The Peptide-Substrate-binding Domain of Human Collagen Prolyl 4-Hydroxylases

BACKBONE ASSIGNMENTS, SECONDARY STRUCTURE, AND BINDING OF PROLINE-RICH PEPTIDES*

Reija Hietan, Liisa Kukkola, Perttu Peräniemi, Päivi Pirilä, Kari I. Kivistö, Ilkka Kilpeläinen, and Johanna Myllyharju

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

The collagens prolyl 4-hydroxylases (C-P4Hs) catalyze the formation of 4-hydroxyproline by the hydroxylation of proline residues in Xaa-Pro-Gly-sequences. The vertebrate enzymes are \( \alpha_1 \beta_1 \) tetramers in which protein-disulfide isomerase serves as the \( \beta \) subunit. Two isoforms of the catalytic \( \alpha \) subunit have been identified and shown to form \( \alpha(1) \) and \( \alpha(2) \) tetramers, the type I and type II C-P4Hs, respectively. The peptide-substrate-binding domain of type I C-P4H has been shown to be located between residues 138 and 244 in the 517-residue \( \alpha(I) \) subunit and to be distinct from the catalytic domain that is located in the C-terminal region. We report here that a recombinant human C-P4H \( \alpha(I) \) polypeptide Phe\(^{144}\)-Ser\(^{244}\) forms a folded domain consisting of five \( \alpha \) helices and one short \( \beta \) strand. This structure is quite different from those of other proline-rich peptide-binding modules, which consist mainly of \( \beta \) strands. Binding of the peptide (Pro-Pro-Gly)\(_x\) to this domain caused major chemical shifts in many backbone amide resonances, the residues showing the largest shifts being mainly hydrophobic, including three tyrosines. The \( K_d \) values determined by surface plasmon resonance and isothermal titration calorimetry for the binding of several synthetic peptides to the \( \alpha(I) \) and the corresponding \( \alpha(II) \) domain were very similar to the \( K_a \) and \( K_v \) values for these peptides as substrates and inhibitors of the type I and type II C-P4H tetramers. The \( K_a \) values of the \( \alpha(I) \) and \( \alpha(II) \) domains for (Gly-Pro-4Hyp)\(_x\) were much higher than those for (Pro-Pro-Gly)\(_x\), indicating a marked decrease in the affinity of hydroxylated peptides for the domain. Many characteristic features of the binding of peptides to the type I and type II C-P4H tetramers can thus be explained by the properties of binding to this domain rather than the catalytic domain.

The prolyl 4-hydroxylases (P4Hs)\(^1\) 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, the resulting 4-hydroxyproline residues being essential for assembly of the triple-helical molecules (1–3). The hypoxia-inducible factor (HIF) P4Hs, a family of cytoplasmic enzymes (4–6), play a key role in the response of cells to hypoxia by catalyzing hydroxylation of the \( \alpha \) subunit of HIF. This subunit is synthesized continuously, and at least one of two critical proline residues becomes hydroxylated under normoxic conditions, the resulting 4-hydroxyproline being essential for rapid degradation of HIF-\( \alpha \) (7–9). In hypoxia this hydroxylation ceases, HIF-\( \alpha \) forms a dimer with HIF-\( \beta \), and the dimer then becomes bound to the HIF-responsive elements in a number of hypoxia-inducible genes.

All vertebrate C-P4Hs are \( \alpha_1 \beta_1 \) tetramers in which the \( \beta \) subunit is identical to the enzyme and chaperone protein-disulfide isomerase (1–3). Two isoforms of the catalytic \( \alpha \) subunit have been characterized from human and mouse tissues and shown to form \( \alpha(I)_x \) and \( \alpha(II)_x \) tetramers, called the type I and type II C-P4Hs, respectively (10, 11). The type I enzyme is the most abundant form in most cells, but type II is the main form in chondrocytes, endothelial cells, and some other cell types (12, 13). The HIF-P4Hs appear to consist of only one type of monomer, the size of which ranges from 239 to 426 residues in the three human isoenzymes (4–6). The C-P4Hs act on Xaa-Pro-Gly- triplets in collagens and more than 15 other proteins with collagen-like sequences (1–3, 14), whereas the HIF-P4Hs hydroxylate Leu-Xaa-Xaa-Leu-Ala-Pro-Tyr- and Leu-Xaa-Xaa-Leu-Ala-Pro- sequences (7–9). All the P4Hs require Fe\(^{2+}\), 2-oxoglutarate, O\(_2\), and ascorbate, the 2-oxoglutarate being stoichiometrically decarboxylated during hydroxylation (1–3). The C-terminal regions of the C-P4H \( \alpha \) subunits and the HIF-P4H monomers contain four conserved residues, two histidines, and one aspartate that bind the Fe\(^{2+}\) atom and a basic residue that binds the C-5 carboxyl group of the 2-oxoglutarate (15, 16). This basic residue is a lysine in all C-P4Hs, whereas it is an arginine in the HIF-P4Hs (4, 5, 16).

\(^*\) This work was supported by Grants 45601 and 200471 from the Health Science Council, Grant 44843 from the Finnish Centre of Excellence Programme 2000–2005 of the Academy of Finland, and grants from Fibrogen Inc. (South San Francisco, CA) (to J. M.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact. ** To whom correspondence should be addressed: Collagen Research Unit, Biocenter Oulu and Dept. of Medical Biochemistry and Molecular Biology, P. O. Box 5000, University of Oulu, FIN-90014 Oulu, Finland. Tel.: 358-8-5375740; Fax: 358-8-5375811; E-mail: johanna.myllyharju@oulu.fi

This is an Open Access article under the CC BY license.

\(^1\) The abbreviations used are: P4H, prolyl 4-hydroxylase, C-P4H, collagen P4H; HIF, hypoxia-inducible factor; HSQC, heteronuclear single-quantum coherence; SPR, surface plasmon resonance; ITC, isothermal titration calorimetry; Biz-Tris, 2-bis(2-hydroxyethylamino)-2-hydroxyethylpropane-1,3-diol; tricine, N-[2-hydroxy-1,1-bis(hydroxymethyl)-ethyl]glycine.
The peptidyl-prolyl cis-trans isomerase activity of the C-P4Hs has recently been shown to be separate from the catalytic domain and to be located between residues 138 and 244 in the pI7-residue human α(I) subunit (17). Its sequence is distinct from those of other proline-rich peptide-binding modules such as the Src homology 3, WW, vav/vaso-oligator-stimulated phosphoprotein, and GYP domains (18-21) and from that of the profilins (22-25). Polypeptide-binding (Pro-Pro-Gly) is an active competitive inhibitor of the type I C-P4H but only a very weak inhibitor of the type II C-P4H, there also being smaller differences between the two isoenzymes in their binding properties with respect to peptide substrates (10, 11, 17). The presence of a glutamate and glutamine in the peptide-substrate-binding domain of the α(I) subunit in the positions corresponding to Ile312 and Tyr393 in the α(II) subunit was found to explain most of these differences (17).

We report here on NMR data indicating that a recombinant human C-P4H α(I) subunit polypeptide Phe144-Ser244 forms a folded domain consisting of five α helices and one short β strand, its secondary structure being distinct from those of other proline-rich peptide-binding modules. Binding of a peptide substrate to this domain was found to cause significant shifts in many backbone amide resonances. The values for this domain and the corresponding α(II) subunit domain for peptide substrates and inhibitors were similar to the Kd and Ki values of these peptides for the type I and type II enzyme tetramers, suggesting that this domain plays a critical role in the binding of peptide substrates to the enzyme.

**EXPERIMENTAL PROCEDURES**

**Protein Expression and Purification**—The recombinant human α(I) Phe144-Ser244 polypeptide was prepared by amplifying a cDNA fragment encoding residues Phe144-Ser244 with an NdeI site and a translation start codon preceding the codon for Phe144 and an XhoI site following the codon for Ser244, using Pfu polymerase (Promega). The template was pET15b-α(I) Gly138-Ser244 (17) in which the codon for Cys138 (TGC) was converted to that for serine (TCC). The C150S mutagenesis was performed using a QuickChange™ site-directed mutagenesis kit according to the manufacturer's instructions (Stratagene). The recombinant human α(II) Met142-Ser242 polypeptide was prepared by amplifying a cDNA fragment encoding residues Met142-Ser242 with an NdeI site preceding the codon for Met142 and an XhoI site following the codon for Ser242, using full-length human C-P4H α(II) cDNA (11) as a template. The PCR for Cys142 (TGC) was converted to that for serine (TCC) in the same PCR amplification step. The PCR products were cloned into NdeI- XhoI-digested pET-22b expression vectors (Novagen) in-frame with the translation start codon preceding the codon for Phe144 and an N-terminal 6 × His tag. Poly(L-proline) binding of the recombinant domains was tested by incubating the pure 13-kDa recombinant protein were pooled, and all experiments were performed using a VP-ITC titration microcalorimeter (Microcal, Northampton, MA). In the individual titrations, 5-μl doses of peptide solutions in 2.5-10 mM concentrations (in 10 mM sodium phosphate buffer, pH 6.8) were injected into a sample cell (volume 1.4476 ml).
containing 0.25–0.5 mM of the peptide-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 one set of sites model (Levenberg-Marquardt algorithm) to determine the dissociation constant \( K_d \), binding stoichiometry \( n \), and change in enthalpy \( \Delta H \) using the Origin software (Microcal).

Other Assays—P4H activity was assayed by a method based on the hydroxylation-coupled decarboxylation of 2-oxo[1-\(^{14}\)C]glutarate (34), and \( K_m \) values were determined as described previously (16).

RESULTS

Optimization of the Recombinant \( \alpha(I) \) Polypeptide—Initial experiments were performed using a purified recombinant human C-P4H \( \alpha(I) \) subunit polypeptide Gly\(^{138}\)-Ser\(^{244} \) (17) in which the only cysteine (Cys\(^{156}\)) was mutated to serine. The quality of the NMR spectra obtained with this \(^{15}\)N-labeled polypeptide was not satisfactory, however, and therefore further experiments were performed with recombinant polypeptides lacking a few residues from the N- or C-terminal end (details not shown). A Phe\(^{144}\)-Ser\(^{244} \) polypeptide was found to give the best spectra and was therefore used for all the experiments reported here. It was purified to homogeneity based on its histidine tag. Limited proteolysis experiments of this purified domain with trypsin (17) demonstrated that it is protease-resistant, except that a minor cleavage product was identified by SDS-PAGE. N-terminal sequencing showed that the major and minor bands had intact N termini. Mass spectrometry analysis indicated that the minor product corresponded to a polypeptide that had lost its C-terminal serine and the histidine tag (details not shown).

CD Analysis of the \( \alpha(I) \) Peptide-binding Domain—Far-UV CD experiments sensitive to protein secondary structure were performed to study the fold and thermal stability of the recombinant \( \alpha(I) \) polypeptide. The CD spectra showed a typical pattern of \( \alpha \)-helical proteins, with negative minima at 208 and 222 nm and a positive maximum at 193 nm (Fig. 1A). A thermal scan from 5 to 95 °C demonstrated that the recombinant polypeptide displays a continuous, gradual decrease in helical content (Fig. 1B) that is almost completely reversible (Fig. 1, A and B). The denaturation data indicate that the recombinant

![Fig. 1. Circular dichroism analysis of the recombinant \( \alpha(I) \) peptide-binding domain. Measurements were performed using 0.1 mg/ml peptide-binding domain in 10 mM sodium phosphate (pH 6.8) in the absence (A and B) and presence (C and D) of 100 mM urea. A and C, the far-UV spectra of a native (black curves) and a renatured (gray curves) domain show negative bands at 208 and 222 nm, typical of \( \alpha \)-helical proteins. Urea causes interference in the spectra below 200 nm, and therefore the curves are not shown in C below 195 nm. B and D, mean residue ellipticity at 222 nm as a function of temperature (T), measured during denaturation from 5 to 95 °C (black curves) and renaturation (gray curves).](image-url)
domain does not have a very compact tertiary structure and behaves more like proteins in a molten globule state. CD spectra were also recorded for a sample containing urea at the same concentration as used in NMR measurements, 100 mM. The shape of the far-UV curve remained unchanged, but a small increase in ellipticity was detected, indicating a slight increase in the compactness of the fold (Fig. 1C). The difference in the ellipticity between the native and thermally denatured proteins was likewise 20% greater in the urea sample than in the sample without urea (Fig. 1, B and D).

**NMR Assignments of the α(I) Peptide-binding Domain—**Uniformly 15N- and 13C/15N-labeled samples of the α(I) peptide-binding domain were prepared for NMR studies, the 1H-15N HSQC spectrum of the 15N-labeled domain being shown in Fig. 2. The spectra of the 15N/13C double-labeled samples had broader lines and poorer dispersion than those of the 15N-labeled samples, but the quality of the spectra could be improved by adding 100 mM urea to the samples. All the spectra were acquired at 15 °C because of the low thermal stability of the domain revealed by the CD and preliminary NMR studies. Backbone assignments were obtained based on the intraregional and sequential Ca/Cβ connectivities observed in the three-dimensional NMR experiments HNCA, HN(CO)CA, CBCA(CO)NH, and CBCA(CO)NH with the exception of the first few N-terminal residues, part of the histidine tag in the C terminus, and the region between the residues Pro219 and Ghu232, where assignments could not be obtained because of spectral overlapping.

The Ca and Cβ chemical shift indices (35) indicate that the α(I) peptide-binding domain is composed of five α helices and a short β strand, Ile182-Asp186, between the second and third helices (Fig. 3). The first four α helices consist of residues Glu148-Thr159, Tyr163-Asp178, Lys187-Gln200, and Leu204-Leu217 (Fig. 3), whereas the fifth is located in the C terminus of the domain, but its length could not be determined because of the missing assignments. Predictions based on the amino acid sequence (36) suggest, however, that this helix may begin at His221. As the predicted locations and sizes of the other four α helices are in good agreement with the data obtained here, the N terminus of helix 5 is shown in Fig. 3 based on such predictions. All these data are in good agreement with the measured dipolar couplings (data not shown).

**NMR Studies of Peptide Binding—**To obtain information on the interaction between the α(I) peptide-binding domain and peptide substrates, NMR was used to study binding of the synthetic peptide (Pro-Pro-Gly)2. Chemical shift changes in backbone amide 1H and 15N resonances were followed by means of 1H-15N-HSQC spectra as the peptide concentration was increased. Several resonances showed significant chemical shift perturbation as a result of the peptide binding (Figs. 4 and 5). The 15 residues showing the largest chemical shift changes were mainly hydrophobic, including 3 tyrosines (Tyr139, Tyr143, and Tyr165), there being only 1 charged residue among the 15 most affected ones (Fig. 5). 6 of the 15 residues are located in helix 3, 3 in helix 4, 2 in helix 1, 1 in helix 2, 1 in the loop between helices 1 and 2, and 1 in the short β strand. None of the 15 residues are located in helix 5, but even this helix had residues showing chemical shift changes exceeding 0.1 ppm, its 3 residues showing the highest chemical shifts being Ile234, Met235, and Tyr233 (Fig. 5A). The residues with chemical shift changes exceeding 0.25 ppm were selected for determination of the dissociation constant, Kd, for (Pro-Pro-Gly)2. The chemical shift changes of these residues were plotted against peptide concentration, and the binding curves were then fitted individually by optimizing the values of the dissociation constant and the maximum chemical shift change. The dissociation constant for (Pro-Pro-Gly)2 obtained with this method was 2600 ± 500 μM (Fig. 5B, Table I).

**SPR Studies of Peptide Binding—**The interaction of the α(I) and the corresponding α(II) peptide-binding domains with synthetic peptides of different lengths was studied by real-time biomolecular interaction analysis using the BLAcore biosensor system, which detects changes in the refractive index at the surface of a sensor chip (37). The peptides became bound to the immobilized domain on the chip with very high on and off rates, giving a square-wave type of binding curve indicative of a low binding affinity (Fig. 6A). Reliable determination of the association and dissociation kinetic rate constants (k on and k off) was thus not possible. Equilibrium dissociation constants (K d) could nevertheless be determined for the peptides by plotting the
steady state values for each peptide analyte concentration as a function of total analyte concentration (Fig. 6B, Table I).

The $K_d$ of the $\alpha(I)$ peptide-binding domain for the peptide (Pro-Pro-Gly)$_2$ was about 1 mM, this value being in relatively good agreement with the $K_d$ of 2.6 mM measured by NMR and the $K_m$ of about 2.6 mM measured for this peptide as a substrate of the type I C-P4H tetramer (Table I). The $K_d$ values measured by SPR decreased with peptide chain length, with the changes being similar to those seen in the $K_m$ values for the same peptides as substrates for the type I C-P4H (Table I). The $K_d$ of the $\alpha(I)$ domain and $K_m$ of the type I enzyme for (Pro-Pro-Gly)$_{10}$ were highly similar, 30 and 20 $\mu$M, respectively (Table I). The SPR measurements further indicated that the $K_d$ of the $\alpha(I)$ domain for (Gly-Pro-4Hyp)$_2$, more than one order of magnitude higher than that for (Pro-Pro-Gly)$_5$ (Table I). The $K_d$ of 4 $\mu$M for poly(L-proline), $M_r = 5000$, was identical to the $K_d$ measured for the enzyme tetramer (Table I).

The $K_d$ of the $\alpha(II)$ peptide-binding domain for the peptide (Pro-Pro-Gly)$_2$ was much higher than that of the $\alpha(I)$ domain and was too high to be determined accurately (Table I). The $K_d$ values of the $\alpha(II)$ domain for (Pro-Pro-Gly)$_5$ and (Pro-Pro-Gly)$_{10}$ were likewise much higher than those of the $\alpha(I)$ domain, with these values being roughly similar to the $K_m$ values of the type II C-P4H tetramer for the same peptides (Table I). The $K_d$ of the $\alpha(II)$ domain for (Gly-Pro-4Hyp)$_2$ was again distinctly higher than that for (Pro-Pro-Gly)$_5$ (Table I). The $K_d$ of the $\alpha(II)$ domain for poly(L-proline), $M_r = 5000$, was much higher than that of the $\alpha(I)$ domain (Table I).

**ITC Studies of Peptide Binding**—Peptide binding to the recombinant $\alpha(I)$ domain was also investigated using isothermal titration calorimetry. Fig. 7 shows the result of an ITC experiment in which the $\alpha(I)$ peptide-binding domain was titrated with the peptide (Pro-Pro-Gly)$_{10}$. The $K_d$ values obtained with this technique were in relatively good agreement with those measured by SPR (Table II). The changes in enthalpy observed upon binding of different peptides were negative indicating exothermic interaction, the change correlating with the increase in peptide length and affinity (Table II).

**DISCUSSION**

The peptide-substrate-binding domain studied here was initially identified as a fragment of a $\alpha(I)$ subunit that became bound to poly(L-proline) after limited proteolysis of the human type I C-P4H tetramer by proteinase K, thermolysin, or trypsin (17). Further experiments demonstrated that a recombinant polypeptide corresponding to the $\alpha(I)$ domain became effectively bound to poly(L-proline) agarose and could be eluted with (Pro-Pro-Gly)$_{10}$ (17). This domain could not be identified simply by sequence criteria, but this region is followed by a region of about 20 residues with a particularly low degree of identity between the $\alpha(I)$ and $\alpha(II)$ subunits suggesting the presence of a variable interdomain region between this domain and the subsequent domain (17).

The data reported here indicate that the recombinant human $\alpha(I)$ polypeptide Phe$_{141}$-Ser$_{244}$ forms a folded domain consisting of five $\alpha$ helices and one short $\beta$ strand. This structure is quite different from those of the other proline-rich peptide-binding domains SH3, WW, EVH1, and GYF (18–21) and from that of...
the proline-rich peptide-binding protein profilin (22, 23), as they all consist mainly of α strands. The other proline-rich peptide-binding modules typically have a hydrophobic path, with critical aromatic residues that are required for the binding of their ligands (18–23). The peptide-binding domain of the C-P4Hs also appears to have these characteristic features, as the 15 residues showing the largest chemical shift changes in NMR analysis upon interaction with (Pro-Pro-Gly)_2 were mainly hydrophobic, including 3 tyrosines. The residues showing the highest chemical shift changes were found especially in α helices 3 and 4, all three most affected tyrosines being located in helix 3. 11 of the 15 residues with the highest chemical shift changes, including all 3 tyrosines, were found to be conserved or replaced with similar residues in the C-P4H isoenzymes I and II and C-P4Hs from various species (Fig. 8).

Table I

| Peptide     | (I) domain | (II) domain | SPR (Kd) | NMR (Kd) | Constant | Kd or Ki |
|-------------|------------|-------------|----------|----------|----------|---------|
| (Pro-Pro-Gly)_2 | >1000^a | 2600 ± 500 | High^6 | K_m | 2600 | High^6 |
| (Pro-Pro-Gly)_3 | 120 ± 15 | 750 ± 80 | K_m | 170 | 490 |
| (Gly-Pro-4Hyp)_5 | 30 ± 8 | 290 ± 15 | K_m | 20 | 100^c |
| Poly(L-Pro), Mr = 5000 | 4 ± 0.9 | <100^c | K_m | 4 | 170 |

^a The K_d value is at least the value given, but could not be determined accurately due to weak affinity.
^b Too high to be accurately measured.
^c The peptide is not a substrate.
^d Ref. 17.
^e The K_d could not be determined accurately due to the high level of unspecific binding to reference surface at high peptide concentrations.

Fig. 5. NMR analysis of the interaction between the α(I) peptide-binding domain and the peptide (Pro-Pro-Gly)₂. A, chemical shift changes observed in the two-dimensional 1H-15N-HSQC spectra for backbone NH groups at a molar peptide to protein ratio of 20:1. The deviations (in ppm) were quantified with the formula δ = √((ΔHN² + ΔN²)) where ΔHN and ΔN are the chemical shift changes (in ppm) of the amide proton and nitrogen, respectively. B, examples of curve fits used for K_d determinations. Weighted chemical shift changes in four residues, Leu₁₅₃ (○), Val₁₉₈ (●), Leu₂₀₈ (■), and Leu₂₁₄ (▲) are plotted against (Pro-Pro-Gly)₂ concentration. The solid lines represent non-linear least squares best fits to the data. All curves were fitted individually by optimizing the values for the maximum chemical shift change and the dissociation constant. The dissociation constant obtained by this method was 2600 µM, calculated as a mean from eight fitted curves.
marked difference in poly(L-proline) binding between the type I and type II C-P4H tetramers can be explained by the presence of a glutamate and glutamine in the subunit in the positions corresponding to those of Ile182 and Tyr233 in the subunit (17). The same amino acid substitutions also explain most of the difference in the binding of peptide substrates, but the difference in the $K_m$ values for (Pro-Pro-Gly)$_{10}$ between the type I and type II C-P4Hs is only 5-fold, whereas that in the $K_i$ values for poly(L-proline), Mr $44,000$, is more than 1000-fold (11). The present NMR analysis indicates that Ile182 showed a negligible chemical shift change of about 0.05 ppm upon interaction with (Pro-Pro-Gly)$_2$, whereas Tyr233 showed a modest change of about 0.1 ppm. It therefore seems possible that these two residues may not be directly involved in the binding of the glutamate and glutamine residues. Nevertheless, it is also possible that (Pro-Pro-Gly)$_2$ was too short to interact with Tyr233 and thus a direct contribution of this residue to the binding of longer peptides and especially of poly(L-proline) cannot be excluded. The (I) domain contains several other conserved aromatic residues in addition to Tyr193, Tyr196, Tyr199, and Tyr233 (Fig. 8), but all except two of them showed only very small chemical shift changes (less than 0.03 ppm) upon the binding of (Pro-Pro-Gly)$_2$, the two exceptions being Phe151 and Trp169, with changes of about 0.11 ppm in both cases (Fig. 5).

The $K_m$ values of the C-P4Hs for peptide substrates and the $K_i$ values for poly(L-proline) decrease markedly with increasing chain length of the substrate or inhibitor (1). These effects are not because of the changes in the N and C termini of the peptides as such charge effects are negligible for substrates longer than hexapeptides and as the length dependence is also seen with peptides having protected N and C termini (38). The length dependence cannot be explained by structural properties such as a higher tendency of longer peptides to form triple helices either, as the C-P4Hs act only on non-triple-helical peptides (1–3). To explain the length dependence mechanisms have been proposed in which the two peptide-binding sites in the enzyme tetramer act processively and bind to the same peptide, which would prevent dissociation of the enzyme-substrate complex between successive interactions with a long peptide having multiple substrate sites (1, 39). The present data indicate, however, that the effects of peptide chain length are seen even with the individual peptide-binding domains; the
$K_d$ values determined for various peptides being very similar to their $K_a$ and $K_c$ values for the enzyme tetramers. It is thus obvious that longer peptides become bound to the peptide-binding domain more effectively than shorter ones even when the domain is not present in the enzyme tetramer and in the absence of any catalysis. The great similarities between the $K_d$ values measured for the recombinant domains and the $K_a$ and $K_c$ values determined for the enzyme tetramers indicate that it is the peptide-binding domain rather than the catalytic domain that appears to be critical for most, if not all, binding properties of various peptide substrates and inhibitors with respect to the enzymes.

The catalytic cycle of P4Hs consists of two half-reactions (1–3), the first half-reaction coinciding with the formation of a ferryl ion bound to His$^{415}$, Asp$^{414}$, and His$^{483}$ (in the case of the human a(I) subunit sequence), while the ferryl ion then hydroxylates a proline residue during the second (1–3). The C-4 of the proline to be hydroxylated must thus point toward the ferryl ion at the catalytic site. The present data indicate that formation of the hydroxy group markedly reduces the affinity of the substrate for its binding domain, as demonstrated by the differences in the $K_d$ values for (Pro-Pro-Gly)$_5$ and (Gly-Pro-4Hyp)$_5$. This suggests that a change in the affinity of the hydroxylated substrate to the peptide-binding domain contributes to the release of the product and prevents the binding of hydroxylated sequences to the enzyme.

The present data thus indicate that the sequence of about 100 residues located between residues 140 and 245 in the C-P4H α subunits forms a folded domain consisting mainly of α helices, possessing many hydrophobic and aromatic residues that show significant changes in chemical shifts in NMR analysis upon the binding of peptide substrates. Many of the characteristic properties found in the binding of peptide substrates and inhibitors to the type I and type II C-P4Hs can be explained by the properties of this domain rather than by those of the catalytic domain. We have recently obtained crystals of this domain that diffract to at least 3 Å resolution (46), and attempts to determine the complete crystal-based structure are now in progress.

Acknowledgments—We thank Johan Kemmink, Leanne Lanjoki, Laura Silvennoinen, and Lloyd Ruddock for valuable suggestions and Merja Niissila, Eeva Lehtimäki, and Jaana Träskelin for expert technical assistance.

REFERENCES

1. Kuivila, K. I., and Pihlajaniemi, T. (1998) Adv. Enzymol. Relat. Areas Mol. Biol. 72, 325–398
2. Kuivila, K. I., and Myllyharju, J. (1998) Matrix Biol. 16, 357–368
3. Myllyharju, J. (2003) Matrix Biol. 22, 15–24
4. Epstein, A. C. R., Gleadle, J. M., McNeill, L. A., Hewitson, K. S., O’Rourke, J., Mole, D. R., Mukherji, M., Metzen, E., Wilson, M. I., Dhandia, A., Tian, Y. M., Masson, N., Hamilton, D. L., Jaakkola, P., Barstead, R., Hodgkin, J., Maxwell, P. H., Pugh, C. W., Schofield, C. J., and Ratcliffe, P. J. (2001) Cell 107, 43–54
5. Bruick, R. K., and McKnight, S. L. (2001) Science 294, 1337–1340
6. Ivan, M., Haberberger, T., Gervasi, D. C., Michelson, K. S., Gunzler, 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. Ivan, M., Kondo, K., Yang, H., Kim, W., Valando, J., Ohh, M., Salie, A., Asara, J. M., Lane, W. S., and Kaelin, W. G., Jr. (2001) Science 292, 464–468
8. Jaakkola, P., Mole, D. R., Tian, Y.-M., Wilson, M. I., Gielbert, J., Gaskell, S. J., Kriegsheim, A. V., Hebestreit, H. F., Mukherji, M., Schofield, C. J., Maxwell, P. H., Pugh, C. W., and Ratcliffe, P. J. (2001) Science 292, 468–472
9. Masson, N., Williams, C., Maxwell, P. H., Pugh, C. W., and Ratcliffe, P. J. (2001) EMBO J. 20, 5197–5206
10. Helaakoski, T., Vuori, K., Myllyla, R., Kivirikko, K. I., and Pihlajaniemi, T. (1989) Proc. Natl. Acad. Sci. U. S. A. 86, 4392–4396
11. Annunen, P., Helaakoski, T., Myllyharju, J., Veijola, J., Pihlajaniemi, T., and Kivirikko, K. I. (1997) J. Biol. Chem. 272, 17342–17348
12. Annunen, P., Autio-Harmainen, H., and Kivirikko, K. I. (1998) J. Biol. Chem. 273, 1143–1153
13. Nissi, R., Autio-Harmainen, H., and Kivirikko, K. I. (1998) J. Biol. Chem. 273, 5989–5992
14. Myllyharju, J., and Kivirikko, K. I. (2001) Proc. Natl. Acad. Sci. U. S. A. 98, 60–66
15. Myllyharju, J., and Kivirikko, K. I. (2001) EMBO J. 16, 1173–1180
16. Myllyharju, J., and Kivirikko, K. I. (1999) EMBO J. 18, 30–37
17. Myllyharju, J., and Kivirikko, K. I. (1999) EMBO J. 18, 30–37
18. Kay, B., Williamson, M. P., and Sudol, M. (2000) FASEB J. 14, 231–241
19. Macias, M. J., Wiesner, S., and Sudol, M. (2002) FEBS Lett. 513, 30–37
20. Prehoda, K. F., Lee, D. J., and Lim, W. A. (1999) Proc. Natl. Acad. Sci. U. S. A. 96, 1285–1289
21. Freund, C., Dote, V., Nishizawa, K., Reinherz, E. L., and Wagner, G. (1999) Nat. Struct. Biol. 6, 595–600
22. Mahoney, N. M., Rozwarski, D. A., Fedorov, E., Fedorov, A. A., and Almo, S. C. (1999) Nat. Struct. Biol. 6, 666–671
23. Nodelman, I. M., Bowman, G. D., Lindberg, U., and Schutt, C. (1999) J. Mol. Biol. 294, 1271–1285
24. Ikura, M. J., Kay, E. L., and Bax, A. (1996) Biochemistry 25, 4659–4667
25. Grzesiek, S., and Bax, A. (1992) J. Magn. Reson. 96, 432–440
26. Bax, A., and Ikura, M. J. (1991) J. Biomol. NMR 1, 99–105
27. Wittekind, M., and Mueller, J. (1991) J. Magn. Reson. 101B, 201–205
28. Morris, B., and Bax, A. (1999) J. Am. Chem. Soc. 121, 6281–6282
29. Kay, L. E., Xu, G. Y., and Yamasaki, T. (1994) J. Magn. Reson. 106A, 129–133
30. Goddard, T. D., and Schutt, C. (1999) SPARKY 3, University of California, San Francisco
31. Kay, L. E., Keifer, P., and Sarafin, T. (1992) J. Am. Chem. Soc. 114, 10663–10665
32. Clubb, R. T., Omichinski, J. G., Clore, G. M., and Gronenborn, A. M. (1994) FEBS Lett. 338, 93–97
33. Gehring, W., Sasaki, T., Heldin, C. H., and Timpl, R. (1998) Eur. J. Biochem. 255, 60–66
34. Myllyharju, J., and Myllyla, R. (1982) Methods Enzymol. 82, 245–304
35. Wishart, D. S., and Sykes, B. D. (1994) J. Biomol. NMR 4, 171–180
36. Rost, B. (1996) Methods Enzymol. 266, 525–529
37. Jonsson, U., Fageret, L., Ivarsson, B., Johnson, B., Karlsson, B., Lundh, K., Loftas, S., Persson, B., Roos, H., Ronnberg, I., Sjolander, S., Stenberg, E., Stahlberg, E., Urbaniczky, S., Ostlin, H., and Malmqvist, M. (1991) Bio-Techniques 11, 620–627
38. Kivirikko, K. I., Prokop, D. J., Lorenzi, G. P., and Blout, E. R. (1969) J. Biol. Chem. 244, 2755–2760
39. De Jong, L., van der Kraan, I., and de Waal, A. (1991) Biochim. Biophys. Acta 1097, 103–111
40. Pekkala, M., Hieti, R., Kursula, P., Kivirikko, K. I., Wierenga, R., and Myllyharju, J. (2003) Acta Crystallogr. Sec. D 59, 940–942
41. Helaakoski, T., Annunen, P., Vuori, K., MacNeil, I. A., Pihlajaniemi, T., and Kivirikko, K. I. (1999) Proc. Natl. Acad. Sci. U. S. A. 92, 4427–4431
42. Hopkinson, I., Smith, S. A., Donne, A., Gregory, H., Franklin, T. J., Grant, M. E., and Rossmund, J. (1994) Gene 213, 391–392
43. Bassuk, J. A., Mao, W.-W., Herr, P., Pedersen, N. L., Sayer, J., DeMartino, J. A., Daugherty, B. L., Mark, G. E., III, and Berg, R. A. (1989) Proc. Natl. Acad. Sci. U. S. A. 86, 7382–7386
44. Veijola, J., Koivunen, P., Annunen, P., Pihlajaniemi, T., and Kivirikko, K. I. (1994) J. Biol. Chem. 269, 26746–26753
45. Winter, A. D., and Page, A. P. (2000) Mol. Cell. Biol. 20, 4084–4093
46. Myllyharju, J., Kukko, L., Winter, A. D., and Page, A. P. (2002) J. Biol. Chem. 277, 29187–29196