crystallization communications

Expression, purification, crystallization and preliminary X-ray studies of a prolyl-4-hydroxylase protein from *Bacillus anthracis*

Megen A. Miller,[a] Emily E. Scott[b]* and Julian Limburg[a]*

[a]Department of Chemistry, The University of Kansas, 1251 Wescoe Hall Drive, Lawrence, KS 66045, USA, and [b]Department of Medicinal Chemistry, The University of Kansas, 1251 Wescoe Hall Drive, Lawrence, KS 66045, USA

Correspondence e-mail: eescott@ku.edu, jlimburg@ku.edu

Collagen prolyl-4-hydroxylase (C-P4H) catalyzes the hydroxylation of specific proline residues in procollagen, which is an essential step in collagen biosynthesis. A new form of P4H from *Bacillus anthracis* (anthrax-P4H) that shares many characteristics with the type I C-P4H from human has recently been characterized. The structure of anthrax-P4H could provide important insight into the chemistry of C-P4Hs and into the function of this unique homodimeric P4H. X-ray diffraction data of selenomethionine-labeled anthrax-P4H recombinantly expressed in *Escherichia coli* have been collected to 1.4 Å resolution.

1. Introduction

Collagen prolyl-4-hydroxylases (C-P4Hs) are essential enzymes in the post-translational modification of procollagen. C-P4Hs catalyze the hydroxylation of proline residues at the Yaa positions of (Gly-Xaa-Yaa)$_n$ repeats in collagen and other proteins containing collagen-like domains, where Xaa is often proline and Yaa is often hydroxyproline (Myllyharju, 2003). The resulting *trans*-4-hydroxyproline residues formed are essential to the stability of the unique collagen triple-helical structure through inter-chain and intra-chain hydrogen bonding (Myllyharju & Kivirikko, 2004). Mature collagen also contains 3-hydroxyproline (Myllyharju, 2005) and recent work has led to the first characterization of recombinant human prolyl-3-hydroxylase isoenzyme 2 (Tiainen *et al.*, 2008).

C-P4Hs are composed of two α subunits and two β subunits. Vertebrate C-P4H is an $\alpha_2\beta_2$ heterotetramer with a total molecular weight of 240 kDa. The α subunit is responsible for the proline-hydroxylation activity and contains an N-terminal peptide-binding domain and a C-terminal catalytic domain. The β subunit is a protein disulfide isomerase, which is responsible for keeping the α subunit properly folded and soluble in the endoplasmic reticulum (Myllyharju, 2003). The α subunit is insoluble in the absence of protein disulfide isomerase and this has hampered attempts to structurally characterize the active-site domain of any of the three isoforms of human C-P4H. Plant and algal forms of C-P4H are soluble monomers and a similar form of C-P4H has also been identified in *Paramecium bursaria* chlorella virus 1 (Eriksson *et al.*, 1999). These monomeric P4Hs share ~30% sequence identity with the C-terminal active-site domain of the α subunit of human type I C-P4H, but lack the peptide-binding domain. We have recently characterized a C-P4H from *Bacillus anthracis*, anthrax-P4H, by expressing it recombinantly in *Escherichia coli* and undertaking thorough kinetic and spectroscopic studies (unpublished work). Like the monomeric forms of C-P4H, anthrax-P4H shares ~30% sequence identity with the active-site domain of type I human C-P4H. In addition to C-P4H, humans also have cytosolic P4Hs that oxidize a specific proline residue in the hypoxia-inducible factor, playing a central role in the hypoxia response (Schofield & Ratcliffe, 2005). Although they catalyze a similar reaction, C-P4Hs and hypoxia-inducible factor P4Hs share little sequence identity apart from the conserved amino-acid residues
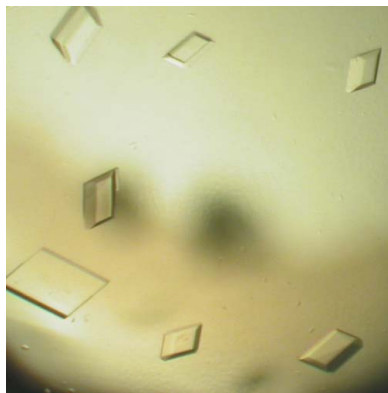
involved in catalysis and are distinct from one another (Myllyharju, 2008).

C-P4Hs belong to a family of α-ketoglutarate mononuclear nonheme iron-dependent dioxygenases which require Fe^{II}, α-ketoglutarate (αKG), O_2 and ascorbate for catalytic activity (Hausinger, 2004). Although this family of enzymes is extensive, they commonly have low sequence identity and catalyze a wide variety of reactions. The general reaction catalyzed by C-P4H is shown in Fig. 1, in which Fe^{II} catalyzes the hydroxylation of proline and αKG undergoes decarboxylation to yield succinate (Myllyla *et al.*, 1984). Similar to other enzymes in the dioxygenase family, the catalytic domain of P4Hs contains sequences consistent with a conserved Fe^{II}-binding site with a characteristic $HXD/E(X)_nH$ sequence as well as a characteristic αKG-binding site. No structures of the tetrameric form ($\alpha_2\beta_2$) of any C-P4H have been reported, but recently the structure of a monomeric C-P4H from the alga *Chlamydomonas reinhardtii* has been determined to 1.9 Å resolution (Koski *et al.*, 2007). This structure was shown to contain an eight-stranded β-helix core fold surrounding the active site (a jelly-roll motif or double-stranded β-helix fold) typical of the αKG-dependent dioxygenases (Clifton *et al.*, 2006). Surrounding the catalytic core are additional α-helices, β-strands and loops that are likely to contribute to the stability of the core.

As predicted from protein-sequence analysis, anthrax-P4H is an αKG-dependent nonheme iron dioxygenase that requires Fe^{II}, O_2 and αKG for activity. The K_m values for αKG and (Gly-Pro-Pro)$_{10}$ for anthrax-P4H are strikingly similar to those reported for type I human C-P4H. Peptidyl hydroxyproline has not previously been observed in prokaryotes, suggesting that anthrax-P4H may play some unique role in the physiology of this pathogen. Elucidation of the anthrax-P4H structure could provide important insights into both substrate binding by C-P4H enzymes in humans and into the function of this unique homodimeric P4H.

2. Materials and methods

2.1. Cloning and expression

Clone BA 4459 containing the 651 bp ORF encoding the putative P4H gene was supplied by The Institute for Genomic Research Pathogen Functional Genomics Resource Center and amplified using the forward primer 5′-CCATGGATGACAAACAACAATCAAA-TAGG-3′ and the reverse primer 5′-GGAATTCCATATGGCCA-ACTTTGTACAAGAAAGC-3′. PCR primers were designed to amplify the gene and to incorporate *Nco*I and *Nde*I sites for subsequent insertion into pET15b (Novagen) in order to express the recombinant protein without an N-terminal His tag. The PCR product was initially ligated into pGEMT-Easy (Promega), sequenced in the forward and reverse directions and then subcloned into pET15b.
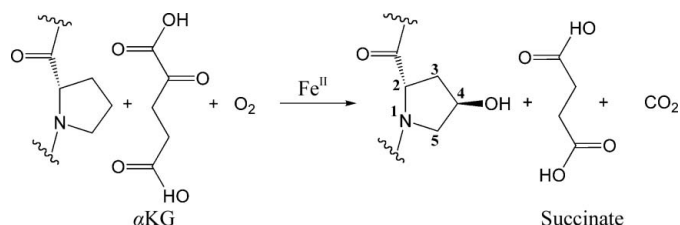
The resulting plasmid was transformed into *E. coli* strain BL21 (DE3) (Novagen) for overexpression. Single colonies from a Luria–Bertani–ampicillin (LB-amp, containing 100 µg l^{-1} ampicillin) agar plate were used to inoculate a 90 ml starter culture (LB-amp broth), which was grown at 310 K overnight. 15 ml aliquots of the overnight culture were used to inoculate 6×1.5 l LB-amp broth. Cultures were grown at 310 K to an OD_{600} of 0.6 and protein expression was induced by the addition of 200 µ*M* isopropyl β-D-1-thiogalacto-pyranoside. The cultures were maintained at 310 K for 3 h after induction, at which point the cells were harvested by centrifugation.

The production of selenomethionine-substituted anthrax-P4H (SeMet-anthrax-P4H) was performed by transforming the expression construct into the methionine-requiring auxotrophic *E. coli* strain B384 (DE3) (Novagen). The bacterial growth was monitored in SelenoMet Medium (Molecular Dimensions Ltd) containing 100 µg l^{-1} ampicillin and 30 µg l^{-1} L-selenomethionine and also induced at an OD_{600} of 0.6 by the addition of 200 µ*M* isopropyl β-D-1-thiogalactopyranoside, similar to the native protein. The SeMet-anthrax-P4H cultures took longer to grow to induction phase and cell growth was maintained at 310 K for 12 h after induction. After 12 h, the cells were harvested by centrifugation similar to native anthrax-P4H.

2.2. Purification

Cells (9 g) were suspended in lysis buffer (50 m*M* Tris–HCl, 1 m*M* EDTA, 5 m*M* β-mercaptoethanol pH 7.4) containing protease inhibitors (1 m*M* PMSF, 1 µ*M* leupeptin, 1 µ*M* antipain, 1 µ*M* pepstatin), lysozyme, DNase and RNase. The cells were lysed using an ultrasonic sonifier (Branson) and then centrifuged at 24 000*g* for 1 h at 277 K. The crude cell lysate was loaded onto a 100 ml DEAE-Sepharose column. The column was washed with three column volumes of lysis buffer and the protein was then eluted with a four column-volume gradient from 0 to 250 m*M* KCl. Fractions containing anthrax-P4H were identified by SDS–PAGE. Those of highest purity (>95%) were pooled, concentrated and loaded onto a HiLoad Superdex 200 16/60 column (GE Healthcare) connected to an ÄKTA FPLC. The protein was eluted with 50 m*M* Tris–HCl, 150 m*M* KCl containing 5 m*M* β-mercaptoethanol pH 7.4 and the fractions containing anthrax-P4H were combined and concentrated to 24 mg ml^{-1}. The protein
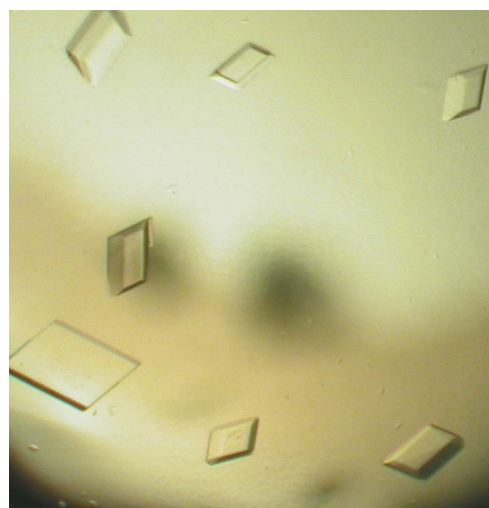
Figure 2
Crystals of anthrax-P4H grown using the hanging-drop method in 16%(*w*/*v*) PEG 8000, 40 m*M* potassium phosphate (monobasic) and 20%(*v*/*v*) glycerol pH 4.0. The approximate dimensions of a typical crystal are $0.8 \times 0.5 \times 0.1$ mm.

Figure 1
Reaction catalyzed by P4H.

Table 1
Summary of crystallographic data for anthrax-P4H.

Values in parentheses are for the highest resolution shell.

	Se peak	Se inflection	Se remote
Space group	$P2_1$	$P2_1$	$P2_1$
Wavelength (Å)	0.9793	0.9795	0.912
Resolution (Å)	53.38–1.40 (1.44–1.40)	53.38–1.40 (1.44–1.40)	53.38–1.40 (1.44–1.40)
Unit-cell parameters (Å, °)	$a = 41.38$, $b = 63.80$, $c = 98.50$, $\beta = 98.74$	$a = 41.38$, $b = 63.80$, $c = 98.50$, $\beta = 98.73$	$a = 41.38$, $b = 63.80$, $c = 98.50$, $\beta = 98.80$
No. of measurements	374798 (27211)	374659 (27152)	376690 (27404)
No. of unique reflections	99723 (7385)	99692 (7369)	99677 (7368)
Redundancy	3.8 (3.7)	3.8 (3.7)	3.8 (3.7)
Completeness (%)	100 (100)	100 (100)	100 (100)
R_{merge}† (%)	4.4 (21.9)	5.0 (22.2)	4.1 (18.8)
Average $I/\sigma(I)$	16.7 (4.8)	14.9 (4.3)	18.1 (5.9)
Wilson B factor (Å^2)	12		

† $R_{merge} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$.

concentration was determined by a Bradford assay using BSA as a standard.

The purification of SeMet-anthrax-P4H was identical to that used for the native form of the enzyme and the overall protein yields of the two forms were comparable. MALDI–TOF mass-spectrometric analysis determined a mean incorporation of three Se atoms per anthrax-P4H monomer, with four methionines present in the sequence including the N-terminus.

2.3. Crystallization

Initial crystallization screening was performed by the hanging-drop vapor-diffusion method using commercially available sparse-matrix screening kits from Hampton Research and Emerald Biosystems. Equal volumes of protein and reservoir solution (1 µl + 1 µl) were mixed and equilibrated against 750 µl reservoir solution at 293 K. Initial rhombic crystals of varying sizes were obtained upon equilibration with 16%(w/v) PEG 8000, 40 mM potassium phosphate

Figure 3
Diffraction pattern of SeMet-anthrax-P4H obtained on beamline BL9-2, SSRL. The resolution of this data set was 1.4 Å.

(monobasic) and 20%(v/v) glycerol. Crystallization conditions were optimized by variation of the pH. Diffraction-quality crystals of anthrax-P4H were formed from 16%(w/v) PEG 8000, 40 mM potassium phosphate (monobasic) and 20%(v/v) glycerol pH 4.0 (Fig. 2). SeMet-anthrax-P4H crystallized under similar conditions where the pH value was 4.2. The crystals used for data collection were plate-shaped and grew within two weeks to approximately 0.8 × 0.5 × 0.1 mm in size.
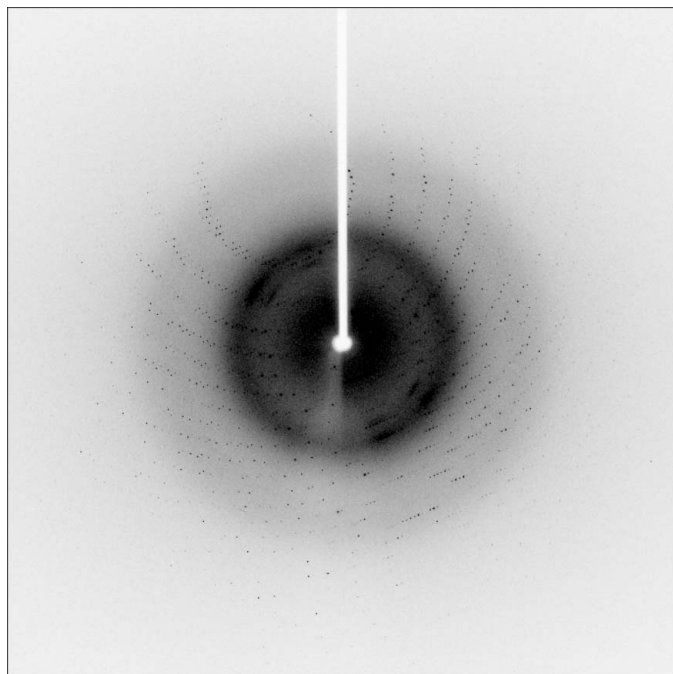
2.4. Data collection and processing

A multiple-wavelength anomalous diffraction (MAD; Hendrickson, 1991) experiment was performed on beamline BL9-2 at the Stanford Synchrotron Radiation Laboratory (SSRL). The SeMet-anthrax-P4H crystals were cryocooled by direct immersion into liquid nitrogen. Because the mother liquor contained high levels of PEG 8000 and glycerol, no additional cryoprotectant was necessary. The harvested crystals were then loaded into a sample cassette designed for use with the Stanford Automated Mounting system and screened for diffraction quality (Fig. 3; Cohen *et al.*, 2002). The wavelengths for anomalous diffraction were determined from the fluorescence absorption spectrum near the Se L_{III} absorption edge. Data collection was performed at 100 K and wavelengths of 0.9793 Å (peak), 0.9795 Å (edge) and 0.912 Å (remote). A total of 180° of data were collected (1° oscillation, 2 s exposure per oscillation) at each wavelength and used for structural determination. The data were processed with the program *MOSFLM* (Leslie, 1999) and scaled with *SCALA* from the *CCP*4 program suite (Collaborative Computational Project, Number 4, 1994).

3. Results and discussion

3.1. Identity, purity and oligomeric state

Recombinant anthrax-P4H could be easily prepared using *E. coli* strain BL21 (DE3), with a typical yield of 10 mg purified protein from a 1 l culture. Anthrax-P4H ran as a single band at ∼27 kDa on SDS–PAGE and eluted from the gel-filtration column consistent with a molecular weight of 54 kDa, suggesting that the protein is a homodimer. As estimated by SDS–PAGE, the purity of the purified protein was >99%. Kinetic studies showed that anthrax-P4H is enzymatically active and turns over in the presence of (Gly-Pro-Pro)$_{10}$ displaying Michalis–Menten kinetics. The K_m values for the peptide and for αKG are very close to those for the human enzyme (unpublished work), substantiating this system as a model for understanding human C-P4H.

3.2. Data collection and preliminary X-ray diffraction analysis

We unsuccessfully attempted to solve the structure of anthrax-P4H using molecular replacement with C-P4H from *C. reinhardtii* (PDB codes 2v4a, 2jig and 2jij); the two proteins share a sequence identity of 30%. We have proceeded to use MAD methods to determine the structure. The statistics of the synchrotron data collection and processing are summarized in Table 1. The presence of two molecules in the asymmetric unit gave a calculated Matthews coefficient (V_M) of 2.62 $\text{Å}^3 \, \text{Da}^{-1}$, which corresponds to a solvent content of 53% (Matthews, 1968). The data were processed using *MOSFLM* and *SCALA* and structure solution by MAD phasing is currently in progress. Comparison of the resulting anthrax-P4H structure with those of other P4Hs, as well as those of other αKG-dependent nonheme iron enzymes, will provide insight into the diversity of these enzymes and provide a structural basis for evaluation of the functionally similar hydroxylation domain of human C-P4H.

References

Clifton, I. J., McDonough, M. A., Ehrismann, D., Kershaw, N. J., Granatino, N. & Schofield, C. J. (2006). *J. Inorg. Biochem.* **100**, 644–669.
Cohen, A. E., Ellis, P. J., Miller, M. D., Deacon, A. M. & Phizackerley, R. P. (2002). *J. Appl. Cryst.* **35**, 720–726.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Eriksson, M., Myllyharju, J., Tu, H., Hellman, M. & Kivirikko, K. I. (1999). *J. Biol. Chem.* **274**, 22131–22134.
Hausinger, R. P. (2004). *Crit. Rev. Biochem. Mol. Biol.* **39**, 21–68.
Hendrickson, W. A. (1991). *Science*, **254**, 51–58.
Koski, M. K., Hieta, R., Bollner, C., Kivirikko, K., Myllyharju, J. & Wierenga, R. K. (2007). *J. Biol. Chem.* **282**, 37112–37123.
Leslie, A. G. W. (1999). *Acta Cryst.* D**55**, 1696–1702.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Myllyharju, J. (2003). *Matrix Biol.* **22**, 15–24.
Myllyharju, J. (2005). *Top. Curr. Chem.* **247**, 115–147.
Myllyharju, J. (2008). *Ann. Med.* **40**, doi:10.1080/07853890801986594.
Myllyharju, J. & Kivirikko, K. I. (2004). *Trends Genet.* **20**, 33–43.
Myllyla, R., Majamma, K., Gunzler, V., Hanauske-Able, H. M. & Kivirikko, K. (1984). *J. Biol. Chem.* **259**, 5403–5405.
Schofield, C. J. & Ratcliffe, P. J. (2005). *Biochem. Biophys. Res. Commun.* **338**, 617–626.
Tiainen, P., Pasanen, A., Sormunen, R. & Myllyharju, J. (2008). *J. Mol. Biol.* **283**, 19432–19439.