Acta Crystallographica Section D
**Biological Crystallography**
ISSN 0907-4449

# Crystallization of the proline-rich-peptide binding domain of human type I collagen prolyl 4-hydroxylase

**Mira Pekkala,[a]† Reija Hieta,[b]† Petri Kursula,[a]‡ Kari I. Kivirikko,[b] Rik K. Wierenga[a]* and Johanna Myllyharju[b]**

[a]Biocenter Oulu and Department of Biochemistry, University of Oulu, PO Box 3000, FIN-90014, Finland, and [b]Biocenter Oulu and Department of Medical Biochemistry and Molecular Biology, University of Oulu, PO Box 5000, FIN-90014, Finland

† These authors contributed equally to the study.
‡ Current address: EMBL Hamburg Outstation, DESY, Notkestrasse 85, Hamburg, Germany.

Correspondence e-mail: rik.wierenga@oulu.fi

Received 26 November 2002
Accepted 6 March 2003

Collagen prolyl 4-hydroxylases catalyze the hydroxylation of -*X*-Pro-Gly- sequences and play an essential role in the synthesis of all collagens. They require $Fe^{2+}$, 2-oxoglutarate, molecular oxygen and ascorbate, and all vertebrate collagen prolyl 4-hydroxylases are $\alpha_2\beta_2$ tetramers. The $\alpha$-subunits contain separate catalytic and peptide substrate-binding domains. Here, the crystallization of the peptide substrate-binding domain consisting of residues 144–244 of the 517-residue human $\alpha$(I) subunit is described. The crystals are well ordered and diffract to at least 3 Å. The space group is $P3_1$ or $P3_2$ and the asymmetric unit most probably contains a dimer.

## 1. Introduction

Collagen prolyl 4-hydroxylases (C-P4Hs; EC 1.14.11.2) catalyze the formation of 4-hydroxyproline in collagens and more than 15 other proteins that have collagen-like domains by the hydroxylation of prolines in -*X*-Pro-Gly- sequences. This modification has a vital role in the synthesis of all collagens, as 4-hydroxyproline residues are essential for the stability of the collagen triple helix. C-P4Hs from all vertebrate species studied are $\alpha_2\beta_2$ tetramers in which the $\beta$-subunit is identical to the enzyme and chaperone protein disulfide isomerase (PDI; for recent reviews, see Kivirikko & Myllyharju, 1998; Kivirikko & Pihlajaniemi, 1998; Myllyharju, 2003). Two isoforms of the catalytic $\alpha$-subunit, $\alpha$(I) and $\alpha$(II), have been cloned and characterized from human and mouse tissues and shown to form $[\alpha(I)]_2\beta_2$ and $[\alpha(II)]_2\beta_2$ tetramers, called type I and type II C-P4Hs, respectively (Helaakoski *et al.*, 1995; Annunen *et al.*, 1997). Data on recombinant coexpression in insect cells strongly argue against the existence of a mixed vertebrate $\alpha(I)\alpha(II)\beta_2$ tetramer (Annunen *et al.*, 1997).

The C-P4H reaction requires $Fe^{2+}$, 2-oxoglutarate, $O_2$ and ascorbate, and involves an oxidative decarboxylation of 2-oxoglutarate (Kivirikko & Myllyharju, 1998; Kivirikko & Pihlajaniemi, 1998; Myllyharju, 2003). Site-directed mutagenesis studies have shown that all critical cosubstrate-binding residues are located in the highly conserved C-terminal region of the $\alpha$-subunit (Lamberg *et al.*, 1995; Myllyharju & Kivirikko, 1997). Poly(L-proline) is an effective competitive inhibitor of vertebrate type I C-P4H but only a very weak inhibitor of the type II enzyme; differences between the isoenzymes also exist in their $K_M$ values for various peptide substrates (Helaakoski *et al.*, 1995; Annunen *et al.*, 1997). It thus seems very likely that distinct differences in the structures of the peptide-binding sites must exist between the two types of C-P4H. The proline-rich-peptide binding domain has recently been shown to be separate from the catalytic domain and to be located between residues Gly138 and Ser244 of the $\alpha$(I) subunit (Myllyharju & Kivirikko, 1999). This domain is distinct from the previously characterized proline-rich-peptide binding domains SH3 (Xu *et al.*, 1997) and WW (Macias *et al.*, 1996) and the polypeptide profilin (Mahoney *et al.*, 1997). No sequence similarity can be detected with sequences of proteins in the PDB database; this domain apparently represents a novel proline-rich-peptide binding module (Myllyharju & Kivirikko, 1999).

Fibrotic diseases constitute a major problem in medicine and C-P4Hs are regarded as particularly suitable targets for antifibrotic therapy. Structural information on the catalytic and peptide substrate-binding sites of C-P4Hs would therefore have a major impact on the rational design of inhibitors. Since crystallization of the vertebrate C-P4H tetramer has so far proved unsuccessful, attempts have been initiated to characterize the structures of the domains of the $\alpha$ and PDI polypeptides (Kemmink *et al.*, 1996, 1997; Dijkstra *et al.*, 1999). The studies reported here are the first structural studies on the $\alpha$-subunit. It is shown that recombinant polypeptide Phe144–Ser244 corresponding to the proline-rich-peptide binding domain of the $\alpha$(I) subunit of human type I C-P4H can be expressed as a functional domain in *Escherichia coli* and can be crystallized into well diffracting crystals.

## 2. Experimental methods

### 2.1. Expression and purification

A human C-P4H $\alpha$(I) subunit cDNA fragment coding for amino acids Phe144–Ser244 with an *Nde*I restriction site and a translation start codon preceding the codon for Phe144 and a stop codon and *Bam*HI restriction site following the codon for Ser244 was amplified using *Pfu* polymerase (Promega). The template was pET15b-$\alpha$(I)Gly138-Ser244 (Myllyharju & Kivirikko, 1999), in which the codon for Cys150 (TGC) was converted to that for serine (TCC), the primers being 5′-GGAATTCCAT**ATG**-TTTCTAACGGCTGAGGACTC-3′ (*Nde*I site underlined, start codon in bold) and 5′-CGGGATCC**TCA**AGACTTATTGACA-TCTTTTTCTTTAG-3′ (*Bam*HI site underlined, stop codon in bold). Cys150Ser mutagenesis was performed using the QuikChange site-directed mutagenesis kit (Stratagene) according to the manufacturer's instructions. The PCR product was cloned into an *Nde*I–*Bam*HI-digested pET-22b expression vector (Novagen) in frame with the sequence encoding a C-terminal histidine tag consisting of six histidines and the plasmid was transformed into the BL21(DE3) *E. coli* host strain (Novagen). A single colony was inoculated in Luria–Bertani (LB) medium containing ampicillin (50 mg ml$^{-1}$) and grown overnight. The culture was diluted 100-fold in 1 l LB medium and the cells were grown at 310 K to an optical density of 0.6–0.8 at 600 nm. Expression was induced by adding isopropyl-$\beta$-D-thiogalactoside to 1 m*M* and was continued for 4 h at 303 K. The cells were harvested by centrifugation, suspended in 1/10 volume of 20 m*M* bis–tris, 0.1 *M* glycine pH 6.8 and disrupted by sonication. The cell lysate was centrifuged (20 min, 20 000*g*, 277 K) and the soluble proteins were applied to a 30 ml Ni–NTA chelating Sepharose (Amersham Biosciences) column equilibrated with 20 m*M* bis–tris, 0.1 *M* glycine pH 6.8. The bound proteins were eluted with a 200 ml linear imidazole gradient (0–0.5 *M*) and the fractions (150 drops) were analyzed by 16.5% Tris–tricine PAGE. The fractions containing the recombinant polypeptide were pooled and concentrated using Ultrafree-15 Biomax 5K (Millipore) concentrators, after which a 1 ml sample was loaded onto a SuperDex 75 HR (Amersham Biosciences) column equilibrated with 20 m*M* bis–tris, 0.1 *M* glycine pH 6.8. 2 ml fractions were collected and analyzed by 16.5% Tris–tricine PAGE and those containing the pure recombinant protein were pooled. A total of 30 mg of protein was obtained from a 1 l culture. Poly(L-proline) binding of the purified recombinant domain was tested as described previously (Myllyharju & Kivirikko, 1999).

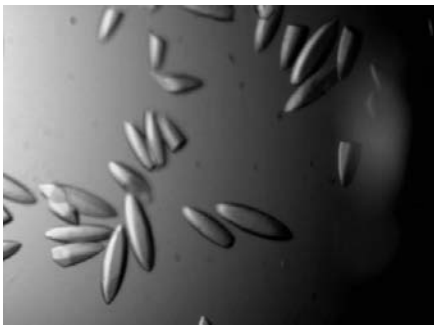

**Figure 1**
Crystals of the peptide substrate-binding domain of human type I collagen prolyl 4-hydroxylase. These crystals were grown in 1.4 *M* ammonium phosphate pH 8.4. The approximate dimensions of a typical crystal are 0.15 × 0.15 × 0.5 mm.

### 2.2. Crystallization and X-ray diffraction analysis

Prior to crystallization, the protein was concentrated to approximately 10 mg ml$^{-1}$ in 20 m*M* bis–tris, 0.1 *M* glycine pH 6.8. A factorial screen (Zeelen *et al.*, 1994) using the hanging-drop vapour-diffusion method was used for initial crystallization trials. Equal volumes of protein and reservoir solution (1 µl + 1 µl) were mixed and equilibrated against 1 ml reservoir solution. Crystals were obtained within a few days at 277 K in the presence of 1.2 *M* ammonium phosphate pH 8.2. The crystallization conditions were then optimized in order to improve the diffraction quality of the crystals. The best crystals were grown using 2 µl + 2 µl drops in the presence of 1.2–1.7 *M* ammonium phosphate pH 8.2–8.6 (Fig. 1). The average crystal size was 0.15 × 0.15 × 0.5 mm.

The crystals were soaked in a cryoprotectant solution containing 1.7 *M* ammonium phosphate pH 8.4 and 20% glycerol and flash-frozen in a stream of nitrogen gas at 100 K prior to data collection. Diffraction data

**Table 1**
Data-collection statistics.

Values in parentheses are for the highest resolution shell.

| | |
|---|---|
| Temperature (K) | 100 |
| Wavelength (Å) | 1.192 |
| Space group | $P3_1$ or $P3_2$ |
| Unit-cell parameters (Å) | $a = b = 55.14$, $c = 105.15$ |
| Resolution (Å) | 10–3.1 (3.25–3.10) |
| Observed reflections | 17739 |
| Unique reflections | 6345 |
| Completeness (%) | 97.5 (96.3) |
| Average $I/\sigma(I)$ | 11.6 (3.0) |
| $R_{\rm sym}$ (%) | 5.7 (36.6) |

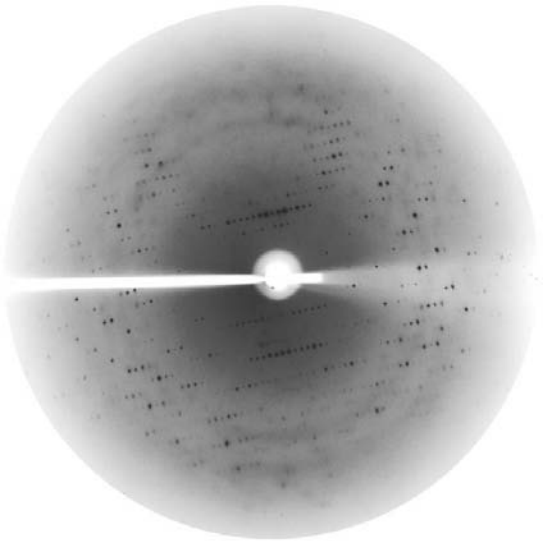

**Figure 2**
X-ray diffraction pattern from a crystal of the peptide substrate-binding domain of human type I prolyl 4-hydroxylase. The resolution at the edge of the detector corresponds to a resolution of 2.7 Å. The image covers an oscillation range of 1°.

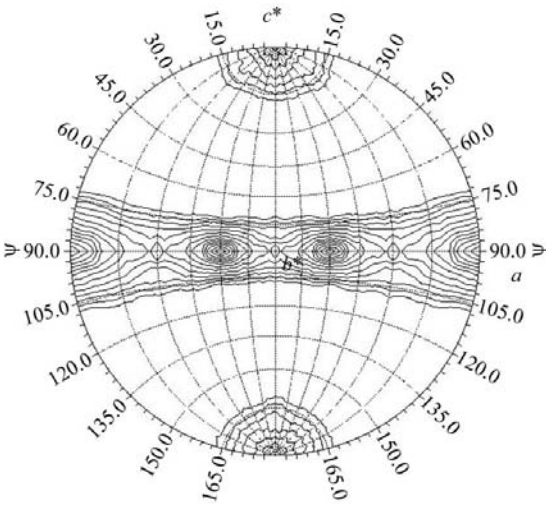


**Figure 3**
The GLRF self-rotation function (Tong & Rossmann, 1990) calculated with data between 10 and 3.1 Å and contoured from 3$\sigma$ above the average with increments of 0.5$\sigma$. The highest peak is at $\varphi = 0$, $\psi = 90°$, indicating that there is a local dimer twofold axis perpendicular to the threefold axis.

were collected at beamline I711 at Max-Lab, Lund, Sweden. The *XDS* program package (Kabsch, 1993) was used to process and scale the data. Further data analysis was carried out using the *CCP*4 package (Collaborative Computational Project, Number 4, 1994). The self-rotation function was calculated with *GLRF* (Tong & Rossmann, 1990).

## 3. Results and discussion

The crystallization experiments were initiated with a recombinant peptide substrate-binding domain consisting of the human P4H $\alpha$(I) subunit residues Gly138–Ser244 and an N-terminal histidine tag (Myllyharju & Kivirikko, 1999). However, preliminary NMR data showed that the N-terminus of the recombinant domain was very flexible and degraded easily (R. Hieta, L. Kukkola, P. Permi, P. Pirilä, K. I. Kivirikko, I. Kilpeläinen and J. Myllyharju, unpublished results). Thus, a shorter construct with a C-terminal histidine tag was generated by cloning a cDNA fragment coding for the $\alpha$(I) subunit amino acids Phe144–Ser244 into the pET-22b vector and high-level expression of the 13 kDa peptide substrate-binding domain was achieved in *E. coli*. This recombinant domain, consisting of an N-terminal methionine, residues Phe144–Ser244 and a six-residue histidine tag at the C-terminus, could be purified to homogeneity by affinity chromatography and gel filtration. It was found that this domain binds efficiently to poly(L-proline) Sepharose (data not shown).

The crystals of this construct diffract to at least 3.0 Å (Fig. 2) and a complete data set from a native crystal was collected to 3.1 Å. It can be concluded from the processed data that the space group of these crystals is $P3_1$ or $P3_2$. The unit-cell parameters are $a = b = 55.1$, $c = 105.2$ Å. The data-processing statistics are shown in Table 1. The self-rotation function calculations with *GLRF* produced no evidence of a local threefold axis, whereas the search for a local twofold axis produced a clear peak (Fig. 3), suggesting that the proline-rich-peptide binding domain has crystallized with a dimer in the asymmetric unit. This is in good agreement with a $V_M$ of 2.4 Å$^3$ Da$^{-1}$, which is in the normal range (Matthews, 1968). The synthesis of a selenomethionine-derivatized domain has been initiated in order to determine the structure.

We thank Merja Nissilä and Eeva Lehtimäki for their expert technical assistance. We gratefully acknowledge the able support of the beamline staff at the Max-Lab synchrotron in Lund. This work was supported by grants from the Health Science Council of the Academy of Finland (45601 and 200471) (JM), the Environmental and Natural Resources Research Council of the Academy of Finland (44198) (RW) and the Finnish Centre of Excellence Programme 2000–2005 (44843) (JM).

## References

Annunen, P., Helaakoski, T., Myllyharju, J., Veijola, J., Pihlajaniemi, T. & Kivirikko, K. I. (1997). *J. Biol. Chem.* **272**, 17342–17348.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Dijkstra, K., Karvonen, P., Pirneskoski, A., Koivunen, P., Kivirikko, K. I., Darby, N. J., van Straaten, M., Scheek, R. M. & Kemmink, J. (1999). *J. Biomol. NMR*, **14**, 195–196.
Helaakoski, T., Annunen, P., Vuori, K., MacNeil, I. A., Pihlajaniemi, T. & Kivirikko, K. I. (1995). *Proc. Natl Acad. Sci. USA*, **92**, 4427–4431.
Kabsch, W. (1993). *J. Appl. Cryst.* **26**, 795–800.
Kemmink, J., Darby, N. J., Dijkstra, K., Nilges, M. & Creighton, T. E. (1996). *Biochemistry*, **35**, 7684–7691.
Kemmink, J., Darby, N. J., Dijkstra, K., Nilges, M. & Creighton, T. E. (1997). *Curr. Biol.* **7**, 239–245.
Kivirikko, K. I. & Myllyharju, J. (1998). *Matrix Biol.* **16**, 357–368.
Kivirikko, K. I. & Pihlajaniemi, T. (1998). *Adv. Enzymol. Related Areas Mol. Biol.* **72**, 325–398.
Lamberg, A., Pihlajaniemi, T. & Kivirikko, K. I. (1995). *J. Biol. Chem.* **270**, 9926–9931.
Macias, M. J., Hyvönen, M., Baraldi, E., Schultz, J., Sudol, M., Saraste, M. & Oschkinat, H. (1996). *Nature (London)*, **382**, 646–649.
Mahoney, N. M., Janmey, P. A. & Almo, S. C. (1997). *Nature Struct. Biol.* **4**, 953–960.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Myllyharju, J. (2003). In the press.
Myllyharju, J. & Kivirikko, K. I. (1997). *EMBO J.* **16**, 1173–1180.
Myllyharju, J. & Kivirikko, K. I. (1999). *EMBO J.* **18**, 306–312.
Tong, L. & Rossmann, M. G. (1990). *Acta Cryst.* A**46**, 783–792.
Xu, W., Harrison, S. C. & Eck, M. J. (1997). *Nature (London)*, **385**, 595–602.
Zeelen, J. Ph., Hiltunen, J. K., Ceska, T. A. & Wierenga, R. K. (1994). *Acta Cryst.* D**50**, 443–447.