The Crystal Structure of an Algal Prolyl 4-Hydroxylase Complexed with a Proline-rich Peptide Reveals a Novel Buried Tripeptide Binding Motif*

M. Kristian Koski†, Reija Hieta§, Maija Hirsiälä§, Anna Rönnkä§, Johanna Myllyharju†, and Rik K. Wierenga††

From the †Biocenter Oulu and Department of Biochemistry and ‡Oulu Center for Cell-Matrix Research, Biocenter Oulu, and Department of Medical Biochemistry and Molecular Biology, University of Oulu, FIN-90014 Oulu, Finland

Plant and algal prolyl 4-hydroxylases (P4Hs) are key enzymes in the synthesis of cell wall components. These monomeric enzymes belong to the 2-oxoglutarate dependent superfamily of enzymes characterized by a conserved jelly-roll framework. This algal P4H has high sequence similarity to the catalytic domain of the vertebrate, tetrameric collagen P4Hs, whereas there are distinct sequence differences with the oxygen-sensing hypoxia-inducible factor P4H subfamily of enzymes. We present here a 1.98-Å crystal structure of the algal *Chlamydomonas reinhardtii* P4H-1 complexed with Zn²⁺ and a proline-rich (Ser-Pro)₃ substrate. This ternary complex captures the competent mode of binding of the peptide substrate, being bound in a left-handed (poly)-proline type II conformation in a tunnel shaped by two loops. These two loops are mostly disordered in the absence of the substrate. The importance of these loops for the function is confirmed by extensive mutagenesis, followed up by enzyme kinetic characterizations. These loops cover the central Ser-Pro-Ser tripeptide of the substrate such that the hydroxylation occurs in a highly buried space. This novel mode of binding does not depend on stacking interactions of the proline side chains with aromatic residues. Major conformational changes of the two peptide binding loops are predicted to be a key feature of the catalytic cycle. These conformational changes are probably triggered by the conformational switch of Tyr¹⁴⁰, as induced by the hydroxylation of the proline residue. The importance of these findings for understanding the specific binding and hydroxylation of (X-Pro-Gly)ₙ sequences by collagen P4Hs is also discussed.

4R-Hydroxyproline (4Hyp)² is an uncommon amino acid produced by the addition of a hydroxyl group to the C4 carbon atom of the proline pyrrolidine ring. 4Hyp residues have an essential role in the extensive collagen family, where they are necessary for the formation of stable triple helical molecules (1–3). 4Hyp residues are also found in many plant and algal hydroxyproline-rich glycoproteins (HRGPs), such as extensins and arabinogalactan proteins, which are the major structural components of their cell walls (4, 5). In addition to these structural roles, 4Hyp has a key role in the regulation of gene expression in an oxygen-dependent manner via the hypoxia-inducible transcription factor (HIF) (6–8).

The formation of 4Hyp in the above proteins is catalyzed by the prolyl 4-hydroxylases (P4Hs), which are 2-oxoglutarate (2OG) dioxygenases and also require Fe²⁺ and O₂ (Fig. 1A) (1–3, 9). Two P4H families, the collagen P4Hs (C-P4Hs) and the HIF-P4Hs, each having three isoenzymes, are responsible for the hydroxylation of collagen and HIF, respectively (1–3, 6–9). The vertebrate C-P4Hs are α₂β₂ tetramers in which the α-subunits are responsible for the hydroxylation and the β-subunits are identical to protein-disulfide isomerase. The tetrameric assembly is required for stability and full activity (1–3). In contrast, the plant P4Hs and probably also the HIF-P4Hs are monomers (2, 10). Plant P4Hs have a ~30% sequence identity to the catalytic domain of the C-P4H α-subunits (11, 12), and they also resemble C-P4Hs in that they hydroxylate proline-rich polypeptides and are located in the lumen of the endoplasmic reticulum (2). HIF-P4Hs, on the other hand, are cytoplasmic and nuclear enzymes (8, 9) that act on proline residues in -Leu-X-Leu-Ala-Pro- motifs in HIF-α (13, 14). C-P4Hs hydroxylate the central proline of the -X-Pro-Gly- repeats of collagen polypeptides, typically generating about 100 4Hyp residues in polypeptides with a length of about 1000 amino acids. Plant P4Hs can also hydroxylate peptides with -X-Pro-Gly- repeats in vitro, but generally much less effectively than peptides representing their physiological substrates, the HRGPs, which are rich in serine and proline and can fold into a left-handed fibrous poly(1-proline) type II (PPII) helix conformation (15, 16). The only known exception is isoenzyme 1 of the large *Arabidopsis thaliana* P4H (At-P4H) family, which efficiently hydroxylates a collagen-like (Pro-Pro-Gly)₁₀ peptide, for which its Kₘ is only 3-fold higher than that of the human C-P4H-I (17). Plant P4Hs accept also poly(1-proline) as a substrate, which is not hydroxylated by any of the C-P4Hs, but instead acts as an efficient competitive inhibitor of C-P4H-I (2).
Several 2OG dioxygenases have now been crystallized and despite their low overall amino acid sequence identity the structures have revealed that the catalytic sites are always located at the same site of a common framework: a double-stranded /H9252-helix (jelly-roll) fold that consists of 8 antiparallel /H9252-strands (18, 19) (Fig. 1B). The catalytic sites of all P4Hs have the conserved -His-X-Asp-…-His- catalytic motif for Fe2+/H11001 coordination and a basic residue that binds the C5 carboxylate moiety of 2OG. The crystal structures of two P4Hs have been solved, namely those of human HIF-P4H-2 (10) and a plant P4H from Chlamydomonas reinhardtii (Cr-P4H-1) (12). The structures of these two enzymes share the jelly-roll core fold preceded by an N-terminal part that contains two long α-helices in both structures, and also three β-strands in Cr-P4H-1 and 2 β-strands in HIF-P4H-2 (10, 12). In each of these two structures the extra β-strands extend the major β-sheet of the jelly-roll fold (Fig. 1B), and the helices of the N-terminal part shield this major sheet from bulk solvent. The proximal histidine and the aspartate of the Fe2+/H11001 coordination motif are located in the /H9252-II-strand of the jelly-roll, the distal histidine of this motif is in the adjacent βVII-strand, whereas the basic 2OG-binding residue (lysine in Cr-P4H-1 and arginine in HIF-P4H-2) is in the βVIII-strand. Both enzymes have been crystallized in the presence of an active site metal ion and a 2OG analogue, but not with a peptide substrate (10, 12).

We present here the first crystal structure of a P4H complexed with a proline-rich peptide substrate. Cr-P4H-1 was co-crystallized with a 10-residue long peptide substrate (Ser-Pro)5 that adopts the PPII helix conformation. The structure reported here reveals an entirely novel binding mode for proline-rich peptide substrates that is also expected to be utilized in C-P4Hs but not in HIF-P4Hs.

**EXPERIMENTAL PROCEDURES**

**Preparation of the Protein Samples and Activity Measurements**—A truncated Cr-P4H-1 (Val29-His253) with an N-terminal His6SUMO fusion partner was expressed in Escherichia coli from Chlamydomonas reinhardtii (Cr-P4H-1) (12). The structures of these two enzymes share the jelly-roll core fold preceded by an N-terminal part that contains two long α-helices in both structures, and also three β-strands in Cr-P4H-1 and 2 β-strands in HIF-P4H-2 (10, 12). In each of these two structures the extra β-strands extend the major β-sheet of the jelly-roll fold (Fig. 1B), and the helices of the N-terminal part shield this major sheet from bulk solvent. The proximal histidine and the aspartate of the Fe2+/H11001 coordination motif are located in the /H9252-II-strand of the jelly-roll, the distal histidine of this motif is in the adjacent βVII-strand, whereas the basic 2OG-binding residue (lysine in Cr-P4H-1 and arginine in HIF-P4H-2) is in the βVIII-strand. Both enzymes have been crystallized in the presence of an active site metal ion and a 2OG analogue, but not with a peptide substrate (10, 12).

We present here the first crystal structure of a P4H complexed with a proline-rich peptide substrate. Cr-P4H-1 was co-crystallized with a 10-residue long peptide substrate (Ser-Pro)5 that adopts the PPII helix conformation. The structure reported here reveals an entirely novel binding mode for proline-rich peptide substrates that is also expected to be utilized in C-P4Hs but not in HIF-P4Hs.

**EXPERIMENTAL PROCEDURES**

**Preparation of the Protein Samples and Activity Measurements**—A truncated Cr-P4H-1 (Val29-His253) with an N-terminal His6SUMO fusion partner was expressed in Escherichia coli and purified to homogeneity as described previously (12). The tag was cleaved by digestion overnight with SUMO protease (12). Mutations were introduced using a QuikChange™ site-
directed mutagenesis kit (Stratagene). The wild-type and mutant variants of Cr-P4H-1 were purified for the kinetic analyses as above, with the exception that the His<sub>6</sub>SUMO fusion partner was not cleaved. The catalytic properties of the wild-type and mutant Cr-P4H-1 were measured using poly(lysine-proline), M<sub>r</sub> 5000, (Ser-Pro)<sub>5</sub> and (Pro-Pro-Gly)<sub>10</sub> as substrates as described previously (11).

**Crystallization**—The purified and tagless Cr-P4H-1 was concentrated to 3 mg/ml in 0.01 M Tris-HCl, 0.1 M NaCl, 0.1 M glycerine, pH 7.8, using a 10,000 molecular weight cutoff Amicon Ultra Centrifugal Filter Device (Millipore). The protein sample was supplemented with 5 mM ZnSO<sub>4</sub>, 2 mM pyridine-2,4-dicarboxylic acid (PDC) (Sigma), and 2.5 mM (Ser-Pro)<sub>5</sub> (Innovent). The microbatch under-oil method and 96-well hydrophilic vapor batch plates (Douglas Instruments) were used to co-crystallize Cr-P4H-1 with Zn<sup>2+</sup>, PDC, and (Ser-Pro)<sub>5</sub>. The crystallization drops and the covering 6-μl paraffin oil layer (Sigma) were pipetted by the Teco Freedom Evo crystallization robot (Teco) via CryoScreen software (Teco). The crystallization drops contained equal volumes (1.5 μl) of the protein sample and the precipitant solution. The precipitant solution contained a 5% (w/v) polyethylene glycol (PEG) mixture with equal amounts of PEG 400, PEG 1000, PEG 1500, PEG 4000, PEG 6000, PEG 8000, PEG monomethyl ether (MME) 550, PEG MME 750, and PEG MME 5000 (this PEG mixture is referred to as PEG Smear (20)) in 0.1M acetate, pH 5.5, and 10 mM zinc acetate. The triangle-like plate crystals appeared in 1 week at +22 °C and grew to a final size of 350 × 350 × 100 μm in one month. The crystals were soaked for ~1 min in a cryosolution of 17.5% (w/v) glycerol, 10% (w/v) PEG Smear, 10 mM zinc acetate, 0.1 M acetate, pH 5.5, and 1 mM (Ser-Pro)<sub>5</sub> and were subsequently flash-frozen in liquid nitrogen.

**Crystal Structure Determination and Validation**—A dataset to 1.98-Å resolution was collected from a Cr-P4H-1 crystal using synchrotron radiation at the X12 beamline, EMBL, DESY, Hamburg at 100 K cryoconditions using a 0.932-Å wavelength. The data were collected using a strategy from BEST (21) based on two reference images collected 90 degrees apart and processed by MOSFLM (22). The final dataset was processed using XDS (23). The data collection statistics are summarized in Table 1. The structure was solved by molecular replacement using Phaser (24) and molecule A of the SeMet(apo) form of Cr-P4H-1 (PDB ID 2V4A) (12), excluding the flexible loop regions, as a model. The structures of Cr-P4H-1 in molecules A-D and of the (Ser-Pro)<sub>5</sub> peptides at the active sites of molecules A and C were built with iterative cycles of manual building using COOT (25). A restrained refinement with translation, libration, and screw-rotation displacement (TLS) (26) was accomplished using Refmac5 (27) from the CCP4 package (28). The (Ser-Pro)<sub>5</sub> at the active site of molecule A consists of residues Pro<sup>5</sup>-Ser<sup>9</sup>, whereas in molecule C the residues Pro<sup>5</sup>-Ser<sup>9</sup> could be built in the electron density. A total of 14 zinc atoms were included in the model, one at each of the four active sites and the rest mostly at the molecule-molecule interfaces. There were strong densities close to the bound Zn<sup>2+</sup> in each molecule in the 2OG-binding pocket, but these were not enough to build the PDC molecule. The acetate ions fitted perfectly in these densities and refined with normal B-values. Finally, 468 water molecules were added to the structure using the ARP/wARP (29) option in Refmac. The final refinement statistics are shown in Table 1. Noncrystallographic symmetry restraints were not used during the refinement. The TLSANL program (30) was used to obtain the isotropic and the anisotropic B-values for each individual atom. Ramachandran plot analysis showed 99.5% of the residues to be in favored regions and 0.5% in allowed regions according to MolProbity (31). The figures were prepared using Pymol (Delano Scientific LLC).

**RESULTS AND DISCUSSION**

**Structure Determination**—A truncated Cr-P4H-1 lacking the 29 N-terminal amino acids, was purified as described (11, 12) and crystallized in the presence of the inhibitors Zn<sup>2+</sup> and PDC, and the peptide substrate (Ser-Pro)<sub>5</sub>. PDC is a homologue of 2OG (Fig. 1A) having two identically placed carboxylate groups, and the (Ser-Pro)<sub>5</sub> peptide is a shortened version of a (Ser-Pro)<sub>19</sub> motif present in the GP1 protein of the C. reinhardtii cell wall, which is a potent substrate for Cr-P4H-1 (11, 16). Four independent molecules A–D are present in the resulting crystal form of Cr-P4H-1, and the final model was refined to
TABLE 2

The four known Zn\(^{2+}\) complex structures of Cr-P4H-1

| Complex structure          | PDB ID/molecule | Metal ion | \(\beta_3-\beta_4\) loop | \(\beta_1-\beta_{II}\) loop | Tyr140\(^a\) | \(\beta_{II}-\beta_{III}\) loop |
|---------------------------|-----------------|-----------|--------------------------|------------------------------|------------|--------------------------|
| Zn-peptide complex        | 3GZE/A, C       | Zn\(^{2+}\) | Ordered, closed           | Ordered                      | In         | Ordered, extended        |
| Zn-PDC complex            | 2JIG/A          | Zn\(^{2+}\) | Ordered, open             | Ordered                      | In         | Ordered, compact         |
| Zn-binary complex         | 3GZE/B, D       | Zn\(^{2+}\) | Disordered                | Disordered                   | Out        | Disordered               |
| Zn-binary complex         | 2JIG/B          | Zn\(^{2+}\) | Disordered                | Disordered                   | Out        | Ordered, extended        |

\(^a\) "In" refers to the active conformation of Tyr\(^{140}\). In the "Out" conformation, the main chain atoms of Tyr\(^{140}\) also change. The \(\beta_{II}\) strand that contains the metal binding -His-X-Asp- motif is always ordered in the Zn\(^{2+}\) complexes and adopts identical conformation in these four structures.

1.98-Å resolution (Table 1). Zn\(^{2+}\) is present at each of the four active sites, while the (Ser-Pro)\(_5\) peptide is found only in molecules A and C (Table 2). The Cr-P4H-1 molecules containing both Zn\(^{2+}\) and the peptide are referred to hereafter as the Zn-peptide complex. PDC is not found in any of the Cr-P4H-1 molecules, but instead an acetate molecule is placed at each active site close to the Zn\(^{2+}\) in the 2OG-binding pocket. The acidic crystallization conditions (pH 5.5) together with the presence of acetate ions in the buffer solution have apparently favored binding of the acetate ion in the 2OG-binding pocket instead of PDC. This is the second crystal form of Cr-P4H-1 obtained in the presence of Zn\(^{2+}\). A Cr-P4H-1 ternary complex with Zn\(^{2+}\) and PDC (referred to as the Zn-PDC complex) was crystallized earlier at pH 8.5 in Tris-HCl buffer (PDB entry 2JIG) (12). The new structural data on the Zn-peptide complex will first be described and compared with the previous Zn-PDC complex. This structure will subsequently be compared with the structures of corresponding ternary complexes of the other superfamily members and discussed also in the context of point mutation studies probing the functional importance of residues in the flexible loops.

The Overall Structure—The (Ser-Pro)\(_5\) peptide adopts a typical PPII helix conformation in the Zn-peptide complex (Fig. 1C). Ser\(^1\) is disordered in molecule C, Ser\(^1\)-Ser\(^3\) in molecule A and Pro\(^{10}\) in both molecules. Molecule C is therefore used as the reference molecule. The middle -Ser\(^5\)-Pro\(^6\)-Ser\(^7\)- tripeptide region of the bound peptide has the lowest B-factors, and its geometry is well defined by the electron density map (Fig. 2). The Pro\(^6\) is deeply buried and points toward the catalytic site region of the bound peptide is completely covered by two loops, the \(\beta_3-\beta_4\) loop (Val\(^{75}\)-Ser\(^{95}\)) and the \(\beta_{II}-\beta_{III}\) loop (Tyr\(^{146}\), Gly\(^{158}\)), which protrude out of the jelly-roll fold (Figs. 1C, 3A, and 4). The \(\beta_3-\beta_4\) loop in particular interacts tightly with the bound peptide via the residues Ser\(^{78}\), Val\(^{79}\), Val\(^{80}\) at the entrance to the loop and Ser\(^{87}\) and Arg\(^{93}\) at the exit (Fig. 3, A and B). These two loops, which are disordered in molecules B and D of this crystal form, are known to have large conformational flexibility (12) (Fig. 5 and Table 2). The \(\beta_3-\beta_4\) loop, which was only seen in an ordered open conformation in the Zn-PDC complex (12), now adopts an ordered closed conformation in which its tip has moved 14 Å toward the tip of the \(\beta_5\) strand of the minor sheet of the jelly-roll fold (Fig. 5). The \(\beta_{II}-\beta_{III}\) loop conformation of the Zn-peptide complex is referred to as an ordered-extended conformation (Table 2), previously also observed in the Zn-binary complex (12). In the other structures, this loop is partially disordered or observed in an ordered, compact conformation (Table 2). In the latter conformation the \(\beta_{II}-\beta_{III}\) loop binds in the peptide-binding groove (Fig. 5).

In the Zn-peptide complex the tips of the \(\beta_3-\beta_4\) and \(\beta_{II}-\beta_{III}\) loops interact with each other, but not with the peptide substrate, as highlighted in Fig. 3A. Both tips adopt an \(\alpha\)-helical-like conformation with a strong hydrogen bond between the main chain oxygen of Asp\(^{81}\) and nitrogen of Gly\(^{85}\) and a weak hydrogen bond between the main chain oxygen of Asp\(^{149}\) and nitrogen of Ala\(^{153}\). Two sequence motifs present in these Cr-P4H-1 loop tips are conserved in plant P4Hs and C-P4Hs, namely -Asp/Asn-X-X-Ser/Thr-Gly- (Asp\(^{81}\)-Gly\(^{85}\)) in the \(\beta_3-\beta_4\) loop of Cr-P4H-1 and -Asp/Glu-X-X-Asp/Asp- (Asp\(^{149}\).
Asn in the βII-βIII loop (Fig. 6). In these motifs, the side chains of Asp and Ser as well as those of Asp and Asn are hydrogen-bonded to each other (Fig. 3A). The side chains of Asp and Ser point into the bulk solvent, whereas the Asp and Asn side chains point to the partner loop (Fig. 3A). Nevertheless, only one direct hydrogen bond exists between the two loop tips, namely Asn-Gly (Fig. 3A). The residue Gly is fully conserved in the plant P4H and C-P4H sequences (Fig. 6), the Phi/Psi values (82°/2°) not favoring a side chain in this position. In addition, there are water (W)-mediated hydrogen bonds between the two loops (Fig. 3A).
bonding interactions between the two loops via three well-defined water molecules W40, W37, and W81 (Fig. 3, A and B). W37, W40, and W81 are also hydrogen-bonded to the (Ser-Pro)₅ peptide, in particular to the side chain oxygen of Ser² (Fig. 3B). There is also a hydrophobic interaction between the two loops, mediated by Val¹⁸⁰ and Phe¹⁴⁷ (Fig. 3, A and B).

The β₃-β₄ and βII-βIII loops have high conformational flexibility (Fig. 5), and the observed conformations will thus be affected by crystal contacts. For this reason a systematic mutagenesis study of the conserved motifs of the two loops was performed to confirm the importance of these two partner loops for the biocatalytic function of Cr-P4H-1 (Table 3). In each of these loop variants, i.e. the D81A, S84A, G85A, D149A, D149N, and N152A mutants, the Kₘ values for the substrate poly(l-proline) were markedly increased and a 4–15-fold decrease was observed in the k₉ values, although none of these residues directly interacts with the substrate. In the N152A variant the only hydrogen bond between the two loops is lost, causing the most drastic effect on the kinetic constants. The mutagenesis data show that modest residue changes, even at the tips of the loops (Table 3), which effect the loop-loop interactions, much reduce the catalytic efficiency.

The bound (Ser-Pro)₅ peptide has adopted the PPII helix conformation, as generally observed in complexes of proline-rich peptides with proteins (32). The interactions between (Ser-Pro)₅ and Cr-P4H-1 are unique in two other respects, however, i.e. (i) the absence of stacking interactions between the proline side chains of the bound peptide and aromatic residues of the protein, and (ii) the -Ser-Pro-Ser- central tripeptide is completely buried in the complex, being shielded from the bulk solvent by the β₃-β₄ and βII-βIII loops (Fig. 4). This substrate-binding mode of Cr-P4H-1 is also unique within the 2OG dioxygenase superfamily. Most of the family members with known structures are microbial enzymes involved in antibiotic biosynthesis that use small molecule substrates (18). Only factor inhibiting HIF (FIH), which hydroxylates a specific asparagine residue in HIF-α, has been co-crystallized with its peptide substrate, but the substrate is not proline-rich and thus has no PPII helix conformation and is not bound in a tunnel (33). A loop region corresponding to the β₃-β₄ loop in Cr-P4H-1 also participates in the peptide binding in FIH (33), however, and it has been proposed that a topologically similar loop may be involved in the substrate binding of HIF-P4H-2 (34), the oxidative DNA/RNA repair enzyme AlkB from E. coli (35) and phytanoyl-CoA hydroxylase (36). However, in the corresponding loops of these homologues the characteristic -Asp/Asn-X-X-Ser/Thr-Gly- sequence motif is not conserved. Moreover, the βII-βIII loop is absent in FIH and in HIF-P4H-2. The fact that the latter loop is not present in any of the three HIF-P4H isoforms (6) indicates that this subfamily of P4H must have a different strategy for peptide substrate binding than plant P4Hs and C-P4Hs, which all have the elongated βII-βIII loop (Fig. 6).
active site and is surrounded by five hydrophobic residues, Val79 and Val80 of the β3-β4 loop, Tyr140 of βII, Phe147 of the βII-βIII loop and Trp243 of βVIII (Figs. 2 and 3, A and B). The side chains of the catalytic -His-X-Asp- motif, the Zn2⁺ ion and two arginines, Arg93 and Arg161, are also located nearby (Figs. 2 and 3). Arg93, which is stacked with the side chain of Tyr140, interacts with the (Ser-Pro)₅ via a water molecule W174, whereas Arg161 forms a direct hydrogen bond with the backbone oxygen of Pro₅ (Fig. 3). In addition, Tyr140 forms a hydrogen bond with the backbone oxygen of Pro₆ of the peptide (Fig. 3, A and B). These three amino acids, Arg₉³, Tyr₁⁴₀, and Arg₁₆₁, are fully conserved in all P4Hs, and mutation of any of these residues to alanine leads to complete inactivation of Cr-P4H-1 (12). Two main chain-main chain hydrogen bonds are formed between the -Ser⁷⁸-Ser⁸⁷ region of the β3-β4 loop of Cr-P4H-1 and the -Ser⁷⁸-Pro⁶-Ser⁷⁸ region of the peptide (Fig. 3). In addition, the side chain oxygens of Ser⁷⁸ and Ser⁸⁷ of the β3-β4 loop are hydrogen-bonded to the side chain oxygen of the peptide Ser⁷ (Fig. 3). Although these four residues of the β3-β4 loop, Ser⁷⁸, Val⁷⁹, Val⁸⁰, and Ser⁸⁷, closely interact with (Ser-Pro)₅, only Val⁷⁹ is highly conserved in the animal C-P4Hs and plant P4Hs (Fig. 6).

The structural data obtained here provide a first rationale for understanding the substrate specificity of the animal C-P4Hs for (X-Pro-Gly)ₙ sequences. Given the notable sequence identity between Cr-P4H-1 and the catalytic domain of C-P4Hs (12), it is predicted that the mode of binding of the peptide in Cr-P4H-1 and C-P4Hs is the same, implying that the bound collagen peptide in the C-P4H complex will also adopt the PPII conformation. As shown in Fig. 7, A and B, both a collagen peptide taken from the structure of a synthetic collagen triple helix and the poly(L-proline) peptide superimpose well with the bound (Ser-Pro)₅ peptide substrate in Cr-P4H-1. In particular the residues of the central tripeptide region, referred to as being at the X, Y, and Z positions (Pro-Hyp-Gly in the collagen peptide and Pro-Pro-Pro in poly-L-proline), superimpose well with the corresponding -Ser⁵-Pro⁶-Ser⁷ tripeptide in both cases. This indicates that the hydrogen bonding interactions that exist between the main chain atoms of the central tripeptide of the (Ser-Pro)₅ substrate and Cr-P4H-1 will also exist between the C-P4Hs and their -X-Pro-Gly- substrates.

The structure presented here shows that each of the three positions of the central tripeptide of the bound (Ser-Pro)₅ has unique features. Firstly, the serine in the Z-position, Ser⁷, points upwards and is hydrogen-bonded to the side chain oxygen of Pro⁶ of the peptide. This hydrogen bond with the backbone oxygen of Pro⁶ of the peptide (Fig. 3, A and B). These three amino acids, Arg₉³, Tyr₁⁴₀, and Arg₁₆₁, are fully conserved in all P4Hs, and mutation of any of these residues to alanine leads to complete inactivation of Cr-P4H-1 (12). Two main chain-main chain hydrogen bonds are formed between the -Ser⁷⁸-Ser⁸⁷ region of the β3-β4 loop of Cr-P4H-1 and the -Ser⁷⁸-Pro⁶-Ser⁷⁸ region of the peptide (Fig. 3). In addition, the side chain oxygens of Ser⁷⁸ and Ser⁸⁷ of the β3-β4 loop are hydrogen-bonded to the side chain oxygen of the peptide Ser⁷ (Fig. 3). Although these four residues of the β3-β4 loop, Ser⁷⁸, Val⁷⁹, Val⁸⁰, and Ser⁸⁷, closely interact with (Ser-Pro)₅, only Val⁷⁹ is highly conserved in the animal C-P4Hs and plant P4Hs (Fig. 6).
TABLE 3
Kinetic studies of wild-type and mutant variants of Cr-P4H-1 using poly(L-proline), M, 5000, as a substrate

| Mutation     | Km (Poly-L-Pro)a | Vmax (Poly-L-Pro)a | Km (Ser-Pro)b | Vmax (Ser-Pro)b | Km (Pro-Pro-Gly)c | Vmax (Pro-Pro-Gly)c |
|--------------|-----------------|-------------------|--------------|----------------|-----------------|-------------------|
| Wild type    | 290 ± 60        | 30                | 380 ± 130    | 19 ± 3         | 930 ± 170       | 7 ± 2             |
| Entrance and exit of the β3-β4 loop |                 |                   |              |                |                 |                   |
| S78T         | 110 ± 20        | 20 ± 3            | 190 ± 50     | 14 ± 2         | 830 ± 120       | 9 ± 2             |
| S87L         | 260 ± 50        | 8 ± 3             | 780 ± 100    | 18 ± 4         | 1050 ± 320      | 6 ± 1             |
| S78T/S87L    | 350 ± 50        | 4 ± 0.1           | 140 ± 50     | 6 ± 2          | 1270 ± 230      | 5 ± 1             |
| Tip of the β3-β4 loop |                 |                   |              |                |                 |                   |
| D81A         | 880 ± 120       | 7 ± 2             |              |                |                 |                   |
| S84A         | 770 ± 40        | 2 ± 0.1           |              |                |                 |                   |
| G85A         | 940 ± 50        | 4 ± 1.5           |              |                |                 |                   |
| Tip of the βII-βIII loop |                 |                   |              |                |                 |                   |
| D149N        | 560 ± 10        | 6 ± 2             |              |                |                 |                   |
| D149A        | 630 ± 60        | 2 ± 0.1           |              |                |                 |                   |
| N152A        | 1590 ± 240      | 2 ± 0.2           |              |                |                 |                   |
| β II-III loop |                 |                   |              |                |                 |                   |
| Y140F        | 340 ± 80        | 3 ± 0.3           |              |                |                 |                   |

a Values are means ± S.D. from at least 3–4 independent experiments.
three protein ligand atoms of the -His-X-Asp-…-His- motif and the two PDC/acetate atoms (Fig. 8A). The remaining sixth coordination position for molecular oxygen is predicted to be opposite the proximal histidine (12). In the Zn-PDC complex, a 31 Å³ cavity is calculated to occur in this region with ICM (MOLSOFT, LLC). This cavity is lined by the carboxylate moiety of Asp145, and by the hydrophobic side chain moieties of Thr164, Leu166, Phe212, Thr241, and Trp243 (Fig. 8A). Each of these residues protrudes out of the major sheet. In the Zn-peptide complex the side chains of Thr241, Trp243, and Leu166 have adopted slightly different orientations (Fig. 8A). Thr241 is hydrogen bonded to the inhibitor molecule in the Zn-PDC complex, but has rotated in the Zn-peptide complex, like the Leu166 and Trp243 side chains, causing the cavity to disappear in this complex. This finding is consistent with the fact that in the catalytic cycle molecular oxygen binds last to the active site (2), implying that both the substrate and the cofactor (2OG) are present when oxygen binds. It has been suggested that the O₂ binding site may also be located opposite the proximal histidine in clavaminate synthase (40) and in the AlkB (35), but interestingly, the currently available structures of 2OG dioxygenases also suggest an alternative site for O₂ binding, opposite to the distal histidine of the -His-X-Asp-…-His-motif (18).

The hydroxylation reaction is known to proceed via a FeIV(=O) ferryl intermediate which is formed after the activated molecular oxygen has reacted with 2OG, by which succinate is also formed (Fig. 1A) (2, 18, 19). The geometry of all available complexes suggest that the oxygen of the FeIV(=O) ferryl intermediate is bound to the metal ion in the position opposite the distal histidine, corresponding to the O1 position in the Cr-P4H-1 inhibitor complex (Fig. 8A). In this catalytic position the ferryl oxygen can abstract the C4 (Pro6) hydrogen atom, as predicted by spectroscopic studies on a P4H from *Paramecium bursaria Chlorella virus 1*, in which an FeII-I(OH) radical and C4 (Pro6) radical are formed (41). Spectroscopic studies performed on taurine dioxygenase (TauD) from *E. coli* have provided very similar information on this mechanism (42). These mechanistic studies therefore suggest that the oxy-ferryl oxygen ligand migrates to its catalytic position after the decarboxylation of 2OG in the P4H reaction. In the next step of the P4H reaction cycle the OH-radical is donated back to the C4 (Pro6) radical, resulting in the hydroxylated product. The geometries of the Cr-P4H-1 Zn-peptide complex and the TauD Fe-2OG-taurine complex (PDB entry 1OS7) are compared in Fig. 8B. The TauD complex includes an Fe²⁺ ion, 2OG and the
Concluding Remarks—The structural and enzymological data on Cr-P4H-1 obtained here show that the β3-β4 and β1I-β3II loops define the substrate binding tunnel. The bound peptide adopts the PPII conformation and the central -Ser-Pro-Ser- tripeptide is bound in this tunnel in such a manner that the first serine (in the X-position) points down toward the β1I-β3II loop, the proline in the Y-position points toward the catalytic site and the serine in the Z-position points up toward the β3-β4 loop. The -Asp/Asn-X-X-Ser/Thr-Gly- and -Asp/Glu-X-X-Asn/Asp- motif at the tips of these loops are important for stabilizing the closed, liganded conformation, whereas the flanking regions of the loops participate in the determination of substrate taurine, but the reaction does not proceed, as the crystals have been grown under anaerobic conditions (43). In both TauD and the P4Hs O₂ binding occurs after binding of the 2OG and the substrate (41, 42). The substrates are bound in topologically identical positions in the TauD and Cr-P4H-1 complexes (Fig. 8B). In Cr-P4H-1 the target C4 of the substrate Pro⁶ points to the acetate oxygen with a distance of 3.8 Å. This acetate oxygen superimposes on the O1-atoms of the PDC and 2OG molecules of the Cr-P4H-1 Zn-PDC complex and the TauD 2OG-taurine complex, respectively (Fig. 8, A and B). This geometry is suitable for the formation of the 4-R-OH product of the proline. The Pro⁶ is in the down-puckering (C⁴-exo) conformation (Fig. 9), which was recently shown to be the favored prolyl substrate conformation for C-P4Hs (44) and HIF-P4Hs (45). However, the preferred conformation of the 4-R-OH-proline is the up-puckering (C⁴-endo) conformation (see PDB entries: 1V4F, 1V6Q, 1V7H, 01YM, 1LM8) (44–48), and therefore it is expected that the 4Hyp will adopt the up-puckering conformation on completion of the chemical conversion. In this conformation it will clash with the side chain of Tyr¹⁴⁰, as is visualized in the superimposition analysis (Fig. 9). Consequently, it can be predicted that the hydroxylation of Pro⁶ will cause Tyr¹⁴⁰ to flip to its out conformation, as seen in the Zn-binary complexes of this study (Table 2) and in our previous study (12). As Tyr¹⁴⁰ in the Zn-peptide complex is stacked with the Arg⁹³ of the β3-β4 loop, the conformational change in Tyr¹⁴⁰ will also directly affect the conformation of that loop. Concomitantly with the flipping out of Tyr¹⁴⁰, causing the β1I-β3II loop to become disordered, the Glu¹⁴¹ side chain flips inwards and overlaps with the position of the β3-β4 loop in the Zn-peptide complex. Interestingly, the Y140A mutation completely inactivates Cr-P4H-1 (12), whereas the mutation to Phe has only minor effects on the kinetic constants (Table 3). Our model predicts that the Y140A mutation will disable the conformational switch of Tyr¹⁴⁰, apparently causing the enzyme to become inactive, whereas the clash with the hydroxylated product in the Y140F variant will be the same as in the wild-type enzyme (Fig. 9). Evolutionary pressure has apparently generated the β3-β4 and β1I-β3II loop-loop interactions to be sufficiently strong to stabilize the protein peptide interactions required for catalysis, while at the same time these loop-loop interactions are not too tight, allowing loop opening on completion of the hydroxylation step and thereby facilitating product release.
The side chain of Tyr140, which will clash with the product, is also shown. The figure also includes a superimposed up-puckered 4-Zn-peptide complex with its (2Fo-Fc) omit density (countered at 1σ). The superimposed complex shows that the structure and sequence comparisons indicate that the PPII mode of binding of the central tripeptide is preserved in the C-P4H enzyme family and it is proposed that this complex enzymatic mechanism is also a common feature of these enzymes.

REFERENCES

1. Kivirikko, K. I., and Pihlajaniemi, T. (1998) Adv. Enzymol. Relat. Areas Mol. Biol. 72, 325–398
2. Myllyharju, J. (2003) Matrix Biol. 22, 15–24
3. Myllyharju, J., and Kivirikko, K. I. (2004) Trends Genet. 20, 33–43
4. Kieliszewski, M. J., and Lamport, D. T. A. (1994) Plant J. 5, 157–172
5. Cassab, G. I. (1998) Annu. Rev. Plant Physiol. Plant Mol. Biol. 49, 281–309
6. Bruick, R. K., and McKnight, S. L. A. (2001) Science 294, 1337–1340
7. Epstein, A. C., Gleadle, J. M., McNeill, L. A., Hewitson, K. S., O’Rourke, J., Mole, D. R., Mukherji, M., Metzen, E., Wilson, M. I., Dhanda, A., Tian, Y. M., Masson, N., Hamilton, D. L., Jaakkola, P., Barstead, R., Hodgkin, J., Maxwell, P. H., Pugh, C. W., Schofield, C. I., and Ratcliffe, P. J. (2001) Cell 107, 43–54
8. Crowdhury, R., Hardy, A., and Schofield, C. J. (2008) Chem. Soc. Rev. 37, 1308–1319
9. Myllyharju, J. (2008) Ann. Med. 40, 402–417
10. McDonough, M. A., Li, V., Flashman, E., Crowdhury, R., Mohr, C., Liénard, B. M., Zondlo, J., Oldham, N. J., Clifton, I. J., Lewis, J., McNeill, L. A., Kurzeja, R. J., Hewitson, K. S., Yang, E., Jordan, S., Syed, R. S., and Schofield, C. J. (2006) Proc. Natl. Acad. Sci. U.S.A. 103, 9814–9819
11. Keskiaho, K., Hietanen, R., Sormunen, R., and Myllyharju, J. (2007) Plant Cell 19, 256–269
12. Koski, M. K., Hietanen, R., Böllner, C., Kivirikko, K. I., Myllyharju, J., and Wierenga, R. K. (2007) J. Biol. Chem. 282, 37112–37123
13. Jaakkola, P., Mole, D. R., Tian, Y. M., Wilson, M. I., Lüettig, J., Gaskell, S. J., Kriegshuis, A., Hebestreit, H. F., Mukherji, M., Schofield, C. I., Maxwell, P. H., Pugh, C. W., and Ratcliffe, P. J. (2001) Science 292, 468–472
14. Ivan, M., Kondo, K., Yang, H., Kim, W., Vålland, J., Ohh, M., Salic, A., Asara, J. M., Lane, W. S., and Kaelin, W. G., Jr. (2001) Science 292, 464–468
15. Adair, W. S., and Snell, W. J. (1990) in Organization and Assembly of Plant and Animal Extracellular Matrix (Adair, W. S., and Mecham, R. P., eds) pp. 15–84, Academic Press, San Diego, CA
16. Ferris, P. J., Woessner, J. P., Waffenschmidt, S., Kilz, S., Drees, J., and Goodenoough, U. W. (2001) Biochemistry 40, 2978–2987
17. Hietanen, R., and Myllyharju, J. (2002) J. Biol. Chem. 277, 23965–23971
18. Clifton, I. J., McDonough, M. A., Ehrismann, D., Kershaw, N. J., Granatino, N., and Schofield, C. I. (2006) J. Inorg. Biochem. 100, 644–669
19. Ozer, A., and Bruick, R. K. (2007) Nat. Chem. Biol. 3, 144–153
20. Newman, J., Egan, D., Walter, T. S., Meged, R., Berry, I., Ben Jelloul, M., Sussman, J. L., Stuart, D. I., and Perrakis, A. (2005) Acta Crystallogr. D Biol. Crystallography 61, 1426–1431
21. Bourenkov, G. P., and Popov, A. N. (2006) Acta Crystallogr. D Biol. Crystallography 62, 58–64
22. Leslie, A. G. W. (1992) Joint CCP4 ESF-EACBM Newsl. Protein Crystallogr. 26, 22–33
23. Kabsch, W. (1993) J. Appl. Crystallogr. 26, 795–800
24. McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C., and Read, R. J. (2005) Acta Crystallogr. D Biol. Crystallography 61, 458–464
25. Emsley, P., and Cowtan, K. (2004) Acta Crystallogr. D Biol. Crystallography 60, 2126–2132
26. Winn, M. D., Isupov, M. N., and Murshudov, G. N. (2001) Acta Crystallogr. D Biol. Crystallography 57, 122–133
27. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Acta Crystallography D Biol. Crystallography 53, 249–255
28. CCP4 (1994) Acta Crystallogr. D Biol. Crystallography 50, 760–763
29. Evrard, G. X., Langer, G. G., Perrakis, A., and Lamzin, V. S. (2007) Acta Crystallography D Biol. Crystallography 63, 108–117
30. Howlin, B., Butler, S. A., Moss, D. S., Harris, G. W., and Driessen, H. P. C. (1993) J. Appl. Crystallogr. 26, 622–624
31. Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, Murray, L. W., Arendall, W. B., 3rd, Snoeijink, J., Richardson, J. S., and Richardson, D. C. (2007) Nucleic Acids Res. 35, W375–W383
32. Ball, L. J., Kühne, R., Schneider-Mergener, J., and Oschkinat, H. (2005) Angew. Chem. Int. Ed. 44, 2852–2869
33. Ellis, J. M., Hewitson, K. S., McNeill, L. A., Seibel, J. F., Schlemminger, I., Pugh, C. W., Ratcliffe, P. J., and Schofield, C. I. (2003) J. Biol. Chem. 278, 1802–1806
34. Flashman, E., Bagg, E. A., Crowdhury, R., Mecinovic, J., Loenarz, C., McDonough, M. A., Hewitson, K. S., and Schofield, C. J. (2008) J. Biol. Chem. 283, 3808–3815
35. Yu, B., Edstrom, W. C., Benach, J., Hamuro, Y., Webster, P. C., Gibney, B. R., and Hunt, J. F. (2006) Nature 439, 879–884
36. McDonough, M. A., Kavanagh, K. L., Butler, D., Sears, T., Oppermann, U., and Schofield, C. J. (2005) J. Biol. Chem. 280, 41101–41110
37. Kaska, D. D., Günzler, V., Kivirikko, K. I., and Myllylä, R. (1987) Biochim. Biophys. Acta 924, 483–490
38. Tiainen, P., Myllyharju, J., and Koivunen, P. (2005) J. Biol. Chem. 280, 1142–1148
39. Annunen, P., Helaakoski, T., Myllyharju, J., Veijola, J., Pihlajaniemi, T., and Kivirikko, K. I. (1997) J. Biol. Chem. 272, 17342–17348
40. Zhang, Z., Ren, J., Harlos, K., McKinnon, C. H., Clifton, I. J., and Schofield,
Prolyl 4-Hydroxylase Complexed with a Peptide Substrate

C. J. (2002) FEBS Lett. 517, 7–12
41. Hoffart, L. M., Barr, E. W., Guyer, R. B., Bollinger, J. M., Jr., and Krebs, C. (2006) Proc. Natl. Acad. Sci. U.S.A. 103, 14738–14743
42. Bollinger, J. M., Jr., Price, J. C., Hoffart, L. M., Barr, E. W., and Krebs, C. (2005) Eur. J. Inorg. Chem. 2005, 4245–4254
43. O’Brien, J. R., Schuller, D. J., Yang, V. S., Dillard, B. D., and Lanzilotta, W. N. (2003) Biochemistry 42, 5547–5554
44. Gorres, K. L., Edupuganti, R., Krow, G. R., and Raines, R. T. (2008) Biochemistry 47, 9447–9455
45. Loenarz, C., Mecinović, J., Chowdhury, R., McNeil, L. A., Flashman, E., and Schofield, C. J. (2009) Angew. Chem. Int. Ed. Engl. 48, 1784–1787
46. Okuyama, K., Hongo, C., Fukushima, R., Wu, G., Narita, H., Noguchi, K., Tanaka, Y., and Nishino, N. (2004) Biopolymers 76, 367–377
47. Schumacher, M., Mizuno, K., and Bächinger, H. P. (2005) J. Biol. Chem. 280, 20397–20403
48. Hon, W. C., Wilson, M. I., Harlos, K., Claridge, T. D., Schofield, C. I., Pugh, C. W., Maxwell, P. H., Ratcliffe, P. J., Stuart, D. I., and Jones, E. Y. (2002) Nature 417, 975–978
The Crystal Structure of an Algal Prolyl 4-Hydroxylase Complexed with a Proline-rich Peptide Reveals a Novel Buried Tripeptide Binding Motif
M. Kristian Koski, Reija Hieta, Maija Hirsilä, Anna Rönkä, Johanna Myllyharju and Rik K. Wierenga

J. Biol. Chem. 2009, 284:25290-25301.
doi: 10.1074/jbc.M109.014050 originally published online June 24, 2009

Access the most updated version of this article at doi: 10.1074/jbc.M109.014050

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 47 references, 15 of which can be accessed free at http://www.jbc.org/content/284/37/25290.full.html#ref-list-1