The Peptide-Substrate-binding Domain of Collagen Prolyl 4-Hydroxylases Is a Tetratricopeptide Repeat Domain with Functional Aromatic Residues*

Received for publication, August 31, 2004, and in revised form, September 24, 2004
Published, JBC Papers in Press, September 28, 2004, DOI 10.1074/jbc.M410007200

Mira Pekkala‡, Reija Hietaš, Ulrich Bergmann‡, Kari I. Kivirikko§, Rik K. Wierenga‡, and Johanna Myllyharju§§

From the ‡Department of Biochemistry and Biocenter Oulu and §Collagen Research Unit, Biocenter Oulu and Department of Medical Biochemistry and Molecular Biology, University of Oulu, FIN-90014 Oulu, Finland

Collagen prolyl 4-hydroxylases catalyze the formation of 4-hydroxyproline in -Pro-Gly- sequences and have an essential role in collagen synthesis. The vertebrate enzymes are αβ2 tetramers in which the catalytic α-subunits contain separate peptide-substrate-binding and catalytic domains. We report on the crystal structure of the peptide-substrate-binding domain of the human type I enzyme refined at 2.3 Å resolution. It was found to belong to a family of tetratricopeptide repeat domains that are involved in many protein-protein interactions and consist of five α-helices forming two tetratricopeptide repeat motifs plus the solvating helix. A prominent feature of its concave surface is a deep groove lined by tyrosines, a putative binding site for proline-rich peptides. Solvent-exposed side chains of three of the tyrosines have a repeat distance similar to that of a poly-L-proline type II helix. The aromatic surface ends at one of the tyrosines, where the groove curves almost 90° away from the linear arrangement of the three tyrosine side chains, possibly inducing a bent conformation in the bound peptide. This finding is consistent with previous suggestions by others that a minimal structural requirement for proline 4-hydroxylation may be a sequence in the poly-L-proline type II conformation followed by a β-turn in the Pro-Gly segment. Site-directed mutagenesis indicated that none of the tyrosines was critical for tetramer assembly, whereas most of them were critical for the binding of a peptide substrate and inhibitor both to the domain and the αβ2 enzyme tetramer.

The prolyl 4-hydroxylases (P4Hs) catalyze the formation of 4-hydroxyproline by the hydroxylation of proline residues in peptide linkages. Two P4H families are known today. The collagen P4Hs (C-P4Hs), enzymes residing within the lumen of the endoplasmic reticulum, have a central role in the synthesis of all collagens, with the resulting 4-hydroxyproline residues being essential for the folding of the newly synthesized collagen polypeptide chains into triple helical molecules (1–3). The P4Hs hydroxylating the hypoxia-inducible factor are cytoplasmic and nuclear enzymes that play a key role in the response of cells to hypoxia (4–6).

All vertebrate C-P4Hs are αβ2 tetramers in which the β-subunit is identical to the enzyme and chaperone protein disulfide isomerase (PDI) (1–3). Three isoforms of the catalytic α-subunit have been characterized from human and mouse sources and shown to form with PDI [α(II)]2β2, [α(III)]2β2, and [α(III)]3β2 tetramers, the type I, II, and III C-P4Hs, respectively (1–3, 7, 8). The type I enzyme is the main form in most cells, but type II is a major form in chondrocytes, osteoblasts, endothelial cells, and some other cell types, whereas type III is expressed in many tissues but at much lower levels than the type I and type II C-P4Hs (7–10). All P4Hs require Fe2+, 2-oxoglutarate, O2, and ascorbate, and their reaction involves a stoichiometric 2-oxoglutarate decarboxylation (1–3). The C-terminal region of the C-P4H α-subunits contains four conserved residues, two histidines and one aspartate that bind the Fe2+ atom and a lysine that binds the C-5 carboxyl group of the 2-oxoglutarate (11).

The C-P4Hs act on -Pro-Gly- triplets in collagens and more than 20 other proteins with collagen-like sequences (1–3, 12). Tripeptides with the structure X-Pro-Gly fulfill a minimum requirement for hydroxylation, given that the residue before the proline cannot be glycine, whereas the residue after the proline must be glycine (1–3). Polytripeptides with the structure (Pro-Pro-Gly)n are good substrates, with only the prolines preceding glycine being hydroxylated and the Ks values decreasing markedly with increasing chain length (1–3). The interaction of peptide substrates with the C-P4Hs is further affected by the amino acid in the X position of the X-Pro-Gly triplet to be hydroxylated, by other nearby amino acids, and by the peptide conformation, with the triple helical conformation of the collagenous peptides completely preventing hydroxylation (1–3). Poly-L-proline is an effective competitive inhibitor of type I C-P4H but only a very weak inhibitor of the type II, there being also smaller differences between these two isoenzymes in the binding properties with respect to peptide substrates, whereas the recently identified type III C-P4H appears to be intermediate between the type I and type II enzymes in this respect (1–3, 7).
The peptide-substrate-binding domain of the C-P4Hs is separate from the catalytic domain and is located between residues 140 and 245 in the 517-residue human α(I)-subunit (13). NMR studies have shown that a recombinant human C-P4H α(I) polypeptide Phe\(^{144}\)Ser\(^{244}\) forms a folded domain consisting of five α-helices and one short putative β-strand between the second and third α-helices (14); this structure is quite different from those of other proline-rich peptide binding modules, which consist mainly of β-strands (15–22). Binding of the peptide (Pro-Pro-Gly)\(_2\) to this domain caused major chemical shifts in five residues in many backbone amides, with the residues showing the largest shifts being located mainly in helix α3 (14). The \(K_d\) values determined by surface plasmon resonance and isothermal titration calorimetry for the binding of several synthetic peptides to the α(I) and corresponding α(II) domains were very similar to the \(K_m\) and \(K_d\) values for these peptides as substrates and inhibitors of the type I and type II C-P4H tetramers. The \(K_d\) values determined for a 4-hydroxyproline-containing peptide indicated a marked decrease in the affinity of hydroxylated peptides for the domain (14). Many characteristic features of peptide binding to the type I and type II C-P4H tetramers can thus be explained by the binding to this domain rather than to the catalytic domain (14).

We recently obtained crystals of the α(I) peptide-substrate-binding domain (23), and we have now obtained crystals that diffract to 2.3 Å resolution. The domain was found to consist of five α-helices and belong to the family of tetratricopeptide repeat (TPR) domains that are involved in many protein-protein interactions (24–27). The peptide substrates and the competitive inhibitor poly-L-proline are suggested to become bound to a groove lined by tyrosines; a critical role for several tyrosines in peptide binding was verified by site-directed mutagenesis studies.

MATERIALS AND METHODS

Recombinant Protein Samples and Site-directed Mutagenesis—The recombinant peptide-substrate-binding domain covers residues 144–244 of the human C-P4H α(I)-subunit. The only cysteine in this domain, Cys\(^{158}\), was mutated to serine, and the domain was extended by a C-terminal tag of six histidines and preceded by a methionine (14). The recombinant domain was expressed in Escherichia coli and purified as described earlier (14, 23). Se-Met derivatization of the protein was performed by expressing it in a methionine-requiring auxotroph E. coli strain B384 (DE3) using induction for 20 h at 20 °C. All tyrosines residues in the peptide-substrate-binding domain and the corresponding wild-type or mutant α(I)-subunits and the PDI polypeptide were converted individually to alanine using a QuikChange™ site-directed mutagenesis kit (Stratagene). The sequences were verified on an automated DNA sequencer (ABI Prism 377; Applied Biosystems). Recombinant type I C-P4H tetramers were expressed in insect cells as described previously (11, 13) by expressing the polypeptide in a medium containing this initiation methionine, Met\(^{170}\) and Met\(^{235}\) (numbered according to the α(I)-subunit residues). It therefore seemed possible to alanine using a QuikChange™ site-directed mutagenesis kit (Stratagene). The sequences were verified on an automated DNA sequencer (ABI Prism 377; Applied Biosystems). Recombinant type I C-P4H tetramers were expressed in insect cells as described previously (11, 13) by coexpressing them with recombinant baculoviruses encoding the wild-type or mutant α(I)-subunits and the PDI polypeptide. The cells were harvested 72 h after infection, homogenized in a Triton X-100 containing buffer, and centrifuged (11, 13). Aliquots of the soluble proteins were analyzed by nondenaturing PAGE, and P4H activity was assayed by a method based on the hydroxylated-coupled decarboxylation of 2-oxo-L\(^{14}\)Cglyceraldehyde (28). \(K_m\) values for (Pro-Pro-Gly)\(_2\) were determined as described (29), and IC\(_{50}\) values for poly-L-proline, \(M_t\) 5000–8000, were measured by adding increasing concentrations of this peptide to the enzyme reaction mixture.

Isothermal Titration Calorimetry—The ITC experiments were performed using a VP-ITC titration microcalorimeter (Microcal, Northampton, MA) as described previously (14). In the individual titrations, 5-μl doses of 1–5 mM solutions of (Pro-Pro-Gly)\(_2\) or poly-L-proline, \(M_t\) 5000–8000 (in 10 mM sodium phosphate buffer, pH 6.8) were injected into sample cell volume 1.4476 ml containing 0.05–0.3 mM of the wild-type or mutant peptide-substrate-binding domain in the above buffer at an interval of 4 min with stirring at 300 rpm. The titration data obtained were corrected by subtracting the heat of dilution of the peptides from the raw data and fitted by the non-linear least squares minimization method using the one set of sites model (Levenberg-Marquardt algorithm) to determine the dissociation constant \(K_d\) using Origin software (Microcal).

RESULTS

Overview of the Structure—The recombinant peptide-substrate-binding domain covers residues Phe\(^{144}\)Ser\(^{244}\) of the 517-residue human C-P4H α(I)-subunit (14) (Fig. 1). The only cysteine in the domain, Cys\(^{158}\), was mutated to serine, and the C terminus contained a six-histidine tag. The recombinant domain produced in E. coli has three methionines, the translation initiation methionine, Met\(^{170}\), and Met\(^{235}\) (numbered according to the α(I)-subunit residues). It therefore seemed possible to resolve the structure by derivatization with Se-Met, performed by expressing the polypeptide in a medium containing this amino acid. The crystal structure of the domain was refined at 2.3 Å resolution with good refinement statistics (Table I). All the residues have \(\phi/\psi\) values in the allowed regions of the Ramachandran plot. It is an all-helical protein with five \(\alpha\)-helices running antiparallel to each other (Fig. 2). There are two molecules/asymmetric unit, with the polypeptide chain of each molecule being well defined from the N terminus until residue 237. There are no regions with high B factors, except for the seven C-terminal residues that are disordered. The two molecules of the asymmetric unit have the same structure, so that on superposition of residues 143–237 of the two molecules, for example, the root mean square difference between the corresponding Ca atoms is 0.2 Å.

The Helical Bundle of the Domain Is a 2.5-TPR Repeat—The five antiparallel \(\alpha\)-helices create a bowl-like surface, with helices α1 and α5 at the rim, α2, α3, and α4 on the bottom, and α3...
Comparison of the fold of the domain with other structures present in the Protein Data Bank, using the DALI server (39), indicated that the C-P4H peptide-substrate-binding domain has a high structural similarity to TPR domains. The TPR motif is a 34-residue repeat consisting of two antiparallel α-helices. Naturally occurring domains that are assembled from TPR motifs often comprise a number of these motifs plus an additional α-helix, referred to as the “solvating” helix (24, 25). The C-P4H domain has two TPR motifs, referred to as TPR-1 and TPR-2, plus the solvating helix (Fig. 2). Most characterized TPR domains have at least three motifs (25). Because the sequence similarity between TPR units is very low, it was not known that the C-P4H domain would have this structure. Nevertheless, the TPR consensus sequence consisting of a conserved pattern of small and large hydrophobic residues (25) is mostly present in both TPR-1 and TPR-2 of the C-P4H peptide-substrate-binding domain. The residues corresponding to the TPR consensus sequence Trp4-Leu7-Gly8-Tyr11-Ala20-Phe24-Ala27-Pro32 (25) are Cys150-Leu153-Gly154-Ala157-Thr166-Met170-Ala173-Asp178 in the C-P4H TPR-1 motif and Leu191-Leu194-Ser195-Val198-Ala207-Thr211-Leu214-Pro219 in C-P4H TPR-2 (Fig. 1).

**Table I**

Data collection and refinement statistics on the C-P4H-I peptide-substrate-binding domain

| Data set name | Peak | Inflection | Remote | Native |
|---------------|------|------------|--------|--------|
| Space group   | P3_1 | P3_2       | P3_2   | P3_2   |
| Unit cell parameters a,b (Å) | 55.50 | 55.53 | 55.55 | 55.98 |
| Temperature (K) | 100 | 100 | 100 | 100 |
| Resolution (Å) | 20–2.7 (2.8–2.7) | 20–2.7 (2.8–2.7) | 20–2.7 (2.8–2.7) | 35–2.3 (2.4–2.3) |
| R_merge (%) | 4.6 (45.0) | 3.9 (41.0) | 4.0 (35.5) | 3.8 (25.8) |
| Completeness (%) | 99.5 (99.9) | 98.8 (99.5) | 99.4 (99.7) | 98.7 (99.2) |
| (I/σ(I)) | 18.6 (3.8) | 12.1 (2.3) | 19.1 (3.9) | 15.7 (3.7) |
| Redundancy | 4.3 | 1.9 | 3.4 | 3.2 |
| Wilson B-factor (Å²) | 52.2 |

Refinement statistics

| Total number of reflections | 35–2.3 |
| Working set: number of reflections | 16190 |
| Test set: number of reflections | 52.2 |
| Protein atoms | 821 |
| Solvent atoms | 25.9 |

Geometry statistics

| Bond distance (Å) | 0.018 |
| Bond angle (°) | 1.7 |
| Main chain bonded atoms (Å²) | 0.7 |
| Side chain bonded atoms (Å²) | 2.4 |
| Ramachandran plot | 92.0 |
| Most favored regions (%) | 92.0 |
| Generously allowed regions (%) | 8.0 |
| Disallowed regions (%) | 0.0 |

*Values in parentheses are for the highest resolution shell.*

**Fig. 1.** Sequence and secondary structure of the C-P4H α(I) peptide-substrate-binding domain. The numbering is according to the α(I)-subunit residues from Phe144, with the starting methionine of the recombinant domain being given the number 143. The C-terminal residues after Lys257 are disordered in the structure and are therefore not shown. The α-helices are indicated by cylinders, the color codes of which identify the two TPR motifs, TPR-1 (green), TPR-2 (orange), and the solvating helix (blue). The tyrosines providing the hydrophobic surface of the groove are indicated by ●. Other side chains involved in defining this postulated binding groove are marked with a ●. The tyrosine residues shown by site-directed mutagenesis data to be highly important, of intermediate importance, and not important for peptide binding are shown by red, cyan, and black stars, respectively.
Peptide-Substrate-binding Domain

Fig. 2. The overall fold of the C-P4H α(I) peptide-substrate-binding domain structure. A side view of the domain. Side chains of all tyrosine residues are shown and labeled. N and C label the N and C termini of the domain, respectively. The three tyrosine residues, Tyr196, Tyr230, and Tyr233, forming a linear array are shown in green.

Complexed with the C-terminal peptide of Hsp70 (27), are shown in Fig. 3. The root mean square distances between corresponding Ca atoms are 1.6 Å for the C-P4H-Pex5 superposition and 1.9 Å for the C-P4H-Hop superposition when the residues of the loop between C-P4H TPR-1 and TPR-2 are excluded from the calculations. The sequence identity between the C-P4H domain and the Pex5 and Hop domains is very low (13% in both cases). Nevertheless, there is remarkable structural similarity between the C-P4H domain and the two other TPR domains (Fig. 3), with a unique feature of the C-P4H domain structure being the longer loop between the TPR-1 and TPR-2 motifs. This loop region, residues 179–186, is reasonably well conserved in the corresponding sequences present in the various C-P4H α-subunit isoforms in different species, except for the domain present in the α(III)-subunit (7, 8) where the loop is five residues longer. This well ordered loop folds back on helix α1 and is stabilized by a hydrophobic contact between the side chains of Ile185 of the loop and Phe151 of helix α1.

The Regular Hydrophobic Surface—A striking feature of the concave surface of the domain is the location and orientation of the side chains of three tyrosines, Tyr233, Tyr230, and Tyr196 (Figs. 2 and 4). Their aromatic rings are nearly parallel to each other and face the bulk solvent (Figs. 2 and 4). The separation of these side chains is almost the same; the distance between the centers of the aromatic rings is 8.1 Å for the Tyr233-Tyr230 pair and 8.7 Å for the Tyr230-Tyr196 pair. Tyr233 is at the edge of the domain, whereas Tyr196 is closest to its center, protruding out of helix α3. The aromatic surface repeat of Tyr233-Tyr230-Tyr196 is linear, but it can be regarded as extending farther at an angle of ~90° if one also considers the surface features defined by Tyr158 and Tyr193. This aromatic surface forms a deep groove, together with Asn227 and Arg223, which are opposite Tyr233-Tyr196 and the Arg223-Asp192 salt bridge, which is opposite Tyr196-Tyr158-Tyr193. This salt bridge completes the groove as it turns 90° away from the linear array (Fig. 4). Below this groove there is a polar pocket filled with a water molecule (Fig. 4A, Wat15), hydrogen bonded to the oxygen side-chain atoms of Asn227, Ser195, and Thr211.

A Possible Mode of Binding of Poly-L-proline and Proline-rich Peptide Substrates—TPR domains are known to be involved in protein-peptide and protein-protein interactions but have not been known to constitute a framework for the binding of proline-rich peptides. Well characterized proline-rich peptide-binding domains include the SH3, WW, EVH1, UEV, and GYF domains and the proline-rich peptide-binding protein profilin.

Proline-rich peptides preferably fold into a poly-L-proline type II (PPII) helix conformation (18, 40–44) and proline-rich peptide ligands are usually bound in this conformation, but several different modes of binding have been observed. A unique feature of the PPII conformation is its ~2-fold axis around a carbonyl CO bond (44). Modes of binding using this property, as seen in the complexes of peptides with the SH3 and WW domains, for example, have also been referred to as recognition of the X-Pro motif. It has been found that proline-rich ligands can become bound to the same binding surface in both directions, from N to C and from C to N (16, 20, 45). In this case the binding groove is rather shallow. A different mode of binding is seen in the case of the EVH1 domain, for example, where two proline side chains of a Pro-X-X-Pro motif bind in a clef lined with aromatic residues (17, 18). A common feature of all characterized binding sites for proline-rich peptides is the presence of solvent-exposed aromatic residues. The prolines and aromatic residues can interact with each other in several ways, for example through van der Waals stacking interactions or CH–π aromatic hydrogen bonding interactions, or both (46).

The aromatic patch on the inner, concave surface of the C-P4H α(I) peptide-substrate-binding domain is dominated by the three parallel, solvent-exposed side chains of Tyr233, Tyr230, and Tyr196 (Figs. 2 and 4). This aromatic surface ends abruptly at Tyr158 (Fig. 4) where the deep groove curves almost 90° away from the linear arrangement of the three tyrosine side chains. The repeat distance of these tyrosine side chains, ~8.5 Å (see above) is similar to the repeat distance of the Pro-X-Pro-X-Pro-X-Pro-X motif, which is ~9.1 Å for the Cγ atoms, of a PPII helix. Rigid docking of a PPII helix onto the aromatic groove indicates that the proline side chains of the Pro-X-X repeat can stack on the Tyr233, Tyr230, and Tyr196 side chains of the C-P4H domain when docked in the same N to C direction as seen for the peptide ligands bound to Pex5 and Hop (Fig. 3). The proposed mode of binding is supported by previous data on the role of Tyr233 in the inhibition of the C-P4H activity by poly-L-proline (13). This peptide is an effective competitive inhibitor of type I C-P4H, which contains Tyr233, but not of the type II or type III C-P4H or of the main Caenorhabditis elegans C-P4H form in which the tyrosine is replaced by a non-aromatic residue (1–3, 13). Replacement of Tyr233 in the human C-P4H α(I)-subunit by a glutamine, a residue present in the corresponding position in the human α(II)-subunit, has been shown to lead to a C-P4H mutant with a markedly increased Kd for poly-L-proline, whereas mutation of the glutamine present in the α(II)-subunit to tyrosine led to a type II C-P4H mutant with a distinctly decreased Kd for poly-L-proline (13). In the predicted mode of binding to the C-P4H α(I) peptide-substrate-binding domain, the proline-rich ligand also interacts with Asn227, Arg223, and Asp192, with these residues also being highly conserved in sequence alignments.

Binding Experiments with Mutant Variants of the Domain and the C-P4H Tetramer—To investigate further the roles of tyrosines, we mutated all the tyrosine residues of the peptide-substrate-binding domain to alanines and determined the apparent Kd values for the binding of the peptide substrate (Pro-Pro-Gly)10 and the inhibitor poly-L-proline by isothermal titration calorimetry. The corresponding mutations were also generated in the full-length C-P4H α(I)-subunit, and their effects on the Kd values for the substrate (Pro-Pro-Gly)10 and the IC50 values for the inhibitor poly-L-proline in the enzyme reaction were analyzed.

The apparent Kd of the wild-type domain for (Pro-Pro-Gly)10 was about 40 μM, whereas the Kd values of the Y196A, Y193A, and Y230A mutants were too high to be measured by isothermal titration calorimetry (Table II). The mutants Y233A,
Y199A, Y158A, and Y163A gave slightly increased \( K_d \) values, whereas Y164A gave the same value as the wild-type domain (Table II). These data suggested that the eight tyrosines can be divided into three categories in terms of their importance (indicated as Groups 1–3 in Table II), namely highly important, of intermediate importance, and not important. The \( K_d \) of the wild-type domain for poly-L-proline, \( M_r 5000–8000 \), was 2 \( \mu M \), whereas the \( K_d \) values of the three Group 1 mutants ranged from 120 to too high to be measured. Those of the four Group 2 mutants ranged from 15 to 60 \( \mu M \) and that of the only Group 3 mutant, Y164A, was only 6 \( \mu M \) (Table II).

The mutant full-length \( \alpha(I) \)-subunits were coexpressed with the PDI polypeptide (i.e. the \( \beta \)-subunit of the C-P4H \( \alpha_2\beta_2 \) tetramer) in insect cells, and the assembly of the mutant tetramers was analyzed by non-denaturing PAGE. Each mutant \( \alpha(I) \)-subunit formed an enzyme tetramer with the PDI polypeptide, indicating that none of the replaced tyrosines played a structural role in the \( \alpha(I) \)-subunit to the extent that its mutation to alanine would impair tetramer assembly (Fig. 5). The \( K_m \) of the wild-type C-P4H for (Pro-Pro-Gly)\(_{10}\) was 30 \( \mu M \), whereas those of the three Group 1 mutants, Y196A, Y193A, and Y230A, had distinctly lower \( K_m \) values, ranging from 40 to 80 \( \mu M \), whereas the Group 3 mutant Y164A showed no increase in \( K_m \) (Table II). The three Group 1 mutants also showed the highest IC\(_{50}\)


## Peptide-Substrate-binding Domain

### Table II

| Group and mutation | Peptide-substrate-binding domain | C-P4H-I tetramer |
|--------------------|----------------------------------|------------------|
|                    | $K_d$ for (Pro-Pro-Gly)$_{10}$ $^a$ | $K_d$ for Poly-L-proline $^a$ | $K_a$ for (Pro-Pro-Gly)$_{10}$ $^b$ | IC$_{50}$ for Poly-L-proline $^b$ |
| Wild-type          | 40                                | 2                | 30                             | 3                              |
| Group 1            | Y196A                             | High $^c$        | 230                            | 350                            |
|                    | Y193A                             | High $^c$        | 120                            | 180                            |
|                    | Y230A                             | High $^c$        | 110                            | 120                            |
| Group 2            | Y233A                             | 80               | 60                             | 20                             |
|                    | Y199A                             | 70               | 30                             | 30                             |
|                    | Y185A                             | 70               | 15                             | 20                             |
|                    | Y163A                             | 60               | 20                             | 2                              |
| Group 3            | Y164A                             | 40               | 6                              | 30                             |

$^a$ Apparent $K_d$ values of the wild-type and mutant peptide-substrate-binding domains for (Pro-Pro-Gly)$_{10}$ and poly-L-proline, $M$, 5000–8000, were determined by isothermal titration calorimetry (14). Similar values were obtained in at least two independent experiments.

$^b$ Similar $K_a$ values of the wild-type and mutant type I C-P4H for (Pro-Pro-Gly)$_{10}$ and IC$_{50}$ values for poly-L-proline, $M$, 5000–8000, were obtained in at least three independent experiments.

$^c$ The value could not be determined accurately because of weak affinity.

**Fig. 5.** Non-denaturing PAGE analysis of C-P4H tetramer formation from the wild-type or mutant α(I) subunits and the wild-type PDI polypeptide (β-subunit) expressed in insect cells by means of recombinant baculoviruses. The cells were homogenized in a buffer containing Triton X-100, and the supernatants were analyzed by nondenaturing PAGE followed by Coomassie Blue staining. Mutant enzymes are indicated by the numbers of the α(I)-subunit tyrosine residues converted to alanine and are shown in the order of importance in the binding of proline-rich peptides (Groups 1–3) defined under “Results” and in Table II. The arrow indicates the position of the $\alpha_1\beta_2$ enzyme tetramer.

values for poly-L-proline, ranging from 120 to 210 $\mu$M, i.e., about 40–70 times the IC$_{50}$ of the wild-type enzyme. The four Group 2 mutants again had intermediate values, whereas the IC$_{50}$ of the Group 3 Y164A mutant was identical to that of the wild-type enzyme (Table II). The $V_{\text{max}}$ values of all the mutant C-P4H-I tetramers were essentially identical to that of the wild-type enzyme (data not shown).

### DISCUSSION

All attempts to crystallize a vertebrate C-P4H $\alpha_1\beta_2$ tetramer or its β-subunit, i.e., the PDI polypeptide, have so far been unsuccessful, but the structures of three of the four PDI domains have been resolved by NMR and shown to have the thioredoxin fold (47–49). The full-length catalytic α-subunit is not amenable to structural studies as such, as it is totally insoluble and non-functional unless assembled with PDI (1–3). Our data on the peptide-substrate-binding domain provide the very first structural information on the catalytic α-subunit of any C-P4H. Structural information on this subunit could potentially have a major impact on the rational design of inhibitors for the treatment of fibrotic diseases caused by excessive collagen accumulation, as they constitute a major problem in medicine and inhibition of C-P4Hs is regarded as a particularly suitable approach for antifibrotic therapy.

The crystal structure of the C-P4H $\alpha(I)$ peptide-substrate-binding domain reported here shows that it is an entirely helical protein consisting of five antiparallel α-helices. Previous NMR characterization of the domain likewise indicated that it consists of five α-helices and one short putative β-strand between helices α2 and α3 (14). The crystal structure shows, however, that residues 179–186 form a loop instead of a β-strand (Fig. 2). The lengths of helices α2, α3, and α4 are identical to those determined by NMR (14), whereas α1 starts one residue earlier, from Ala$^{147}$. The N terminus of α5 could not be determined by NMR because of missing assignments, but based on the predicted secondary structure it was expected to start from His$^{221}$ (14). The crystal structure shows, however, that it starts one residue later, from Glu$^{222}$.

Our structural data indicate that the C-P4H α(I) peptide-substrate-binding domain belongs to the family of TPR domains, consisting of two TPR motifs plus the solvating helix. The most prominent surface feature of the domain is a deep groove on its concave side, lined with tyrosine residues. All eight tyrosines of the domain are at or near its surface, and none of them belongs to the TPR consensus sequence. At least one tyrosine, Tyr$^{164}$, is far away from the proposed binding groove and completely exposed to solvent, whereas Tyr$^{163}$ and Tyr$^{199}$ are much closer to the other five tyrosines, Tyr$^{237}$, Tyr$^{239}$, Tyr$^{196}$, Tyr$^{158}$, and Tyr$^{193}$, all of which participate directly in shaping the aromatic groove. Because the structure of the domain in itself does not indicate which surface features are important for binding of the peptide substrates and which features are important for the assembly of the domain in the full-length tetramer, it is important to note that all eight tyrosine to alanine mutant full-length α-subunits formed the $\alpha_1\beta_2$ tetramer with the PDI polypeptide. Thus, none of the tyrosines was critical for tetramer assembly. Consequently, the mutagenesis data strengthen the hypothesis that the complete groove may also be important for peptide binding in the C-P4H tetramer. This hypothesis is further strengthened by the finding that the binding properties of the domain and the tetramer were highly similar. Furthermore, the only mutant that had essentially no effect was Y164A, which concerns a solvent-exposed side chain far away from the binding groove. The combined mutational data suggest that the three Group 1 residues, Tyr$^{196}$, Tyr$^{193}$, and Tyr$^{230}$, are very important for efficient binding, whereas the four Group 2 residues, Tyr$^{237}$, Tyr$^{199}$, Tyr$^{158}$, and Tyr$^{163}$, are of intermediate importance. This result strongly implicates the deep binding groove be-
between Tyr\textsuperscript{233} and Tyr\textsuperscript{193} (Fig. 4) in the binding of the proline-rich peptides, such that Tyr\textsuperscript{196} and Tyr\textsuperscript{193} are the most important sites for interaction. It is of interest that these two tyrosines, like Asp\textsuperscript{192} and Tyr\textsuperscript{196}, protrude out of helix α3. Previous NMR studies also showed that binding of (Pro-Pro-Gly)\textsubscript{2} had the greatest effect on the environment of the main chain of this helix (14). Of the seven tyrosines implicated as being important for peptide binding, only Tyr\textsuperscript{199} and Tyr\textsuperscript{160} do not line the aromatic groove. Because the side-chain hydroxyls of Tyr\textsuperscript{199} and Tyr\textsuperscript{230} are hydrogen-bonded to each other (Fig. 4), Tyr\textsuperscript{199} may be important for fixing Tyr\textsuperscript{230} in the right position. Tyr\textsuperscript{196} sits in a surface groove near Tyr\textsuperscript{199} and Tyr\textsuperscript{158} (Fig. 4), and consequently the Y163A mutation may induce structural rearrangements near Tyr\textsuperscript{196} and Tyr\textsuperscript{158} that weaken the peptide-protein interactions. Altogether, the mutagenesis data agree very well with the structure-based hypothesis that the proline-rich peptides become bound to the deep groove on the concave surface of the domain.

Previous analyses of conformational features of synthetic peptide substrates and their interaction with the C-P4H tetramer have suggested that a minimal structural requirement for proline 4-hydroxylation may be a sequence in the PPII helix (50). The PPII helix conformation may be necessary for effective interaction at the substrate binding sites, whereas the β-turn may be essential for hydroxylation at the catalytic site (50). Interestingly, the structure of the peptide-substrate-binding domain shows that its concave surface contains three parallel, solvent-exposed side chains of Tyr\textsuperscript{233}, Tyr\textsuperscript{230}, and Tyr\textsuperscript{196} (Fig. 4) with a repeat distance similar to that of a PPII helix. Furthermore, the aromatic surface ends at Tyr\textsuperscript{158} (Fig. 4) where the groove curves almost 90° away from the linear arrangement of the three tyrosine side chains, possibly inducing a bent conformation in the bound peptide.

Crystallographic binding studies of the domain have been initiated, but the quality of the current crystals deteriorates greatly when they are soaked with several proline-rich peptides. In this crystal form the side chains of Tyr\textsuperscript{233} and Tyr\textsuperscript{230} are both involved in crystal contacts in both molecules. Although the observed crystal damage further confirms the importance of these tyrosines for the binding of proline-rich peptides, it also shows that a further search for suitable crystallization conditions for the peptide-domain complex is required to study the interactions in detail. The present data highlight the importance and versatility of the TPR unit for the formation of peptide-protein complexes, and it also emphasizes the importance of aromatic residues for the formation of binding grooves for proline-rich peptides.

Acknowledgments—We gratefully acknowledge the skilful support of the staff at the synchrotron beamlines I11 at the Maxlab (Lund), providing the 2.3 Å native data set and BW7A, at the European Molecular Biology Laboratory outstation (German Synchrotron Research Centre, Hamburg), providing the multiwavelength anomalous dispersion data set. We thank Tanja Aatsinki, Raija Juntunen, Eeva Lehtimäki, Merja Nissila, Ville Ratase, and Eeva-Liisa Stefanius for excellent technical assistance, Petri Kursula for important contributions at the initial stages of this work, and Kristian Koski and Jukka Taskinen for help with preparation of the figures.

REFERENCES

1. Kivirikko, K. I., and Nilsson, B. (1972) Adv. Enzymol. Relat. Areas Mol. Biol. 37, 325–398
2. Kivirikko, K. I., and Myllyharju, J. (1991) Matrix Biol. 10, 357–368
3. Myllyharju, J. (2003) Matrix Biol. 22, 15–24
4. Bruick, R. K., and McKnight, S. L. (2001) Science 294, 1337–1340
5. Epstain, A. C. R., Gleedale, J. M., McNeill, L. A., Hewitson, K. S., O’ourke, J., Mole, D. R., Mukherji, M., Metzen, E., Wilson, M. I., Dhanda, A., Tian, Y.-M., Masson, N., Hamilton, D. L., Jayakumar, P., Barstead, R., Hodgkin, J., Maxwell, P. H., Pugh, C. W., Schofield, C. J., and Ratcliffe, P. J. (2001) Cell 107, 43–54
6. Ivan, M., Haberberger, T., Gervasi, D. C., Michelsen, K. S., Günzler, V., Kondo, K., Yang, H., Sorokin, I., Conaway, R. C., Conaway, J. W., and Kaelin, W. G., Jr. (2002) Proc. Natl. Acad. Sci. U. S. A. 99, 13459–13464
7. Kukkola, L., Hieti, R., Myllyharju, K. I., and Myllyharju, J. (2003) J. Biol. Chem. 278, 47685–47693
8. Van Den Dresder, C., Papay, B., Bolender, Z., Brown, A., and Pickering, G. (2003) Circulation 108, 508–511
9. Annunen, P., Autio-Harjanne, H., and Kivirikko, K. I. (1991) J. Biol. Chem. 266, 5980–5992
10. Niss, R., Autio-Harjanne, H., Marttila, P., Sormunen, R., and Kivirikko, K. I. (2001) J. Histochem. Cytochem. 49, 1143–1153
11. Myllyharju, J., and Kivirikko, K. I. (1997) EMBO J. 16, 1173–1180
12. Myllyharju, J., and Kivirikko, K. I. (2004) Trends Genet. 20, 53–43
13. Myllyharju, J., and Kivirikko, K. I. (1999) EMBO J. 18, 306–312
14. Hieta, R., Kukkola, L., Permi, P., Pirilä, P., Kivirikko, K. I., Kilpeläinen, I., and Myllyharju, J. (2003) J. Biol. Chem. 278, 34966–34974
15. Kay, B., Williamson, M. P., and Sudol, M. (2000) FEBS Lett. 414, 231–241
16. Macias, M. J., Wiesner, S., and Sudol, M. (2002) FEBS Lett. 513, 30–37
17. Prehoda, K. E., Lee, D. J., and Lim, W. A. (1999) Cell 97, 471–480
18. Bell, D. J., Jarchau, T., Oeschkinat, H., and Walter, U. (2002) FEBS Lett. 513, 45–52
19. Freund, C., Kühne, R., Yang, H., Park, S., Reinerz, E., and Wagner, G. M. (2002) EMBO J. 21, 5884–5895
20. Mahoney, N. M., Rozanski, D. A., Fedorov, E., Fedorov, A. A., and Almo, S. C. (1999) Nat. Struct. Biol. 6, 666–671
21. Nodelman, I. M., Bowman, G. D., Lindberg, U., and Schutt, C. (1999) J. Mol. Biol. 294, 1271–1285
22. Pernillos, O., Alam, S. L., Davis, R., and Sundquist, W. J. (2002) Nat. Struct. Biol. 9, 812–817
23. Pekkola, M., Hieti, R., Kurusla, P., Kivirikko, K. I., Vieregna, R. K., and Myllyharju, J. (2003) Acta Crystallogr. Sect. D Biol. Crystallogr. 59, 940–942
24. Main, E. R. G., Xiong, Y., Coco, M. J., d’Andrea, L., and Regan, L. (2003) Structure 11, 497–508
25. d’Andrea, L., and Regan, L. (2003) Trends Biochem. Sci. 28, 655–662
26. Gatto, G. J., Jr., Geisbrecht, B. V., Gould, S. J., and Berg, J. M. (2000) Nat. Struct. Biol. 7, 1091–1095
27. Schreiber, C., Brinker, A., Beurenken, G., Pegoraro, S., Moroder, L., Bartunik, H., Hartl, F. U., and Moarefi, I. (2000) Cell 101, 199–210
28. Kivirikko, R. I., and Myllyla, M. (1982) Methods Enzymol. 82, 245–304
29. Myllyla, R., Tuderman, L., and Kivirikko, K. I. (1977) Eur. J. Biochem. 80, 349–357
The Peptide-Substrate-binding Domain of Collagen Prolyl 4-Hydroxylases Is a Tetratricopeptide Repeat Domain with Functional Aromatic Residues

Mira Pekkala, Reija Hieta, Ulrich Bergmann, Kari I. Kivirikko, Rik K. Wierenga and Johanna Myllyharju

*J. Biol. Chem.* 2004, 279:52255-52261.
doi: 10.1074/jbc.M410007200 originally published online September 28, 2004

Access the most updated version of this article at doi: 10.1074/jbc.M410007200

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 49 references, 11 of which can be accessed free at
http://www.jbc.org/content/279/50/52255.full.html#ref-list-1