ABSTRACT

Proline hydroxylation is the most prevalent post-translational modification in collagen. The resulting product trans-4-hydroxyproline (Hyp) is of critical importance for the stability and thus function of collagen, with defects leading to several diseases. Prolyl 4-hydroxylases (P4Hs) are mononuclear non-heme iron α-ketoglutarate (αKG)-dependent dioxygenases that catalyze Hyp formation. While animal and plant P4Hs target peptidyl proline, prokaryotes have been known to use free L-proline as a precursor to form Hyp. The P4H from Bacillus anthracis (BaP4H) has been postulated to act on peptidyl proline in collagen peptides, making it unusual within the bacterial clade, but its true physiological substrate remains enigmatic. Here we use mass-spectrometry, fluorescence binding, X-ray crystallography, and docking experiments to confirm that BaP4H recognizes and acts on peptidyl substrates but not free proline, using elements characteristic of an Fe(II)/αKG dependent dioxygenases. We further show that BaP4H can hydroxylate unique peptidyl proline sites in collagen-derived peptides with asymmetric hydroxylation patterns. The cofactor-bound crystal structures of BaP4H reveal active site conformational changes that define open and closed forms and mimic “ready” and “product-released” states of the enzyme in the catalytic cycle. These results help to clarify the role of BaP4H as well as provide broader insights into human collagen P4H and proteins with poly-L-proline type II helices.

INTRODUCTION

Prolyl 4-hydroxylases (P4Hs) are ubiquitous enzymes that catalyze the post-translational hydroxylation of proline residues forming 4-hydroxyproline (Hyp) (1). The P4Hs belong to the mononuclear non-heme iron α-ketoglutarate (αKG) family of oxygenases requiring Fe(II), αKG, and O2 for catalysis (2,3). These enzymes are characterized by 2His-1-carboxylate (Asp/Glu) facial triad residues that coordinate to the active site iron center (4-7). The common co-substrate αKG binds to iron via its oxo and carboxylate groups and provides reducing equivalents for these enzymes to reduce O2 to a hydroxyl group (2). The oxidation of various substrates containing prolyl residues is accomplished via the activation of a C-H bond of an unactivated alkyl group (8-10). While one atom from molecular oxygen is incorporated into the substrate, the second atom is utilized in decarboxylation of αKG to form the by-product succinate.

The P4Hs are remarkable in their ability to recognize varied substrates with diverse sequences motifs and are central to the growth, assembly, and function of the host organism (1,11-13). While these enzymes exhibit substrate diversity, they share a general reaction mechanism and a conserved fold at their core. A double-stranded β-helix (DSBH) comprising of eight anti-parallel β-strands contains the conserved active site residues required for iron and αKG binding (Fig. 1) (14,15). Though the DSBH motif is highly conserved in all P4Hs and oxygenases in general, the overall sequence identity is low (11) and the area surrounding the...
DSBH is the region that gives specificity to each enzyme and substrate (14).

In humans, there are two types of P4Hs – a collagen P4H (C-P4H) and hypoxia inducible factor-1α (HIF1α-prolyl hydroxylase domain proteins (PHDs) (13,16-18). The HIF1α-PHDs are involved in the post-translational hydroxylation of a LXXLAP motif in HIF1α and play a key role in cellular response to hypoxia (19-21). Whereas, C-P4H catalyzes the hydroxylation of proline residues at the Y position of (X-Y-G)n repeats (where, n denotes the number of repeated sequences and X is frequently proline) in collagen and other collagenous domain containing proteins (17,22). The (P-P-G)n repeats found in collagen adopt a poly-L-proline type II (PPII) conformation in each strand of the collagen triple helix. The PPII helical residues are known to interact with proline recognition domains (PRDs) and are involved in a variety of processes including bacterial and viral pathogenesis, elasticity, transcription, cell motility, and self-assembly (23,24). The hydroxylation of Y proline residues in collagen is of utmost importance in providing stability to the collagen helical structure. Defects in collagen stability due to insufficient hydroxylation or excessive C-P4H activity can cause diseases like scurvy and fibrosis respectively (1). Since C-P4H is the rate-limiting enzyme in collagen biosynthesis, therefore, it is an attractive target for developing corresponding mechanism-based activators and inhibitors for scurvy and fibrotic disorders.

Three C-P4H isoenzymes exist in vertebrates and these are all αβ2 tetramers. The β-subunit is identical to protein disulfide isomerase and is necessary for preventing the aggregation of α-subunit and to retain it in the endoplasmic reticulum (22,25). The catalytic α-subunit consists of 3 domains: an N-domain responsible for tetramer assembly, a peptide-substrate binding (PSB) domain, and a C-terminal catalytic domain containing the active site (22,26). The crystal structures of the N- and PSB- domains have been determined in presence of a C-P4H inhibitor (poly-(L)-proline) and (P-P-G)3 substrate (26,27). The two peptides adopt a PPII helical conformation and bind to the PSB domain in a groove lined by conserved tyrosines. The PSB domain of C-P4H consists of α-helices that forms tetratricopeptide repeat (TPR) motifs. The TPR motifs are promiscuous in mediating protein-protein interactions and recognize many ligands including proline-rich peptides (28). Since C-P4H acts on X-P-G repeats in collagen and other proteins with collagen-like sequences, the substrate specificity of this enzyme must arise from residues beyond the conserved region of TPR. An open question, which is mainly due to the lack of a structure of C-P4H with an intact catalytic domain, is how the PSB and catalytic domains communicate to position a substrate around the active site for site-specific prolyl hydroxylation. The poor solubility of the α-subunit containing the catalytic site has proved challenging for structural studies and there has been an ongoing effort in pursuit of an improved model for C-P4H.

Previous structures of P4Hs have provided initial insights into enzyme structure and function. A viral collagen P4H (vCPH) has been shown to hydroxylate poly-(L)-proline and prolines in proteins containing (P-A-P-K)n proline-rich repeats (29,30). The crystal structures of vCPH with manganese-bound and a zinc-αKG complex show residues involved in metal and αKG binding and related conformational changes when αKG binds (30). Docking simulations using a PKPAPK peptide showed possible interactions with vCPH and the peptide model forms a PPII helix in the PSB groove of this enzyme (30). Plants and green algae have collagenous domain containing glycoproteins with proline-rich motifs (PRMs) where Hyp is important for cell-wall assembly (31,32). A structure of an algal P4H from Chlamydomonas reinhardtii (CrP4H) complexed with a (S-P)5 peptide is one of the few P4H crystal structure solved to date with a substrate bound in the active site (33). The structure illustrates various enzyme-substrate interactions in the binding groove and in the substrate binding loop (SBL) region adjacent to the active site, with the peptide adopting the PPII helix conformation (34,35). These highly mobile SBL regions are conserved through P4Hs and in the absence of a substrate one of the loops is usually disordered (30,36,37).

Similar to other organisms bacteria contain Hyp, but a distinctive characteristic of bacterial
Collagen peptide hydroxylation by BaP4H

P4Hs is to produce Hyp from free L-proline instead of peptidyl proline (38). On the other hand, *Bacillus anthracis* P4H (BaP4H) is unique among bacterial P4Hs in that it is reported to bind the human collagen peptide (P-P-G)$_n$ with an affinity similar to that of human C-P4H (1,36,39). A sequence alignment of BaP4H with human C-P4H shows ~30 % sequence identity. Structural alignment of BaP4H with CrP4H and vCPH structures shows conserved secondary structure similarities between various P4Hs (Fig. 1). Examining the proteome of *Bacillus anthracis* reveals collagen-like proteins that contain stretches of (G-X-Y) repeats. A genome-based study of collagen structural motifs (CSMs) in bacterial and viral proteins showed that bacterial CSMs favor a (G-P-T) triplet unit rather than the second most prevalent (G-P-Q) repeat unit (40). Since the proteome of *Bacillus anthracis* contains (G-P-T) repeats in some proteins, it has been proposed that a possible physiological substrate for BaP4H could be a protein consisting of (G-P-T)$_n$ repeats (36,41). However, the substrate for BaP4H is not yet fully established.

The previously reported crystal structure of BaP4H is an apo-form of the enzyme lacking necessary cofactors (36,41). To gain insight into the site-specific prolyl hydroxylation activity of BaP4H, it is important to understand how the active site residues are organized to assemble the cofactors, iron and aKG, for substrate recognition.

In the current study we examine cofactor and substrate binding, including site-specific hydroxylation activity of BaP4H by a combination of biochemical, mass-spectrometry, and X-ray crystallographic methods. Attempts towards determining the crystal structure of a substrate bound form of BaP4H has not been fully successful, therefore molecular docking was used to simulate the binding mode of a collagen-like proline-rich (P-P-G)$_3$ peptide. Here we describe the first biochemical and structural analysis of the cofactor-bound forms of BaP4H, and identified collagen-derived (P-P-G)$_{3,10}$ peptides as substrates with hydroxylation of proline residues occurring in both X and Y positions. Furthermore, the crystal structures reveal conformational changes upon cofactor binding, and represent “ready” and “product-released” states of the enzyme in the catalytic mechanism (42).

**EXPERIMENTAL PROCEDURES**

*Materials*— All buffers, media ingredients, and other reagents were acquired from Sigma-Aldrich (St. Louis, MO) and, unless otherwise stated, were of the highest purity available. Solutions were prepared using Nanopure deionized water. N$_2$ (99.98 %), NO (>99 %), and liquid helium (>99 %) were obtained from Praxair (Cedar Rapids, IA).

*Protein Expression and Purification*— The BaP4H gene was synthesized by Genscript (Piscataway, NJ), and was then sub-cloned into the pET28a expression vector using restriction sites NcoI and NdeI. The BaP4H plasmid was transformed into *E. coli* BL21 star DE3 pLysS cells (Life Technologies, Grand Island, NY). Apo-BaP4H was expressed and purified with slight modification as reported previously (41). The BaP4H-(P-P-G)$_3$ construct was made by PCR using appropriate primers to extend the C-terminus end to contain a (P-P-G)$_3$ peptide. The PCR product was cut with NcoI and BamHI and ligated back into pET28a. This new plasmid was transformed, expressed and purified as above.

*UV-visible Spectroscopy Measurements*— To prevent any turnover, all UV-visible measurements were performed in an anaerobic chamber (Vacuum Atmospheres, Hawthorne, CA) using an Ocean Optics DH-2000-BAL light source (Ocean Optics, Dunedin, FL). Anaerobic solutions were made from dry stocks in a glovebox. Spectra were measured between 250-800 nm in a buffer containing 50 mM Tris pH 7.4, 30 mM NaCl in a 100 µL quartz cuvette.

The spectrum of Fe(II)-BaP4H was measured after adding equimolar concentrations of Fe(NH$_4$)$_2$(SO$_4$)$_2$ and 776 µM apo-BaP4H. To detect any spectral changes associated with aKG binding, stoichiometric aliquots of Fe(II) and aKG were added to a 776 µM solution of apo-BaP4H and spectra were recorded after each addition. The difference spectrum of Fe(II)-BaP4H-aKG ternary complex (solid line, Fig. 2A), was obtained by subtracting the Fe(II)-BaP4H spectrum (dashed line, Fig. 2A) from the original ternary complex spectrum to highlight the absorption maximum at 520 nm.
characteristic of αKG-dependent non-heme iron dioxygenases (43,44).

To determine the metal stoichiometry, a 100 µL solution containing equimolar concentration of BaP4H and αKG (800 µM) was titrated with a 2 mM stock solution of Fe(NH4)2(SO4)2 anaerobically in a glovebox. Aliquots of Fe(II) were added to this solution of BaP4H/αKG until the absorption at 520 nm had saturated.

Electron paramagnetic resonance spectroscopy (EPR)—X-band EPR spectra were recorded on a Bruker EMX spectrometer (Bruker Biospin Corp., Billerica, MA), equipped with an Oxford ITC4 temperature controller, a Hewlett-Packard model 5340 automatic frequency counter and Bruker gaussmeter. The EPR samples were prepared anaerobically with an enzyme concentration of 100 µM in 50 mM Tris pH 7.4, 150 mM KCl, 5 mM β-mercaptoethanol. Reduced Fe(II)-BaP4H samples were prepared by adding sodium dithionite to a final concentration of 1mM. NO gas was bubbled into the protein solutions with cofactors in an anaerobic sealed vial and the solution transferred to an EPR tube via Hamilton syringe. All spectra were collected at microwave frequency, 9.43 GHz; receiver gain, 2 × 10^4; modulation frequency, 100 kHz; temperature, 4 K; microwave power, 10 milliwatts; modulation amplitude, 10 G; sweep time, 83.89 s; and 16 scans. The Fe(II)-BaP4H spectrum (data not shown) was recorded using 1 scan keeping all other parameters the same.

Fluorescence Binding Titrations—Fluorescence spectra were recorded at room temperature using a Cary Eclipse Fluorescence Spectrophotometer (Agilent Technologies, Santa Clara, CA). Fluorescence titration was used to determine the binding affinity of substrates and cofactors to BaP4H. The decrease in intensity of intrinsic tryptophan fluorescence at 295 nm served as an indicator of cofactor binding. Titration experiments were performed using an enzyme concentration of 1 µM in 50 mM Tris pH 7.4, 30 mM NaCl. All titrations were performed anaerobically using solutions that were prepared in a glove box. In a typical anaerobic titration, a degassed protein solution (3 mL) was sealed in a screw top quartz cuvette with a septum to prevent diffusion of O2 into the anaerobic solution. Anaerobic cofactor solutions were added in aliquots via syringe to the quartz cuvette. The wavelength used for excitation was 295 nm and the emission was monitored from 310-380 nm using excitation and emission slit widths of 5 nm. The signal intensity was recorded at the maxima, 329 nm.

Hydroxylation Product Identification by Liquid Chromatography Tandem Mass Spectrometry—All mass spectra LC-MS/MS data were collected on an Orbitrap Fusion mass spectrometer (Thermo Fisher Scientific) equipped with an Easy nLC 1000 for sample handling and peptide separations. Approximately 200 fmol of peptide resuspended in 5% formic acid + 5% acetonitrile was loaded onto a 125 µm inner diameter fused-silica microcapillary with a needle tip pulled to an internal diameter less than 5 µm. The column was packed in-house to a length of 15 cm with a C18 reverse phase resin (120 Å pore size, 5.0 µm particle size, GP-C18, SePax Technologies). The peptides were separated using a 20 min linear gradient from 3% to 35% buffer B (100% ACN + 0.125% formic acid) equilibrated with buffer A (3% ACN + 0.125% formic acid) at a flow rate of 600 nL/min across the column. The Orbitrap Fusion mass spectrometer was operated in the data-dependent positive ion mode using the top speed strategy. In brief, the scan sequence for the Fusion Orbitrap began with an MS1 spectrum (Orbitrap analysis, resolution 120,000, 800-1000 m/z scan range for (P-P-G)10, 600-800 m/z scan range for (P-P-G)5), AGC target 5 × 10^5, maximum injection time 100 ms, dynamic exclusion of 5 seconds). ‘Top speed’ (2 seconds) was selected for MS2 analysis, which consisted of HCD (quadrupole isolation set at 2.0 Da and ion trap analysis, AGC 1 × 10^5, collision energy 25, maximum injection time 250 ms). MS2 fragment ions were analyzed in the Orbitrap, resolution 15,000. A suite of in-house software tools were used for .RAW file processing and controlling peptide false discovery rates (45). MS/MS spectra were searched against a custom database containing the (P-P-G)10 sequence and 100 common contaminants with both the forward and reverse sequences. Database search criteria are as follows: tryptic with two missed cleavages, a precursor mass tolerance of 1 Da, fragment ion mass tolerance of 0.1 Da, and variable
hydroxylation of proline (15.99491 Da). Peptides were filtered to a 1% false discovery rate using linear discriminate analysis (45). We used a modified version of the Ascore algorithm to quantify the confidence with which each hydroxylation site could be assigned to a particular residue. Hydroxylation sites with Ascore values >13 (P ≤ 0.05) were considered confidently localized to a particular residue (45). The enzymatic assays to verify peptide hydroxylation were performed aerobically in 100 mM Tris pH 7.4, using (10 µM or 50 µM) apo-BaP4H pre-incubated anaerobically for 30 min with (10 µM or 50 µM) Fe(II). A reaction mixture containing 500 µM αKG, ascorbate (10 µM or 50 µM), and 100 µM of either (P-P-G)₅ or (P-P-G)₁₀ were prepared at 20 °C. The reactions were quenched in 0.5 % TFA and vacuum concentrated (SpeedVac Concentrator, Savant, SPD131DDA). The dried powder containing the product was collected and stored at 4 °C until it was analyzed by mass spectrometry.

Data Analysis and Equations— Data in Fig. 2 were fit with Sigma Plot 12.0 (Systat Software Inc., Point Richmond, CA) using a ligand-binding model (Eq. (1)) as described previously (46,47),

\[
f = f_0 + \left( f_m - f_0 \right) \left( \frac{nP + x + K_d}{2nP} \right)^2 - 4nPx
\]

where f is the fluorescence signal resulting from binding of the metal cofactor, f₀ is the signal from a protein solution in the absence of any cofactors, fₘ corresponds to the maximal quenched fluorescence intensity when a cofactor was bound, Kₐ is the dissociation constant, P and x are total protein and added metal ion/ligand concentrations respectively, n is the number of binding sites. A better approximation of the binding affinity can be obtained by using Eq. (1) in cases where the protein concentration is higher than the estimated Kₐ (46,47).

Crystallization of apo-BaP4H— Initial crystals of apo-BaP4H were obtained using the hanging-drop vapor-diffusion method with optimization around a previously reported condition (41) by varying pH and PEG concentrations. Crystal drops were set up at 20 °C with 1 µL of 24 mg/mL apo-BaP4H mixed with 1 µL of precipitant solution containing 0.04 M potassium phosphate monobasic pH 4.0-7.0, 13-18% PEG 8000, and 20% glycerol. Crystals would appear in 1-2 days and the best crystal used for data collection was obtained from 0.04 M potassium phosphate monobasic pH 6.0, 14% PEG 8000, and 20% glycerol. Crystals were taken from the precipitant solution and directly frozen in liquid nitrogen.

Crystallization of Co-BaP4H-MLI— A solution of 9 mg/mL apo-BaP4H, 1 mM CoCl₂, and 1 mM L-Pro was incubated on ice for 30 minutes before setting crystal drops. Crystals were obtained using hanging-drop vapor-diffusion at 20 °C from the PACT screen (Molecular Dimensions) by mixing 1 µL of the above protein solution with 1 µL of reservoir solution. Initial crystals appeared in 0.1 M malonate-imidazole-boric acid (MIB) pH 6.0, 25% PEG 1500 and were optimized by varying pH and precipitant concentration. Diffraction quality crystals were obtained from 0.1 M MIB pH 6.5, 18% PEG 1500. The crystals of Co-BaP4H-MLI were plate shaped and grew in about 2 days. The crystals were soaked in a cryoprotectant solution of 0.1 M MIB pH 6.5, 18% PEG 1500, and 20% glycerol for 5 minutes and submerged in liquid nitrogen.

Crystallization of Co-BaP4H-PPG— Crystals of BaP4H fused with (P-P-G)₃ peptide to the C-terminus were obtained from the JCSG screen (Qiagen) using sitting-drop vapor-diffusion at 20 °C. A solution of 12 mg/mL BaP4H-(PPG)₃, 1 mM CoCl₂, and 1 mM αKG was incubated on ice for 30 minutes. Crystal drops were set up by mixing equal volumes of above protein solution with reservoir solution. Crystals were obtained in 0.15 M KBr and 30% PEG 2000 MME. Plate-shaped crystals would appear in about 2 days with dimensions of ~70 × 70 × 20 µm. The crystals were soaked in cryoprotectant solution containing 0.15 M KBr, 30% PEG 2000 MME, and 20% glycerol for 5 minutes, prior to freezing in liquid nitrogen.

Data Collection and Analysis— Data sets for apo-BaP4H and Co-BaP4H-PPG were collected
Collagen peptide hydroxylation by BaP4H

at 100 K at the Advanced Light Source, beamline 4.2.2. Data collection parameters were: apo-BaP4H, 0.5° oscillation, 4 s exposure time, 120 mm detector distance; and Co-BaP4H-PPG, 0.1° oscillation, 0.4 s exposure time, 250 mm detector distance. The data set for Co-BaP4H-MLI was collected at 100 K at the Advanced Photon Source, beamline 19-ID-D using 0.5° oscillation, 0.6 s exposure time, and 182 mm detector distance. The data set for Co-BaP4H-MLI was collected at 100 K at the Advanced Photon Source, beamline 19-ID-D using 0.5° oscillation, 0.6 s exposure time, and 182 mm detector distance. The data were indexed and integrated using the XDS package (48) and merged and scaled using SCALA (49) (CCP4 suite) (50). Fluorescence scans were taken for cobalt containing crystals and a peak for cobalt was observed at 6.92 KeV. Data collection and refinement statistics are summarized in Table 2.

Structure Determination—Molecular replacement for apo-BaP4H was done with PHASER (51) using 3ITQ as a starting model. Phases for Co-BaP4H-MLI and Co-BaP4H-PPG were obtained by molecular replacement with our apo-BaP4H structure. Matthews coefficients and estimated solvent contents for apo-BaP4H, Co-BaP4H-MLI, and Co-BaP4H-PPG were 2.15 Å³/Da, 43%; 1.96 Å³/Da, 37%; and 2.03 Å³/Da, 39%, respectively. Refinement was carried out using phenix.refine from the PHENIX software package (52). Structures containing metals had bond length and angle parameters generated by phenix.metal_coordination. Following iterative rounds of refinement using simulated annealing, energy minimization, real space refinement, and B-factor refinement, model building was done in COOT (53). Addition of ligands into distinct positive electron density was based on a simulated annealing omit map generated by omitting the ligand. Water molecules were added into clear densities in later rounds of refinement and composite omit maps were generated to verify the structures. All structures contain a disordered N-terminus and flexible loop region that varies slightly for each model. The disordered residues that were not modeled for each structure and chain are as follows: apo-BaP4H chain A 1-10, chain B 1-11, 73-82; Co-BaP4H-MLI chain A: 1-11, 72-80, chain B: 1-11, 71-81; Co-BaP4H-PPG chain A: 1-11, 70-73, 79-83, chain B: 1-11, 70-83. The 2nd and 3rd repeat of (P-P-G)₃ in the Co-BaP4H-PPG structure were also disordered and not included in the final model. In Co-BaP4H-MLI, no density for proline appeared in the structure and instead density for malonate (from crystallization solution) was observed in the final model. The Ramachandran geometries of each structure were analyzed by PROCHECK (54,55) (Table 2). All structural figures were made using PyMOL.

RESULTS

Analysis of cofactor and substrate binding to BaP4H—To probe the nature of spectral changes associated with cofactor and substrate binding to BaP4H, we employed UV-visible and EPR in conjunction with fluorescence saturation studies. The UV-visible spectrum of as-isolated BaP4H exhibited no absorption features above 280 nm. Anaerobic addition of Fe(II) had no effect on the absorption spectrum. However, when \(\alpha\)KG was added to complex of Fe(II)-BaP4H the absorbance in the 300-500 nm region developed a broad shoulder with the absorption maximum at 520 nm (Fig. 2A). This feature is due to a charge transfer transition of the \(\alpha\)KG-Fe(II)-Ba4H ternary complex and is characteristic of \(\alpha\)KG-dependent non-heme iron dioxygenases (43,44,59). The absorbance at 520 nm was monitored to determine the stoichiometry of Fe(II) binding to the enzyme. Titrating Fe(II) into an anaerobic solution of BaP4H in the presence of \(\alpha\)KG resulted in an increase in absorption at 520 nm and showed a
clear inflection point at 1:1 ratio of Fe(II) to BaP4H monomer (data not shown).

To investigate the interaction of ferrous iron with dioxygen in BaP4H we used nitric oxide (NO), which serves as a surrogate of dioxygen and assessed the NO-Fe(II)-BaP4H complex by EPR. BaP4H with reduced Fe(II) exists in an EPR-silent \( S = 2 \) spin state, similar to other non-heme iron proteins (data not shown). When NO interacts with the Fe(II) center, the spin state is changed from \( S = 2 \) into an EPR-active \( S = 3/2 \) species. EPR signals associated with this species are \( g = 3.94, 1.97 \) (Fig. 2B, top) (60-63). Anaerobic addition of NO to the reduced \( \alpha \)KG-Fe(II)-BaP4H complex diminished the intensity of the signal at \( g = 3.91 \) and the nitrosyl signal at \( g = 1.98 \) predominates (Fig. 2B, bottom). In addition, a signal due to oxidized ferric iron is detectable in both NO spectra at \( g = 4.31 \). These results suggest that \( \alpha \)KG binds to the Fe(II) center and alters the electronic environment without displacing bound NO, as indicated by the small changes in the observed \( g \)-values.

To determine the binding affinities of various cofactor and potential substrates with BaP4H, intrinsic tryptophan fluorescence served as a probe for ligand binding. Since \((P-P-G)_{10}\) was suggested to bind to BaP4H with an affinity similar to that of C-P4H (36) we sought to determine the binding affinities of \((P-P-G)_{10}\) and \((P-P-G)_{5}\) peptides in addition to L-proline. By measuring the decreasing tryptophan fluorescence intensity at 329 nm while titrating in the appropriate ligand to \( \alpha \)KG-Fe(II)-BaP4H ternary complex, the binding affinities were determined. The \( K_d \) of \((P-P-G)_{10}\) and \((P-P-G)_{5}\) peptides were calculated to be \( 0.74 \pm 0.11 \) \( \mu M \) and \( 0.3 \pm 0.02 \) \( \mu M \) respectively (Fig. 2C). While free L-proline binds to Fe(II)-BaP4H (\( K_d \) of 7.9 \( \pm 2.5 \) \( \mu M \)), the affinity is an order of magnitude lower than the peptides (Figs. 2C, 2D). The binding affinity of a known P4H inhibitor malonate with Fe(II)-BaP4H complex was determined to be \( 4.2 \pm 0.5 \) \( \mu M \) and Co(II) binds to apo-BaP4H with a \( K_d \) of \( 0.81 \pm 0.09 \) \( \mu M \) (Figs. 2E, 2F).

**Hydroxylation of human collagen-like (P-P-G)\(_{5,10}\) peptides by BaP4H**— It has been proposed that BaP4H hydroxylates peptidyl proline but not free L-proline, a common substrate for bacterial P4Hs. To establish the specific targets recognized by BaP4H, we conducted enzymatic assays using \((P-P-G)_{5}\), \((P-P-G)_{10}\), and free L-proline. While BaP4H catalyzes the hydroxylation of both \((P-P-G)_{5}\) and \((P-P-G)_{10}\), it is unreactive towards free L-proline.

To identify the hydroxylated peptides from BaP4H catalyzed reactions, liquid chromatography mass spectrometry (LC-MS/MS) was performed on products isolated from reactions carried out under steady state and single turn-over conditions using either \((P-P-G)_{10}\) or \((P-P-G)_{5}\) peptides. The extracted ion chromatogram (XIC) of the \((P-P-G)_{5}\) peptide eluted as a sharp peak around 13.3 min (Fig. 3A, black trace) and the hydroxylated-(P-P-G)\(_{5}\) ((P-P-G)\(_{5}\)-OH) eluted as a major peak at 12.3 min and a shoulder at 11.8 min (Fig. 3A, green trace). The major peak consists primarily of \((P-P-G)_{5}\)-OH, with prolines hydroxylated in the \( Y \) position of second and third repeats (positions 8 and 5, Table 1). The shoulder consists of \((P-P-G)_{5}\)-OH, with hydroxylated prolines in the fourth repeat (positions 11 and 10, Table 1). The XIC of the \((P-P-G)_{10}\) peptide eluted as a sharp peak around 16.3 mins and the hydroxylated-(P-P-G)\(_{10}\) ((P-P-G)\(_{10}\)-OH) eluted as a broad peak from 15.5 to 17.3 min with multiple shoulders (Fig. 3B). Unlike \((P-P-G)_{5}\)-OH we are uncertain of the exact hydroxylated species of \((P-P-G)_{10}\)-OH that corresponds to each of the shoulders in Fig. 3B. This is due to the fact that a small modification such as hydroxylation relative to a large molecule like \((P-P-G)_{10}\) molecule has little effect on the retention time of the modified peptide. Therefore, it made it difficult to resolve the different hydroxylated forms under our experimental conditions. While hydroxylated \((P-P-G)_{5}\) appears at \( m/z \) 645.8298 that corresponds to a doubly charged species, the spectrum of \((P-P-G)_{10}\) contained a peak with \( m/z \) 849.4352, corresponding to a triply charged single hydroxylation of \((P-P-G)_{10}\) (Figs. 3C and 3D).

Under steady state conditions hydroxylation sites were localized to both \( X \) and \( Y \) prolines on the third and fourth repeats of \((P-P-G)_{5}\) peptide (Table 1). Whereas, with \((P-P-G)_{10}\) although hydroxylated peptides were identified, due to low localization scores we were unable to assign specific hydroxylation sites.

The single turn-over conditions uncovered additional hydroxylation sites on both the \((P-P\-
G)₅ and (P-P-G)₁₀ peptides. The MS2 data of (P-P-G)₅ identified two additional peptides: one with a hydroxyl contained on the Y proline of the second repeat; and a second peptide that is doubly hydroxylated with the hydroxyl groups located on adjacent prolines in the fifth repeat (Table 1). The MS2 fragment ion pattern for the Y proline on the fourth repeat is shown in Fig. 4A. Interestingly, hydroxylation localized to the X proline of the fourth repeat was also observed (Table 1). With (P-P-G)₁₀, hydroxylated peptides were indeed detected with acceptable localization site scores (>15), which irrevocably allowed assignment of hydroxylation sites. Preferential hydroxylation of proline residues towards the C-terminus end of the (P-P-G)₁₀ peptide were observed, with peptides containing a hydroxyl group on the X proline of fifth and seventh repeats, and the Y proline of the ninth repeat (Table 1). The MS2 fragment ion pattern for the X proline on the fifth repeat is shown in Fig. 4B. In addition, a doubly hydroxylated peptide containing hydroxyl groups on adjacent prolines of the seventh repeat was observed (Table 1). While small hydroxylation of the proline residues in the X positions of (P-P-G)₁₀ were observed previously with Arabidopsis thaliana P4H (At-P4H-1), there was greater uncertainty associated with the true degrees of hydroxylation (31). The MS results presented here irrefutably revealed the peptides recognized by BaP4H, including specific prolyl hydroxylation sites.

**BaP4H Structural Overview**— To establish how cofactors and substrate peptides are assembled, two structures of BaP4H in the presence of Co(II) were determined: one in complex with αKG and a (P-P-G) peptide; and the other structure in complex with a P4H inhibitor, malonate (Figs. 5, 6, and Table 2). In addition, a structure of apo-BaP4H was obtained (Fig. 7, Table 2) near physiological pH and in a space group different from the previously published apo-BaP4H structure (36). To obtain a peptide-bound structure of BaP4H, a shorter (P-P-G) peptide was incorporated into the C-terminus end of the protein close to the predicted substrate binding groove. The resulting structure had clear density only for one repeat that is located close to the active site of the symmetry related molecule (Fig. 5). Attempts to determine Fe(II)-BaP4H structure have failed owing to the poor diffraction quality of the resultant crystals.

All three structures solved contain two molecules in the asymmetric unit, therefore, for simplicity all discussion hereafter will be based on chain A from each structure. The overall structures are similar, with root-mean-square deviations (rmsd) of 0.21-0.25 Å over almost 155 Ca atoms. With the exception of chain A in apo-BaP4H, all other chains in the BaP4H structures have about 10-15 disordered residues encompassing the region between residues 70-83, which we refer to as the substrate binding loop (SBL). Although there is electron density around the SBL region in both Co-bound BaP4H structures indicating the presence of flexible amino acids, the density quality is too poor to model and the structures re-converge at residue 84.

Structural alignment between BaP4H and an algal P4H (CrP4H), with an rmsd of 0.60 Å over 117 Ca atoms, shows that two short β-strands (β4 and β5) should be located in the disordered SBL region and are flanked between two α-helices (α1 and α2) and the β3 and β6 strands (Fig. 1). Although the apo-BaP4H structure is the only one with a complete loop, subtle differences were seen in the two cobalt bound structures with partially complete SBLs. The SBL in the apo-form is positioned in an open conformation perpendicular and away from the active site. Although Co-BaP4H-MLI and Co-BaP4H-PPG structures do not have a completely ordered SBL, however, it appears to be adopting a slightly more closed conformation with the SBL oriented towards the protein surface, which could be due to crystallographic contacts with residues that are ordered in the loop region. A truly closed conformation of the SBL would not be expected until a peptide that is not crystallographically constrained is bound.

**Active site features of Co-BaP4H-PPG and Co-BaP4H-MLI complexes**— Structural comparison of Co(II)-bound forms of BaP4H with our apo-BaP4H shows that regions and residues around the active site within the core β-barrel motif become more organized upon binding of cofactors and/or peptide (Figs. 1, 5-8). Strands βII and βVII bring triad residues His 127 (2.1 Å), Asp 129 (2.1 Å), and His 193 (2.0 Å) into proximity of coordination to the cobalt.
atom. The remaining cobalt coordination sites are occupied by either αKG or malonate along with water molecules, resulting in a six-coordinate complex (Fig. 6).

In Co-BaP4H-PPG, αKG coordinates cobalt in a planar fashion via its C1 carboxylate (2.0 Å) and the C2 ketone (2.1 Å) opposite of Asp 129 (Figs. 6A and 6B). The C5 carboxylate of αKG is anchored in the active site by hydrogen bonding interactions with Tyr 118 (2.6 Å), Lys 203 (2.9 Å), and Thr 159 (2.6 Å). One of the C1 carboxylates is hydrogen bonded to Thr 207 (2.6 Å) (Fig. 6B). The bidentate binding of αKG to metal center and stabilization by lysine, arginine, and threonine interactions is conserved in the Fe(II)/αKG-dependent enzymes, although the positioning of αKG differs among various enzymes (2). Surprisingly, in the Co-BaP4H-PPG structure, the peptide was located in a symmetry related molecule adjacent to the active site (Fig. 5). Although the (PPG)_3 peptide was disordered with density existing for only one repeat, however, the structure displays that the central Y Pro of the peptide points to the cobalt center and is at a distance of 5.2 Å (Fig. 6C), which is slightly longer than that observed for CrP4H (30,34). In Co-BaP4H-MLI, malonate occupies the αKG binding space in the active site, with its C1 carboxylate located in the same position as the C2 ketone of αKG (Figs. 6 and 8). Malonate coordinates cobalt in a monodentate mode with the C1 carboxylate trans to Asp 129 at a distance of 2.6 Å from the cobalt center. The C3 carboxylate forms a salt bridge with Lys 203 and is hydrogen bonded to Thr 159 (Figs. 6D and 6E).

In both Co(II)-bound structures, in addition to the catalytic triad residues and αKG or, malonate that coordinate cobalt, the remaining coordinate sites are occupied by water molecules thus forming a six-coordinate cobalt complex. In the αKG bound Co-BaP4H-PPG, a water molecule is located 4 Å away from cobalt nearly trans to His 193 in both chains in the structure, which is likely the site where dioxygen would bind with iron present (Fig. 6B). In Co-BaP4H-MLI, two water molecules are coordinating cobalt at distances of ~2.5 Å (Figs. 6D and 6E). The apo-BaP4H structure has multiple water molecules spread through the active site (Fig. S2). One of the waters (W1) occupies the space where metal would bind and is hydrogen bonded to Asp 129 and His 193 (Fig. 7A). Two of the other water molecules (W2 and W3) are hydrogen bonding with Thr 159 and Lys 203, while a fourth water (W4) is hydrogen bonding with W2 and W3 (Fig. 7B).

**Structures of BaP4H reveal open and closed active site conformations—** In the Co-BaP4H-PPG structure, binding of αKG and peptide induces a conformational change in the β-domain involving strands βI, βII, and βIV. These strands are relatively disordered in the apo- and malonate-bound structures, but become more ordered in the Co-BaP4H-PPG structure. Furthermore, the backbone between residues 122-125 in the apo-BaP4H structure is shifted out away from the active site by about 2 Å compared to both cobalt-bound structures (Fig. 8), which indicates that regions of the active site must have greater conformational flexibility.

In the BaP4H structures, we speculate that the residues important for tuning the active site include Tyr 118, Tyr 124, and Phe 160. These residues are located on the ‘tunable’ face of the main β-barrel motif and are around 5-10 Å away from the metal center (Figs. 6B, 6E, and 8). Tyr 124 exhibits different conformations in the active site. In the apo-form, Tyr 124 is 10 Å (measured from the OH to water located at the metal site) away from the metal site and flipped upwards about 90° from the carbon backbone (Figs. 7 and 8). In Co-BaP4H-PPG, this tyrosine is closer to the active site at approximately 5.3 Å from cobalt and is oriented such that the plane of the aromatic ring is parallel to the metal-binding site (Figs. 6B and 8). In Co-BaP4H-MLI, malonate binding induces a conformational change causing movement within key active site residues (Tyr 118, Tyr 124, and Phe 160). The aromatic ring plane of Tyr 124 is oriented perpendicular to the metal binding site and the hydroxyl group is about 7 Å away from cobalt (Figs. 6E and 8). The different open and closed conformations of Tyr 124 suggest that it is presumably a catalytically important residue.

Similarly, Tyr 118 has multiple conformations and is important in αKG and malonate binding. The apo-form has Tyr 118 oriented perpendicular to the αKG site. In Co-BaP4H-PPG, it has a similar orientation but is shifted down by 1.6 Å to position itself about
2.6 Å from the C5 carboxylate of αKG. However, in Co-BaP4H-MLI Tyr 118 is swung outwards away from the active site, at distances of 4.9 Å and 6.5 Å from malonate C3 and C2 carboxylate ends respectively. Another residue that is more open in the apo structure is Phe 160, located in strand βIV. It protrudes out in apo-BaP4H but is oriented more towards the active site when cobalt is bound (Fig. 8).

Docking of (P-P-G)₅ peptide with Co-BaP4H-PPG structure—In the absence of a true peptide bound crystal structure of BaP4H, (P-P-G)₅ was modeled into the active site of the αKG-bound Co-BaP4H complex. The model shows that the (P-P-G)₅ peptide forms an elongated PPII helical conformation similar to that observed in CrP4H (34,35) (Fig. 9A). The peptide binds BaP4H between β₄, β₅ (missing loops), and β₆ loops and the region between α₁, βI, and βII. It interacts with the protein through hydrophobic residues Phe 85 from β₆ and Phe 131 (Fig. 9B). It forms hydrogen bonds with protein residues Glu 111, Arg 142, Lys 163, and Tyr 215 through backbone carbonyl atoms and sidechains. A proline from the second repeat (P₃) is involved in hydrogen bonding with R142. The peptide is positioned within the active site of BaP4H with the C4 carbon atom of Y proline of second repeat (P₄) directed towards the metal center at a distance of 4.6 Å (Fig. 9B), thus indicating that it is a probable hydroxylation site. This model is supported by our mass spectrometry data showing that the Y proline residue in the second repeat of (P-P-G)₅ peptide (P₄) is hydroxylated by BaP4H (Table 1).

**DISCUSSION**

The high-resolution MS results demonstrate for the first time that BaP4H can hydroxylate proline residues in both X and Y positions of collagen-like (P-P-G)₅ and (P-P-G)₁₀ peptides. Animal CP4H is known to solely hydroxylate proline in the Y position (64-67). Earlier studies have demonstrated that Hyp in the Y position of collagen repeats contributes greater stability to the collagen triple helix through stereoelectronic effects of the inserted oxygen atom (1). While preferential hydroxylation of a proline in the X position of (P-P-G)₁₀ has been observed with CrP4H-1A (32), At-P4H-1 acts preferentially but not exclusively on proline in the Y position (31). Interestingly, the MS data revealed asymmetric hydroxylation of (P-P-G)₅,₁₀ peptides by BaP4H (Table 1). Both C-P4H (65,67) and At-P4H-1 (31) displayed an asymmetrical hydroxylation preference in (P-P-G)₅,₁₀ peptides, whereas, CrP4H-1A (32) evenly hydroxylated the Pro residues in (P-P-G)₁₀. Furthermore, BaP4H hydroxylates the Y proline in the ninth triplet of (P-P-G)₁₀ in addition to other sites (Table 1), which is consistent with the hydroxylation pattern observed previously using ¹⁴C-labeled (P-P-G)₁₀ peptide and either C-P4H (67) or At-P4H-1 (31). With (P-P-G)₅, BaP4H catalyzes the hydroxylation of at least one proline residue in the triplets, with the exception of the first repeat (Table 1). Hydroxylation of both X and Y proline residues were identified, which is quite distinct from the hydroxylation pattern of (P-P-G)₅ by C-P4H (65). Using ¹⁴C-labeled (P-P-G)₅, previous studies have shown that C-P4H preferentially hydroxylates the Y proline in the fourth triplet of (P-P-G)₅ (65).

Docking studies with (P-P-G)₅ illustrates that the peptide is located over the active site with the Y proline residue of the second repeat (P₄) in proximity of the metal (Fig. 9). The C-4 carbon atom of this proline is positioned optimally to the metal center for C-H bond activation and subsequent hydroxylation, via a reactive Fe(IV)-oxo intermediate. Although modeling displayed P₄ in the proper orientation for hydroxylation, alternatives with prolines from other repeat units of the peptide in correct conformation for hydroxylation are feasible. The MS data revealed that proline residues of second, fourth, and fifth repeats of the (P-P-G)₅ peptide are indeed hydroxylated by BaP4H (Table 1).

Based on the crystal structures of BaP4H, the surface contains a groove across the active site comprising of aromatic and other conserved residues found in Src Homology 3 (SH3) domain proteins known to recognize PRMs (68). The SH3 domain-like proteins have been reported to occur in prokaryotes (23). Based on structural studies with many protein families, the mechanism of proline-rich peptide recognition is converging, which suggests that a groove composed of aromatic residues is responsible for proline recognition (68). However, the
enzymes have been solved. These are HIF
BaP4H with bound cofactors display structural
feature characteristic of Fe(II)/apo-BaP4H (36). The crystal structures of
(PPHD) (37), CrP4H (34,35), vCPH (30), and
PHD2 (70),
hand has an affinity for
residues, BaP4H can adopt alternate
conformations when bound to different ligands.
The residue that undergoes the largest
movement is Tyr 124. While Tyr 124 does not
directly interact with αKG, it serves as a lid to
keep the active site closed when αKG is present.
This tyrosine is conserved across P4Hs and
similar movement has been observed in both
vCPH and CrP4H (30,34,35). It was proposed
that this tyrosine could serve as a
“conformational switch” to modulate loop
movement and substrate binding or release,
which is further supported in the BaP4H
structures (30,34,35). Thus the Co-BaP4H-PPG
structure with αKG bound represents the “ready”
or pre-catalytic state of the enzyme.

In BaP4H αKG coordinates cobalt in the “in
line” binding mode observed in many Fe(II)/αKG-dependent dioxygenases (2), with
the C-1 carboxylate opposite of His 127 and
carbonyl oxygen opposite of Asp129, indicating
that substrate will be positioned near the open
site trans to His 193 for productive catalysis
(Figs. 6, 8 and 9). It is evident
that this tyrosine could serve as a
“conformational switch” to modulate loop
movement and substrate binding or release,
which is further supported in the BaP4H
structures (30,34,35). Thus the Co-BaP4H-PPG
structure with αKG bound represents the “ready”
or pre-catalytic state of the enzyme.

In BaP4H αKG coordinates cobalt in the “in
line” binding mode observed in many Fe(II)/αKG-dependent dioxygenases (2), with
the C-1 carboxylate opposite of His 127 and
carbonyl oxygen opposite of Asp129, indicating
that substrate will be positioned near the open
site trans to His 193 for productive catalysis
(Figs. 6, 8 and 9). It is evident
that this tyrosine could serve as a
“conformational switch” to modulate loop
movement and substrate binding or release,
which is further supported in the BaP4H
structures (30,34,35). Thus the Co-BaP4H-PPG
structure with αKG bound represents the “ready”
or pre-catalytic state of the enzyme.

In BaP4H αKG coordinates cobalt in the “in
line” binding mode observed in many Fe(II)/αKG-dependent dioxygenases (2), with
the C-1 carboxylate opposite of His 127 and
carbonyl oxygen opposite of Asp129, indicating
that substrate will be positioned near the open
site trans to His 193 for productive catalysis
(Figs. 6, 8 and 9). It is evident
that this tyrosine could serve as a
“conformational switch” to modulate loop
movement and substrate binding or release,
which is further supported in the BaP4H
structures (30,34,35). Thus the Co-BaP4H-PPG
structure with αKG bound represents the “ready”
or pre-catalytic state of the enzyme.

In BaP4H αKG coordinates cobalt in the “in
line” binding mode observed in many Fe(II)/αKG-dependent dioxygenases (2), with
the C-1 carboxylate opposite of His 127 and
carbonyl oxygen opposite of Asp129, indicating
that substrate will be positioned near the open
site trans to His 193 for productive catalysis
(Figs. 6, 8 and 9). It is evident
that this tyrosine could serve as a
“conformational switch” to modulate loop
movement and substrate binding or release,
which is further supported in the BaP4H
structures (30,34,35). Thus the Co-BaP4H-PPG
structure with αKG bound represents the “ready”
or pre-catalytic state of the enzyme.

In BaP4H αKG coordinates cobalt in the “in
line” binding mode observed in many Fe(II)/αKG-dependent dioxygenases (2), with
the C-1 carboxylate opposite of His 127 and
carbonyl oxygen opposite of Asp129, indicating
that substrate will be positioned near the open
site trans to His 193 for productive catalysis
(Figs. 6, 8 and 9). It is evident
that this tyrosine could serve as a
“conformational switch” to modulate loop
movement and substrate binding or release,
which is further supported in the BaP4H
structures (30,34,35). Thus the Co-BaP4H-PPG
structure with αKG bound represents the “ready”
or pre-catalytic state of the enzyme.

In BaP4H αKG coordinates cobalt in the “in
line” binding mode observed in many Fe(II)/αKG-dependent dioxygenases (2), with
the C-1 carboxylate opposite of His 127 and
carbonyl oxygen opposite of Asp129, indicating
that substrate will be positioned near the open
site trans to His 193 for productive catalysis
(Figs. 6, 8 and 9). It is evident
that this tyrosine could serve as a
“conformational switch” to modulate loop
movement and substrate binding or release,
which is further supported in the BaP4H
structures (30,34,35). Thus the Co-BaP4H-PPG
structure with αKG bound represents the “ready”
or pre-catalytic state of the enzyme.

In BaP4H αKG coordinates cobalt in the “in
line” binding mode observed in many Fe(II)/αKG-dependent dioxygenases (2), with
the C-1 carboxylate opposite of His 127 and
carbonyl oxygen opposite of Asp129, indicating
that substrate will be positioned near the open
site trans to His 193 for productive catalysis
(Figs. 6, 8 and 9). It is evident
that this tyrosine could serve as a
“conformational switch” to modulate loop
movement and substrate binding or release,
which is further supported in the BaP4H
structures (30,34,35). Thus the Co-BaP4H-PPG
structure with αKG bound represents the “ready”
or pre-catalytic state of the enzyme.

In BaP4H αKG coordinates cobalt in the “in
line” binding mode observed in many Fe(II)/αKG-dependent dioxygenases (2), with
the C-1 carboxylate opposite of His 127 and
carbonyl oxygen opposite of Asp129, indicating
that substrate will be positioned near the open
site trans to His 193 for productive catalysis
(Figs. 6, 8 and 9). It is evident
that this tyrosine could serve as a
“conformational switch” to modulate loop
movement and substrate binding or release,
which is further supported in the BaP4H
structures (30,34,35). Thus the Co-BaP4H-PPG
structure with αKG bound represents the “ready”
or pre-catalytic state of the enzyme.

In BaP4H αKG coordinates cobalt in the “in
line” binding mode observed in many Fe(II)/αKG-dependent dioxygenases (2), with
the C-1 carboxylate opposite of His 127 and
carbonyl oxygen opposite of Asp129, indicating
that substrate will be positioned near the open
site trans to His 193 for productive catalysis
(Figs. 6, 8 and 9). It is evident
that this tyrosine could serve as a
“conformational switch” to modulate loop
movement and substrate binding or release,
which is further supported in the BaP4H
structures (30,34,35). Thus the Co-BaP4H-PPG
structure with αKG bound represents the “ready”
or pre-catalytic state of the enzyme.

In BaP4H αKG coordinates cobalt in the “in
line” binding mode observed in many Fe(II)/αKG-dependent dioxygenases (2), with
the C-1 carboxylate opposite of His 127 and
carbonyl oxygen opposite of Asp129, indicating
that substrate will be positioned near the open
site trans to His 193 for productive catalysis
(Figs. 6, 8 and 9). It is evident
that this tyrosine could serve as a
“conformational switch” to modulate loop
movement and substrate binding or release,
αKG binding space in the active site and mimics binding of the αKG decarboxylation product succinate. The malonate bound form of BaP4H appears to be more open than the αKG-BaP4H-PPG bound form since it simulates a product state of the enzyme post-hydroxylation, where an open form is necessary for product release. The C-1 carboxylate of malonate is coordinated to the metal trans to Asp 129 (Figs. 6 and 8). The binding mode is similar for this family of enzymes in that the keto group of αKG always binds trans to the carboxylate residue in the catalytic triad (2). Even with two fewer residues responsible for malonate binding to BaP4H compared to αKG, malonate still binds BaP4H with high affinity (Kd ~4 μM) (Fig. 2E). The slightly lower binding affinity of malonate with BaP4H compared to αKG (Kd ~1 μM) is somewhat surprising. However, this could lend some evidence that similar metabolites, such as succinate, could provide a feedback type of inhibition to BaP4H. It has been suggested previously that concentrations of αKG could serve as an indicator of a cell’s metabolic status and that the Fe(II)/αKG dependent family of enzymes could possess a wider role in regulation of cellular processes (74,75). This leads to the conclusion that there does not appear to be an obvious trend in analyzing binding affinities of αKG for different enzymes. A more complex combination of local cellular environment and structural dynamics is most likely involved.

In conclusion, the results described here uncovered that BaP4H can target proline-rich human collagen-like peptides for site-specific hydroxylation. The BaP4H structure-function studies have broader implications in deciphering molecular mechanisms that regulate the PSB domains in P4Hs and PPII substrate binding enzymes, including substrate specificity and the overall reaction mechanisms. Further studies to determine the X-ray crystal structure of active Fe(II)-BaP4H with a peptide substrate bound are underway, in addition to site-directed mutagenesis to confirm the catalytic residues, which altogether would allow the identification of the molecular determinants of substrate binding and site-specific catalysis.

**Acknowledgements:** We thank the University of Iowa College of Liberal Arts and Sciences for funding this work. We are grateful to Dr. Ryan Kunz at the Thermo Fisher Scientific Center for Multiplexed Proteomics at Harvard Medical School for helping us with the mass spectrometry data collection and analysis. X-ray data were collected at the Advanced Light Source (ALS) and the Advanced Photon Source (APS). Support for the synchrotron sources is provided by the U.S. Department of Energy.

The coordinates and structure factors of apo-BaP4H, Co(II)-BaP4H-MLI, and Co(II)-BaP4H-PPG complex structures have been deposited in the RCSB Protein Data Bank with accession codes 5IAT, 5IAV, and 5IAX.

**Conflict of interest:** The authors declare that they have no conflicts of interest with the contents of this article.

**Author contributions:** N.J.S. and M.D. designed the experiments, performed the experiments and data analysis, and wrote the manuscript.

**REFERENCES**

1. Gorres, K. L., and Raines, R. T. (2010) Prolyl 4-hydroxylase. Crit. Rev. Biochem. Mol. Biol. 45, 106-124
2. Hausinger, R. P. (2004) FeII/α-ketoglutarate-dependent hydroxylases and related enzymes. Crit. Rev. Biochem. Mol. Biol. 39, 21-68
Collagen peptide hydroxylation by BaP4H

3. Koehntop, K. D., Emerson, J. P., and Que, L., Jr. (2005) The 2-His-1-carboxylate facial triad: a versatile platform for dioxygen activation by mononuclear non-heme iron(II) enzymes. *J. Biol. Inorg. Chem.* **10**, 87-93

4. Bollinger, J. M., Jr., and Krebs, C. (2006) Stalking intermediates in oxygen activation by iron enzymes: motivation and method. *J. Biol. Inorg. Chem.* **100**, 586-605

5. Bruijin, P. C., van Koten, G., and Klein Gebbink, R. J. (2008) Mononuclear non-heme iron enzymes with the 2-His-1-carboxylate facial triad: recent developments in enzymology and modeling studies. *Chem. Soc. Rev.* **37**, 2716-2744

6. Costas, M., Mehn, M. P., Jensen, M. P., and Que, L. (2004) Dioxygen Activation at Mononuclear Nonheme Iron Active Sites: Enzymes, Models, and Intermediates. *Chem. Rev.* **104**, 939-986

7. Lange, S. J., Que, L., Jr. (1998) Oxygen activating nonheme iron enzymes. *Curr. Opin. Chem. Biol.* **2**, 159-172

8. Bollinger, J. M., Jr., and Broderick, J. B. (2009) Frontiers in enzymatic C-H-bond activation. *Curr. Opin. Chem. Biol.* **13**, 51-57

9. Price, J. C., Barr, E. W., Tirupati, B., Bollinger, J. M., Jr., and Krebs, C. (2003) The first direct characterization of a high-valent iron intermediate in the reaction of an alpha-ketoglutarate-dependent dioxygenase: a high-spin FeIV complex in taurine/alpha-ketoglutarate dioxygenase (TauD) from Escherichia coli. *Biochemistry* **42**, 7497-7508

10. Hoffart, L. M., Barr, E. W., Guyer, R. B., Bollinger, J. M., Jr., and Krebs, C. (2006) Direct spectroscopic detection of a C-H-cleaving high-spin Fe(IV) complex in a prolyl-4-hydroxylase. *Proc. Natl. Acad. Sci. U S A* **103**, 14738-14743

11. Loenarz, C., and Schofield, C. J. (2008) Expanding chemical biology of 2-oxoglutarate oxygenases. *Nat. Chem. Biol.* **4**, 152-156

12. Loenarz, C., and Schofield, C. J. (2011) Physiological and biochemical aspects of hydroxylations and demethylations catalyzed by human 2-oxoglutarate oxygenases. *Trends Biochem. Sci.* **36**, 7-18

13. Myllyharju, J. (2008) Prolyl 4-hydroxylases, key enzymes in the synthesis of collagens and regulation of the response to hypoxia, and their roles as treatment targets. *Ann. Med.* **40**, 402-417

14. Clifton, I. J., McDonough, M. A., Ehrismann, D., Kershaw, N. J., Granatino, N., and Schofield, C. J. (2006) Structural studies on 2-oxoglutarate oxygenases and related double-stranded beta-helix fold proteins. *J. Inorg. Biochem.* **100**, 644-669

15. Aik, W., McDonough, M. A., Thalhammer, A., Chowdhury, R., and Schofield, C. J. (2012) Role of the jelly-roll fold in substrate binding by 2-oxoglutarate oxygenases. *Curr. Opin. Struct. Biol.* **22**, 691-700

16. Myllyharju, J. (2013) Prolyl 4-hydroxylases, master regulators of the hypoxia response. *Acta Physiol.* **208**, 148-165

17. Myllyharju, J., and Kivirikko, K. I. (2004) Collagens, modifying enzymes and their mutations in humans, flies and worms. *Trends Genet.* **20**, 33-43

18. Schofield, C. J., and Ratcliffe, P. J. (2004) Oxygen sensing by HIF hydroxylases. *Nat. Rev. Mol. Cell. Biol.* **5**, 343-354

19. Aprelikova, O., Chandramouli, G. V., Wood, M., Vasselli, J. R., Riss, J., Maranchie, J. K., Linehan, W. M., and Barrett, J. C. (2004) Regulation of HIF prolyl hydroxylases by hypoxia-inducible factors. *J. Cell. Biochem.* **92**, 491-501

20. Kaelin, W. G., Jr., and Ratcliffe, P. J. (2008) Oxygen sensing by metazoans: the central role of the HIF hydroxylase pathway. *Mol. Cell.* **30**, 393-402

21. Pappalardi, M. B., McNulty, D. E., Martin, J. D., Fisher, K. E., Jiang, Y., Burns, M. C., Zhao, H., Ho, T., Sweitzer, S., Schwartz, B., Annan, R. S., Copeland, R. A., Tummino, P. J., and Luo, L. (2011) Biochemical characterization of human HIF hydroxylases using HIF protein substrates that contain all three hydroxylation sites. *Biochem. J.* **436**, 363-369

22. Hieta, R., Kukkola, L., Permi, P., Pirilä, P., Kivirikko, K. I., Kilpeläinen, I., and Myllyharju, J. (2003) The peptide-substrate-binding domain of human collagen prolyl 4-hydroxylases.
Backbone assignments, secondary structure, and binding of proline-rich peptides. J. Biol. Chem. 278, 34966-34974

23. Adzhubei, A. A., Sternberg, M. J., and Makarov, A. A. (2013) Polyproline-II helix in proteins: structure and function. J. Mol. Biol. 425, 2100-2132

24. Berisio, R., and Vitagliano, L. (2012) Polyproline and triple helix motifs in host-pathogen recognition. Curr. Protein Pept. Sci. 13, 855-865

25. Myllyharju, J. K., K. I. (1999) Identification of a novel proline-rich peptide-binding domain in poly 4-hydroxylase. EMBO J. 18, 306-312

26. Pekkala, M., Hieta, R., Bergmann, U., Kivirikko, K. I., Wierenga, R. K., and Myllyharju, J. (2004) The peptide-substrate-binding domain of collagen prolyl 4-hydroxylases is a tetrapeptide repeat domain with functional aromatic residues. J. Biol. Chem. 279, 52255-52261

27. Anantharajan, J., Koski, M. K., Kursula, P., Hieta, R., Bergmann, U., Myllyharju, J., and Wierenga, R. K. (2013) The structural motifs for substrate binding and dimerization of the alpha subunit of collagen prolyl 4-hydroxylase. Structure 21, 2107-2118

28. Cerveny, L., Straskova, A., Dankova, V., Hartlova, A., Ceckova, M., Staud, F., and Stulik, J. (2013) Tetrapeptide repeat motifs in the world of bacterial pathogens: role in virulence mechanisms. Infect. Immun. 81, 629-635

29. Eriksson, M., Myllyharju, J., Tu, H., Hellman, M., and Kivirikko, K. I. (1999) Evidence for 4-hydroxyproline in viral proteins. Characterization of a viral poly 4-hydroxylase and its peptide substrates. J. Biol. Chem. 274, 22131-22134

30. Longbotham, J. E., Levy, C., Johannissen, L. O., Tarhonskaya, H., Jiang, S., Loenarz, C., Flashman, E., Hay, S., Schofield, C. J., and Scrutton, N. S. (2015) Structure and Mechanism of a Viral Collagen Polyly Hydroxylase. Biochemistry 54, 6093-6105

31. Hietanen, R., Myllyharju, J. (2002) Cloning and Characterization of a Low-Molecular Weight Prolyl 4-Hydroxylase from Arabidopsis thaliana. Effective hydroxylation of proline-rich, collagen-like, and hypoxia-inducible transcription factor alpha-like peptides. J. Biol. Chem. 277, 23965-23971

32. Keskiaho, K., Hieta, R., Sormunen, R., and Myllyharju, J. (2007) Chlamydomonas reinhardtii has multiple prolyl 4-hydroxylases, one of which is essential for proper cell wall assembly. The Plant cell 19, 256-269

33. Horita, S., Scotti, J. S., Thinnes, C., Mottaghi-Taromsari, Y. S., Thalhammer, A., Ge, W., Aik, W., Loenarz, C., Schofield, C. J., and McDonough, M. A. (2015) Structure of the ribosomal oxygenase OGFOD1 provides insights into the regio- and stereoselectivity of prolyl hydroxylases. Structure 24, 639-652

34. Koski, M. K., Hieta, R., Bollner, C., Kivirikko, K. I., Myllyharju, J., and Wierenga, R. K. (2007) The active site of an algal prolyl 4-hydroxylase has a large structural plasticity. J. Biol. Chem. 282, 37112-37123

35. Koski, M. K., Hieta, R., Hirsila, M., Ronka, A., Myllyharju, J., and Wierenga, R. K. (2009) The crystal structure of an algal prolyl 4-hydroxylase complexed with a proline-rich peptide reveals a novel buried tripeptide binding motif. J. Biol. Chem. 284, 25290-25301

36. Culpepper, M. A., Scott, E. E., and Limburg, J. (2010) Crystal structure of prolyl 4-hydroxylase from Bacillus anthracis. Biochemistry 49, 124-133

37. Scotti, J. S., Leung, I. K., Ge, W., Bentley, M. A., Paps, J., Kramer, H. B., Lee, J., Aik, W., Choi, H., Paulsen, S. M., Bowman, L. A., Loik, N. D., Horita, S., Ho, C. H., Kershaw, N. J., Tang, C. M., Claridge, T. D., Preston, G. M., McDonough, M. A., and Schofield, C. J. (2014) Human oxygen sensing may have origins in prokaryotic elongation factor Tu prolyl-hydroxylase. Proc. Natl. Acad. Sci. USA 111, 13331-13336

38. Adefarati, A. A., Hensens, O. D., Jones, E. T., and Tkacz, J. S. (1992) Pneumocandins from Zalerion arboricola. V. Glutamic acid- and leucine-derived amino acids in pneumocandin A0 (L-671,329) and distinct origins of the substituted proline residues in pneumocandins A0 and B0. J. Antibiot. 45, 1953-1957
Collagen peptide hydroxylation by BaP4H

39. Kukkola, L., Koivunen, P., Pakkanen, O., Page, A. P., and Myllyharju, J. (2004) Collagen prolyl 4-hydroxylase tetramers and dimers show identical decreases in Km values for peptide substrates with increasing chain length: mutation of one of the two catalytic sites in the tetramer inactivates the enzyme by more than half. *J. Biol. Chem.* **279**, 18656-18661

40. Rasmussen, M., Jacobsson, M., and Bjorek, L. (2003) Genome-based identification and analysis of collagen-related structural motifs in bacterial and viral proteins. *J. Biol. Chem.* **278**, 32313-32316

41. Miller, M. A., Scott, E. E., and Limburg, J. (2008) Expression, purification, crystallization and preliminary X-ray studies of a prolyl-4-hydroxylase protein from Bacillus anthracis. *Acta Crystallogr. F Struct. Biol. Cryst. Commun.* **64**, 788-791

42. Robert, X., and Gouet, P. (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* **42**, W320-324

43. Ryle, M. J., Liu, A., Muthukumaran, R. B., Ho, R. Y., Koehntop, K. D., McCracken, J., Que, L., Jr., and Hausinger, R. P. (2003) O2- and a-Ketoglutarate-Dependent Tyrosyl Radical Formation in TauD, an R-Keto Acid-Dependent Non-Heme Iron Dioxygenase. *Biochemistry* **42**, 1854-1862

44. Ryle, M. J., Koehntop, K. D., Liu, A., Que, L., Jr., and Hausinger, R. P. (2003) Interconversion of two oxidized forms of taurine/alpha-ketoglutarate dioxygenase, a non-heme iron hydroxylase: evidence for bicarbonate binding. *Proc. Natl. Acad. Sci. U S A* **100**, 3790-3795

45. Huttlin, E. L., Jedrychowski, M. P., Elias, J. E., Goswami, T., Rad, R., Beausoleil, S. A., Villen, J., Haas, W., Sowa, M. E., and Gygi, S. P. (2010) A tissue-specific atlas of mouse protein phosphorylation and expression. *Cell* **143**, 1174-1189

46. Brummett, A. E., Schnicker, N. J., Crider, A., Todd, J. D., and Dey, M. (2015) Biochemical, Kinetic, and Spectroscopic Characterization of Ruegeria pomeroyi DddW-A Mononuclear Iron-Dependent DMSP Lyase. *PloS one* **10**, e0127288

47. Knauer, S. H., Hartl-Spiegelhauer, O., Schwarzingerr, S., Hanzelmann, P., and Dobbeck, H. (2012) The Fe(II)/alpha-ketoglutarate-dependent taurine dioxygenases from Pseudomonas putida and Escherichia coli are tetramers. *FEBS J.* **279**, 816-831

48. Kabsch, W. (2010) Xds. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125-132

49. Evans, P. R. (2011) An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 282-292

50. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments. *Acta Cryst. D* **67**, 235-242

51. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658-674

52. Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H., and Adams, P. D. (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 352-367

53. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126-2132

54. Laskowski, R. A., Macarthur, M. W., Moss, D. S., and Thornton, J. M. (1993) Procheck - a Program to Check the Stereochemical Quality of Protein Structures. *J Appl Crystallogr* **26**, 283-291

55. Ramachandran, G. N., Ramakrishnan, C., and Sasisekharan, V. (1963) Stereochemistry of Polypeptide Chain Configurations. *J. Mol. Biol.* **7**, 95-99

56. Pierce, B. G., Wiehe, K., Hwang, H., Kim, B. H., Vreven, T., and Weng, Z. (2014) ZDOCK server: interactive docking prediction of protein-protein complexes and symmetric multimers. *Bioinformatics* **30**, 1771-1773
Collagen peptide hydroxylation by BaP4H

57. London, N., Raveh, B., Cohen, E., Fathi, G., and Schueler-Furman, O. (2011) Rosetta FlexPepDock web server-high resolution modeling of peptide-protein interactions. *Nucleic Acids Res.* 39, W249-253

58. Raveh, B., London, N., and Schueler-Furman, O. (2010) Sub-angstrom modeling of complexes between flexible peptides and globular proteins. *Proteins* 78, 2029-2040

59. Grzyska, P. K., Ryle, M. J., Monterosso, G. R., Liu, J., Ballou, D. P., and Hausinger, R. P. (2005) Steady-State and Transient Kinetic Analyses of Taurine/α-Ketoglutarate Dioxygenase: Effects of Oxygen Concentration, Alternative Sulfonates, and Active-Site Variants on the FeIV-oxo Intermediate. *Biochemistry* 44, 3845-3855

60. Coufal, D. E., Tavares, P., Pereira, A. S., Hyunh, B. H., Lippard, S. J. (1999) Reactions of Nitric Oxide with the Reduced Non-Heme Diiron Center of the Soluble Methane Monoxygenase Hydroxylase. *Biochemistry* 38, 4504-4513

61. Gaffney, B. J. (2009) EPR of Mononuclear Non-Heme Iron Proteins. *Biol. Magn. Reson.* 28, 233-268

62. Lancaster, J. R., Jr., Langrehr, J. M., Bergonia, H. A., Murase, N., Simmons, R. L., Hoffman, R. A. (1992) EPR Detection of Heme and Nonheme Iron-containing Protein Nitrosylation by Nitric Oxide during Rejection of Rat Heart Allograft. *J. Biol. Chem.* 267

63. Orville AM, L. J. (1993) Simultaneous Binding of Nitric Oxide and Isotopically Labeled Substrates or Inhibitors 3,4-Dioxygenase by Reduced Protocatechuate. *J. Biol. Chem.* 268, 8596-8607

64. Hutton, J. J., Jr., Kaplan, A., and Udenfriend, S. (1967) Conversion of the amino acid sequence gly-pro-pro in protein to gly-pro-hyp by collagen proline hydroxylase. *Arch. Biochem. Biophys.* 121, 384-391

65. Kivirikko, K. I., Suga, K., Kishida, Y., Sakakibara, S., and Prockop, D. J. (1971) Asymmetry in the hydroxylation of (Pro-Pro-Gly) 5 by protocollagen proline hydroxylase. *Biochem. Biophys. Res. Commun.* 45, 1591-1596

66. Berg, R. A., and Prockop, D. J. (1973) Purification of (14C) protocollagen and its hydroxylation by prolyl-hydroxylase. *Biochemistry* 12, 3395-3401

67. Berg, R. A., Kishida, Y., Sakakibara, S., and Prockop, D. J. (1977) Hydroxylation of (Pro-Pro-Gly)5 and (Pro-Pro-Gly)10 by prolyl hydroxylase. Evidence for an asymmetric active site in the enzyme. *Biochemistry* 16, 1615-1621

68. Zarrinpar, A., and Lim, W. A. (2000) Converging on proline: the mechanism of WW domain peptide recognition. *Nat. Struct. Biol.* 7, 611-613

69. Adefarati, A. A., Giacobbe, R., Hensens, O. D., Tkacz, J. S. (1991) Biosynthesis of L-671,329, an Echinocandin-Type Antibiotic Produced by Zalerion arboricola: Origins of Some of the Unusual Amino Acids and the Dimethylmyristic Acid Side Chain. *J. Am. Chem. Soc.* 113, 3542-3545

70. McDonough, M. A., Li, V., Flashman, E., Chowdhury, R., Mohr, C., Lienard, B. M., Zondlo, J., Oldham, N. J., Clifton, I. J., Lewis, J., McNeill, L. A., Kurzeja, R. J., Hewitson, K. S., Yang, E., Jordan, S., Syed, R. S., and Schofield, C. J. (2006) Cellular oxygen sensing: Crystal structure of hypoxia-inducible factor prolyl hydroxylase (PHD2). *Proc. Natl. Acad. Sci. U S A* 103, 9814-9819

71. Schnicker, N. J., and Dey, M. (2016) Structural analysis of cofactor binding for a prolyl 4-hydroxylase from the pathogenic bacterium *Bacillus anthracis*. *Acta Cryst.* D72, doi:10.1107/S2059798316004198

72. McNeill, L. A., Flashman, E., Buck, M. R., Hewitson, K. S., Clifton, I. J., Jeschke, G., Claridge, T. D., Ehrismann, D., Oldham, N. J., and Schofield, C. J. (2005) Hypoxia-inducible factor prolyl hydroxylase 2 has a high affinity for ferrous iron and 2-oxoglutarate. *Mol. Biosyst.* 1, 321-324

73. Majamaa, K., Hanauiske-Abel, H. M., Gunzler, V., Kivirikko, K. I. (1984) The 2-oxoglutarate binding site of prolyl4-hydroxylase Identification of distinct subsites and evidence for 2-
oxoglutarate decarboxylation in a ligand reaction at the enzyme-bound ferrous ion. Eur. J. Biochem. 138, 239

74. Yuan, H. X., Xiong, Y., and Guan, K. L. (2013) Nutrient sensing, metabolism, and cell growth control. Molecular cell 49, 379-387

75. Ploumakis, A., and Coleman, M. L. (2015) OH, the Places You'll Go! Hydroxylation, Gene Expression, and Cancer. Molecular cell 58, 729-741

FOOTNOTES
This work was supported by the University of Iowa College of Liberal Arts and Sciences.

The abbreviations used are: Hyp, hydroxyproline; αKG, alpha-ketoglutarate; MLI, malonate; PPII, poly-L-proline type II; DSBH, double-stranded beta-helix; PRM, proline rich motif; PRD, proline recognition domain; PSB, peptide-substrate binding; SBL, substrate binding loop; TPR, tetratricopeptide repeat; PCR, polymerase chain reaction; EPR, electron paramagnetic resonance; LC-MS/MS, liquid chromatography mass spectrometry
FIGURE LEGENDS

FIGURE 1. Sequence alignment of BaP4H with vCPH, human C-P4H, and CrP4H and secondary structures defined above the sequence. Alpha helices are indicated by helical lines with α labels, beta strands are indicated by arrows with β labels. Active site residues are highlighted in magenta. The missing loop in BaP4H comprises of residues 70-83. The 3_10 helices have not been labeled. The sequences were aligned by their secondary structures using ESPript 3.

FIGURE 2. Spectral properties and substrate/cofactor binding affinities (K_d) of Fe(II)-BaP4H. (A) UV-visible spectra of the reaction of 776 µM apo-BaP4H in the presence of Fe(II) and αKG. Apo-BaP4H (dotted line); Fe(II)-BaP4H (dashed line); Fe(II)-BaP4H-αKG ternary complex (solid line). (B) EPR spectra of BaP4H in the presence of NO: (top) 100 µM apo-BaP4H with equimolar concentration of Fe(NH4)2(SO4)2; (bottom) NO-Fe(II)-BaP4H with 100 µM αKG. The intrinsic tryptophan fluorescence was monitored as corresponding peptides were titrated into an anaerobic solution containing 1 µM apo-BaP4H, 3 µM Fe(NH4)2(SO4)2, and 100 µM αKG. The binding affinities for the (P-P-G)_10 and (P-P-G)_5 peptides were determined to be 0.74 ± 0.11 µM (circles and solid line) and 0.3 ± 0.02 µM (triangles and dashed line), respectively. (D) The K_d for L-proline was determined to be 7.9 ± 2.5 µM and was titrated into an anaerobic solution containing 1 µM BaP4H, 3 µM Fe(NH4)2(SO4)2, and 100 µM αKG. (E) The K_d (4.2 ± 0.5 µM) for malonate was obtained by titrating it into an anaerobic solution containing 1 µM BaP4H and 3 µM Fe(NH4)2(SO4)2. (F) The K_d for Co(II) was determined to be 0.81 ± 0.09 µM. CoCl2 was titrated into an aerobic solution containing 1 µM BaP4H.

FIGURE 3. Hydroxylated peptide product identification by LC-MS. (A) XIC Peak profiles of [(P-P-G)_5 +OH]^2+ (green) and (P-P-G)_5 (black) as separated on a C18 column. (B) XIC peak profiles for [(P-P-G)_10 +OH]^3+ (green) and (P-P-G)_10 (black) separated on a C18 column. (C) MS analysis of (P-P-G)_5 before and after hydroxylation. The top panel shows [(P-P-G)_5 +OH]^2+, the isotopic mass envelope starts at m/z 645.820. The bottom panel shows the substrate peptide [(P-P-G)_5], with an isotopic mass envelope starting at m/z 637.8345. Hydroxylation of (P-P-G)_5 results in a m/z change of 7.9945 for the double charged, this is equivalent to a 15.989 Da increase to the peptide mass, which corresponds to incorporation of a single oxygen atom. (D) MS analysis of (P-P-G)_10 before and after hydroxylation. The top panel shows [(P-P-G)_10 +OH]^3+, the isotopic mass envelope starts at m/z 849.4371. The bottom panel shows the substrate peptide [(P-P-G)_10], with an isotopic mass envelope starting at m/z 844.1041. Hydroxylation of (P-P-G)_10 results in a m/z change of 5.3323 for the triply charged ion, this is equivalent to a 15.989 Da increase to the peptide mass, which corresponds to incorporation of a single oxygen atom.

FIGURE 4. MS2 fragment ion pattern of hydroxylated (P-P-G)_5 and (P-P-G)_10 peptides. Assignment of b (blue) and y (red) fragment ion series in the MS2 scans from [(P-P-G)_5 +OH]^2+ (A) and [(P-P-G)_10 +OH]^3+ (B) precursor ions. Using a variation of the Ascore algorithm, the hydroxylation was localized with >99% confidence to the Y position in the 4th repeat for (P-P-G)_5 and to the X position in the 5th repeat for (P-P-G)_10. The precursor peptides were selected from the single-turnover experiments.
**FIGURE 5.** Structure of the Co(II)-BaP4H-PPG monomer (green) showing the C-terminus of a symmetry related molecule (cyan) interacting with the active site of chain A (green). The β-strands of the DSBH fold (light blue) are labeled βI-βVIII. Other β strands (light orange) are labeled β0, β1, β2, β6, β7, and β8. Missing residues in the substrate binding loop region are indicated by a dashed line. The metal binding residues, His 127, Asp 129, and His 193, αKG (yellow), and the fused P-P-G peptide are shown as sticks (oxygen, red; nitrogen, blue). Cobalt is shown as a magenta sphere.

**FIGURE 6.** Active site of BaP4H with bound cofactors. (A) Bidentate binding mode for αKG in Co(II)-BaP4H-PPG structure. (B) Active site residues of Co(II)-BaP4H-PPG involved in hydrogen bonding interactions between protein, αKG, and water are indicated by black dashes with corresponding distances. The distance between Tyr 124 and W1 is 2.8 Å and was excluded from the picture for clarity. (C) The PPG peptide (green) at the C-terminus interacts with the active site of a symmetry related molecule (cyan). The Y proline (P2) is oriented with the C-4 carbon 5.2 Å away from the cobalt metal center. The 2nd and 3rd repeats of (PPG)3 were disordered. (D) Monodetate binding mode for malonate in Co(II)-BaP4H-MLI structure. (E) Active site residues of Co(II)-BaP4H-MLI showing hydrogen bonding interactions between protein, malonate, and water with labeled distances. Co-substrate and residues are shown as sticks: carbon (protein backbone color), oxygen (red), nitrogen (blue), water molecules as spheres (cyan), cobalt as a sphere (magenta). The 2F_o-F_c composite omit maps (blue mesh) contoured at 1.0 σ are shown for protein residues in (A), (C), and (D). The F_o-F_c omit maps (also shown as blue mesh) for αKG and MLI are contoured at 3.0 σ in (A) and (D).

**FIGURE 7.** (A) Active site of apo-BaP4H structure with no cofactors bound. (B) Hydrogen bonding interactions of active site residues of apo-BaP4H between protein and water are indicated by black dashes with corresponding distances. The metal binding residues (His 127, Asp 129, and His 193) and residues involved in hydrogen bonds are shown as sticks: carbon (protein backbone color), oxygen (red), nitrogen (blue). The 2F_o-F_c composite omit maps (blue mesh) are contoured at 1.0 σ around the active site residues.

**FIGURE 8.** BaP4H structures highlighting conformational changes involving residues Tyr 118, Tyr 124, and Phe 160 upon metal and αKG or malonate binding. Active site comparison between the Co-BaP4H-PPG “closed” active site (green), Co-BaP4H-MLI “partly open” active site (pink), and apo-BaP4H “fully open” active site (light blue). Cobalt is shown as a magenta sphere.

**FIGURE 9.** Model of (P-P-G)5 peptide (cyan) into the αKG bound structure of Co-BaP4H-PPG (green). (A) Surface representation of BaP4H monomer showing the peptide binding groove residues colored in grey and the overall position of (P-P-G)5. The peptide is shown as cyan sticks with oxygen (red), nitrogen (blue). The active site region is colored orange. The residues representing the ends of the missing loop are colored violet. Cobalt and αKG are shown as a magenta sphere and yellow sticks respectively. (B) Location of the (P-P-G)5 peptide with potential interacting BaP4H residues (green sticks). The C-4 position of the Y proline in the second repeat of (P-P-G)5 is located 4.6 Å away from the metal center.
TABLES

Table 1. Amino acid sequence of (P-P-G)\textsubscript{5} and (P-P-G)\textsubscript{10} with the unique hydroxylation sites shown. Xcorr values were calculated by the SEQUEST algorithm. The hydroxylation sites and localization scores were determined by a modified version of the Ascore algorithm. The table shows only peptides that have one or more hydroxylation site(s) localized to >99\% confidence. The first three hydroxylated peptides for (P-P-G)\textsubscript{5} were observed under both steady state and single-turn over conditions and the last two peptides were observed only in the single turnover condition. All peptides listed for (P-P-G)\textsubscript{10} were discovered under single turnover conditions. RT, retention time.

| Peptide | Xcorr | Sequence | Observed m/z | Precursor intensity | RT (min) | Site 1 Position | Localization score | Site 2 Position | Localization score |
|---------|-------|----------|--------------|--------------------|----------|-----------------|--------------------|-----------------|--------------------|
| (P-P-G)\textsubscript{5} | 4.3 Nter-PPGPPGPPGPPGPPGPPG-Cter | 645.82804 | 5.99E+05 | 11.82 | 11 | 40 | - | - |
| | 4.8 Nter-PPGPPGPPGPPGPPGPPG-Cter | 645.83138 | 2.64E+06 | 12.09 | 10 | 40 | - | - |
| | 4.2 Nter-PPGPPGPPGPPGPPGPPG-Cter | 645.83033 | 1.17E+06 | 12.39 | 6 | 40 | - | - |
| | 4.2 Nter-PPGPPGPPGPPGPPGPPG-Cter | 645.82847 | 7.49E+05 | 12.31 | 5 | 32.3 | - | - |
| | 1.8 Nter-PPGPPGPPGPPGPPGPPG-Cter | 653.82512 | 3.07E+05 | 12.98 | 13 | 25.7 | 14 | 25.7 |
| (P-P-G)\textsubscript{10} | 7.4 Nter-PPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG-Cter | 849.43506 | 7.78E+05 | 15.72 | 19 | 36.9 | - | - |
| | 7.1 Nter-PPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG-Cter | 849.4339 | 1.08E+05 | 15.62 | 26 | 32.3 | - | - |
| | 7.8 Nter-PPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG-Cter | 849.43425 | 3.83E+05 | 15.8 | 13 | 20.8 | - | - |
| | 5.6 Nter-PPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG-Cter | 854.79881 | 9.60E+04 | 16.05 | 19 | 19.6 | 20 | 19.6 |
Table 2. Data Collection and Refinement Statistics.

|                     | Apo-BaP4H | Co-BaP4H-MLI | Co-BaP4H-PPG/αKG |
|---------------------|-----------|--------------|------------------|
| **Data Collection** |           |--------------|------------------|
| Space Group         | P 2₁,2₁,2₁| P 1 2₁ 1     | P 2₁,2₁,2₁       |
| Unit Cell Dimensions|           |              |                  |
| a b c (Å)           | 51.6 80.1 103.8 | 42.8 41.5 111 | 50.8 75.6 105.8  |
| α β γ (deg)         | 90 90 90 | 90 97 90    | 90 90 90        |
| Beam line           | ALS 4.2.2| APS 19-ID-D  | ALS 4.2.2       |
| Wavelength (Å)      | 1.00      | 0.979       | 1.00             |
| Resolution Range (Å) | 33.3-1.67 | 55.2-1.70  | 61.5-2.10       |
|                      | (1.76-1.67)| (1.79-1.70)| (2.21-2.10)     |
| Observations a       | 347255 (32419) | 155721 (23012) | 111271 (15669) |
| Unique Reflections a | 50418 (6990) | 42773 (6257) | 24447 (3527)    |
| Redundancy a         | 6.9 (4.6) | 3.6 (3.7)   | 4.6 (4.4)       |
| Completeness (%) a   | 99.5 (96.5) | 99.6 (100)  | 99.8 (100)      |
| I/σ (I) a            | 27.1 (3.2) | 25.5 (6.6)  | 13.1 (3.0)      |
| R_merge (%) a,b      | 4.2 (46.0) | 2.6 (16.3)  | 6.8 (49.6)      |
| R_meas (%) a,b,c     | 5.0 (65.6) | 3.7 (22.2)  | 8.8 (60.7)      |
| CC1/2                | 1.000 (0.661) | 0.999 (0.967) | 0.998 (0.911)  |
| **Refinement**       |           |              |                  |
| R_work R_free (%)    | 17.5 (21.1) | 21.2 (22.9) | 19.7 (23.4)     |
| Resolution Range (Å) | 33.3-1.67 | 29.7-1.70  | 61.5-2.10       |
| No. of Reflections   | 50232     | 42742       | 24291           |
| Number of Atoms:     |           |              |                  |
| Protein              | 3165      | 3004        | 3063            |
| Co                   | -         | 2           | 2               |
| αKG                  | -         | -           | 20              |
| PPG                  | -         | -           | 36              |
| Malonate             | -         | 14          | -               |
| Water                | 396       | 254         | 154             |
| Average B-Factor (Å²) |           |              |                  |
| Protein              | 16.8      | 20.0        | 29.0            |
| Co                   | -         | 25.4        | 19.4            |
| αKG                  | -         | -           | 24.0            |
| PPG                  | -         | -           | 46.2            |
| Malonate             | -         | 26.3        | -               |
| Water                | 33.3      | 33.0        | 34.8            |
| RMSD Bond Lengths (Å)| 0.010     | 0.020       | 0.011           |
| RMSD Bond Angles (deg)| 1.2     | 1.1         | 1.1             |
| Ramachandran favored (%) | 97.5 | 97.7        | 97.6            |
| Ramachandran allowed (%) | 2.5  | 2.3         | 2.4             |
| Clash score          | 4.5       | 3.4         | 3.2             |
Values in parentheses are the highest-resolution shell. $^bR_{merge} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$, where $I_i(hkl)$ is the $i^{th}$ measured diffraction intensity and $\langle I(hkl) \rangle$ is the mean intensity for the miller index (hkl).

$^cR_{meas} = \frac{\sum_{h,k,l} \left( \frac{n(hkl)}{n_hk_l} - 1 \right)^{1/2} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{h,k,l} \sum_i I_i(hkl)}$
Collagen peptide hydroxylation by BaP4H

Fig. 3
Collagen peptide hydroxylation by BaP4H

Fig. 4

A

Relative Abundance

b_1 b_2 b_3 b_4 b_5 b_6 b_7 b_8 b_9 b_10 b_11 b_12

m/z

200 400 600 800 1000

B

Relative Abundance

y_1 y_2 y_3 y_4 y_5 y_6 y_7 y_8 y_9 y_10 y_11 y_12

m/z

200 400 600 800 1000 1200 1400
Fig. 5
Collagen peptide hydroxylation by BaP4H

Fig. 6
Fig. 7
Fig. 8
Fig. 9
Bacillus anthracis Prolyl 4-Hydroxylase Modifies Collagen-like Substrates in Asymmetric Patterns
Nicholas J. Schnicker and Mishtu Dey

J. Biol. Chem. published online April 21, 2016

Access the most updated version of this article at doi: 10.1074/jbc.M116.725432

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC’s e-mail alerts