The EMBO Journal Vol.18 No.2 pp.306–312, 1999

Identification of a novel proline-rich peptide-binding domain in prolyl 4-hydroxylase

Johanna Myllyharju and Kari I.Kivirikko[1]

Collagen Research Unit, Biocenter and Department of Medical Biochemistry, University of Oulu, Kajaanintie 52 A, FIN-90220 Oulu, Finland

[1]Corresponding author
e-mail: kkivirik@lbiok-1.oulu.fi

Prolyl 4-hydroxylase (EC 1.14.11.2) catalyzes the hydroxylation of -X-Pro-Gly- sequences and plays a central role in the synthesis of all collagens. The $[\alpha(I)]_2\beta_2$ type I enzyme is effectively inhibited by poly(L-proline), whereas the $[\alpha(II)]_2\beta_2$ type II enzyme is not. We report here that the poly(L-proline) and $(Pro\text{-}Pro\text{-}Gly)_{10}$ peptide substrate-binding domain of prolyl 4-hydroxylase is distinct from the catalytic domain and consists of ~100 amino acids. Peptides of 10–19 kDa beginning around residue 140 in the 517 residue α(I) subunit remained bound to poly(L-proline) agarose after limited proteolysis of the human type I enzyme tetramer. A recombinant polypeptide corresponding to the α(I) subunit residues 138–244 and expressed in *Escherichia coli* was soluble, became effectively bound to poly(L-proline) agarose and could be eluted with $(Pro\text{-}Pro\text{-}Gly)_{10}$. This polypeptide is distinct from the SH3 and WW domains, and from profilin, and thus represents a new type of proline-rich peptide-binding module. Studies with enzyme tetramers containing mutated α subunits demonstrated that the presence of a glutamate and a glutamine in the α(II) subunit in the positions corresponding to Ile182 and Tyr233 in the α(I) subunit explains most of the lack of poly(L-proline) binding of the type II prolyl 4-hydroxylase.

Keywords: collagen/dioxygenases/peptide-binding domain/proline-rich/prolyl hydroxylase

Introduction

Prolyl 4-hydroxylase (EC 1.14.11.2) catalyzes the formation of 4-hydroxyproline in collagens and >10 additional proteins with collagen-like sequences. This co-translational and post-translational modification has a central role in the synthesis of all collagens, as the hydroxy groups of the 4-hydroxyproline residues are essential for the formation of the collagen triple helix at body temperature (for recent reviews, see Prockop and Kivirikko, 1995; Kivirikko and Myllyharju, 1998; Kivirikko and Pihlajaniemi, 1998).

The prolyl 4-hydroxylases from all vertebrate sources studied are $\alpha_2\beta_2$ tetramers, in which the β subunit is identical to protein disulfide isomerase (PDI, EC 5.3.4.1; Pihlajaniemi *et al*., 1987). Recently an isoform of the α subunit termed the α(II) subunit was cloned and characterized from mouse (Helaakoski *et al*., 1995) and human (Annunen *et al*., 1997, 1998) tissues. Correspondingly, the previously known vertebrate α subunit is now called the α(I) subunit, the $[\alpha(I)]_2\beta_2$ and $[\alpha(II)]_2\beta_2$ tetramers being called the type I and type II enzymes, respectively. Data on coexpression in insect cells strongly argue against the existence of a mixed $\alpha(I)\alpha(II)\beta_2$ tetramer (Annunen *et al*., 1997).

Prolyl 4-hydroxylase requires Fe^{2+}, 2-oxoglutarate, O_2 and ascorbate and acts on proline residues in -X-Pro-Gly- sequences (see Kivirikko and Myllyharju, 1998). Poly(L-proline) is a highly effective competitive inhibitor of the vertebrate type I enzyme but only a very weak inhibitor of the type II enzyme (Helaakoski *et al*., 1995; Annunen *et al*., 1997). Recent site-directed mutagenesis studies and other work have identified five residues, all of which are located in the highly conserved C-terminal region of the α subunit, that are critical at the cosubstrate-binding sites (Myllylä *et al*., 1992; Lamberg *et al*., 1995; Myllyharju and Kivirikko, 1997). His412, Asp414 and His483 [numbered according to the human α(I) subunit sequence] provide the three ligands required for the binding of Fe^{2+}, while Lys493 binds the C-5 carboxyl group of the 2-oxoglutarate (Lamberg *et al*., 1995; Myllyharju and Kivirikko, 1997). His501 is the fifth critical residue, probably being involved in both the orientation of the C-1 carboxyl group of 2-oxoglutarate to the active iron center and in the decarboxylation of this cosubstrate (Myllyharju and Kivirikko, 1997). Based on these and other data it is assumed that the highly conserved C-terminal region of the α subunit, involving ~100–120 residues, represents the catalytic domain of the enzyme tetramer (Kivirikko and Myllyharju, 1998; Kivirikko and Pihlajaniemi, 1998). No data are currently available to indicate whether the binding site for the peptide substrates and their competitive inhibitors such as poly(L-proline) in the case of the type I enzyme is also located in this domain or whether the enzyme has a separate proline-rich peptide-binding domain, nor are any data available to indicate why poly(L-proline) is only a very weak inhibitor of the type II enzyme.

We report here that the peptide-binding domain of prolyl 4-hydroxylase is distinct from the catalytic domain and consists of ~100 amino acid residues. This domain is also distinct from the previously characterized proline-rich peptide-binding domains such as the SH3 (Noble *et al*., 1993; Yu *et al*., 1994; Xu *et al*., 1997) and WW (Sudol *et al*., 1995; Macias *et al*., 1996) domains and the polypeptide profilin (Björkegren *et al*., 1993; Mahoney *et al*., 1997), and thus represents a new type of proline-rich peptide-binding module. We further report that substitutions in two amino acid positions within this domain explain most of the difference in poly(L-proline) binding between the type I and type II prolyl 4-hydroxylases.

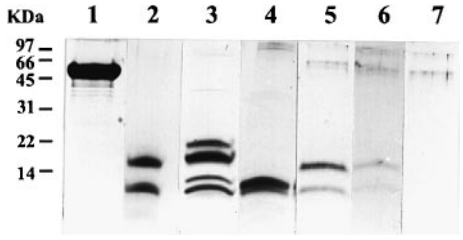

Fig. 1. Peptides of 10–19 kDa from the α(I) subunit remain bound to poly(L-proline) agarose after limited proteolysis of the type I prolyl 4-hydroxylase tetramer. Purified recombinant human type I prolyl 4-hydroxylase tetramer was incubated with poly(L-proline) agarose, and the bound enzyme was subjected to limited proteolysis with proteinase K, thermolysin or trypsin for 30 min at 4°C as described in Materials and methods. The samples were washed three times, the pelleted poly(L-proline) agarose with the bound material was boiled in a Tris–tricine sample buffer, and after centrifugation the supernatant was analyzed by 16.5% Tris–tricine PAGE and Coomassie Blue staining. Lane 1 shows the undigested control sample and lanes 2–4 the samples digested with proteinase K, thermolysin or trypsin, respectively. The protease-resistant peptides bound to poly(L-proline) agarose were eluted with 3 mg/ml poly(L-proline) (mol. wt 8000) (lane 5), 5 mg/ml $(Pro\text{-}Pro\text{-}Gly)_{10}$ (lane 6), or with the buffer alone (lane 7) (shown only for the proteinase K-resistant peptides).

Results

Peptides of 10–19 kDa from the α subunit remain bound to poly(L-proline) agarose after limited proteolysis of the type I prolyl 4-hydroxylase tetramer

In order to identify the poly(L-proline)-binding domain in the human type I prolyl 4-hydroxylase tetramer, the purified recombinant enzyme was incubated with poly(L-proline) agarose for 2 h at 4°C, and the bound enzyme was subjected to limited proteolysis with proteinase K, thermolysin or trypsin for 30 min at 4°C. The samples were then washed three times, the pelletted poly(L-proline) agarose was boiled in a Tris–tricine sample buffer, and after centrifugation the supernatant was analyzed by 16.5% Tris–tricine PAGE. Two peptides of 15 and 10.5 kDa, four peptides of 19, 15.5, 12 and 10.5 kDa, and two peptides of 11.5 and 10 kDa remained bound to the poly(L-proline) agarose after digestion with proteinase K, thermolysin and trypsin, respectively (Figure 1, lanes 2–4). All these protease-resistant peptides could be eluted from the poly(L-proline) agarose with either poly (L-proline) or $(Pro\text{-}Pro\text{-}Gly)_{10}$ (as shown for proteinase K-resistant peptides in Figure 1, lanes 5 and 6).

N-terminal sequencing of the peptides showed that they all originated from the α(I) subunit. The N-terminal amino acid of the proteinase K-resistant peptides was either His141 or Ser143, while all four thermolysin-resistant peptides had the same N-terminal amino acid, Val139, and the trypsin-resistant peptides began at Ser143 (Figure 2).

Recombinant α(I) subunit peptide Gly138–Ser244 expressed in Escherichia coli is soluble and binds effectively to poly(L-proline)

In order to express the poly(L-proline)-binding domain of the human α(I) subunit in *E.coli*, a recombinant pET-15b vector coding for the α(I) subunit amino acids Gly138–Ser244 (Figure 2), with an N-terminal histidine tag, was constructed and transformed into the BL21(DE3) strain. The N-terminus of this peptide was chosen based on the N-termini of the protease-resistant poly(L-proline)-binding peptides and the presence of an exon–intron boundary between nucleotides coding for amino acids Pro137 and Gly138 in the human α(I) subunit gene (Helaakoski *et al*., 1994). The C-terminus was chosen based on the sizes of the smallest protease-resistant poly(L-proline)-binding peptides and the presence of a non-homologous stretch of 22 amino acids in the human α(I) and α(II) subunits and the *Caenorhabditis elegans* α subunit beginning from Val241 in the human α(I) subunit (Figure 2) and possibly representing a variable inter-domain region.

Expression of the histidine-tagged α(I) subunit peptide Gly138–Ser244 was induced with isopropyl-β-D-galacto-pyranoside (IPTG) and the cells were incubated at 37°C for 2 h. The cells were then harvested, sonicated in a Tris–HCl buffer pH 8.0, and the soluble and insoluble fractions were analyzed by 16.5% Tris–tricine PAGE. The vast majority of the recombinant peptide was found in the soluble fraction (Figure 3, lanes 2 and 3). An aliquot of the soluble fraction was incubated with poly(L-proline) agarose, and samples of the bound and unbound proteins were analyzed by Tris–tricine PAGE. Essentially all of the Gly138–Ser244 peptide was found in the poly (L-proline)-bound fraction (Figure 3, lane 4), and only a very small amount was found in the unbound fraction (Figure 3, lane 5). The Gly138–Ser244 peptide could be eluted from the poly(L-proline) agarose with either poly(L-proline) or $(Pro\text{-}Pro\text{-}Gly)_{10}$ (Figure 3, lanes 6 and 7). Studies by NMR spectroscopy and urea gradient gel electrophoresis indicate that the recombinant peptide is folded (J.Kemmink, R.Ruotsalainen, L.Kukkola, J.Myllyharju and K.I.Kivirikko, unpublished data).

Poly(L-proline)-binding studies with enzyme tetramers containing hybrid α(I)/α(II) subunits localize a critical region between Asp162 and Leu266

As an independent approach and in order to further narrow down the critical region for proline-rich peptide binding in the α(I) subunit, several recombinant baculovirus expression vectors coding for hybrid α(I)/α(II) subunits were constructed in which part of the α(I) subunit sequence was replaced with that of the α(II) subunit (Figure 4). Recombinant viruses expressing such hybrid α(I)/α(II) subunits were generated and used to infect High Five insect cells together with a baculovirus coding for the PDI polypeptide (i.e. the β subunit). The cells were harvested 72 h after infection, homogenized in a buffer containing Triton X-100 and centrifuged. Aliquots of the Triton X-100-soluble fraction were incubated with poly(L-proline) agarose at 4°C for 4–6 h, centrifuged, and samples of the supernatants containing the unbound proteins were analyzed by non-denaturing 8% PAGE (Figure 5, lanes U) together with the original Triton X-100-soluble fractions of the cell homogenates (Figure 5, lanes T).

All hybrid α(I)/α(II) subunits formed an enzyme tetramer with the PDI polypeptide (Figure 5). Tetramers in which the α(I) subunit sequence covered only residues from the N-terminus to Ala161 or from Pro267 to the C-terminus showed no binding to poly(L-proline) agarose, whereas those in which the α(I) subunit sequence covered residues from the N-terminus to Leu266 or from Asp162 to the C-terminus became bound as effectively as the

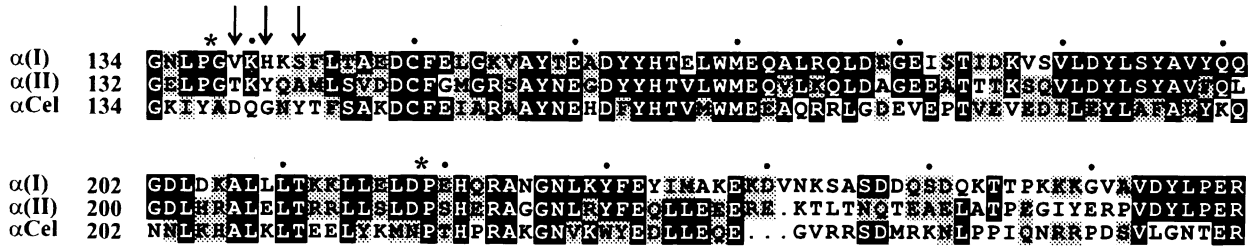

Fig. 2. Alignment of amino acid residues in the α(I) and α(II) subunits of human type I and type II prolyl 4-hydroxylases and the α subunit of *C.elegans* prolyl 4-hydroxylase. Gaps (·) were introduced for maximal alignment of the polypeptides. White letters on a black background indicate identity and black letters on a gray background indicate similarity. The sequence is shown only for residues corresponding to 134–269 in the 517-residue α(I) subunit. Dots (●) indicate every tenth amino acid. The arrows indicate the N-terminal amino acids of the peptides digested with proteinase K (His141 or Ser143), thermolysin (Val139) or trypsin (Ser143). The exon–intron boundaries in the 5′ and 3′ ends of exon 6 in the human α(I) subunit gene are indicated by asterisks (*). The sequences of the human α(I) [α(I)] and α(II) [α(II)] subunits and the *C.elegans* α subunit (αCel) are from Helaakoski *et al.* (1989), Annunen *et al.* (1997) and Veijola *et al.* (1994), respectively.

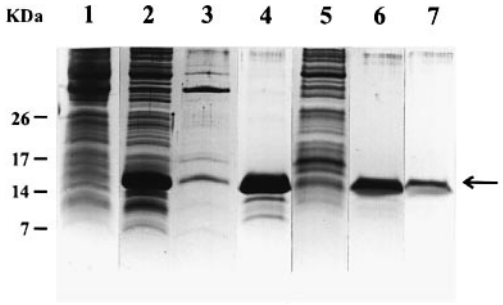

Fig. 3. Analysis of the expression of the α(I) subunit peptide Gly138–Ser244 in *E.coli*. Expression was induced by the addition of IPTG and the cells were incubated at 37°C for 2 h. The cells were then harvested and sonicated, and the soluble fractions from uninduced cells (lane 1), and the soluble (lane 2) and insoluble (lane 3) fractions from induced cells were boiled in a Tris–tricine sample buffer and analyzed by 16.5% Tris–tricine PAGE and Coomassie Blue staining. An aliquot of the soluble fraction from induced cells was incubated with poly(L-proline) agarose, and samples of the bound (lane 4) and unbound (lane 5) proteins were analyzed. The proteins bound to poly(L-proline) agarose were eluted with poly(L-proline) (lane 6) or $(Pro\text{-}Pro\text{-}Gly)_{10}$ (lane 7). The arrow indicates the position of the Gly138–Ser244 peptide.

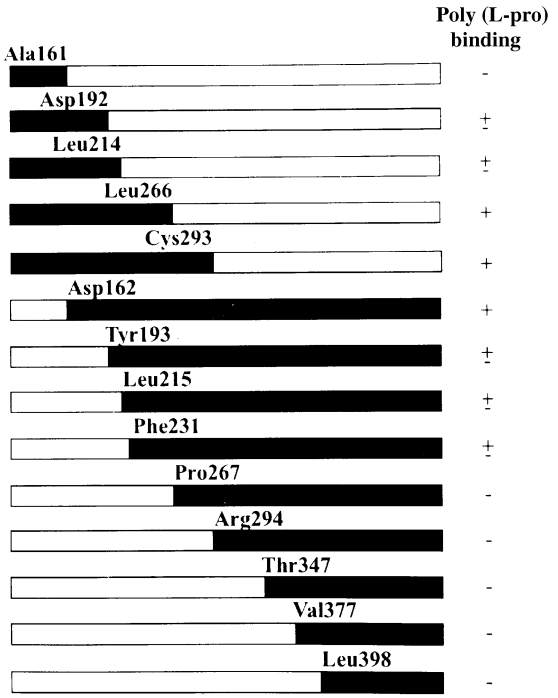

Fig. 4. Schematic representation of the hybrid prolyl 4-hydroxylase α(I)/α(II) subunits expressed in insect cells together with the PDI polypeptide, and binding of the corresponding tetramers to poly(L-proline) agarose. The α(I) and α(II) subunit amino acids are shown in black and white, respectively, and the last or first α(I) amino acid at the junction point is indicated. Effective binding of the tetramer to poly(L-proline) agarose is indicated by '+' and lack of binding by '–', while partial binding is indicated by '±'. Experimental details are given in Figure 5.

type I prolyl 4-hydroxylase tetramer (Figures 4 and 5). The critical region for poly(L-proline) binding in the α(I) subunit thus appears to be located between Asp162 and Leu266. Tetramers in which the α(I) subunit sequence ended at Asp192 or Leu214 or began at Tyr193, Leu215 or Phe231 became bound less effectively than the type I enzyme tetramer but more effectively than the type II tetramer (Figures 4 and 5). The degrees of inhibition of these five tetramers with 25 μM poly(L-proline), mol. wt 13 000, were likewise between those of the type I and type II enzymes (Table I), whereas all other mutant tetramers shown in Figures 4 and 5 were inhibited to the same extent as either the type I or the type II enzyme (details not shown). The data thus suggest that critical residues may be present both between Asp162 and Asp192 and between Phe231 and Leu266.

Mutations Ile182Glu and Tyr233Gln in the α(I) subunits abolish binding of the type I prolyl 4-hydroxylase tetramer to poly(L-proline)

Comparison of the amino acid sequence between residues Asp162 and Leu266 in the human α(I) subunit to the corresponding sequences in the human α(II) subunit and the *C.elegans* α subunit indicates the presence of only a few distinct differences (Figure 2). The *C.elegans* prolyl 4-hydroxylase is similar to the human type II enzyme in that it is inhibited by poly(L-proline) only in high concentrations (Veijola *et al.*, 1994, 1996). In order to identify amino acids in the α(II) subunit that abolish binding of the type II prolyl 4-hydroxylase to poly (L-proline) agarose, mutations Glu167Val, Ile182Glu, Val188Ser, Ser189Gln, Leu209Glu, Glu216Ser and Tyr233Gln were made individually to the human α(I) subunit by site-directed mutagenesis. The mutant α subunits were expressed together with the PDI polypeptide in insect cells and the poly(L-proline) binding of the mutant tetramers was analyzed by non-denaturing PAGE as above.

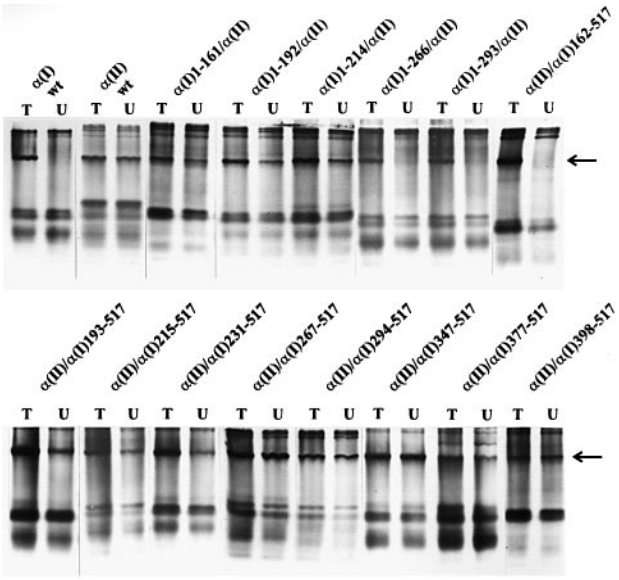

Fig. 5. Non-denaturing PAGE analysis of the binding to poly (L-proline) agarose of prolyl 4-hydroxylase tetramers containing hybrid α(I)/α(II) subunits. Recombinant baculoviruses expressing various hybrid α(I)/α(II) subunits shown in Figure 4 were generated. Viruses expressing the wild-type (wt) α(I) or α(II) or hybrid α(I)/α(II) subunits were used to infect insect cells together with a virus expressing the wild-type PDI polypeptide. The cells were harvested 72 h after infection and extracted with a buffer containing 0.1% Triton X-100. Samples of the Triton X-100-soluble fraction were then incubated with poly(L-proline) agarose, centrifuged, and samples of the original Triton X-100-soluble fractions (T) and the supernatants containing proteins that remain unbound after the incubation (U) were analyzed by non-denaturing 8% PAGE and Coomassie Blue staining. Enzymes containing hybrid α(I)/α(II) subunits (Figure 4) are indicated as in Table I. The arrow indicates the position of the enzyme tetramer.

Table I. Inhibition of prolyl 4-hydroxylase tetramers containing certain hybrid α(I)/α(II) subunits by 25 μM poly(L-proline), mol. wt 13 000

Prolyl 4-hydroxylase[a]	Degree of inhibition (%)
Type I, wild-type	86
Type II, wild-type	28
α(I)1–192/α(II)	55
α(I)1–214/α(II)	66
α(II)/α(I)193–517	59
α(II)/α(I)215–517	64
α(II)/α(I)231–517	62

The degree of inhibiton was calculated as a percentage of the prolyl 4-hydroxylase activity obtained without poly(L-proline).
[a]The enzymes containing hybrid α subunits are indicated by the residues of the α(I) subunit sequence present. The details are shown in Figure 2.

Prolyl 4-hydroxylase tetramers containing the Glu167-Val, Val188Ser, Ser189Gln, Leu209Glu or Glu216Ser α(I) subunits became bound to poly(L-proline) as effectively as the wild-type type I enzyme, whereas tetramers containing the Ile182Glu or Tyr233Gln mutant α(I) subunits did not become bound at all (Figure 6). The K_m values of the Ile182Glu and Tyr233Gln mutant enzyme tetramers for $(Pro\text{-}Pro\text{-}Gly)_{10}$ and the K_i value of the Tyr233Gln mutant tetramer for poly(L-proline), mol. wt 5700, were between those of the type I and type II prolyl 4-hydroxylases, whereas the K_i of the Ile182Glu mutant tetramer for poly(L-proline) was as high as that of the type II enzyme (Table II).

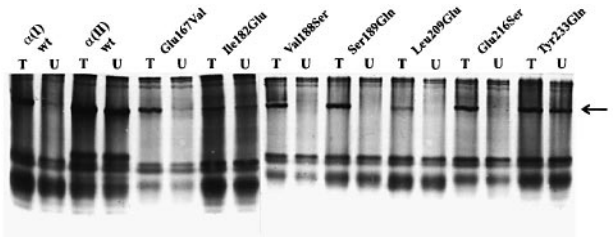

Fig. 6. Non-denaturing PAGE analysis of the binding to poly(L-proline) agarose of type I prolyl 4-hydroxylase tetramers containing mutant α(I) subunits. The mutant α subunits were expressed in insect cells together with the PDI polypeptide, and binding of the mutant tetramers to poly(L-proline) agarose was analyzed as in Figure 5. The Triton X-100-soluble fractions before incubation with poly(L-proline) agarose are indicated by T, and proteins that remain unbound after incubation are indicated by U. Mutant enzymes are indicated by the number of the α(I) subunit residues converted. The arrow indicates the position of the enzyme tetramer.

Table II. K_m values of certain mutant prolyl 4-hydroxylase tetramers for $(Pro\text{-}Pro\text{-}Gly)_{10}$ and K_i values for poly(L-proline), mol. wt 5700

Prolyl 4-hydroxylase[a]	K_m for $(Pro\text{-}Pro\text{-}Gly)_{10}$ (μM)	K_i for poly(L-proline) (μM)
Type I, wild-type	20	1
Type II, wild-type	100	130
Type I, α(I)Ile182Glu	90	130
Type I, α(I)Tyr233Gln	50	60
Type II, α(II)Glu180Ile,Gln231Tyr	30	10

All K_m and K_i values were determined using Triton X-100 extracts from cells expressing either a wild-type or mutant prolyl 4-hydroxylase as sources of the enzyme.
[a]Mutant enzymes are indicated by the α subunit residues converted.

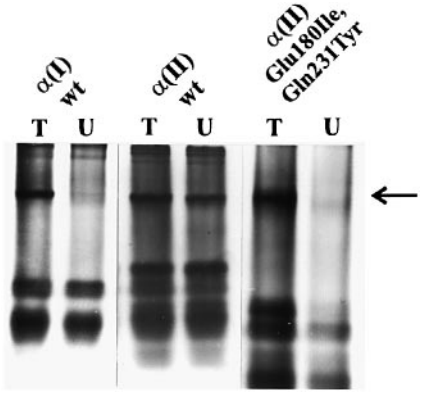

Fig. 7. Non-denaturing PAGE analysis of the binding to poly-(L-proline) agarose of a type II prolyl 4-hydroxylase tetramer containing Glu180Ile,Gln231Tyr double-mutant α(II) subunits. Experimental details and symbols are as in Figures 5 and 6. The arrow indicates the position of the enzyme tetramer.

Type II prolyl 4-hydroxylase tetramer containing Glu180Ile,Gln231Tyr double-mutant α(II) subunits becomes effectively bound to poly(L-proline) agarose

Based on the above data, a Glu180Ile,Gln231Tyr double-mutant α(II) subunit was generated and expressed together with the PDI polypeptide in insect cells. An enzyme tetramer containing these double-mutant α(II) subunits became effectively bound to poly(L-proline) agarose (Figure 7). The K_m of this mutant type II tetramer for $(Pro\text{-}Pro\text{-}Gly)_{10}$ was much lower than that of the wild-

type type II enzyme and only slightly higher than that of the type I enzyme (Table II). The K_i of the mutant tetramer for poly(L-proline), mol. wt 5700, was more than one order of magnitude lower than that of the type II enzyme but was still distinctly higher than that of the type I enzyme (Table II).

Discussion

The present data demonstrate that the peptide substrate binding domain in the α subunits of prolyl 4-hydroxylases is distinct from the catalytic domain which is located in their C-terminal region (Kivirikko and Myllyharju, 1998; Kivirikko and Pihlajaniemi, 1998). Limited proteolysis experiments indicated that the N-terminus of the peptide-binding domain is around residue 140 and corresponds closely to the beginning of sequences coded by exon 6 in the respective human gene (Helaakoski *et al*., 1994). The C-terminus is likely to be around residue 240, i.e. the beginning of a non-homologous stretch of 22 amino acids from Val241 in the human α(I) subunit. The degree of amino acid sequence identity between the human α(I) and α(II) subunits within this 22-residue region is only ~5% and that between the human α(I) subunit and the *C.elegans* α subunit is ~9% (Figure 2), suggesting that this stretch may represent a variable inter-domain region. The size of the peptide-binding domain therefore appears to be ~100 residues.

Independent experiments with prolyl 4-hydroxylase tetramers containing hybrid α(I)/α(II) subunits localized the difference in poly(L-proline) binding between the type I and type II enzymes to a region in the α subunits extending from residue 162 to 266 [numbered according to the α(I) subunit sequence]. However, the data do not exclude the possibility that some residues between positions 140 and 162 may also contribute to the binding in both types of enzyme tetramer. On the other hand, the fact that the recombinant α(I) subunit peptide Gly138–Ser244 became effectively bound to poly(L-proline), as well as the sequence comparisons discussed above, suggest that residues between positions 240 and 266 are unlikely to have any significant role in the binding. The partial binding observed with tetramers containing some of the hybrid α(I)/α(II) subunits together with the above considerations concerning the C-terminal end of the peptide-binding domain suggest that differences in at least two distinct regions may account for differences in poly (L-proline) binding between the two types of enzyme tetramer, one region being located between the α(I) subunit residues 162 and 192 and the other between residues 231 and 240. Finally, studies with the type I and type II enzyme tetramers containing mutated α subunits indicated that replacement of Ile182 and Tyr233 in the α(I) subunit by glutamate and glutamine, respectively, in the α(II) subunit appears to explain most of the lack of poly(L-proline) binding of the type II prolyl 4-hydroxylase tetramer. Nevertheless, these two substitutions do not explain all of the difference, as the type II enzyme that contained the Glu180Ile,Gln231Tyr double-mutant α(II) subunits and became effectively bound to poly(L-proline) agarose had a K_i for poly(L-proline) which was distinctly higher than that of the type I enzyme.

Although the peptide-binding domain of prolyl 4-hydroxylase was identified here largely by experiments based on binding to poly(L-proline) agarose, the same domain is likely to be involved in the binding of the peptide substrate. Numerous previous studies have demonstrated that poly(L-proline) inhibits prolyl 4-hydroxylase competitively with respect to the peptide substrate (see Kivirikko and Pihlajaniemi, 1998), thus indicating binding to the same site on the enzyme. The present data indicate that the peptides obtained by limited proteolysis from the type I enzyme tetramer, and the recombinant α(I) subunit peptide Gly138–Ser244, could be eluted from poly (L-proline) agarose with the peptide substrate (Pro-Pro-Gly)$_{10}$. Furthermore, the K_m values of the Ile182Glu and Tyr233Gln mutant type I enzyme tetramers for (Pro-Pro-Gly)$_{10}$ were distinctly higher than that of the wild-type enzyme, while the corresponding K_m of the Glu180 Ile,Gln231Tyr double-mutant type II enzyme was much lower than that of the wild-type enzyme and was almost as low as that of the type I enzyme. Thus these two residues are also involved in the binding of the peptide substrate.

The sequence of the peptide-binding domain identified here shows no similarity to those of the SH3 (Noble *et al*., 1993; Yu *et al*., 1994; Xu *et al*., 1997) and WW (Sudol *et al*., 1995; Macias *et al*., 1996) domains or of profilin (Björkegren *et al*., 1993; Mahoney *et al*., 1997), and no homologous sequences were found in the databases. Thus, this domain represents a new type of proline-rich peptide-binding module. Profilin has five highly conserved aromatic residues which appear to be involved in poly (L-proline) binding (Björkegren *et al*., 1993; Mahoney *et al*., 1997), and aromatic residues together with polar interactions are also involved in the binding of proline-rich peptides by the SH3 (Noble *et al*., 1993; Yu *et al*., 1994; Xu *et al*., 1997) and WW (Sudol *et al*., 1995; Macias *et al*., 1996) domains. The 101-amino acid sequence from residue 140 to 240 in the human α(I) subunit has 15 aromatic residues, most of which are conserved in the α(II) subunit and the *C.elegans* α subunit sequences (Figure 2). It thus seems possible that aromatic residues may also be critically involved in the binding of various proline-rich peptide substrates by the prolyl 4-hydroxylase α subunits. Our data indicating that the recombinant α(I) subunit peptide Gly138–Ser244 was soluble, became effectively bound to poly(L-proline) agarose, could be eluted with either poly(L-proline) or (Pro-Pro-Gly)$_{10}$ and was folded when studied by NMR and urea gradient gel electrophoresis (J.Kemmink, R.Ruotsalainen, L.Kukkola, J.Myllyharju and K.I.Kivirikko, unpublished data) suggest that this peptide should be suitable for detailed structural studies on the peptide-binding domain.

Materials and methods

Protease digestion of prolyl 4-hydroxylase bound to poly(L-proline) agarose

Purified recombinant human type I prolyl 4-hydroxylase tetramer, 200 μg (Vuori *et al*., 1992), was incubated with 100 μl of poly(L-proline) agarose (Kivirikko and Myllylä, 1982) in a 0.1 M NaCl, 0.1 M glycine, 10 μM dithiothreitol, and 10 mM Tris pH 7.8 buffer for 2 h at 4°C. The poly(L-proline) agarose was pelleted by centrifugation, washed three times with 1 ml of the above buffer, and 2 μg of proteinase K (Boehringer Mannheim), 80 μg of thermolysin (Boehringer Mannheim) or 10 μg of trypsin (Sigma) was added in 100 μl of the above buffer. The samples

were digested for 30 min at 4°C, and the digestion was stopped by the addition of Pefabloc SC (Boehringer Mannheim) or EDTA (for thermolysin) to 3 mM. After a 2 min incubation, the samples were washed three times with 1 ml of the above buffer. Poly(L-proline) agarose with the bound material was then boiled directly in a Tris–tricine sample buffer for 3 min, and after centrifugation the supernatant was analyzed by 16.5% Tris–tricine PAGE. Alternatively, the peptides bound to poly(L-proline) agarose were eluted with 3 mg/ml poly(L-proline) (mol. wt 8000, Sigma) or 5 mg/ml $(Pro\text{-}Pro\text{-}Gly)_{10}$ (Peptide Institute) in the above buffer.

Expression and analysis of the α(I) subunit peptide Gly138–Ser244 in E.coli

An α(I) subunit cDNA sequence coding for amino acids Gly138–Ser244 with a *Nde*I site preceding the codon for Gly138, and a stop codon and a B*am*HI site following the codon for Ser244 was synthesized by PCR, cloned into *Nde*I–*Bam*HI-digested pET-15b expression vector (Novagen) and transformed into the BL21(DE3) (Novagen) bacterial host strain. The cells were grown to an optical density of 0.4–1.0 at 600 nm and expression was induced by the addition of IPTG to 1 mM for 2 h at 37°C. The cells were then harvested, suspended in 1/10 volume of 50 mM Tris–HCl pH 8.0, sonicated, and the soluble and insoluble fractions were boiled in a Tris–tricine sample buffer for 3 min and analyzed by 16.5% Tris–tricine PAGE. Aliquots of 100 µl of the soluble fraction from induced cells were incubated with 100 µl of poly(L-proline) agarose at 4°C for 2 h, samples were centrifuged and the bound and unbound proteins were analyzed by 16.5% Tris–tricine PAGE. Alternatively, the poly(L-proline) agarose-bound proteins were eluted with 3 mg/ml poly(L-proline) (mol. wt 8000) or 5 mg/ml $(Pro\text{-}Pro\text{-}Gly)_{10}$ as above.

Expression and analysis of wild-type and mutant recombinant prolyl 4-hydroxylases in insect cells

Baculovirus transfer vectors coding for hybrid prolyl 4-hydroxylase α subunits were constructed by preparing two PCR fragments using kinase-treated primers and *Pfu* DNA polymerase (Stratagene) with pVLα59 (Vuori *et al*., 1992) and pVLαII (Annunen *et al*., 1997) as templates. To generate the constructs coding for hybrid α subunits in which the N-terminal half consists of α(I) subunit amino acids, the α(I) subunit PCR fragments were amplified from an internal *Afl*II site to the codons for α(I) Ala161, Asp192, Leu214, Leu266 or Cys293, and the α(II) subunit PCR fragments were amplified from the codons for α(II) Asp160, Tyr191, Leu213, Pro265 or Arg292 to the pVL1392 *Bam*HI site, and the fragments were ligated into *Afl*II–*Bam*HI-digested pVLα59. To generate the constructs coding for hybrid α subunits in which the N-terminal half consists of α(II) subunit amino acids, the α(II) subunit PCR fragments were amplified from the pVL1392 *Bgl*II site to the codons for α(II) Gly159, Asp190, Leu212, Tyr228, Leu263, Cys290, Ala341, Val373 or Leu394, and the α(I) subunit PCR fragments were amplified from the codons for α(I) Asp162, Tyr193, Leu215, Phe231, Pro267, Arg294, Thr347, Val377 or Leu398 to the internal *Pst*I site, and the fragments were ligated into *Bgl*II–*Pst*I-digested pVLα59.

Site-directed mutagenesis was carried out using an oligonucleotide-directed *in vitro* system based on the unique site elimination procedure (Pharmacia Biotech Inc.). To generate the mutant α(I) subunits mutagenesis reactions were performed in a pBluescript vector (Stratagene) containing the full-length cDNA clone (PA-59) for the human prolyl 4-hydroxylase α(I) subunit at the *Sma*I site (Helaakoski *et al*., 1989), after which the plasmids were digested with *Afl*II and *Bam*HI. The resulting 1.8 kb *Afl*II–*Bam*HI cDNA fragments containing the mutant sites were then cloned into the *Afl*II–*Bam*HI-digested pVLα59. The double-mutant α(II) subunit was constructed in two steps by first creating the single mutants into the full-length cDNA clone for the human prolyl 4-hydroxylase α(II) subunit at the *Not*I–*Eco*RV site of pBluescript (Annunen *et al*., 1997), generating pBSα(II)Glu180Ile and pBSα(II)Gln231Tyr. The full-length Gln231Tyr mutant α(II) cDNA was digested with *Not*I and *Eco*RV and cloned into *Not*I–*Sma*I-digested pVL1392 (Invitrogen) to generate pVLα(II)Gln231Tyr. A *Not*I–*Sac*I fragment from pBSα(II)Glu180Ile was then ligated into *Not*I–*Sac*I-digested pVLα(II)Gln231Tyr. The nucleotide sequences were verified by the dideoxynucleotide chain termination method (Sanger *et al*., 1977) with the Sequenase enzyme (United States Biochemical) in an automated DNA sequencer (Applied Biosystems).

The recombinant baculovirus transfer vectors were cotransfected into *Spodoptera frugiperda* Sf9 cells with a modified *Autographa californica* nuclear polyhedrosis virus DNA (PharMingen) by calcium-phosphate transfection, and the recombinant viruses were selected (Gruenwald and Heitz, 1993).

High Five insect cells (Invitrogen) were cultured as monolayers in TNM-FH medium (Sigma) supplemented with 10% fetal bovine serum (BioClear) at 27°C. Cells seeded at a density 5×10^6 per 100 mm plate were infected at a multiplicity of 5 with the viruses coding for the wild-type or mutant α subunit together with a virus coding for the PDI polypeptide (Vuori *et al*., 1992). The cells were harvested 72 h after infection, washed with a solution of 0.15 M NaCl and 0.02 M phosphate pH 7.4, homogenized in a 0.1 M NaCl, 0.1 M glycine, 10 µM dithiothreitol, 0.1% Triton X-100 and 10 mM Tris buffer pH 7.8, and centrifuged at 10 000 *g* for 20 min. Aliquots of 200 µl of the Triton X-100-soluble fraction of the cell homogenates were then incubated with 200 µl of poly(L-proline) agarose at 4°C for 4–6 h, centrifuged, and 20 µl samples of the supernatants containing the unbound proteins and the original Triton X-100-soluble fractions were analyzed by 8% non-denaturing PAGE.

Other assays

Prolyl 4-hydroxylase activity was assayed by a method based on the hydroxylation-coupled decarboxylation of 2-oxo[1-^{14}C]glutarate (Kivirikko and Myllylä, 1982). The reaction was performed in a final volume of 1.0 ml, which contained 10 µl of the Triton X-100 extract as the source of the enzyme, 0.1 mg $(Pro\text{-}Pro\text{-}Gly)_{10}\cdot 9H_2O$ as substrate, 0.05 µmol $FeSO_4$, 0.1 µmol 2-oxo[1-^{14}C]glutarate (100 000 d.p.m.), 1 µmol ascorbate, 0.3 mg catalase (Sigma), 0.1 µmol dithiothreitol, 2 mg bovine serum albumin (Sigma) and 50 µmol Tris–HCl buffer adjusted to pH 7.8 at 25°C. K_m and K_i values were determined as described previously (Myllylä *et al*., 1977).

N-terminal sequencing was performed in an Applied Biosystems 477A pulse-liquid protein sequencer.

Acknowledgements

We thank Eeva Lehtimäki for excellent technical assistance and Hongmin Tu for performing the N-terminal sequencing. This study was supported by grants from the Health Sciences Council of the Academy of Finland, from FibroGen Inc. (South San Francisco, CA) and from the European Commission (BIO4-CT96-0436).

References

Annunen,P., Helaakoski,T., Myllyharju,J., Veijola,J., Pihlajaniemi,T. and Kivirikko,K.I. (1997) Cloning of the human prolyl 4-hydroxylase α subunit isoform α(II) and characterization of the type II enzyme tetramer. The α(I) and α(II) subunits do not form a mixed $\alpha(I)\alpha(II)\beta_2$ tetramer. *J. Biol. Chem.*, **272**, 17342–17348.

Annunen,P., Autio-Harmainen,H. and Kivirikko,K.I. (1998) The novel type II prolyl 4-hydroxylase is the main enzyme form in chondrocytes and capillary endothelial cells, whereas the type I enzyme predominates in most cells. *J. Biol. Chem.*, **273**, 5989–5992.

Björkegren,C., Rozycki,M., Schutt,C.E., Lindberg,U. and Karlsson,R. (1993) Mutagenesis of human profilin locates its poly(L-proline)-binding site to a hydrophobic patch of aromatic amino acids. *FEBS Lett.*, **333**, 123–126.

Gruenwald,S. and Heitz,J. (1993) *Baculovirus Expression Vector System: Procedures and Methods Manual*. Pharmingen, San Diego, CA.

Helaakoski,T., Vuori,K., Myllylä,R., Kivirikko,K.I. and Pihlajaniemi,T. (1989) Molecular cloning of the α-subunit of human prolyl 4-hydroxylase: The complete cDNA-derived amino acid sequence and evidence for alternative splicing of RNA transcripts. *Proc. Natl Acad. Sci. USA*, **86**, 4392–4396.

Helaakoski,T., Veijola,J., Vuori,K., Rehn,M., Chow,L.T., Taillon-Miller,P., Kivirikko,K.I. and Pihlajaniemi,T. (1994) Structure and expression of the human gene for the α subunit of prolyl 4-hydroxylase. The two alternatively spliced types of mRNA correspond to two homologous exons the sequences of which are expressed in a variety of tissues. *J. Biol. Chem.*, **269**, 27847–27854.

Helaakoski,T., Annunen,P., Vuori,K., MacNeil,I.A., Pihlajaniemi,T. and Kivirikko,K.I. (1995) Cloning, baculovirus expression, and characterization of a second mouse prolyl 4-hydroxylase α-subunit isoform: Formation of an $\alpha_2\beta_2$ tetramer with the protein disulfide-isomerase/β subunit. *Proc. Natl Acad. Sci. USA*, **92**, 4427–4431.

Kivirikko,K.I. and Myllylä,R. (1982) Posttranslational enzymes in the biosynthesis of collagen: intracellular enzymes. *Methods Enzymol.*, **82**, 245–304.

Kivirikko,K.I. and Myllyharju,J. (1998) Prolyl 4-hydroxylases and their protein disulfide isomerase subunit. *Matrix Biol.*, **16**, 357–368.

Kivirikko,K.I. and Pihlajaniemi,T. (1998) Collagen hydroxylases and the protein disulfide isomerase subunit of prolyl 4-hydroxylases. *Adv. Enzymol. Related Areas Mol. Biol.*, **72**, 325–398.

Lamberg,A., Pihlajaniemi,T. and Kivirikko,K.I. (1995) Site-directed mutagenesis of the α subunit of human prolyl 4-hydroxylase. Identification of three histidine residues critical for catalytic activity. *J. Biol. Chem.*, **270**, 9926–9931.

Macias,M.J., Hyvönen,M., Baraldi,E., Schultz,J., Sudol,M., Saraste,M. and Oschkinat,H. (1996) Structure of the WW domain of a kinase-associated protein complexed with a proline-rich peptide. *Nature*, **382**, 646–649.

Mahoney,N.M., Janmey,P.A. and Almo,S.C. (1997) Structure of the profilin–poly-L-proline complex involved in morphogenesis and cytoskeletal regulation. *Nature Struct. Biol.*, **4**, 953–960.

Myllyharju,J. and Kivirikko,K.I. (1997) Characterization of the iron- and 2-oxoglutarate-binding sites of human prolyl 4-hydroxylase. *EMBO J.*, **16**, 1173–1180.

Myllylä,R., Tuderman,L. and Kivirikko,K.I. (1977) Mechanism of the prolyl hydroxylase reaction. 2. Kinetic analysis of the reaction sequence. *Eur. J. Biochem.*, **80**, 349–357.

Myllylä,R., Günzler,V., Kivirikko,K.I. and Kaska,D.D. (1992) Modification of vertebrate and algal prolyl 4-hydroxylases and vertebrate lysyl hydroxylase by diethyl-pyrocarbonate: evidence for histidine residues in the catalytic site of 2-oxoglutarate coupled dioxygenases. *Biochem. J.*, **286**, 923–927.

Noble,M.E.M., Musacchio,A., Saraste,M., Courtneidge,S.A. and Wierenga,R.K. (1993) Crystal structure of the SH3 domain in human Fyn; comparison of the three-dimensional structures of SH3 domains in tyrosine kinases and spectrin. *EMBO J.*, **12**, 2617–2624.

Pihlajaniemi,T., Helaakoski,T., Tasanen,K., Myllylä,R., Huhtala,M.-L., Koivu,J. and Kivirikko,K.I. (1987) Molecular cloning of the β-subunit of human prolyl 4-hydroxylase. This subunit and protein disulphide isomerase are products of the same gene. *EMBO J.*, **6**, 643–649.

Prockop,D.J. and Kivirikko,K.I. (1995) Collagens: Molecular biology, diseases, and potentials for therapy. *Annu. Rev. Biochem.*, **64**, 403–434.

Sanger,K., Nicklen,S. and Coulson,A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl Acad. Sci. USA*, **74**, 5463–5467.

Sudol,M., Chen,H.I., Bougeret,C., Einbond,A. and Bork,P. (1995) Characterization of a novel protein-binding module—the WW domain. *FEBS Lett.*, **369**, 67–71.

Veijola,J., Koivunen,P., Annunen,P., Pihlajaniemi,T. and Kivirikko,K.I. (1994) Cloning, baculovirus expression, and characterization of the α subunit of prolyl 4-hydroxylase from the nematode *Caenorhabditis elegans*. This α subunit forms an active αβ dimer with the human protein disulfide isomerase/β subunit. *J. Biol. Chem.*, **269**, 26746–26753.

Veijola,J., Annunen,P., Koivunen,P., Page,A.P., Pihlajaniemi,T. and Kivirikko,K.I. (1996) Baculovirus expression of two protein disulphide isomerase isoforms from *Caenorhabditis elegans* and characterization of prolyl 4-hydroxylases containing one of these polypeptides as their β subunit. *Biochem. J.*, **317**, 721–729.

Vuori,K., Pihlajaniemi,T., Marttila,M. and Kivirikko,K.I. (1992) Characterization of the human prolyl 4-hydroxylase tetramer and its multifunctional protein disulfide-isomerase subunit synthesized in a baculovirus expression system. *Proc. Natl Acad. Sci. USA*, **89**, 7467–7470.

Xu,W., Harrison,S.C. and Eck,M.J. (1997) Three-dimensional structure of the tyrosine kinase c-Src. *Nature*, **385**, 595–602.

Yu,H., Chen,J.K., Feng,S., Dalgarno,D.C., Brauer,A.W. and Schreiber,S.L. (1994) Structural basis for the binding of proline-rich peptides to SH3 domains. *Cell*, **76**, 933–945.

Received November 4, 1998; revised November 13, 1998; accepted November 23, 1998