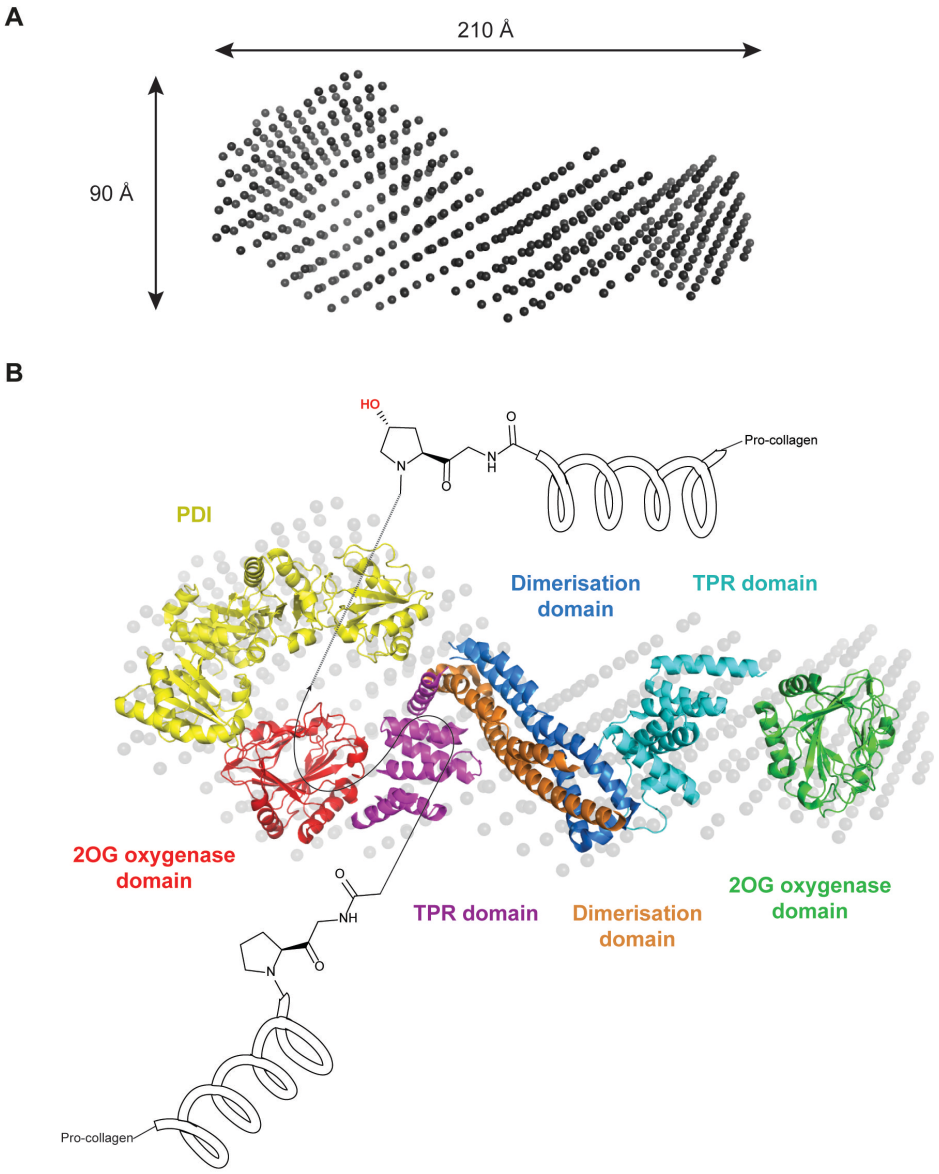


Figure 5.10. SAXS envelope derived CP4H model and predicted substrate shuttling mechanism. (**A**) Averaged SAXS envelope for CP4H generated by the DAMMIN software package at the beamline under the supervision of Dr. Ed Lowe (*43*). The envelope dimensions are consistent with a large protein complex. (**B**) PDI (yellow) (PDB ID: 4EL1), one CP4H monomer consisting of an *N*-terminal dimerisation domain (orange) (PDB ID: 4BTA), TPR (PSB) domain (magenta) (PDB ID: 4BTA) and 2OG oxygenase domain (red) (PDB ID: 3GZE). A second CP4H domain was modelled into the envelope consisting of an *N*-terminal dimerisation domain (blue) (PDB ID: 4BTA), TPR (PSB) domain (cyan) (PDB ID: 4BTA) and 2OG oxygenase domain (green) (PDB ID: 3GZE).

It is difficult to speculate whether CP4H:PDI protein thus far referred to as, and assumed to be, a $\alpha_2\beta_2$ tetramer was in fact an $\alpha_2\beta$ trimer. Despite the wealth of literature on CP4H:PDI $\alpha_2\beta_2$ tetramer purification (*4, 5, 8, 44*), there

has not been conclusive demonstration that the 'tetramer' did in fact contain two β subunits. As described in this chapter, previous characterisation of the $\alpha_2\beta_2$ tetramer typically involved native PAGE.

Notably, the manner in which a protein migrates through a native PAGE gel is reflective of not only its molecular weight (as is the case in SDS-PAGE), but also its 3-dimensional structure. Globular proteins such as BSA migrate well on native PAGE and as such are used as standards. It is possible that the CP4H:PDI $\alpha_2\beta_2$ tetramer (or $\alpha_2\beta$ trimer), given its oblong shape, would migrate through the gel at a significantly slower rate (and so appear at a relatively higher molecular weight) than would otherwise be expected from a globular protein and, thus, direct comparison to BSA tetramer using a native PAGE gel is not indicative of true CP4H:PDI $\alpha_2\beta_2$ tetramer, but rather $\alpha_2\beta$ trimer. Revisiting the highest resolution crystallographic dataset (3.8 Å, see Table 5.8), which had previously assumed the asymmetric unit contained an $\alpha_2\beta_2$ tetramer, revealed that an $\alpha_2\beta$ trimer could be accommodated in the asymmetric unit with roughly ~50% solvent content. However, repeated attempts to solve the structure of the $\alpha_2\beta$ trimer have yet to be successful.

Nonetheless, the αβ CP4H:PDI heterodimer region of the SAXS-derived model suggests a mechanism for collagen substrate binding and turnover. The molecular envelope revealed a depression, or possibly a tunnel, in the centre of the αβ heterodimer region encircled by PDI, the CP4H 2OG oxygenase domain, and the CP4H PSB domain. Further inspection of the PSB domain revealed that the face on which collagen-like substrate was bound in the structure of the isolated *N*-terminal dimerisation and PSB domains (PDB ID: 4BTA) was positioned close to the active site of the CP4H 2OG oxygenase domain (*35*). These

observations suggest a substrate shuttling, or 'assembly line', mechanism for CP4H:PDI substrate turnover whereby binding of substrate to the PSB domain organises it for hydroxylation and subsequent exit through the hole projecting from the opposite face of the 2OG oxygenase domain active site. It is possible that an 'assembly line' mechanism, which permits high-throughput catalysis with minimal major enzyme/substrate conformational changes (as observed in the PPHD:EF-Tu complex), evolved to allow CP4H to efficiently catalyse the hydroxylation of prolyl residues in many sequence contexts along the elongated (>1000 residue) pro-collagen chain.

5.9 Conclusions

The identification and characterisation of human collagen prolyl-hydroxylase nearly 50 years ago marked the first discovery of an enzyme in a family now known as the 2OG oxygenases (*1, 45*). Despite its critical role in human collagen biosynthesis (*46*), studies on CP4H have been limited primarily due to difficulties in its recombinant production (*47*), and its 3-dimensional structure remains unknown. This chapter reports a novel purification of active CP4H and the first direct characterisation of CP4H-catalysed hydroxylation of peptides, including those with identical sequence to naturally occurring human collagen. The results will be useful in the design of selective inhibitors of CP4H and in the validation of existing inhibitors of the human PHDs, including compounds in phase III clinical trials, which have not yet been profiled against CP4H. Most importantly, the first known crystallisation and preliminary data collection of a CP4H:PDI $\alpha_2\beta_2$ tetramer (or possibly $\alpha_2\beta$ trimer) is described, and will be of immediate interest for future attempts at a structure solution.

5.10 Experimental procedures

5.10.1 Plasmid design and construction

The DNA sequences for human CP4H (Uniprot ID: P13674, residues 18-534, including a *N*-terminal hexahistidine tag and 5′ NdeI and 3′ KpnI restriction sites) and human PDI (Uniprot ID: P07237, residues 18-508 with 5′ EcoRI and 3′ NotI restriction sites) lacking their respective *N*-terminal transmembrane localisation domains were designed, including codon optimisation for *E. coli*, and obtained from GeneArt (Invitrogen) The final DNA sequence for each gene codon optimised for *E. coli* is shown below.

Human CP4H (residues 18-534) DNA fasta sequence:

CCATGGCATATGGGCAGCAGCCATCACCATCATCACCACCATCCGGGTTTTTTTACCAGCATTGGTCAG
ATGACCGATCTGATTCATACCGAAAAAGATCTGGTTACCAGCCTGAAAGATTATATCAAAGCCGAAGA
AGATAAACTGGAACAAATCAAAAAATGGGCAGAGAAACTGGATCGTCTGACCAGCACCGCAACCAAAG
ATCCGGAAGGTTTTGTTGGTCATCCGGTTAATGCATTTAAACTGATGAAACGTCTGAACACCGAATGGT
CAGAACTGGAAAATCTGGTTCTGAAAGATATGAGCGACGGCTTTATTAGCAATCTGACAATTCAGCGTC
AGTATTTTCCGAACGATGAAGATCAGGTTGGTGCAGCAAAAGCACTGCTGCGTCTGCAGGATACCTATA
ATCTGGATACCGATACCATTAGCAAAGGTAATCTGCCTGGTGTGAAACATAAAAGTTTTCTGACCGCAG
AAGATTGCTTTGAACTGGGTAAAGTTGCATATACCGAGGCCGATTATTATCATACAGAACTGTGGATG
GAACAGGCACTGCGTCAGCTGGATGAAGGTGAAATCAGCACCATTGATAAAGTTAGCGTTCTGGATTA
TCTGAGCTATGCAGTTTATCAGCAGGGTGATCTGGATAAAGCCCTGCTGCTGACCAAAAAACTGCTGGA
ACTGGATCCGGAACATCAGCGTGCCAATGGTAATCTGAAATACTTCGAATACATCATGGCCAAAGAAAA
AGATGTGAACAAAAGCGCAAGTGATGATCAGAGCGATCAGAAAACCACCCCGAAAAAAAAAGGTGTTG
CAGTTGATTATCTGCCGGAACGTCAGAAATATGAAATGCTGTGTCGTGGTGAAGGCATTAAAATGACA
CCGCGTCGTCAGAAAAAACTGTTTTGTCGTTATCATGATGGCAACCGCAATCCGAAATTTATCCTGGCA
CCGGCAAAACAAGAGGATGAATGGGATAAACCGCGTATTATTCGCTTTCACGATATTATCAGTGATGCC
GAAATCGAAATTGTTAAAGACCTGGCAAAACCGCGTCTGCGTCGTGCAACCATTAGTAATCCGATTACA
GGCGATCTGGAAACCGTTCATTATCGTATTAGCAAAAGCGCCTGGCTGAGCGGTTATGAAAATCCGGTT
GTTAGCCGCATTAATATGCGTATTCAGGATCTGACCGGTCTGGATGTTAGCACAGCAGAAGAACTGCAG
GTTGCAAATTATGGTGTTGGTGGTCAGTATGAACCGCATTTTGATTTTGCACGTAAAGATGAACCGGA
TGCCTTTAAAGAACTGGGCACCGGTAATCGTATTGCCACCTGGCTGTTTTATATGAGTGATGTTAGTGC
CGGTGGTGCAACCGTTTTTCCGGAAGTTGGTGCCAGCGTTTGGCCGAAAAAAGGCACCGCAGTGTTTTG
GTATAACCTGTTTGCAAGCGGTGAAGGTGATTATAGCACCCGTCATGCAGCATGTCCGGTTCTGGTTGG
TAATAAATGGGTTAGCAATAAATGGCTGCATGAACGCGGTCAAGAATTTCGTCGTCCGTGTACCCTGAG
CGAACTGGAATAAGGTACCAAGCTT

Human PDI (residues 18-508) DNA fasta sequence:

GAATTCGGATGCACCGGAAGAAGAAGATCACGTTCTGGTTCTGCGTAAAAGCAATTTTGCAGAAGCACT
GGCAGCACATAAATATCTGCTGGTTGAGTTTTATGCACCGTGGTGTGGTCATTGTAAAGCCCTGGCACC
GGAATATGCAAAAGCAGCAGGTAAACTGAAAGCAGAAGGTAGCGAAATTCGTCTGGCAAAAGTTGATG
CAACCGAAGAAAGCGATCTGGCACAGCAGTATGGTGTTCGTGGTTATCCGACCATTAAATTCTTTCGTA
ATGGTGATACCGCAAGCCCGAAAGAATATACCGCAGGTCGTGAAGCAGATGATATTGTTAATTGGCTG

AAAAAACGTACCGGTCCGGCAGCAACCACCCTGCCGGATGGTGCAGCAGCAGAAAGCCTGGTTGAAAGC
AGCGAAGTTGCAGTTATTGGCTTTTTCAAAGATGTGGAAAGCGATAGCGCAAAACAGTTTCTGCAGGC
AGCCGAAGCAATTGATGATATTCCGTTTGGTATTACCAGCAACAGTGATGTGTTTAGCAAATACCAGCT
GGATAAAGATGGTGTGGTGCTGTTCAAAAAATTCGATGAAGGTCGCAATAACTTTGAAGGCGAAGTGA
CCAAAGAAAACCTGCTGGATTTTATCAAACACAATCAGCTGCCGCTGGTGATTGAATTTACCGAACAGA
CCGCACCGAAAATCTTTGGTGGTGAAATCAAAACCCATATCCTGCTGTTTCTGCCGAAAAGCGTTAGCG
ATTATGATGGTAAACTGTCCAACTTTAAAACCGCAGCCGAAAGCTTTAAAGGCAAAATCCTGTTTATCT
TCATCGATAGCGATCACACCGATAATCAGCGTATCCTGGAATTTTTCGGTCTGAAAAAAGAAGAATGTC
CGGCAGTTCGTCTGATTACCCTGGAAGAAGAGATGACCAAATACAAACCGGAAAGCGAAGAACTGACCG
CAGAACGTATTACCGAATTTTGTCATCGTTTTCTGGAAGGTAAAATCAAACCGCATCTGATGAGCCAAG
AACTGCCTGAAGATTGGGATAAACAGCCGGTTAAAGTTCTGGTGGGTAAAAACTTTGAGGATGTGGCC
TTTGACGAGAAAAAAAACGTGTTCGTCGAGTTCTATGCCCCTTGGTGCGGACATTGCAAACAGCTGGCA
CCGATTTGGGACAAACTGGGTGAAACCTATAAAGATCATGAGAACATTGTGATCGCCAAAATGGATAG
CACCGCAAATGAAGTTGAAGCCGTTAAAGTGCATAGCTTTCCGACACTGAAATTTTTCCCTGCAAGCGC
AGATCGTACCGTGATTGATTATAATGGTGAACGTACCCTGGATGGCTTCAAAAAATTCCTGGAAAGCGG
TGGTCAGGATGGTGCCGGTGATGATGATGACCTGGAAGATCTGGAAGAGGCCGAAGAACCGGATATGG
AAGAGGATGATGATCAGAAAGCAGTGAAAGATGAACTGTAAGCGGCCGC

The pETduet-1 vector was obtained from Novagen. CP4H and PDI genes were subcloned into pETduet-1 to generate the pETduet-1_CP4H_PDI expression vector. For a plasmid map, see Figure 5.1.

5.10.2 CP4H protein purification

E. coli Origami 2 (DE3) cells were transformed with the pETduet-1_CP4H_PDI (encoding for protein with an *N*-terminal hexahistidine tag) and grown (37 °C; 180 rpm) to an OD_{600} of 0.8; recombinant protein production was induced with 0.5 mM β-D-1-thiogalactopyranoside (IPTG). The cells were then grown at 18°C overnight, harvested by centrifugation (10,000x*g*; 7 min), and stored at -80 °C. Cell pellets were resuspended in 50 mM Tris-HCl pH 7.5, 500 mM NaCl, 20 mM imidazole, one EDTA-free protease inhibitor tablet (Roche), and approximately 1 mg DNAseI (bovine pancreas, grade II, Roche) at room temperature with gentle stirring. Cells were lysed on ice by sonication and the lysate was cleared by centrifugation (50,000x*g*; 20 min). The supernatant was then loaded onto a 5 mL HisTrap HP column pre-equilibrated with the resuspension buffer and purified using an AKTA FPLC system (GE Healthcare).

The column was washed with 10 column volumes of 50 mM Tris-HCl pH 7.5, 500 mM NaCl, 40 mM imidazole, and protein was eluted with 50 mM Tris-HCl pH 7.5, 500 mM NaCl, 250 mM imidazole. The purified sample was then immediately buffer exchanged to 25 mM Tris-HCl pH 7.5 (crystallisation buffer) using a PD-10 column (Millipore). The sample was then further purified by cation exchange chromatography using a HiTrap Heparin HP column (GE Healthcare), pre-equilibrated with crystallisation buffer, and eluted in a gradient of crystallisation buffer plus 1 M NaCl. The sample was then buffer exchanged back into crystallisation buffer and subjected to further purification by anion exchange chromatography using a 20 mL MonoQ column, which was pre-equilibrated with crystallisation buffer. The protein was eluted in crystallisation buffer plus 1 M NaCl. The protein was then subjected to further purification using a Superdex 200 size exclusion column (GE Healthcare) pre-equilibrated in crystallisation buffer. Proteins were eluted using the same buffer. CP4H:PDI-containing fractions were pooled, concentrated to 5 mg/mL by diafiltration. For crystallisation, protein was not frozen, but instead directly used in crystallisation experiments. For activity assays, protein was aliquoted (70 µL), flash frozen in liquid N_2, and stored at -80 °C.

5.10.3 CP4H MALDI assay platform

Standard hydroxylation assays for CP4H:PDI tetramer were carried out by preparing a reaction mix of 1 µM CP4H:PDI tetramer, 100 µM peptide substrate (either HCOL1, HCOL3, or HCOL5), 50 µM Fe(II)(prepared from $(NH_4)_2Fe(SO_4)_2$), 400 µM 2OG, 1 mM sodium L-ascorbate in 25 mM Tris-HCl pH 7.5. HCOL3 and HCOL5 were derived from human collagen α-1(I) (Uniprot ID:

P02452). For determinations of apparent K_m of Fe(II), 2OG, ascorbate, each cofactor concentration was varied and reactions were incubated at room temperature for 20 minutes. For determination of apparent K_m of HCOL5 peptide, time courses at various concentrations of HCOL5 were carried out and steady-state reaction rate was determined. All reactions were quenched with an equal volume of 20% (v/v) formic acid. The extent of substrate hydroxylation was analysed by MALDI-TOF-MS (matrix-assisted laser-desorption ionisation–time-of-flight mass spectrometry). Recrystallised CHCA (α-cyano-4-hydroxycinnamic acid) MALDI matrix (1 μL) and quenched assay solution (1 μL) were spotted onto a 96-well MALDI sample plate and analysed using a Waters Micromass™ MALDI micro MX™ mass spectrometer in the positive ion reflectron mode with the following settings as described (*16*): laser energy 100-150, pulse 1950, detector 2350, suppression 700. Raw data was further analysed with MassLynx™ version 4.1. Data were fitted to the Michaelis-Menten equation using GraphPad Prism and kinetic parameters were determined from the fit. All assays were performed in triplicate and plotted as mean ± standard deviation.

5.10.4 Differential scanning fluorimetry

Differential scanning fluorimetry was performed using a BioRad MiniOpticon real-time PCR detection system and MJ Mini thermal cycler. 5,000x SYPRO Orange dye (Invitrogen) was used for nonspecific binding to hydrophobic protein residues. Final concentrations were, where applicable, 2 μM CP4H:PDI tetramer, 50 μM MnCl2, 100 μM NOG, 100 μM HCOL1, 100 μM HCOL3, and 100 μM HCOL5.

The fluorescence readings were taken every 1 °C from 25 – 95 °C with a linear temperature increase of 1 °C/min. In order to determine accurate T_m values, the data were exported to and analysed on GraphPad Prism software. The midpoint, or T_m, was calculated by fitting a Boltzmann curve between the minimum and maximum fluorescence intensities. The conditions that gave the highest ΔT_m compared to the reference indicated those that best stabilised CP4H:PDI tetramer.

5.10.5 X-ray crystallography

Protein solution for X-ray crystallography experiments typically contained 5 mg/mL CP4H:PDI tetramer, 0.5 mM $MnCl_2$, 1 mM NOG, and 1 mM HCOL5 in 25 mM Tris-HCl pH 7.5. Crystals of CP4H:PDI tetramer in complex with $MnCl_2$, NOG, and HCOL5 were grown in sitting drops using the vapour diffusion method (drop size: 200-300 nL) at 293 K in 96-well Intelliplates (Art Robbins). Crystals were cryo-protected by transfer to 25% (v/v) glycerol in well solution and then harvested in nylon loops (Hampton Research) and cryo-cooled by plunging in liquid nitrogen. Data were collected at 100 K using single crystals at Diamond Light Source beamline I02/I04/I04-1 all equipped with a Pilatus 6M-F detector. Data were then indexed, integrated, and scaled using HKL3000 (*48*). Attempts to determine the structure by molecular replacement (MR) were performed using the MR-PHASER (*32*) subroutine of PHENIX (*33*) using *C. reinhardtii* P4H (PDB ID: 3GZE)(*34*), *H. sapiens* human CP4H *N*-terminal dimerisation domain and PSB domain (PDB ID: 4BTA)(*35*), and *H. sapiens* oxidised PDI (PDB ID: 4EL1)(*36*) as search models.

5.10.6 Small-angle X-ray scattering (SAXS)

Protein solution was identical to that used in X-ray crystallographic experiments, which was then subjected to a serial 2x dilution. SAXS data were collected at the European Synchrotron Radiation Facility BioSAXS beamline BM29 equipped with a Pilatus 1M detector. Data processing, analysis, and molecular modelling were performed using programs of the ATSAS package (*49*) available at the beamline. *Ab initio* models were built using DAMMIN (*43*), also available at the beamline. Crystal structures of the components of CP4H α and PDI β subunits were manually superimposed onto an *ab initio* model generated by DAMMIN to obtain a functional model for CP4H $\alpha_2\beta_2$ tetramer (or $\alpha_2\beta$ trimer).

5.11 References

1. J. J. Hutton, Jr., A. Kaplan, S. Udenfriend, Conversion of the amino acid sequence gly-pro-pro in protein to gly-pro-hyp by collagen proline hydroxylase. *Arch Biochem Biophys.* **121**, 384-391 (1967).
2. K. I. Kivirikko, J. Myllyharju, Prolyl 4-hydroxylases and their protein disulfide isomerase subunit. *Matrix Biology.* **16**, 357-368 (1998).
3. R. A. Berg, D. J. Prockop, Affinity column purification of protocollagen proline hydroxylase from chick embryos and further characterization of the enzyme. *J Biol Chem.* **248**, 1175-1182 (1973).
4. T. Helaakoski, P. Annunen, K. Vuori, I. A. MacNeil, T. Pihlajaniemi, K. I. Kivirikko, Cloning, baculovirus expression, and characterization of a second mouse prolyl 4-hydroxylase alpha-subunit isoform: formation of an alpha 2 beta 2 tetramer with the protein disulfide-isomerase/beta subunit. *Proc Natl Acad Sci U S A.* **92**, 4427-4431 (1995).
5. A. Neubauer, P. Neubauer, J. Myllyharju, High-level production of human collagen prolyl 4-hydroxylase in Escherichia coli. *Matrix Biol.* **24**, 59-68 (2005).
6. N. R. Rose, M. A. McDonough, O. N. F. King, A. Kawamura, C. J. Schofield, Inhibition of 2-oxoglutarate dependent oxygenases. *Chemical Society Reviews.* **40**, 4364-4397 (2011).
7. A. Neubauer, J. Soini, M. Bollok, M. Zenker, J. Sandqvist, J. Myllyharju, P. Neubauer, Fermentation process for tetrameric human collagen prolyl 4-hydroxylase in Escherichia coli: improvement by gene optimisation of the PDI/beta subunit and repeated addition of the inducer anhydrotetracycline. *J Biotechnol.* **128**, 308-321 (2007).
8. E. A. Kersteen, J. J. Higgin, R. T. Raines, Production of human prolyl 4-hydroxylase in Escherichia coli. *Protein Expr Purif.* **38**, 279-291 (2004).
9. D. C. John, N. J. Bulleid, Prolyl 4-hydroxylase: defective assembly of alpha-subunit mutants indicates that assembled alpha-subunits are intramolecularly disulfide bonded. *Biochemistry.* **33**, 14018-14025 (1994).
10. A. Lamberg, T. Pihlajaniemi, K. I. Kivirikko, Site-directed mutagenesis of the alpha subunit of human prolyl 4-hydroxylase. Identification of three histidine residues critical for catalytic activity. *J Biol Chem.* **270**, 9926-9931 (1995).
11. J. J. Nietfeld, I. Van der Kraan, A. Kemp, Dissociation and reassociation of prolyl 4-hydroxylase subunits after cross-linking of monomers. *Biochim Biophys Acta.* **661**, 21-27 (1981).
12. A. Bentolila, I. Vlodavsky, C. Haloun, A. J. Domb, Synthesis and heparin-like biological activity of amino acid-based polymers. *Polym Advan Technol.* **11**, 377-387 (2000).
13. W. Aik, M. Demetriades, M. K. Hamdan, E. A. Bagg, K. K. Yeoh, C. Lejeune, Z. Zhang, M. A. McDonough, C. J. Schofield, Structural basis for inhibition of the fat mass and obesity associated protein (FTO). *J Med Chem.* **56**, 3680-3688 (2013).
14. W. Aik, J. S. Scotti, H. Choi, L. Gong, M. Demetriades, C. J. Schofield, M. A. McDonough, Structure of human RNA N(6)-methyladenine demethylase

ALKBH5 provides insights into its mechanisms of nucleic acid recognition and demethylation. *Nucleic Acids Res.* **42**, 4741-4754 (2014).
15. I. Wittig, H. P. Braun, H. Schagger, Blue native PAGE. *Nat Protoc.* **1**, 418-428 (2006).
16. L. Kukkola, P. Koivunen, O. Pakkanen, A. P. Page, J. Myllyharju, Collagen prolyl 4-hydroxylase tetramers and dimers show identical decreases in Km values for peptide substrates with increasing chain length: mutation of one of the two catalytic sites in the tetramer inactivates the enzyme by more than half. *J Biol Chem.* **279**, 18656-18661 (2004).
17. R. Hieta, J. Myllyharju, Cloning and characterization of a low molecular weight prolyl 4-hydroxylase from Arabidopsis thaliana. Effective hydroxylation of proline-rich, collagen-like, and hypoxia-inducible transcription factor alpha-like peptides. *J Biol Chem.* **277**, 23965-23971 (2002).
18. K. Vuori, T. Pihlajaniemi, M. Marttila, K. I. Kivirikko, Characterization of the human prolyl 4-hydroxylase tetramer and its multifunctional protein disulfide-isomerase subunit synthesized in a baculovirus expression system. *Proc Natl Acad Sci U S A.* **89**, 7467-7470 (1992).
19. J. J. Hutton, Jr., A. L. Trappel, S. Udenfriend, Requirements for alpha-ketoglutarate, ferrous ion and ascorbate by collagen proline hydroxylase. *Biochem Biophys Res Commun.* **24**, 179-184 (1966).
20. J. J. Hutton, A. L. Tappel, Udenfrie.S, Cofactor and Substrate Requirements of Collagen Proline Hydroxylase. *Archives of Biochemistry and Biophysics.* **118**, 231-& (1967).
21. K. J. Carpenter, *The history of scurvy and vitamin C*. (Cambridge University Press, Cambridge, 1986), pp. viii, 288 p.
22. J. J. Hutton, A. L. Tappel, Udenfrie.S, A Rapid Assay for Collagen Proline Hydroxylase. *Anal Biochem.* **16**, 384-& (1966).
23. L. A. McNeill, E. Flashman, M. R. Buck, K. S. Hewitson, I. J. Clifton, G. Jeschke, T. D. Claridge, D. Ehrismann, N. J. Oldham, C. J. Schofield, Hypoxia-inducible factor prolyl hydroxylase 2 has a high affinity for ferrous iron and 2-oxoglutarate. *Mol Biosyst.* **1**, 321-324 (2005).
24. E. Flashman, E. A. Bagg, R. Chowdhury, J. Mecinovic, C. Loenarz, M. A. McDonough, K. S. Hewitson, C. J. Schofield, Kinetic rationale for selectivity toward N- and C-terminal oxygen-dependent degradation domain substrates mediated by a loop region of hypoxia-inducible factor prolyl hydroxylases. *J Biol Chem.* **283**, 3808-3815 (2008).
25. E. Flashman, S. L. Davies, K. K. Yeoh, C. J. Schofield, Investigating the dependence of the hypoxia-inducible factor hydroxylases (factor inhibiting HIF and prolyl hydroxylase domain 2) on ascorbate and other reducing agents. *Biochem J.* **427**, 135-142 (2010).
26. J. Myllyharju, K. I. Kivirikko, Collagens, modifying enzymes and their mutations in humans, flies and worms. *Trends Genet.* **20**, 33-43 (2004).
27. J. Newman, D. Egan, T. S. Walter, R. Meged, I. Berry, M. Ben Jelloul, J. L. Sussman, D. I. Stuart, A. Perrakis, Towards rationalization of crystallization screening for small- to medium-sized academic laboratories: the PACT/JCSG+ strategy. *Acta Crystallogr D Biol Crystallogr.* **61**, 1426-1431 (2005).

28. A. McPherson, A comparison of salts for the crystallization of macromolecules. *Protein Sci.* **10**, 418-422 (2001).
29. C. Grimm, A. Chari, K. Reuter, U. Fischer, A crystallization screen based on alternative polymeric precipitants. *Acta Crystallogr D Biol Crystallogr.* **66**, 685-697 (2010).
30. S. Radaev, S. Li, P. D. Sun, A survey of protein-protein complex crystallizations. *Acta Crystallogr D Biol Crystallogr.* **62**, 605-612 (2006).
31. A. McPherson, B. Cudney, Searching for silver bullets: An alternative strategy for crystallizing macromolecules. *J Struct Biol.* **156**, 387-406 (2006).
32. A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, R. J. Read, Phaser crystallographic software. *J Appl Crystallogr.* **40**, 658-674 (2007).
33. P. D. Adams, P. V. Afonine, G. Bunkoczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, P. H. Zwart, PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr.* **66**, 213-221 (2010).
34. M. K. Koski, R. Hieta, M. Hirsila, A. Ronka, J. Myllyharju, R. K. Wierenga, The crystal structure of an algal prolyl 4-hydroxylase complexed with a proline-rich peptide reveals a novel buried tripeptide binding motif. *J Biol Chem.* **284**, 25290-25301 (2009).
35. J. Anantharajan, M. K. Koski, P. Kursula, R. Hieta, U. Bergmann, J. Myllyharju, R. K. Wierenga, The structural motifs for substrate binding and dimerization of the alpha subunit of collagen prolyl 4-hydroxylase. *Structure.* **21**, 2107-2118 (2013).
36. C. Wang, W. Li, J. Q. Ren, J. Q. Fang, H. M. Ke, W. M. Gong, W. Feng, C. C. Wang, Structural Insights into the Redox-Regulated Dynamic Conformations of Human Protein Disulfide Isomerase. *Antioxid Redox Sign.* **19**, 44-53 (2013).
37. Z. Dauter, M. Dauter, K. R. Rajashankar, Novel approach to phasing proteins: derivatization by short cryo-soaking with halides. *Acta Crystallogr D.* **56**, 232-237 (2000).
38. Z. Dauter, M. Dauter, Entering a new phase: Using solvent halide ions in protein structure determination. *Structure.* **9**, R21-R26 (2001).
39. T. Beck, A. Krasauskas, T. Gruene, G. M. Sheldrick, A magic triangle for experimental phasing of macromolecules. *Acta Crystallogr D.* **64**, 1179-1182 (2008).
40. T. Beck, C. E. da Cunha, G. M. Sheldrick, How to get the magic triangle and the MAD triangle into your protein crystal. *Acta Crystallogr F.* **65**, 1068-1070 (2009).
41. M. V. Petoukhov, D. I. Svergun, Applications of small-angle X-ray scattering to biomacromolecular solutions. *Int J Biochem Cell Biol.* **45**, 429-437 (2013).
42. C. E. Blanchet, D. I. Svergun, Small-angle X-ray scattering on biological macromolecules and nanocomposites in solution. *Annu Rev Phys Chem.* **64**, 37-54 (2013).

43. D. I. Svergun, Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys J.* **76**, 2879-2886 (1999).
44. A. Vuorela, J. Myllyharju, R. Nissi, T. Pihlajaniemi, K. I. Kivirikko, Assembly of human prolyl 4-hydroxylase and type III collagen in the yeast Pichia pastoris: formation of a stable enzyme tetramer requires coexpression with collagen and assembly of a stable collagen requires coexpression with prolyl 4-hydroxylase. *Embo Journal.* **16**, 6702-6712 (1997).
45. C. Loenarz, C. J. Schofield, Expanding chemical biology of 2-oxoglutarate oxygenases. *Nat Chem Biol.* **4**, 152-156 (2008).
46. C. L. Jenkins, R. T. Raines, Insights on the conformational stability of collagen. *Nat Prod Rep.* **19**, 49-59 (2002).
47. J. Myllyharju, Prolyl 4-hydroxylases, the key enzymes of collagen biosynthesis. *Matrix Biol.* **22**, 15-24 (2003).
48. W. Minor, M. Cymborowski, Z. Otwinowski, M. Chruszcz, HKL-3000: the integration of data reduction and structure solution - from diffraction images to an initial model in minutes. *Acta Crystallogr D.* **62**, 859-866 (2006).
49. P. V. Konarev, M. V. Petoukhov, V. V. Volkov, D. I. Svergun, ATSAS 2.1, a program package for small-angle scattering data analysis. *Journal of Applied Crystallography.* **39**, 277-286 (2006).

6 Conclusions: insights into the evolution and distribution of prolyl-hydroxylases in life

6.1 Summary of results presented in this thesis

For more than 30 years after their discovery, the collagen prolyl-hydroxylases were assumed to be the only prolyl-hydroxylase subfamily. The discovery of the HIF prolyl-hydroxylases more than a decade ago as a subfamily distinct from the CP4Hs in both structure and function prompted renewed interest in prolyl-hydroxylases as both mediators of intracellular signalling and as potential therapeutic targets (*1, 2*). The collective findings sparked further interest in the evolutionary origins of these enzymes.

The determination of the crystal structure of human PHD2, the first of a prolyl-4-hydroxylase, provided structural conformation of previously performed bioinformatic analyses, which suggested that homologues of the PHDs were present in organisms that lacked HIF-α, such as *Pseudomonas spp.* (*3, 4*).

Embarking from these preliminary findings, Chapter 2 of this thesis details the identification and biochemical characterisation of PPHD in *Pseudomonas spp*. PPHD is revealed to be the first bacterial *bona fide* prolyl-hydroxylase that, like the PHDs and CP4Hs, catalyses the *trans*-4 hydroxylation of prolyl-residues. EF-Tu, an elongation factor universally conserved in prokaryotic organisms and known for its critical role in bacterial translation, is confirmed as a PPHD substrate both *in vitro* and *in vivo*. While kinetic studies predict no specific effect of prolyl-hydroxylation on bacterial translation, it is impossible to rule out the possibility that PPHD-catalysed hydroxylation of EF-

Tu Pro54 plays other roles in the cell, such as in cell stability or in yet undiscovered signalling pathways.

The observations that i) a PPHD deleted strain of *P. aeruginosa* overproduces the virulence factor pyocyanin and ii) that transfection of PPHD in human cells upregulates HIF-α combine to imply that PPHD and human prolyl-hydroxylases may be involved in 'cross talk' during *P. aeruginosa* mediated infection. Further work will be necessary to deduce the intracellular role of Pro54 hydroxylation in *Pseudomonas*.

Chapter 3 presents the crystal structures of PPHD, EF-Tu, and a PPHD:EF-Tu protein-protein complex. Comparisons of the PPHD structure with those of PHD2 reveal a striking structural homology, thereby further cementing genetic and biochemical evidence that PPHD is indeed a PHD homologue. The structure of a PPHD:EF-Tu complex, the first of any 2OG oxygenase in complex with its full-length protein substrate, reveals that PPHD binding to EF-Tu is mediated by induced fit. Comparisons of substrate recognition elements from PPHD:EF-Tu, PHD2:HIF-α CODD, and CrP4H:proline-rich peptide structures reveals that prolyl-4-hydroxylases share conserved substrate binding mechanisms, such as the conformational flexibility of the β2-β3 finger loop. Finally, structural analyses of the PPHD, EF-Tu, and PPHD:EF-Tu structures will be of immediate interest in the design of novel PHD2 (and other 2OG oxygenase) inhibitors based on different conformational states rather than active site iron chelators, which comprise the majority of 2OG oxygenase inhibitors including those of the PHDs currently in clinical trials (*5*).

The crystal structure of OGFOD1, the first of a human prolyl-3-hydroxylase, was determined in Chapter 4. Together with crystal structures of *S.*

cerevisiae Tpa1 in complex with inhibitors, the structures reveal similarities and important differences between the RPS23 hydroxylases and the PHDs and will be useful in the development of selective inhibitors.

Finally, Chapter 5 details a novel expression and purification protocol for human CP4H $\alpha_2\beta_2$ tetramer and presents the first known reports of its crystallisation and diffraction in nearly 50 years of study. Further attempts at a structure solution are of immediate importance, and will most likely involve obtaining a crystal form that diffracts to higher resolution, possibly by making use of different constructs or CP4H isoforms.

6.2 Insights into the evolutionary origins of prolyl-hydroxylases

Taken together, the results presented in this thesis suggest that prolyl-hydroxylases are ancient and have evolved to accept a variety of substrates ranging from EF-Tu in bacteria to, at least, collagen, HIF-α, and RPS23 in eukaryotes.

Crystal structures of PPHD, PHD2, OGFOD1, and Tpa1 are useful for performing structurally guided bioinformatic analyses on the distribution of prolyl-hydroxylases with a high confidence level. Topological comparisons of PPHD, PHD2, CrP4H, and Tpa1 reveal conservation not only of the DSBH, but also of other important elements in substrate recognition, such as the β2-β3 finger loop, and even α-helices that pack against the DSBH (Figure 6.1). Remarkably, enzymes in the prolyl-hydroxylase family that seem to possess similar structures are observed to catalyse different reactions (i.e. prolyl-3- vs. prolyl-4-hydroxylation) and accept different substrates, raising the question as to their evolutionary origin.

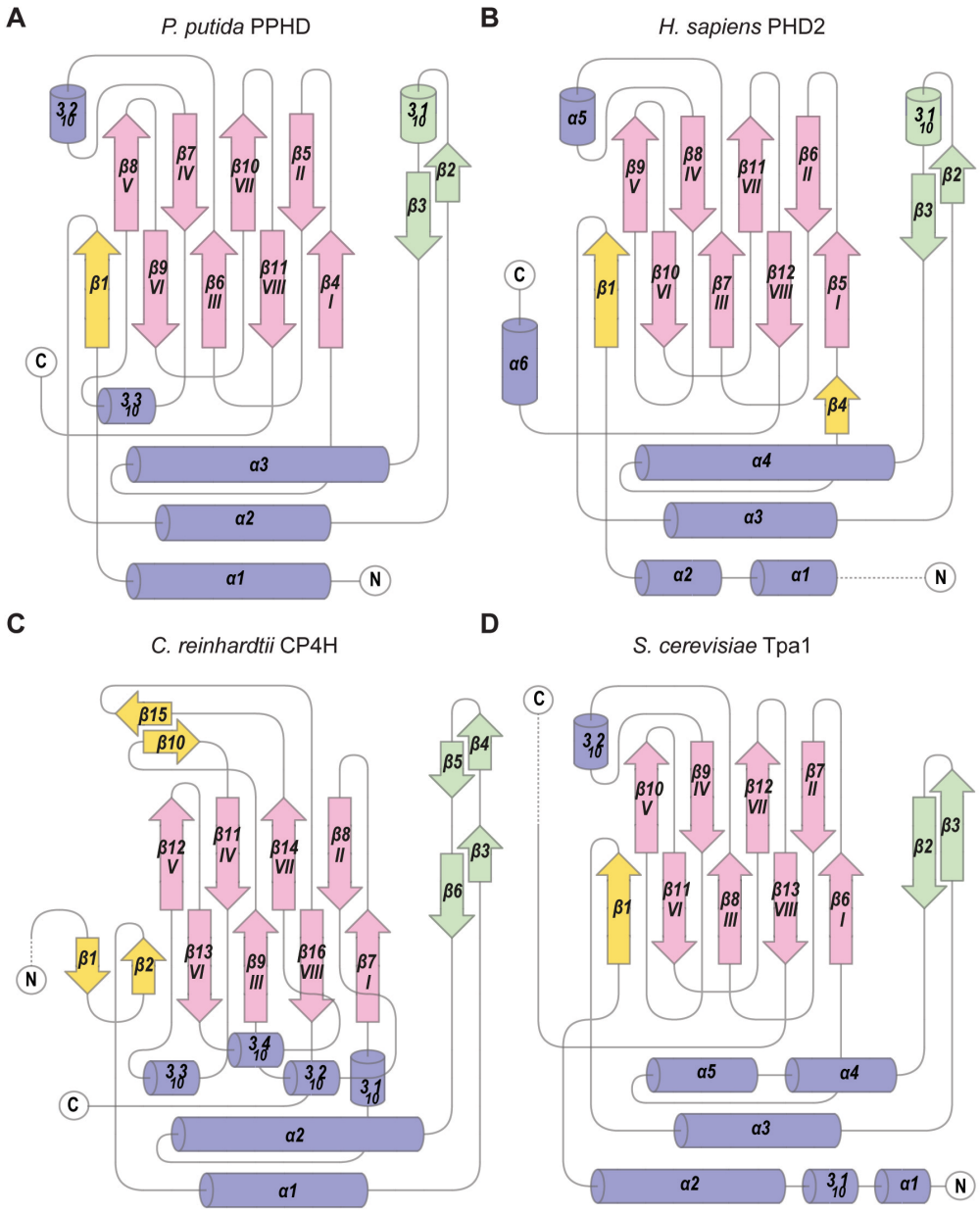


Figure 6.1. Comparison of the catalytic domain of prolyl-4- and prolyl-3-hydroxylases reveals similar overall topologies. 2-dimensional topology diagrams of representative prolyl-hydroxylases (**A**) PPHD, (**B**) PHD2, (**C**) CrP4H, (**D**) Tpa1 from different subfamilies reveal secondary structure conservation between catalytic domains. Colours represent structural homology to $PPHD_{putida}$; α- and 3_{10}-helices (blue), DSBH β-strands (pink), β2-β3 finger loop (green), other β-strands (yellow). Only the *N*-terminal catalytic domain of *S. cerevisiae* Tpa1 is shown. Figure and figure legend are adapted with permission from J.S. Scotti *et al*, Human oxygen sensing may have origins in prokaryotic elongation factor Tu prolyl-hydroxylation. *Proc Nat Acad Sci USA*. In press (2014).

Structurally guided bioinformatic analyses were performed in order to map the distribution of the CP4Hs, PHDs/PPHDs, Leprecans (collagen prolyl-3-hydroxylases), and OGFOD1s across all Kingdoms of life (Figure 6.2 and Figure 6.3 and for more detail see Figure 6.5). The results revealed that all subtypes of prolyl-hydroxylases are present in metazoa, and only the Leprecans are absent from choanoflagellates. Putative PHD/PPHD homologues are distributed widely both in the *Pseudomonas* genus and within gammaproteobacteria, including in the well-known human pathogen *Vibrio cholerae* and the facultative *Shewanella oneidensis*. Proteomic MS studies on *S. oneidensis* report that EF-Tu is modified by a +16 mass shift, suggesting that *S. oneidensis* may contain a putative EF-Tu hydroxylase similar to PPHD (*6*). On the whole, however, prolyl-hydroxylase homologues are more sporadically distributed in prokaryotes and likely absent in archaea.

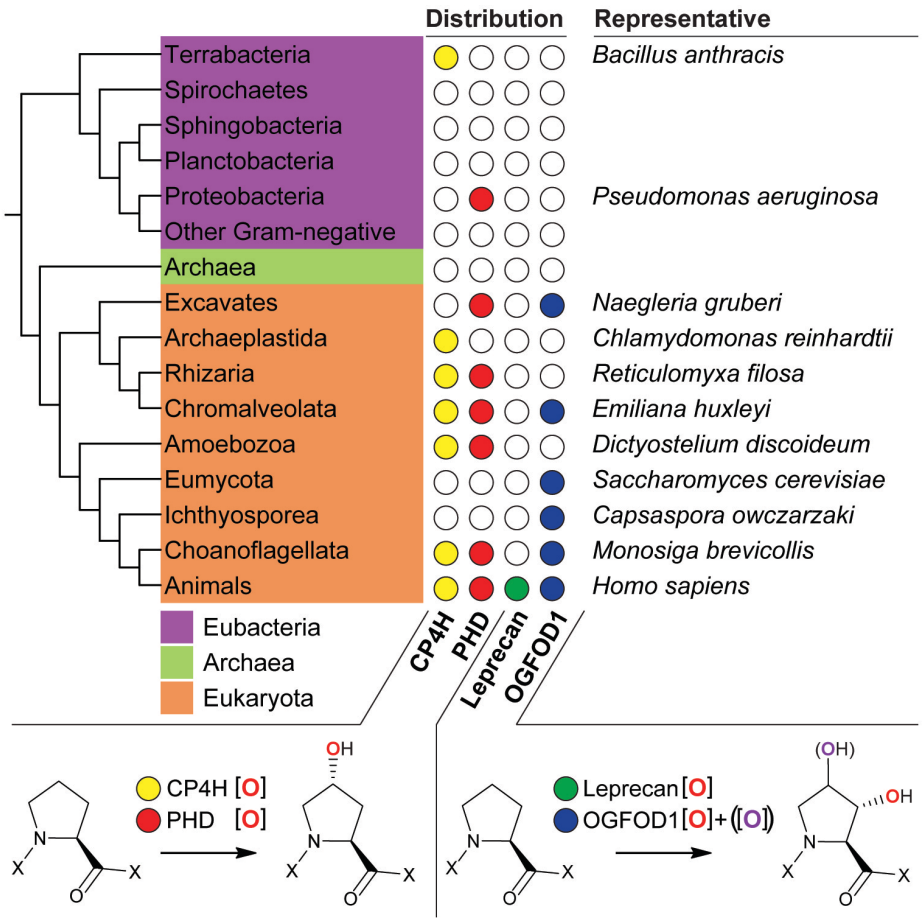


Figure 6.2. Prolyl-hydroxylases are sporadically distributed in Kingdoms of life. Tracing putative homologues of the CPHs, PHDs, Leprecans, and OGFOD1s in representative organisms throughout the Kingdoms of life (for a wider distribution, see Figure 6.3). Bioinformatics analyses performed in collaboration with Mr. Michael Bentley (Plant Sciences, University of Oxford) and Dr. Jordi Paps (Zoology, University of Oxford). Figure and figure legend are adapted with permission from J.S. Scotti *et al*, Human oxygen sensing may have origins in prokaryotic elongation factor Tu prolyl-hydroxylation. *Proc Nat Acad Sci USA*. In press (2014).

Of particular interest is the presence of CP4H and PHD homologues in organisms that lack collagen or HIF-α. CP4H homologues are present in the model organism *Arabidopsis thaliana* and the algae *Chlamydomonas reinhardtii*, where they have been reported to catalyse the hydroxylation of collagen-like sequences (*7*), yet neither of these organisms is predicted to contain a homologue of the PHDs. Inversely, the protozoan *Naegleria gruberi* contains a putative PHD homologue and lacks a CP4H homologue.

The observation that algal CrP4H, human PHD2, and *Pseudomonas* PHD share conserved substrate recognition machinery (i.e. substantial substrate-induced conformational changes in the β2-β3 finger loop) may reflect their divergence from a common prolyl-hydroxylase ancestor. The current structural and bioinformatic data can only lead to speculations as to whether ancestral prolyl-hydroxylases catalysed *trans* prolyl-4- or prolyl-3-hydroxylation, yet the presence and absence of prolyl-4- and prolyl-3-hydroxylases in prokaryotes, respectively, may indicate that prolyl-4-hydroxylation is more ancient. On the contrary, bacterial prolyl-4-hydroxylases may have emerged as a result of a horizontal gene transfer event from the eukaryotes to the bacteria. Indeed, many of the bacteria predicted to contain putative prolyl-hydroxylase homologues are also opportunistic human pathogens, such as *P. aeruginosa*, *V. cholerae*, and *Bacillus anthracis* and are thus in constant contact with eukaryotic organisms.

The RPS23 hydroxylases, including OGFOD1 and Tpa1, likely evolved in eukaryotes from single domain prokaryotic, or ancient eukaryotic, prolyl-hydroxylases through a gene duplication event, resulting in their characteristic double DSBH domain architecture. Having been eventually rendered void of their catalytic potential, the additional DBSH-containing domain likely retained function in substrate recognition and/or oligomerisation, which complemented that of the catalytically viable domain. The potential additional functions of these second domains, aside from oligomerisation in Tpa1, are still unknown, but may implicate in the substrate differentiation observed between OGFOD1, PHD2, and CP4H, and, more generally, prolyl-3- and prolyl-4-hydroxylases.

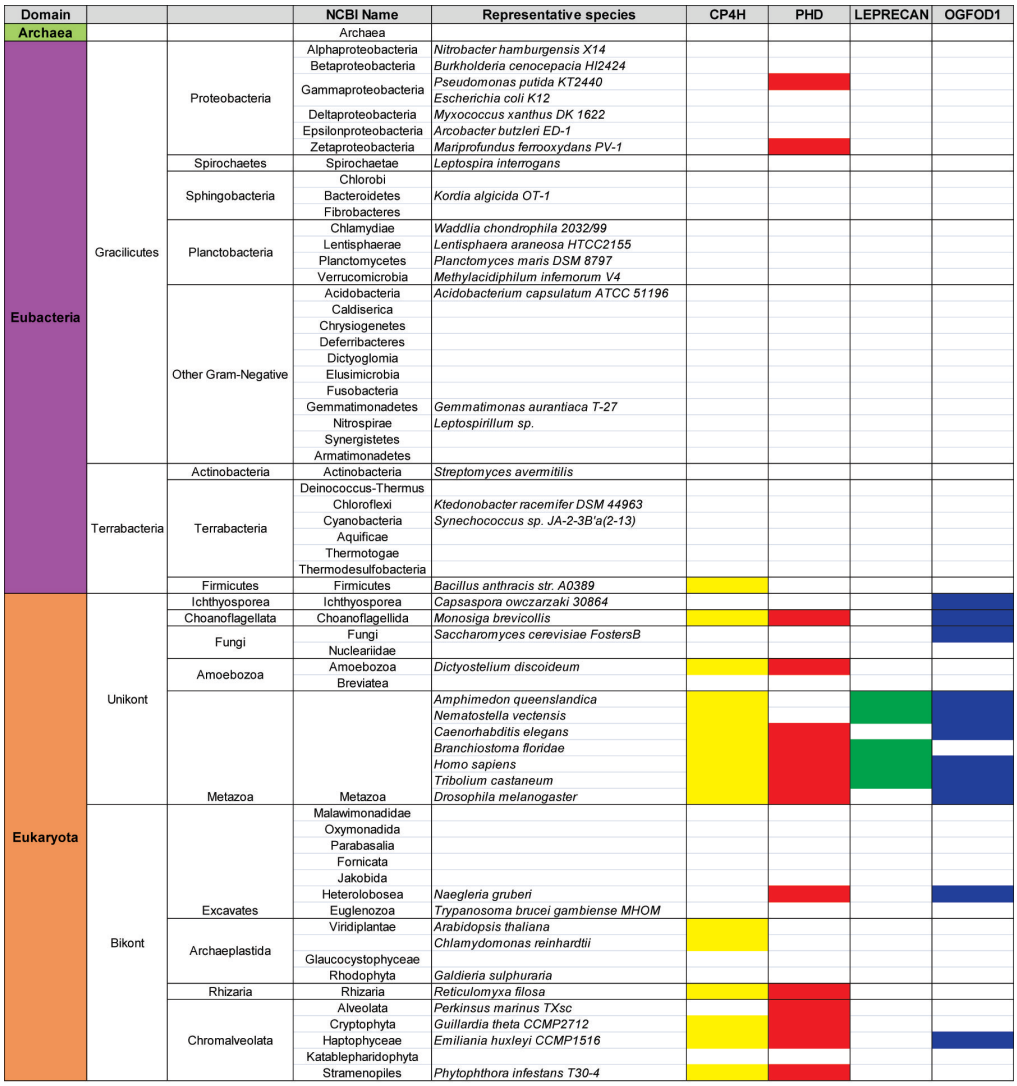


Figure 6.3. Prolyl-hydroxylases are sporadically distributed in Kingdoms of life (wider distribution). Tracing putative homologues of the CPHs, PHDs, Leprecans, and OGFOD1s in representative organisms throughout the Kingdoms of life. Bioinformatics analyses performed in collaboration with Mr. Michael Bentley (Plant Sciences, University of Oxford) and Dr. Jordi Paps (Zoology, University of Oxford).

The combined structural and bioinformatic data suggest two opposing theories regarding the evolution of prolyl-hydroxylases: i) prolyl-hydroxylases may have evolved in bacteria from earlier, small-molecule hydroxylases, such as proline hydroxylases (*8*), to catalyse the hydroxylation of conserved regions of abundant proteins, such as the switch I loop of EF-Tu, or ii) prolyl-hydroxylases

may have evolved in eukaryotes and emerged in various prokaryotic organisms via horizontal gene transfer.

In this regard it is notable that hydroxyproline residues are present in many small-molecule metabolites including α-amanitin and phalloidin, which are toxins produced by the *Amanita* genus of poisonous mushrooms that function by binding to RNA polymerase II and F-actin, respectively (Figure 6.4) (*9-12*). It is possible that proline residues are particularly prone to oxidation because of their over-represented presence in secondary structure elements often associated with functional and/or signalling roles, such as β-turns or the apex of loops (e.g. the location of Pro62 in RPS23 and Pro54 in EF-Tu).

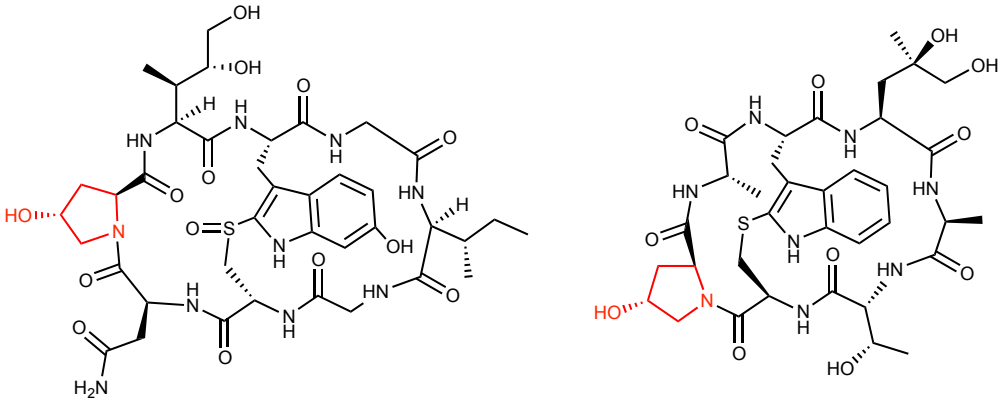


Figure 6.4. Chemical structures of hydroxyproline containing toxins α-amanitin (left) and phalloidin (right). The hydroxyproline moiety is shown in red.

Although the wider distribution of prolyl-hydroxylases in eukaryotes suggests that horizontal gene transfer from eukaryotes to prokaryotes may be likely, it is worth noting that eukaryotic prolyl-hydroxylases have been studied for decades, whereas work on prokaryotic prolyl-hydroxylases is still in its infancy. It is possible that future studies on prokaryotic prolyl-hydroxylases, led by those on PPHD, will uncover new roles and types of bacterial prolyl-hydroxylases and provide a more thorough answer as to their origins.

Irrespective of their ancestral roots, prolyl-hydroxylases likely evolved divergently in eukaryotes to catalyse the hydroxylation of a variety of substrates involved in both translational and transcriptional regulation, including RPS23, Skp1, collagen, and HIF-α (*13-16*). Their function in HIF-α hydroxylation emerged in animals, possibly via fusion of a bHLH domain containing transcription factor with an ancestral EF-Tu derived substrate. In this respect, the fact that PHD2 has been shown to interact with, but not hydroxylate, eEF2, the eukaryotic functional homologue of EF-G, an EF-Tu homologue (*17*), is of interest as it suggests an ancestral relationship between the PHDs/PPHDs and EF-Tu.

It is with hope that the research presented in this thesis will spur continued interest in prolyl-hydroxylases in all Kingdoms of life and in their constantly expanding biological roles, many of which likely extend beyond collagen stabilisation and/or hypoxic sensing in an isolated organism.

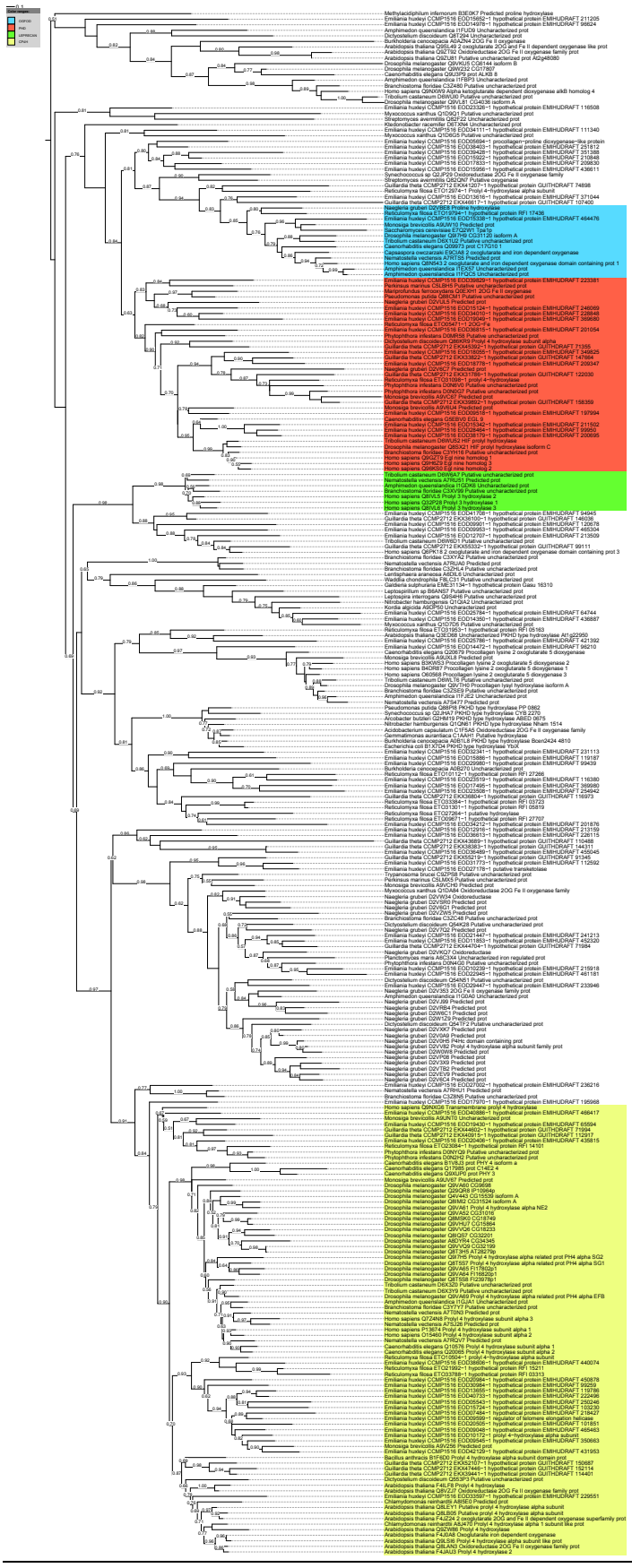


Figure 6.5. Phylogenetic inference of putative prolyl-hydroxylases from selected taxa from major taxonomic groups in life. FastTree inference of 278 protein 2OG-Fe(II) domain protein sequences under the WAG model of protein evolution reveals putative prolyl-hydroxylase clades corresponding to CP4H (yellow), PHD (red), Leprecan (green) and OGFOD1 (blue). Sequences were only considered in the OGFOD1 clade if they were also predicted to contain the Pfam domain Ofd1_CTDD (PF10637). SH support values of equal or greater than 50% are shown for each node. The scale bar represents 10% estimated sequence divergence. The tree is arbitrarily rooted in the Alpha-ketoglutarate-dependent dioxygenase (AlkB) group (*18, 19*), a putative outgroup, to ease reading of the tree. Bioinformatics analyses performed in collaboration with Mr. Michael Bentley (Plant Sciences, University of Oxford) and Dr. Jordi Paps (Zoology, University of Oxford). Figure and figure legend are adapted with permission from J.S. Scotti *et al*, Human oxygen sensing may have origins in prokaryotic elongation factor Tu prolyl-hydroxylation. *Proc Nat Acad Sci USA*. In press (2014).

6.3 Experimental Procedures

6.3.1 Identification of putative prolyl-hydroxylase orthologs, multiple-sequence alignment and phylogenetic tree construction

Using human prolyl-hydroxylase genes *EGLN1* (UniProt accession number Q9GZT9), *LEPRE1* (Q32P28), *OGFOD1* (Q8N543), and *P4HA1* (P13674) and lysyl-hydroxylase *PLOD1* (Q02809) as query sequences, Uniprot (http://www.uniprot.org/) and Genbank (https://www.ncbi.nlm.nih.gov/genbank/) were searched for putative prolyl-hydroxylase orthologs in the major taxonomic groups from the tree of life. Two conserved Pfam (*20*) domains, 2OG-FeII_Oxy (PF03171) and 2OG-FeII_Oxy_3 (PF13640), were identified among the queried sequences and the protein set containing these domains was downloaded from one representative species for each major taxonomic group across the Tree of Life using the Uniprot Batch tool. These results were validated by searching Genbank using Delta BLAST (default parameters) with the same query sequences (*21*). In general, there was agreement between Uniprot-identified and Genbank-identified sequences for most groups, but four additional phyla containing putative orthologous sequences were identified using Delta BLAST: Cryptophyta, Haptophyceae, Rhizaria and Rhodophyta. To isolate putative orthologous sequences from these groups, the proteomes from whole-genome sequenced organisms corresponding to each group were obtained: *Guillardia theta* (Genbank accession number PRJNA223305), *Emilianaia huxleyi* (PRJNA222302), *Reticulomyxa filosa* (PRJNA29155) and *Galdieria sulphuraria* (PRJNA221242). The Pfam Batch

search tool, using the default Gathering Threshold, was used to identify and download sequences containing the aforementioned domains.

The hidden-markov model (HMM) profiles were obtained for both domains, 2OG-FeII_Oxy (PF03171) and 2OG-FeII_Oxy_3 (PF13640), and HMMER3 hmmalign (http://hmmer.janelia.org/) was used to perform a domain alignment on the corresponding proteins sets. The domain-alignments were trimmed to include all domain-aligned amino acid residues in addition to 20aa up- and down-stream residues. To minimise homology misalignments, sequences not containing the two conserved histidine residues characteristic of the 2OG-Fe(II) oxygenase family members metal-binding HXD…H motif were removed (thus, we may have excluded 2OG oxygenases not containing these motifs). Both domain-alignments were then combined in a single FASTA file and all sequences were realigned using MAFFT (Parameters: L-INS-i, BLOSUM45, Gap opening penalty 2, Offset value 0.5, Unalign level 0.2)(*22*). To remove regions of ambiguous alignment, Trimal (*23*) was used to retain only those sites in which greater than or equal to 90% of the sequences contained an amino acid. Non-unique sequences were removed using ElimDupes (http://hcv.lanl.gov/content/sequence/ELIMDUPES/elimdupes.html). The resulting alignment contained 89 sites across 278 sequences.

The alignment was used to infer a gene tree using the program FastTree (*24*), under the WAG model of amino acid evolution. Branch support values were obtained with the Shimodaira-Hasegawa test (*25*). The tree was edited with iTOL (*26*) and arbitrarily rooted in the Alpha-ketoglutarate-dependent dioxygenase (AlkB) group to ease reading of the tree.

6.4 References

1. A. C. R. Epstein, J. M. Gleadle, L. A. McNeill, K. S. Hewitson, J. O'Rourke, D. R. Mole, M. Mukherji, E. Metzen, M. I. Wilson, A. Dhanda, Y. M. Tian, N. Masson, D. L. Hamilton, P. Jaakkola, R. Barstead, J. Hodgkin, P. H. Maxwell, C. W. Pugh, C. J. Schofield, P. J. Ratcliffe, C-elegans EGL-9 and mammalian homologs define a family of dioxygenases that regulate HIF by prolyl hydroxylation. *Cell.* **107**, 43-54 (2001).
2. W. G. Kaelin, P. J. Ratcliffe, Oxygen sensing by metazoans: The central role of the HIF hydroxylase pathway. *Molecular Cell.* **30**, 393-402 (2008).
3. L. Aravind, E. V. Koonin, The DNA-repair protein AlkB, EGL-9, and leprecan define new families of 2-oxoglutarate- and iron-dependent dioxygenases. *Genome Biol.* **2**, RESEARCH0007 (2001).
4. M. A. McDonough, V. Li, E. Flashman, R. Chowdhury, C. Mohr, B. M. Lienard, J. Zondlo, N. J. Oldham, I. J. Clifton, J. Lewis, L. A. McNeill, R. J. Kurzeja, K. S. Hewitson, E. Yang, S. Jordan, R. S. Syed, C. J. Schofield, Cellular oxygen sensing: Crystal structure of hypoxia-inducible factor prolyl hydroxylase (PHD2). *Proc Natl Acad Sci U S A.* **103**, 9814-9819 (2006).
5. N. R. Rose, M. A. McDonough, O. N. F. King, A. Kawamura, C. J. Schofield, Inhibition of 2-oxoglutarate dependent oxygenases. *Chemical Society Reviews.* **40**, 4364-4397 (2011).
6. N. Gupta, S. Tanner, N. Jaitly, J. N. Adkins, M. Lipton, R. Edwards, M. Romine, A. Osterman, V. Bafna, R. D. Smith, P. A. Pevzner, Whole proteome analysis of post-translational modifications: applications of mass-spectrometry for proteogenomic annotation. *Genome Res.* **17**, 1362-1377 (2007).
7. J. Myllyharju, Prolyl 4-hydroxylases, the key enzymes of collagen biosynthesis. *Matrix Biol.* **22**, 15-24 (2003).
8. R. P. Hausinger, Fe(II)/alpha-ketoglutarate-dependent hydroxylases and related enzymes. *Crit Rev Biochem Mol.* **39**, 21-68 (2004).
9. D. A. Bushnell, P. Cramer, R. D. Kornberg, Structural basis of transcription: alpha-Amanitin-RNA polymerase II cocrystal at 2.8 A resolution. *Proc Natl Acad Sci U S A.* **99**, 1218-1222 (2002).
10. F. Brueckner, P. Cramer, Structural basis of transcription inhibition by alpha-amanitin and implications for RNA polymerase II translocation. *Nature Structural & Molecular Biology.* **15**, 811-818 (2008).
11. J. Vandekerckhove, A. Deboben, M. Nassal, T. Wieland, The Phalloidin Binding-Site of F-Actin. *Embo Journal.* **4**, 2815-2818 (1985).
12. T. Oda, K. Namba, Y. Maeda, Position and orientation of phalloidin in F-actin determined by X-ray fiber diffraction analysis. *Biophys J.* **88**, 2727-2736 (2005).
13. R. S. Singleton, P. Liu-Yi, F. Formenti, W. Ge, R. Sekirnik, R. Fischer, J. Adam, P. J. Pollard, A. Wolf, A. Thalhammer, C. Loenarz, E. Flashman, A. Yamamoto, M. L. Coleman, B. M. Kessler, P. Wappner, C. J. Schofield, P. J. Ratcliffe, M. E. Cockman, OGFOD1 catalyzes prolyl hydroxylation of RPS23 and is involved in translation control and stress granule formation. *Proc Natl Acad Sci U S A.* **111**, 4031-4036 (2014).

14. C. Loenarz, R. Sekirnik, A. Thalhammer, W. Ge, E. Spivakovsky, M. M. Mackeen, M. A. McDonough, M. E. Cockman, B. M. Kessler, P. J. Ratcliffe, A. Wolf, C. J. Schofield, Hydroxylation of the eukaryotic ribosomal decoding center affects translational accuracy. *Proc Natl Acad Sci U S A.* **111**, 4019-4024 (2014).
15. M. J. Katz, J. M. Acevedo, C. Loenarz, D. Galagovsky, P. Liu-Yi, M. Perez-Pepe, A. Thalhammer, R. Sekirnik, W. Ge, M. Melani, M. G. Thomas, S. Simonetta, G. L. Boccaccio, C. J. Schofield, M. E. Cockman, P. J. Ratcliffe, P. Wappner, Sudestada1, a Drosophila ribosomal prolyl-hydroxylase required for mRNA translation, cell homeostasis, and organ growth. *Proc Natl Acad Sci U S A.* **111**, 4025-4030 (2014).
16. H. van der Wel, A. Ercan, C. M. West, The Skp1 prolyl hydroxylase from Dictyostelium is related to the hypoxia-inducible factor-alpha class of animal prolyl 4-hydroxylases. *Journal of Biological Chemistry.* **280**, 14645-14655 (2005).
17. A. Romero-Ruiz, L. Bautista, V. Navarro, A. Heras-Garvin, R. March-Diaz, A. Castellano, R. Gomez-Diaz, M. J. Castro, E. Berra, J. Lopez-Barneo, A. Pascual, Prolyl hydroxylase-dependent modulation of eukaryotic elongation factor 2 activity and protein translation under acute hypoxia. *J Biol Chem.* **287**, 9651-9658 (2012).
18. S. C. Trewick, T. F. Henshaw, R. P. Hausinger, T. Lindahl, B. Sedgwick, Oxidative demethylation by Escherichia coli AlkB directly reverts DNA base damage. *Nature.* **419**, 174-178 (2002).
19. P. O. Falnes, R. F. Johansen, E. Seeberg, AlkB-mediated oxidative demethylation reverses DNA damage in Escherichia coli. *Nature.* **419**, 178-182 (2002).
20. M. Punta, P. C. Coggill, R. Y. Eberhardt, J. Mistry, J. Tate, C. Boursnell, N. Pang, K. Forslund, G. Ceric, J. Clements, A. Heger, L. Holm, E. L. Sonnhammer, S. R. Eddy, A. Bateman, R. D. Finn, The Pfam protein families database. *Nucleic Acids Res.* **40**, D290-301 (2012).
21. G. M. Boratyn, A. A. Schaffer, R. Agarwala, S. F. Altschul, D. J. Lipman, T. L. Madden, Domain enhanced lookup time accelerated BLAST. *Biol Direct.* **7**, 12 (2012).
22. K. Katoh, D. M. Standley, MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* **30**, 772-780 (2013).
23. S. Capella-Gutierrez, J. M. Silla-Martinez, T. Gabaldon, trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics.* **25**, 1972-1973 (2009).
24. M. N. Price, P. S. Dehal, A. P. Arkin, FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS One.* **5**, e9490 (2010).
25. H. Shimodaira, M. Hasegawa, Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular biology and evolution.* **16**, 1114-1116 (1999).
26. I. Letunic, P. Bork, Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic acids research.* **39**, W475-W478 (2011).

Appendix A: FIH catalyses the hydroxylation of D-amino acids

A.1 FIH catalyses the hydroxylation of ankyrin repeat proteins

FIH plays a critical role in the human hypoxic response, where it catalyses the hydroxylation of a conserved asparagine residue (Asn803) in the *C*-terminal transactivation domain (CTAD) of HIF-1α, so preventing HIF-1α interaction with p300/CBP coactivator proteins (*1-5*). FIH, like the PHDs, is an Fe(II) and 2OG dependent dioxygenase, yet belongs to a 2OG oxygenase subfamily distinct from that of the PHDs or CP4Hs and has different catalytic properties (*5, 6*).

In addition to its role in the human hypoxic response, FIH was subsequently found to catalyse the hydroxylation of asparagine residues in ankyrin repeat (AR) domain (ARD) proteins (*7-10*). ARDs are predicted to be present in over 300 human proteins and consist of a variable number of approximately 33 residue repeats, which individually fold into antiparallel α-helices connected by a β-hairpin and can act as a protein interaction surface (*8, 11, 12*). Crystallographic studies of hydroxylated ARDs revealed no significant conformational changes compared to their unhydroxylated forms (*8*). Although the precise physiological role of ARD hydroxylation remains unclear, biophysical analyses revealed that asparaginyl β-hydroxylation was shown to stabilise the ARD fold (*13, 14*).

A.2. FIH catalyses the β-hydroxylation of L-amino acids other than asparagine

Further studies revealed that FIH catalyses the β-hydroxylation of residues other than asparagine, including aspartate in human cytoskeletal ARDs and histidine in human tankyrase-2 (*15, 16*). Using a consensus ankyrin peptide (CAP: **c**onsensus **a**nkyrin **p**eptide; sequence: H_2N-HLEVVKLLLEHGADV**N**AQDK-$CONH_2$) as a substrate (*13*), in which the asparagine residue (bold red) subject to FIH-catalysed modification was substituted with all 20 L-amino acids, FIH was demonstrated to catalyse modification of leucine, isoleucine, tryptophan, and serine residues, in addition to those already reported, namely asparagine, aspartate, and histidine residues (Table A.1)(*17*). These results led to the proposal that FIH catalysis displays a high degree of promiscuity with respect to side chain selectivity and suggested that it may be possible to engineer the FIH active site for tailored selectivities.

Table A.1. **Comparison of HIF-1α CTAD and CAP derived FIH substrates.** X can be substituted for residues other than asparagine, such as histidine, serine, tryptophan, leucine, and isoleucine (*17*).

Peptide Name	Peptide Sequence
HIF-1α CTAD (residues 788-806)	DESGLPQLTSYDCEV**N**API
CAP(X)	HLEVVKLLLEHGADV**X**AQDK

Crystal structures of FIH in complex with CAP(L-leucine) (PDB ID: 4B7E) and CAP(L-serine) (PDB ID: 4B7K) and of FIH in complex with CTAD (PDB ID: 1H2K) revealed that CAP and CTAD peptides exhibit similar overall binding modes, yet differ in the position of active site residues (Figure A.2)(*17, 18*). In all

available structures, the hydroxylated residue is positioned at the apex of a γ-turn. In the FIH:HIF-1α CTAD structure, the asparagine amide side chain is positioned to hydrogen bond to the side chain of Arg238; whereas in the FIH:CAP(L-leucine) and FIH:CAP(L-serine) structures the Arg238 side chain either pivots to provide space for the hydrophobic leucine or is too distant (4.2 Å) to hydrogen bond to the serine hydroxyl group (although electrostatics are still possible) (Figure A.2). The likely interactions with either the Arg238 side chain or the active site metal may possibly explain the observation that the serine side chain is positioned in two equal occupancy conformations (Figure A.2*C*).

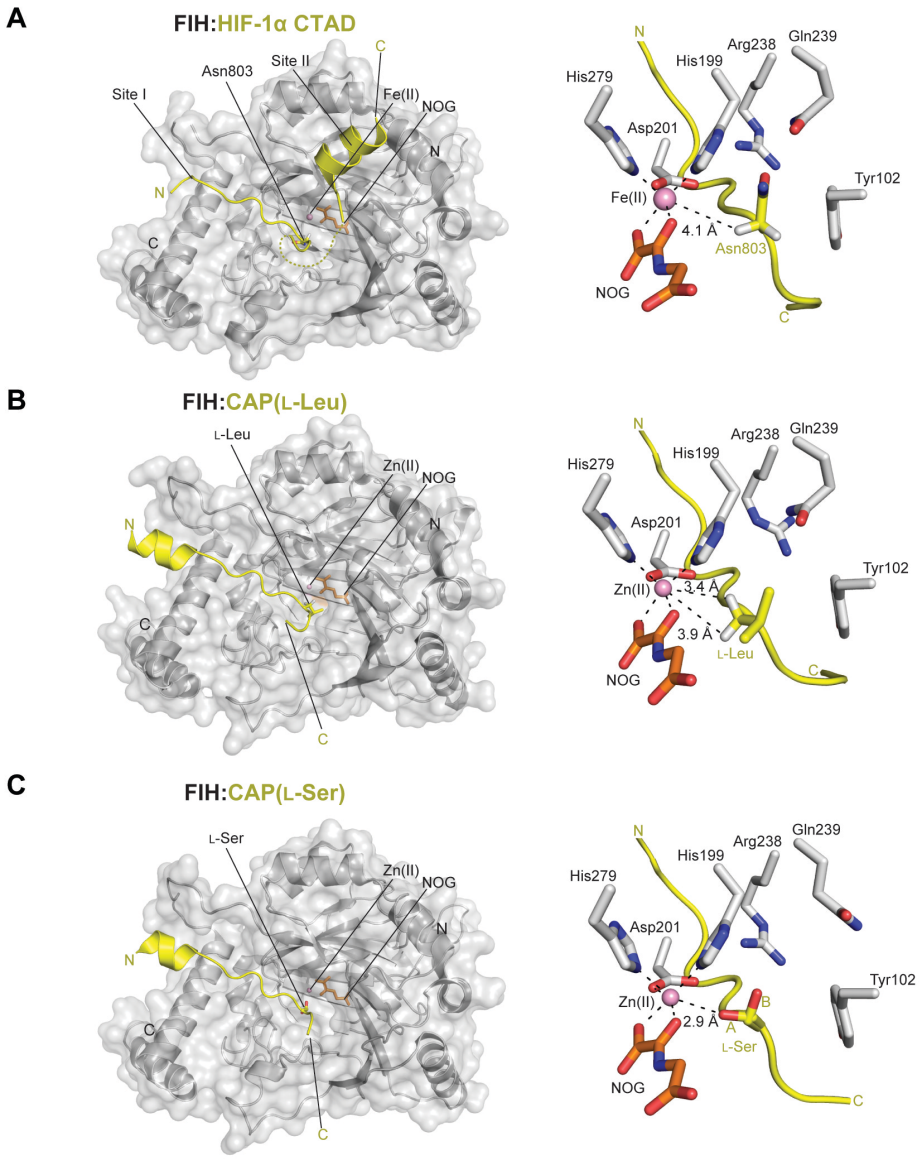


Figure A.2. Views from crystal structures of FIH in complex with CTAD and consensus ankyrin repeat peptide (CAP) reveals similar substrate binding modes. (**A**) A crystal structure of FIH in complex with HIF-1α CTAD reveals that the asparagine C_β pro-3*S* hydrogen is positioned 4.1 Å from the metal, poised for hydroxylation (PDB ID: 1H2K)(*18*). (**B**) A crystal structure of FIH in complex with CAP(L-leucine) reveals that the both leucine C_β pro-3*S* and pro-3*R* hydrogens are positioned close to the metal (3.9 Å and 3.3 Å, respectively) (PDB ID: 4B7E)(*17*). (**C**) A crystal structure of FIH in complex with CAP(L-serine) reveals that serine is positioned in two equal occupancy conformations. Conformation B positions the pro-3*S* hydrogen for hydroxylation, whereas conformation A positions the side chain hydroxyl for metal chelation (2.9 Å) (PDB ID: 4B7K)(*17*).

The FIH:HIF-1α CTAD structure reveals that the asparagine pro-3*S* hydrogen is positioned towards the metal (4.1 Å), consistent with formation of a

2*S*,3*S* β-hydroxyasparagine product (Figure A.2*A*)(*18*). Of the two serine side chain conformations, one (B) is positioned similarly to that of asparagine such that a pro-3*S* hydrogen is pointed towards the metal (2.9 Å)(Figure A.2*C*). In contrast, the positions of the two leucine C_β hydrogens are such that the stereochemical outcome of the reaction is difficult to predict; the distance from the metal to the leucine C_β pro-3*R* hydrogen and pro-3*S* hydrogen is 3.4 Å and 3.9 Å, respectively (Figure A.2*B*). Further analyses revealed that leucyl-hydroxylation results in the formation of the 2*S*,3*R* product (thus modification of the pro-3*R* hydrogen, 3.4 Å from the metal) and proceeds with the same relative stereochemistry as asparaginyl-hydroxylation (*17*). The results suggest that the FIH active site can accommodate hydrophobic residues, and raised the question of the extent of promiscuity in FIH-catalysed protein modification.

A.3 FIH catalyses the hydroxylation of D-amino acids

A.3.1 FIH-catalysed modification of D-leucine, D-histidine, and D-allylglycine residues

In an attempt to further probe the FIH substrate profile, the CAP hydroxylated residue (**X**) of all known L-residue substrates was substituted for unnatural D-amino acids. It was discovered that FIH catalyses the hydroxylation of D-leucine, D-histidine, and D-allylglycine residues (Dr. Adam Hardy, Chemistry, unpublished data). Interestingly, hydroxylation is not observed for L-allylglycine. Further, D-leucine was shown to undergo an unprecedented dihydroxylation (not observed with L-leucine) (Dr. Adam Hardy, Chemistry, unpublished data).

A.3.2 Crystallisation of FIH in complex with of D-amino acid containing CAP

It was of interest to determine crystal structures of FIH in complex with D-amino acids in order to probe the structural basis for the observed differential reactivity. Purified, recombinant FIH was obtained from Dr. Wei Ge (Chemistry) or Dr. Rashed Chowdhury (Chemistry). To the purified FIH (12 mg/mL) was added ZnOAc (1 mM), NOG (2 mM), and the CAP peptide (5 mM) containing either D-leucine, D-β-(2*R*,3*S*)-hydroxyleucine (the predicted product of D-leucine monohydroxylation), D-histidine, or D-allylglycine residues [all peptides synthesised by Dr. Wei Ge (Chemistry), except for the D-β-(2*R*,3*S*)-hydroxyleucine peptide provided by Dr. Hwanho Choi (Chemistry)], and crystallisation trials were performed using an FIH-specific optimisation screen (Dr. Rashed Chowdhury, Chemistry). Final crystal conditions for all structures consist of 1.6 M $(NH_4)_2SO_4$, 6% (v/v) PEG 400, and 0.1 M HEPES pH 7.5.

Crystals for FIH in complex with of D-leucine, D-β-(2*R*,3*S*)-hydroxyleucine, and D-allylglycine CAP were harvested and a complete data set was collected for each crystal to 2.4 Å, 2.2 Å, and 2.7 Å, respectively. The structures were solved by molecular replacement using the MR-PHASER (*19*) subroutine of the PHENIX (*20*) crystallographic software package using the FIH:HIF-1α CTAD complex (PDB ID: 1H2K)(*18*) as a search model (note that CTAD peptide was manually removed prior to molecular replacement to eliminate model bias).

A.3.3 Overall structures of FIH in complex with D-leucine, D-β-(2R,3S)-hydroxyleucine, and D-allylglycine CAP

The overall structures of FIH in complex with D-leucine, D-β-(2*R*,3*S*)-hydroxyleucine, and D-allylglycine CAP revealed that the peptide main chains

bound to FIH in a similar conformation (Figure A3). Importantly, in the case of the FIH:CAP(D-β-(2*R*,3*S*)-hydroxyleucine) structure, the 2.2 Å resolution dataset was sufficient to clearly view the β-hydroxyl group of the active site D-leucine within the electron density map (Figure A.3*B*).

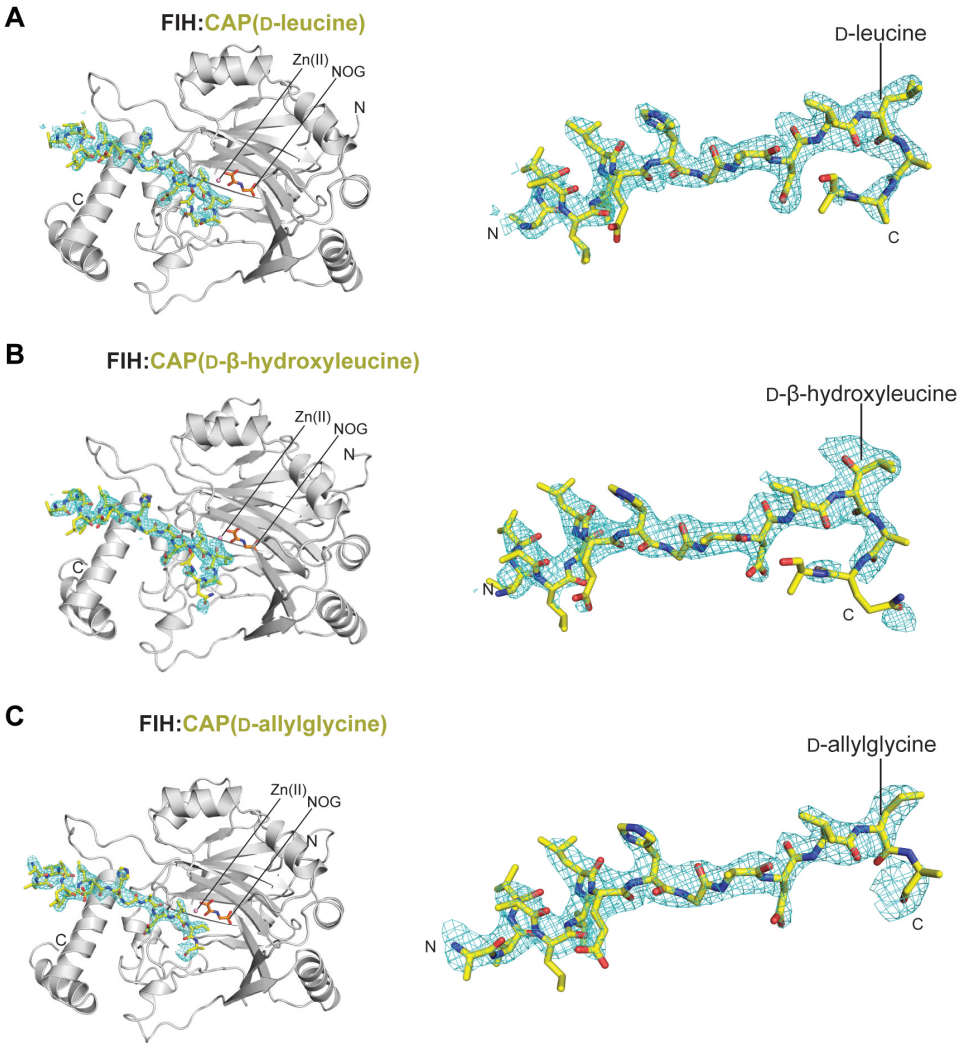


Figure A.3. Views from crystal structures of FIH in complex with CAP peptide containing an unnatural D-amino acid. The overall structure of FIH in complex with (**A**) CAP(D-leucine), (**B**) CAP(D-β-(2*R*,3*S*)-hydroxyleucine), and (**C**) CAP(D-allylglycine); the electron density OMIT $|mF_o - DF_c|$ (shown in cyan mesh) are contoured to 3.0 σ.

A.3.4 Structural insights into FIH-catalysed D-allylglycine hydroxylation

The observation that FIH catalyses the modification of D-, and not L-allylglycine was of interest from a mechanistic perspective. Inspection of the conformation of the D-allylglycine residue within the FIH reveals that the C_β pro-3*S* hydrogen is positioned 3.2 Å from the metal (Figure A.4). Oxidation at the C_β pro-3*S* hydrogen would result in a *trans* allylic alcohol.

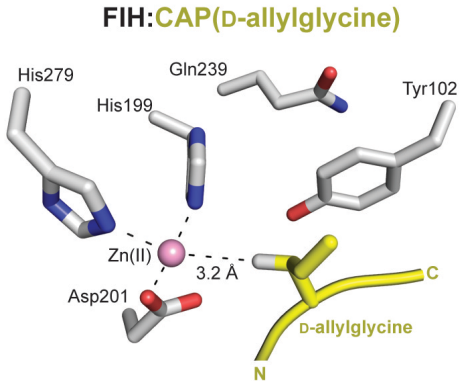


Figure A.4. View from a crystal structure of FIH in complex with CAP(D-allylglycine).

In addition, the structure also reveals that the D-allylglycine allyl group is positioned away from Tyr102. Although it is difficult to predict the origin of FIH D- vs L-allylglycine specificity based on the structural data alone, it is possible that L-allylglycine would adopt an unfavourable conformation in the FIH active site, in which the allyl group is positioned facing the opposite direction compared to that of D-allylglycine, resulting in a steric clash with the side chain of Tyr102.

A.3.5 Structural insights into FIH-catalysed L/D-leucine hydroxylation

Comparison of the crystal structures of FIH in complex with CAP(L-leucine), CAP(D-leucine), and CAP(D-β-(2*R*,3*S*)-hydroxyleucine) provides insight into the structural basis underlying the observation that FIH catalyses

dihydroxylation of D-leucine, yet monohydroxylation of L-leucine (Figure A.5). In the case of L-leucine, hydroxylation proceeds via the C_β pro-3*R* hydrogen (3.4 Å from metal)(Figure A.5*A*). In contrast, D-leucine occupies an inverted position in the active site such that hydroxylation likely proceeds via the C_β pro-3*S* hydrogen (3.3 Å from metal; stereochemistry has yet to be fully confirmed by amino acid analysis)(Figure A5*B*). The structure of FIH in complex with D-β-(2*R*,3*S)*-hydroxyleucine (the likely product of D-leucine monohydroxylation) reveals that the β-3*S*-hydroxyl group is positioned to hydrogen bond to the carboxylate of Asp199 (2.6 Å), the other oxygen of which is positioned to chelate active site metal (Figure A.5*C*). This hydrogen bond likely stabilises the conformation of D-β-(2*R*,3*S)*-hydroxyleucine in the active site, in which the side chain isopropyl group is rotated 180° relative to that observed in the L- and D-leucine structures, so positioning the tertiary C_γ hydrogen 3.4 Å from the active site metal for likely hydroxylation (Figure A.5*C*)(the next closest hydrogen is derived from the isopropyl-associated methyl group and is positioned 4.5 Å from metal; conformation of the stereochemistry of the second hydroxylation will require additional experimental evidence).

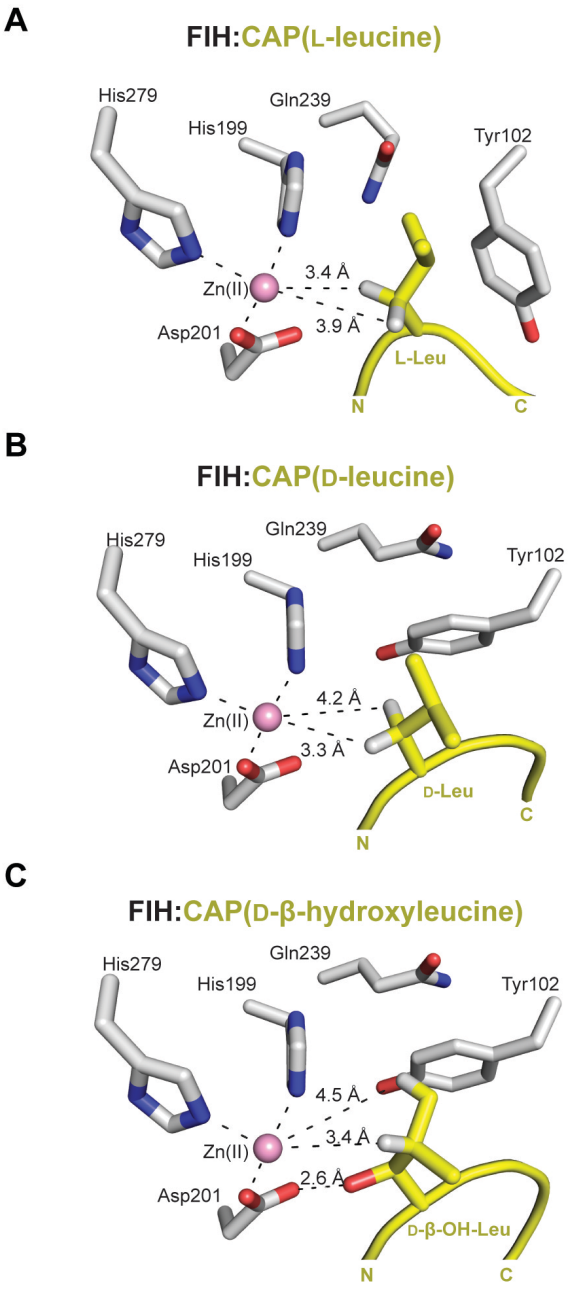


Figure A.5. Views from crystal structures of FIH in complex with CAP peptide containing L-leucine, D-leucine, or D-β-(2*R*,3*S*)-hydroxyleucine provides insights into the structural basis of D-leucine dihydroxylation. Views of the FIH active site in complex with (**A**) L-leucine, (**B**) D-leucine, and (**C**) D-β-(2*R*,3*S*)-hydroxyleucine.

Given the structural evidence, two opposing explanations for the observed absence of FIH-catalysed L-leucine dihydroxylation are possible: i) the conformation of the isopropyl group of the L-leucine hydroxylation product, L-β-

(2*S*,3*R*)-hydroxyleucine, likely inverted relative to that of D-β-(2*R*,3*S)*-hydroxyleucine, would preclude formation of an analogous hydrogen bond to the Asp199 carboxylate, thus sufficiently destabilising binding to FIH, or ii) the conformation of the main chain residues *C*-terminal to the hydroxylated residue may obstruct binding of an L-β-(2*S*,3*R*)-hydroxyleucine containing peptide. Indeed, all available structures of FIH in complex with D-amino acids show a characteristic movement of the Tyr102 and Gln239 side chains relative to their positions in L-amino acids structures (Figure A2 and Figure A5). In D-amino acid structures, the peptide main chain occupies a position further into the FIH active site, necessitating the rotation of the Tyr102 side chain in order to prevent a steric clash (Figure A2 and Figure A5). This rotation preorganises the Tyr102 side chain for π-stacking with the Gln239 side chain amide (Figure A2 and Figure A5). It is possible that this conformation preferentially stabilises the binding of D-amino acid containing peptides in the FIH active site.

At this stage, there is no evidence that the hydroxylation of D-amino acids as catalysed by FIH is of any biological relevance. However, it does suggest the possibility that FIH, or various other oxygenases, may catalyse oxidative epimerisation, which may in turn regulate expression (e.g. the epimerisation of an L- to D-amino acid in a transcription factor regulates gene expression) in a similar way to characterised HIF-CTAD hydroxylation.

A.4 Experimental Procedures

A.4.1 X-ray crystallography

Crystals of FIH in complex with Zn(II), NOG and CAP peptides were grown in sitting drops using the vapour diffusion method (drop size: 200-300 nL) at 293 K in 96-well Intelliplates (Art Robbins). Crystals were cryo-protected by transfer to 25% (v/v) glycerol in well solution and then harvested in nylon loops (Hampton Research) and cryo-cooled by plunging in liquid nitrogen. Data were collected at 100 K using single crystals at Diamond Light Source beamline I04-1 [FIH:CAP(D-leucine) and FIH:CAP(D-allylglycine)] with a MarMosaic 300 mm CCD detector, and Diamond Light Source beamline I04 [FIH:CAP(D-β-hydroxyleucine)] with a Pilatus 6M-F detector. Data were then indexed, integrated, and scaled using SCALA [FIH:CAP(D-leucine) and FIH:CAP(D-allylglycine)](*21*) and HKL3000 [FIH:CAP(D-β-hydroxyleucine)](*22*). All structures were determined by molecular replacement (MR) using the MR-PHASER (*19*) subroutine of PHENIX (*20*) using *H. sapiens* FIH (PDB ID: 1H2K)(*18*) as the search model. Model building and refinement were performed iteratively using COOT (*23*) and PHENIX until converging *R* and *R*free no longer decreased. Mn(II), NOG, and CAP were modelled in the final stages of refinement based on the $F_{obs} - F_{calc}$ electron density map.

Table A.6. Crystallographic data and refinement statistics.

	FIH:CAP(D-leucine)	FIH:CAP(D-allylglycine)	FIH:CAP(D-β-hydroxyleucine)
X-ray source	Diamond Light Source beamline I04-1	Diamond Light Source beamline I04-1	Diamond Light Source beamline I04
Wavelength (Å)	0.91730	0.91730	0.97949
PDB Acquisition Code	4JAA	4NR1	-
Resolution (Å)	2.39 (2.55–2.39)§	2.68 (2.75–2.68)§	2.20 (2.28-2.20)§
Space group	$P\,4_1\,2_1\,2$	$P\,4_1\,2_1\,2$	$P\,4_1\,2_1\,2$
Unit Cell Dimensions (*a* Å, *b* Å, *c* Å)	86.08, 86.08, 147.03	86.37, 86.37, 148.72	86.42, 86.42, 146.70
Molecules per a.u.	1	1	1
Total Number of Reflections Observed	112791	164722	640987
Number of Unique Reflections	22451 (3937)§	16485 (1178)§	29287 (2878)§
Redundancy	5.0 (4.6)§	10.0 (10.8)§	21.9 (21.1)§
Completeness (%)	99.3 (98.7)§	99.9 (99.9)§	100.0 (100.0)§
Wilson B	44.6	67.0	53.9
I/σ(I)	9.3 (2.6)§	17.3 (2.9)§	17.9 (2.1)§
§§R_{merge}	0.089	0.100	0.184
*R_{cryst}	0.184	0.164	0.213 (unrefined)
†R_{free}	0.219	0.210	0.247 (unrefined)
¶RMS deviation	0.01 (1.1°)	0.01 (1.2°)	(unrefined)
Average *B* factors (Å²)	41.7	63.4	(unrefined)
Number of Water Molecules	120	54	(unrefined)

§ Parentheses indicate high resolution shell

§§ $R_{merge} = \sum_j \sum_h | I_{hj} - \langle I_h \rangle | / \sum_j \sum_h \langle I_h \rangle \times 100$

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

† R_{free}, based on 2-5% of the total reflections

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

A.5 References

1. N. L. Sang, J. Fang, V. Srinivas, I. Leshchinsky, J. Caro, Carboxyl-terminal transactivation activity of hypoxia-inducible factor 1 alpha is governed by a von Hippel-Lindau protein-independent, hydroxylation-regulated association with p300/CBP. *Molecular and Cellular Biology.* **22**, 2984-2992 (2002).
2. P. C. Mahon, K. Hirota, G. L. Semenza, FIH-1: a novel protein that interacts with HIF-1 alpha and VHL to mediate repression of HIF-1 transcriptional activity. *Gene Dev.* **15**, 2675-2686 (2001).
3. D. Lando, D. J. Peet, J. J. Gorman, D. A. Whelan, M. L. Whitelaw, R. K. Bruick, FIH-1 is an asparaginyl hydroxylase enzyme that regulates the transcriptional activity of hypoxia-inducible factor. *Gene Dev.* **16**, 1466-1471 (2002).
4. L. A. McNeill, K. S. Hewitson, T. D. Claridge, J. F. Seibel, L. E. Horsfall, C. J. Schofield, Hypoxia-inducible factor asparaginyl hydroxylase (FIH-1) catalyses hydroxylation at the beta-carbon of asparagine-803. *Biochemical Journal.* **367**, 571-575 (2002).
5. W. G. Kaelin, P. J. Ratcliffe, Oxygen sensing by metazoans: The central role of the HIF hydroxylase pathway. *Molecular Cell.* **30**, 393-402 (2008).
6. P. Koivunen, M. Hirsila, V. Gunzler, K. I. Kivirikko, J. Myllyharju, Catalytic properties of the asparaginyl hydroxylase (FIH) in the oxygen sensing pathway are distinct from those of its prolyl 4-hydroxylases. *Journal of Biological Chemistry.* **279**, 9899-9904 (2004).
7. M. E. Cockman, D. E. Lancaster, I. P. Stolze, K. S. Hewitson, M. A. McDonough, M. L. Coleman, C. H. Coles, X. Yu, R. T. Hay, S. C. Ley, C. W. Pugh, N. J. Oldham, N. Masson, C. J. Schofield, P. J. Ratcliffe, Posttranslational hydroxylation of ankyrin repeats in IkappaB proteins by the hypoxia-inducible factor (HIF) asparaginyl hydroxylase, factor inhibiting HIF (FIH). *Proc Natl Acad Sci U S A.* **103**, 14767-14772 (2006).
8. M. L. Coleman, M. A. McDonough, K. S. Hewitson, C. Coles, J. Mecinovic, M. Edelmann, K. M. Cook, M. E. Cockman, D. E. Lancaster, B. M. Kessler, N. J. Oldham, P. J. Ratcliffe, C. J. Schofield, Asparaginyl hydroxylation of the Notch ankyrin repeat domain by factor inhibiting hypoxia-inducible factor. *J Biol Chem.* **282**, 24027-24038 (2007).
9. M. E. Cockman, J. D. Webb, H. B. Kramer, B. M. Kessler, P. J. Ratcliffe, Proteomics-based Identification of Novel Factor Inhibiting Hypoxia-inducible Factor (FIH) Substrates Indicates Widespread Asparaginyl Hydroxylation of Ankyrin Repeat Domain-containing Proteins. *Mol Cell Proteomics.* **8**, 535-546 (2009).
10. M. E. Cockman, J. D. Webb, P. J. Ratcliffe, FIH-Dependent Asparaginyl Hydroxylation of Ankyrin Repeat Domain-Containing Proteins. *Ann Ny Acad Sci.* **1177**, 9-18 (2009).
11. J. Schultz, F. Milpetz, P. Bork, C. P. Ponting, SMART, a simple modular architecture research tool: identification of signaling domains. *Proc Natl Acad Sci U S A.* **95**, 5857-5864 (1998).
12. L. K. Mosavi, T. J. Cammett, D. C. Desrosiers, Z. Y. Peng, The ankyrin repeat as molecular architecture for protein recognition. *Protein Science.* **13**, 1435-1448 (2004).

13. L. Kelly, M. A. McDonough, M. L. Coleman, P. J. Ratcliffe, C. J. Schofield, Asparagine beta-hydroxylation stabilizes the ankyrin repeat domain fold. *Mol Biosyst.* **5**, 52-58 (2009).
14. A. P. Hardy, I. Prokes, L. Kelly, I. D. Campbell, C. J. Schofield, Asparaginyl beta-hydroxylation of proteins containing ankyrin repeat domains influences their stability and function. *J Mol Biol.* **392**, 994-1006 (2009).
15. M. Yang, R. Chowdhury, W. Ge, R. B. Hamed, M. A. McDonough, T. D. W. Claridge, B. M. Kessler, M. E. Cockman, P. J. Ratcliffe, C. J. Schofield, Factor-inhibiting hypoxia-inducible factor (FIH) catalyses the post-translational hydroxylation of histidinyl residues within ankyrin repeat domains. *Febs J.* **278**, 1086-1097 (2011).
16. M. Yang, W. Ge, R. Chowdhury, T. D. W. Claridge, H. B. Kramer, B. Schmierer, M. A. McDonough, L. Z. Gong, B. M. Kessler, P. J. Ratcliffe, M. L. Coleman, C. J. Schofield, Asparagine and Aspartate Hydroxylation of the Cytoskeletal Ankyrin Family Is Catalyzed by Factor-inhibiting Hypoxia-inducible Factor. *Journal of Biological Chemistry.* **286**, 7648-7660 (2011).
17. M. Yang, A. P. Hardy, R. Chowdhury, N. D. Loik, J. S. Scotti, J. S. O. McCullagh, T. D. W. Claridge, M. A. McDonough, W. Ge, C. J. Schofield, Substrate Selectivity Analyses of Factor Inhibiting Hypoxia-Inducible Factor. *Angew Chem Int Edit.* **52**, 1700-1704 (2013).
18. J. M. Elkins, K. S. Hewitson, L. A. McNeill, J. F. Seibel, I. Schlemminger, C. W. Pugh, P. J. Ratcliffe, C. J. Schofield, Structure of factor-inhibiting hypoxia-inducible factor (HIF) reveals mechanism of oxidative modification of HIF-1 alpha. *J Biol Chem.* **278**, 1802-1806 (2003).
19. A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, R. J. Read, Phaser crystallographic software. *J Appl Crystallogr.* **40**, 658-674 (2007).
20. P. D. Adams, P. V. Afonine, G. Bunkoczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, P. H. Zwart, PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr.* **66**, 213-221 (2010).
21. M. D. Winn, C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin, K. S. Wilson, Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr.* **67**, 235-242 (2011).
22. Z. Otwinowski, W. Minor, Processing of X-ray diffraction data collected in oscillation mode. *Method Enzymol.* **276**, 307-326 (1997).
23. P. Emsley, K. Cowtan, Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr.* **60**, 2126-2132 (2004).

Appendix B: Standard Operating Procedures

B.1 pH measurements

The pH of solutions was measured using a Jenway pH meter 3305 coupled to a glass/calomel electrode (Aldrich). The pH meter was calibrated with phthalate pH 4.0, phosphate pH 7.0, and borate pH 10.0 (Fisher Scientific), immediately prior to use. NaOH or HCl were added to adjust the pH of buffer solutions as necessary.

B.2 Agarose gel electrophoresis

DNA was analysed by agarose gel electrophoresis on a 1% (w/v) agarose gel containing SYBR Safe DNA gel stain (Invitrogen). Samples were diluted in 5x DNA loading buffer (Table B.1 and Table B.2) prior to loading on the gel and GeneRuler 1 kb DNA ladder (Fermentas) was used as a reference. The gel was run in 1x TAE buffer at a constant 90 V (BioRad).

Table B.1. 50x TAE buffer recipe.

Reagent	Per 1 L
Tris	242 g
Glacial acetic acid	57 mL
0.5 M EDTA pH 8.0	100 mL

Table B.2. 5x DNA loading buffer recipe.

Reagent	Per 20 mL
Glycerol	6 mL
Bromophenol blue	50 mg
Xylene cyanol FF	50 mg
MilliQ-H_2O	14 mL

B.3 Polymerase Chain Reaction (PCR)

PCR was used to amplify the DNA sequence of interest for use in various downstream biochemical experiments (Table B.3). *Pseudomonas putida KT2440* genomic DNA (a generous gift from Dr. Gail Preston, Plant Sciences, University of Oxford) was used as a template.

Table B.3. Representative PCR reaction.

Reagent	Stock	Volume added (µl)
10x Pfu buffer (Stratagene)	10x	5
Forward primer (Sigma)	125 ng/µL	1
Reverse primer (Sigma)	125 ng/µL	1
DNA template	50 ng/µL	1
dNTPs (Stratagene)	100mM (25mM each)	1
Pfu turbo (Stratagene)	2.5 U/µL	1
MilliQ-H_2O		40
Total		50

50 µL PCR reactions were run in a TC-312 thermocycler (TECHNE) according to the standard conditions in Table B.4.

Table B.4. PCR thermocycler conditions.

Segment	Cycles	Temperature (°C)	Time
Initial denaturation	1	95	5 min
Denaturation		95	40 s
Annealing	37	55	1 min
Extension		72	1 min 30 s
Final extension	1	72	7 min
Final hold		4	~

A sample (3 µL) of each reaction was analysed by 1% TAE-agarose gel electrophoresis for the presence of the amplified DNA. DNA was isolated by spin column purification using a PCR Purification Kit (Fermentas), or using a GeneJET Gel Extraction Kit (Fermentas) according to the manufacturer's specifications (http://www.thermoscientificbio.com/uploadedFiles/Resources/k069-product-information.pdf). Purified DNA was stored at -20 °C.

B.4 Restriction enzyme digests

DNA digests were performed using commercially available restriction enzymes (New England Biolabs and Promega) under the following recommended conditions (Table B.5) by incubation at 37 °C for between 10 min and 4 h.

Table B.5. Restriction enzyme digestions.

Reagent	Stock	Volume added (µL)
New England Biolabs buffer #(1-4)	10x	5
Purified vector or insert DNA, up to 10 µg	~50-300 ng/µl	1-42
Restriction enzyme A	20,000 U/mL	1
Restriction enzyme B	20,000 U/mL	1
BSA (New England Biolabs)	50x	1
MilliQ-H_2O		Up to 50

High-fidelity restriction enzymes (New England Biolabs) and NEB (New England Biolabs) buffer 4 were used whenever possible. Reactions were analysed for digested DNA by 1% agarose gel electrophoresis, and digested DNA fragments were isolated by spin column purification using a GeneJET Gel Extraction Kit (Fermentas).

B.5 DNA ligations

Purified, double digested DNA fragments bearing sticky overhangs (2–4 nucleotides) were ligated using T4 DNA ligase (New England Biolabs). Ligation reactions typically contained ~50 ng purified vector backbone, a 3-fold molar excess of purified insert to vector, 2 µL of 10× ligation buffer (New England Biolabs), and 1 µL of T4 DNA ligase in a total volume of 20 µL. Ligations were incubated for 1 h at room temperature or overnight at 16 °C. Self-ligation controls were performed without insert DNA.

B.6 Transformations

Transformations into XL10-Gold Ultracompetent Cells (Agilent) or BL21(DE3) Competent Cells (Agilent) were performed according to the following steps: i) 1 µL of plasmid DNA (up to 5 µl of a ligation mixture) and 50 µL of XL10/BL21(DE3) cells were incubated in a 15 mL Falcon polypropylene tube for 15 mins on ice, ii) the mixture was then subjected to a 30 s heat shock at 42 °C in a water bath, iii) the mixture was incubated on ice for an additional 2 mins, iv) LB medium (450 µL) was added and the cells were allowed to grow at 37 °C for 1 hr, v) cells (100 µL) were selected on LB plates with the relevant antibiotic (100 µg/mL ampicillin or 30 µg/mL kanamycin) and growth at 37 °C overnight.

B.7 Colony PCR

Colonies were analysed for the presence of inserts by colony PCR using 2x Biomix Red (Bioline) using the same thermocycler conditions already described (Table B.4), followed by 1% agarose gel electrophoresis. Plasmid DNA of successful recombinants was isolated by Miniprep (Fermentas) and subsequently verified by Sanger sequencing (SourceBioscience) using the T7 promoter and terminator primer, or the M13 promoter and terminator primer, as appropriate, and stored at -20 °C.

B.8 Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

SDS-PAGE was used as the primary method for analysing the purity of protein purifications. SDS-PAGE separating gels were prepared according to the recipe in Table A6. Standard protocol involved up to 12 µL of a column fraction to be analysed diluted to 15 µL with 5x SDS-PAGE loading buffer (250 mM Tris pH 6.8, 5% β-mercaptoethanol, 30% glycerol, 5% SDS, 0.02% (w/v)

bromophenol blue). Boiling at 100 °C for 5 min in a hot block denatures the protein and a subsequent quick centrifugation step was found to increase resolution of protein bands on a gel due to limiting the amount of aggregate present that could hinder proper separation of proteins.

Table B.6. SDS-PAGE gel composition.

Reagent	Separating gel		Stacking gel	
	2 gels (mL)	4 gels (mL)	2 gels (mL)	4 gels (mL)
MilliQ-H_2O	0.79	1.6	2.3	4.6
1.5 M Tris-HCl pH 8.8	6.0	12		
0.5 M Tris-HCl pH 6.8			0.40	0.80
10% SDS	0.12	0.24	0.032	0.064
10% ammonium persulfate	0.20	0.40	0.060	0.12
30% acrylamide	5.0	10.	0.32	0.64
TMED	0.020	0.040	0.010	0.020

The gel was run at 190 V in 1x running buffer (25 mM Tris pH 8.3, 192 mM glycine, 0.1% SDS) until the smallest band in the ladder was visibly approaching the bottom. The gel was washed with water and stained for 10 min with shaking in stain buffer (0.25% (w/v) Coomassie Brilliant Blue, 10% acetic acid, 50% methanol in MilliQ-H_2O). The gel was destained using 10% acetic acid, 50% methanol in MilliQ-H_2O.

B.9 Growth media

Typically, growth media were prepared according to the following recipes (Table A7 and Table A8) and autoclaved by Mr. Herminio Manso Jubier.

Tables B.7 and B.8. Bacterial growth media compositions.

2x Tryptone / Yeast extract (2TY) medium	Per liter (g)
Bacto Tryptone	16
Yeast Extract	10
NaCl	5

Luria Broth (LB) medium	Per liter (g)
Bacto Tryptone	10
Yeast Extract	5
NaCl	10

B.10 Glycerol stock preparation

When desired, bacterial cultures were reserved for further use by the addition of 300 µL glycerol to 700 µL cell culture with gentle mixing. The cultures were frozen on dry ice and stored at -80 °C.

NATURE PUBLISHING GROUP LICENSE TERMS AND CONDITIONS

Aug 23, 2014

This is a License Agreement between John S Scotti ("You") and Nature Publishing Group ("Nature Publishing Group") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Nature Publishing Group, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number	3454731326112
License date	Aug 23, 2014
Licensed content publisher	Nature Publishing Group
Licensed content publication	Nature
Licensed content title	What recent ribosome structures have revealed about the mechanism of translation
Licensed content author	T. Martin Schmeing and V. Ramakrishnan
Licensed content date	Oct 18, 2009
Volume number	461
Issue number	7268
Type of Use	reuse in a dissertation / thesis
Requestor type	academic/educational
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables /illustrations	1
High-res required	no
Figures	Thesis figure 2.15 corresponds to Nature Schmeing et al 2009 Figure 2
Author of this NPG article	no
Your reference number	None
Title of your thesis / dissertation	Structural, Functional and Evolutionary Studies on Prolyl-Hydroxylases
Expected completion date	Sep 2014
Estimated size (number of pages)	300

Total 0.00 USD

Terms and Conditions

Terms and Conditions for Permissions

Nature Publishing Group hereby grants you a non-exclusive license to reproduce this material for this purpose, and for no other use,subject to the conditions below:

1. NPG warrants that it has, to the best of its knowledge, the rights to license reuse of this material. However, you should ensure that the material you are requesting is original to Nature Publishing Group and does not carry the copyright of another entity (as credited in the published version). If the credit line on any part of the material you have requested indicates that it was reprinted or adapted by NPG with permission from another source, then you should also seek permission from that source to reuse the material.

2. Permission granted free of charge for material in print is also usually granted for any electronic version of that work, provided that the material is incidental to the work as a whole and that the electronic version is essentially equivalent to, or substitutes for, the print version.Where print permission has been granted for a fee, separate permission must be obtained for any additional, electronic re-use (unless, as in the case of a full paper, this has already been accounted for during your initial request in the calculation of a print run).NB: In all cases, web-based use of full-text articles must be authorized separately through the 'Use on a Web Site' option when requesting permission.

3. Permission granted for a first edition does not apply to second and subsequent editions and for editions in other languages (except for signatories to the STM Permissions Guidelines, or where the first edition permission was granted for free).

4. Nature Publishing Group's permission must be acknowledged next to the figure, table or abstract in print. In electronic form, this acknowledgement must be visible at the same time as the figure/table/abstract, and must be hyperlinked to the journal's homepage.

5. The credit line should read:
   Reprinted by permission from Macmillan Publishers Ltd: [JOURNAL NAME] (reference citation), copyright (year of publication)
   For AOP papers, the credit line should read:
   Reprinted by permission from Macmillan Publishers Ltd: [JOURNAL NAME], advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].XXXXX)

   Note: For republication from the *British Journal of Cancer,* the following credit lines apply.
   Reprinted by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: [JOURNAL NAME] (reference citation), copyright (year of publication)For AOP papers, the credit line should read:
   Reprinted by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: [JOURNAL NAME], advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].XXXXX)

6. Adaptations of single figures do not require NPG approval. However, the adaptation should be credited as follows:

Adapted by permission from Macmillan Publishers Ltd: [JOURNAL NAME] (reference citation), copyright (year of publication)

Note: For adaptation from the *British Journal of Cancer,* the following credit line applies.
Adapted by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: [JOURNAL NAME] (reference citation), copyright (year of publication)

7. Translations of 401 words up to a whole article require NPG approval. Please visit http://www.macmillanmedicalcommunications.com for more information.Translations of up to a 400 words do not require NPG approval. The translation should be credited as follows:

   Translated by permission from Macmillan Publishers Ltd: [JOURNAL NAME] (reference citation), copyright (year of publication).

   Note: For translation from the *British Journal of Cancer,* the following credit line applies.
   Translated by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: [JOURNAL NAME] (reference citation), copyright (year of publication)

We are certain that all parties will benefit from this agreement and wish you the best in the use of this material. Thank you.

Special Terms:

v1.1

You will be invoiced within 48 hours of this transaction date. You may pay your invoice by credit card upon receipt of the invoice for this transaction. Please follow instructions provided at that time.
To pay for this transaction now; please remit a copy of this document along with your payment. Payment should be in the form of a check or money order referencing your account number and this invoice number RLNK501384578.
Make payments to "COPYRIGHT CLEARANCE CENTER" and send to:

Copyright Clearance Center
Dept 001
P.O. Box 843006
Boston, MA 02284-3006
Please disregard electronic and mailed copies if you remit payment in advance.
Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

Gratis licenses (referencing $0 in the Total field) are free. Please retain this printable license for your reference. No payment is required.