Introduction

Most polypeptides are built from 20 amino-acid building blocks. Even greater molecular diversity is attainable by chemical modification of these building blocks after their condensation (Walsh et al., 2005; Walsh, 2006). Some modifications are permanent, whereas others are reversible. A protein can be modified by cleavage of the polypeptide chain, or by appending molecules or functional groups. The appendages range in size from an entire protein, carbohydrate, or lipid, as in ubiquitination, glycosylation, and farnesylation, to but a few atoms, as in phosphorylation, sulfation, acetylation, methylation, and carboxylation.

Here, we focus on prolyl 4-hydroxylase (P4H), which catalyzes the single most prevalent posttranslational modification in humans – the formation of (2S,4R)-4-hydroxyproline (Hyp) (Figure 1A). Hyp was discovered in gelatin hydrolysates by Emil Fischer (Fischer, 1902; Vickery and Schmidt, 1931). The abundance of Hyp among the residues in animal proteins is ∼4%, a value calculated from the abundance of collagen amongst animal proteins (1/3) and that of Hyp within collagen (∼38% × 1/3) (Ramshaw et al., 1998). Thus, Hyp is more abundant in animals than seven “common” amino-acid residues: Cys, Gln, His, Met, Phe, Trp, and Tyr (McCaldon and Argos, 1988).

The hydroxylation of proline residues is also amongst the most subtle of posttranslational modifications, adding merely 16 atomic mass units to a protein. That small perturbation, along with the instability of radioactive isotopes of oxygen (e.g. t½ = 122 s for 15O), has made the detection of Hyp problematic in intact proteins. Only recently has high-resolution mass spectrometry revealed

Prolyl 4-hydroxylase

Kelly L. Gorres1, and Ronald T. Raines1,2

1Department of Biochemistry, University of Wisconsin–Madison, Madison, WI, USA, and 2Department of Chemistry, University of Wisconsin–Madison, Madison, WI, USA

Abstract

Posttranslational modifications can cause profound changes in protein function. Typically, these modifications are reversible, and thus provide a biochemical on-off switch. In contrast, proline residues are the substrates for an irreversible reaction that is the most common posttranslational modification in humans. This reaction, which is catalyzed by prolyl 4-hydroxylase (P4H), yields (2S,4R)-4-hydroxyproline (Hyp). The protein substrates for P4Hs are diverse. Likewise, the biological consequences of prolyl hydroxylation vary widely, and include altering protein conformation and protein–protein interactions, and enabling further modification. The best known role for Hyp is in stabilizing the collagen triple helix. Hyp is also found in proteins with collagen-like domains, as well as elastin, conotoxins, and argonaute 2. A prolyl hydroxylase domain protein acts on the hypoxia inducible factor α, which plays a key role in sensing molecular oxygen, and could act on inhibitory κB kinase and RNA polymerase II. P4Hs are not unique to animals, being found in plants and microbes as well. Here, we review the enzymic catalysts of prolyl hydroxylation, along with the chemical and biochemical consequences of this subtle but abundant posttranslational modification.

Keywords: Non-heme iron dioxygenase; proline; hydroxyproline; posttranslational modification; collagen

Abbreviations: ER, endoplasmic reticulum; Flp, (2S,4R)-4-fluoroproline; flp, (2S,4S)-4-fluoroproline; HIF, hypoxia inducible factor; Hyp or O, (2S,4R)-4-hydroxyproline; hyp, (2S,4S)-4-hydroxyproline; PDB, protein data bank; P4H, prolyl 4-hydroxylase; PDI, protein disulfide isomerase; PHD, prolyl 4-hydroxylase domain protein; Pro or P, (2S)-proline.
its widespread occurrence. Ensuing analyses are being facilitated by new expression systems (Kerstean et al., 2004; Neubauer et al., 2005) and activity assays (Gorres and Raines, 2009) for P4H.

Herein, we review the biological chemistry of prolyl 4-hydroxylases. We emphasize the similarities and differences among these fascinating enzymes in the context of their varied substrates. We note that prolyl 4-hydroxylases are generating much interest as drug targets, a topic that was reviewed recently elsewhere (Myllyharju, 2008; Fraisl et al., 2009).

**Chemical consequences of prolyl hydroxylation**

Oxygen is a highly electronegative element (Pauling, 1939). For that reason alone, hydroxylation alters fundamental properties of proline. The simplest manifestation of this electronegativity is the through-bond inductive effect that lowers the nitrogen pKᵦ value in the free amino acid from 10.8 in ProOH to 9.68 in HypOH (Figure 1B) (Eberhardt et al., 1996). This inductive effect also diminishes amidic resonance within a prolyl peptide bond, making the prolyl nitrogen more pyramidal (Panasik et al., 1994) and increasing the rate of cis–trans prolyl bond isomerization (Eberhardt et al., 1996).

Installing an electronegative substituent at the 4R position of proline affects the pucker of its pyrrolidine ring (Figure 1B). This consequence arises from the gauche effect, which refers to the tendency of a molecule to adopt the conformation that has the maximum number of adjacent polar bonds (here, Cᵦ–Oᵦ and Cᵦᵦ–N) with a gauche (that is, ±60°) dihedral angle (Eberhardt et al., 1996). The gauche effect endows Hyp with a strong preference for the Cᵦexo conformation, whereas Pro has a slight preference for the Cᵦendo pucker (Bretscher et al., 2001; DeRider et al., 2002). The gauche effect is manifested in Hyp despite its orienting the hydroxyl group in the more constrained pseudo-axial position. The collagen triple helix is stabilized by installing a proline derivative that favors the Cᵦendo pucker in the first (Xaa) position or one that favors the Cᵦexo pucker in the second (Yaa) position of its Xaa–Yaa–Gly triplet (Shoulders et al., 2006; 2010).

Finally, substitutions on the pyrrolidine ring influence the transcis ratio of the prolyl peptide bond (Figure 1B). Proline is unique among amino acids in forming a tertiary amide, which is found often in both trans and cis conformations in folded proteins. The peptide bonds in collagen are all in the trans conformation, and inhibition of peptidyl-prolyl cis-trans isomerase decreases collagen production (Bächinger, 1987; Steinmann et al., 1991). Electronegative substituents in the 4R position, as in Hyp, lead to a high transcis ratio because the Cᵦexo ring pucker, enforced by the gauche effect, enables a strong n→π* interaction between the oxygen of the prolyl peptide bond (Oᵦ) and the prolyl carbonyl group (Cᵦᵦ=O) (Bretscher et al., 2001; DeRider et al., 2002). This interaction can only occur when the prolyl peptide bond is in its trans conformation, and thus stabilizes that conformation. Electronegative substituents in the 4S position of proline enforce the Cᵦendo ring pucker, which suppresses the n→π* interaction and leads to a low transcis ratio. The n→π* interaction also leads to pyramidalization of Cᵦ of proline in the direction of Oᵦᵦ (Choudhary et al., 2009).

**Prolyl 4-hydroxylase**

**Collagen as a substrate**

Collagen is the most abundant protein in animals, and the major component of connective tissue (Shoulders...
and Raines, 2009). The strands within the most common collagen, type I, each contain 338 Xaa–Yaa–Gly triplets. Pro is the amino acid found most commonly in the Xaa position, whereas Hyp is most often in the Yaa position. The Pro–Hyp–Gly sequence occurs in 10.5% of collagen triplets (Ramshaw et al., 1998).

Collagen has a characteristic triple-helical supersecondary structure (Figure 2) (Bella et al., 1994; Kramer et al., 1999; Nagarajan et al., 1999; Berisio et al., 2001). The triple helix consists of three left-handed helical chains in a right-handed supercoil. The presence of Hyp is required for collagen stability at physiological temperatures, as demonstrated by the difference in melting temperature \(T_m\) which is the temperature at the midpoint of the thermal transition) between a fully hydroxylated type I collagen \(T_m = 43°C\) and its unhydroxylated form \(T_m = 27°C\), which is below physiological temperature (Berg and Prockop, 1973a). Stabilization of the triple helix by the presence of Hyp has been studied extensively using peptide mimics of collagen (Table 1) (Sakakibara et al., 1973; Holmgren et al., 1998).

The key to the impact of Hyp on collagen stability are the stereoelectronic effects mediated by its hydroxyl oxygen. This fact was made clear by examining a collagen-related peptide in which the Hyp hydroxyl group is replaced with a fluoro group (Holmgren et al., 1998). Fluorine is even more electronegative than oxygen (Pauling, 1939), but organic fluorine has a low tendency to form hydrogen bonds. Peptides containing (2S,4R)-4-fluoroproline (Flp) in the Yaa position, \((\text{Pro–Flp–Gly})_{10}\), assemble into triple helices, and the \(T_m\) of these helices is 91°C, which is –20°C greater than that of \((\text{Pro–Hyp–Gly})_{10}\) triple helices (Table 1). Analogous results were obtained with (2S,4R)-4-chloroproline (Shoulders et al., 2008). These results indicate that the stabilizing effect of Hyp is due to the inductive effect generated from the electron-withdrawing substituent on the proline ring, rather than hydrogen bonding (Holmgren et al., 1998, 1999). The effect depends on stereochemistry, as replacing Pro with (2S,4S)-4-fluoroproline (Flp) destabilizes the triple helix (Bretscher et al., 2001) and 4,4-difluoroproline confers no extra stability (Shoulders et al., 2009). In essence, the C-\(\text{exo}\) ring pucker and high \(\text{trans/cis}\) ratio of Hyp preorganize collagen strands in the conformation found in the triple helix (Jenkins and Raines, 2002; Raines, 2006; Shoulders and Raines, 2009).

**Catalysis of Hyp formation**

The biosynthesis of fibrillar collagens entails a series of posttranslational modifications. One of the first is the hydroxylation of specific proline residues catalyzed by P4H (EC 1.14.11.2). The catalytic activity of P4H was first demonstrated in microsomal fractions (Peterkofsky and Udenfriend, 1965), and the enzyme was purified subsequently from chick embryos (Kivirikko and Prockop, 1967; Halme et al., 1970). P4H activity is critical for the proper folding of collagen, and P4H activity is necessary for the viability of the nematode Caenorhabditis elegans (Friedman et al., 2000; Winter and Page, 2000; Myllyharju et al., 2002) and the mouse Mus musculus (Holster et al., 2007).

Mammalian P4H is an \(\alpha_\beta_2\) tetramer (Berg and Prockop, 1973b; Nietfeld and Kemp, 1981; Koivu and Myllylä, 1986) in which the 59-kDa \(\alpha\) subunit (P4H\(\alpha\)) contains the peptide-substrate-binding domain and the enzymic active site (Prockop and Juva, 1965; Hutton et al., 1966; Helaakoski et al., 1989). Three isoforms of the P4H\(\alpha\)

<table>
<thead>
<tr>
<th>Peptide</th>
<th>(T_m) (°C)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>((\text{Pro–Flp–Gly})_{10})</td>
<td>91</td>
<td>Holmgren et al. (1999)</td>
</tr>
<tr>
<td>((\text{Pro–Hyp–Gly})_{10})</td>
<td>61–69</td>
<td>Holmgren et al. (1999)</td>
</tr>
<tr>
<td>((\text{Pro–Pro–Gly})_{10})</td>
<td>31–41</td>
<td>Holmgren et al. (1999)</td>
</tr>
<tr>
<td>((\text{Pro–Hyp–Gly})_{10})</td>
<td>No helix</td>
<td>Inouye et al. (1976)</td>
</tr>
<tr>
<td>((\text{Pro–Flp–Gly})_{10})</td>
<td>45</td>
<td>Bretscher et al. (2001)</td>
</tr>
<tr>
<td>((\text{Pro–Hyp–Gly})_{10})</td>
<td>36</td>
<td>Bretscher et al. (2001)</td>
</tr>
<tr>
<td>((\text{Pro–Pro–Gly})_{10})</td>
<td>No helix</td>
<td>Hodges and Raines (2005)</td>
</tr>
<tr>
<td>((\text{Pro–Flp–Gly})_{10})</td>
<td>No helix</td>
<td>Bretscher et al. (2001)</td>
</tr>
</tbody>
</table>
subunit, α(I), α(II), and α(III), have been identified in vertebrates, with α(I) being the most prevalent (Helaakoski et al., 1989; 1995; Annunen et al., 1997; Kukkola et al., 2003). All of the isoforms associate in an αβ₂ tetrameric form. Most of the conserved amino-acid residues occur in the C-terminal region, proximal to the active-site residues. P4Hs from other animals, such as C. elegans and the fly Drosophila melanogaster, have been characterized, and those from C. elegans (PHY-1 and PHY-2) can assemble either with a single β subunit to form dimers or as a mixed PHY-1/PHY-2/β₂ tetramer (Myllyharju et al., 2002; Myllyharju and Kivirikko, 2004).

The 55-kDa β subunit functions independently as protein disulfide isomerase (PDI) (Koivu and Myllylä, 1986; Pihlajaniemi et al., 1987; Kersteen and Raines, 2003). As a P4H subunit, PDI retains the enzyme in the lumen of the endoplasmic reticulum (ER) through its C-terminal KDEL retention signal and maintains the α subunit in a soluble and active form (Vuori et al., 1992a; 1992b). In the absence of PDI, the α subunit is insoluble and cannot be refolded in vitro (Nietfeld and Kemp, 1981). Recombinant P4H tetramers have been produced by co-production of the α subunit and PDI in mammalian, plant, insect, and yeast cells, as well as Escherichia coli expression systems (Kersteen et al., 2004; Neubauer et al., 2005).

P4H is a member of the non-heme iron(II), α-ketoglutarate-dependent dioxygenase family. Molecular oxygen (O₂), α-ketoglutarate and iron(II) are required for activity (Hutton and Udenfriend, 1966). During the reaction, α-ketoglutarate is decarboxylated oxidatively to produce succinate and CO₂ (Figure 1A) (Rhoads and Udenfriend, 1968).

The putative mechanism for prolyl hydroxylation by P4H (Figure 3) is based on studies of related dioxygenases (Costas et al., 2004). The reaction occurs in two stages. The first involves the formation of a highly reactive Fe(IV)=O species without the direct participation of the proline substrate. In the second stage, this species abstracts the pro-R hydrogen atom from C-4 of the proline substrate (Fujita et al., 1964), and the ensuing radicals combine to yield Hyp (Groves and McClusky, 1976).

Ascorbate (that is, vitamin C) is linked to catalysis by P4H (Myllylä et al., 1978; Nietfeld and Kemp, 1981). P4H can catalyze the decarboxylation of α-ketoglutarate without effecting the hydroxylation of proline, leading to an uncoupling of co-substrate turnover (Counts et al., 1978; Rao and Adams, 1978). The uncoupled reaction leads to inactivation of the enzyme that can be overcome by ascorbate (Myllylä et al., 1984). Ascorbate rescues the enzyme by reducing the inactive iron(III) state to the active iron(II) state (de Jong et al., 1982; de Jong and Kemp, 1984). A deficiency of ascorbate leads to scurvy (Lind, 1753; De Vreese, 2008), a disease caused by collagen instability (Carpenter, 1986).

![Figure 3](image-url)
Substrate recognition

P4H catalyzes hydroxylation of Pro residues in the Yaa position of the Xaa–Yaa–Gly triplets within collagen strands (Hutton et al., 1967). The hydroxylation reaction is performed on individual protocollagen chains but not triple helices (Berg and Prockop, 1973c). Proline itself is not hydroxylated by P4H (Cardinale and Udenfriend, 1974). The minimum substrate required for hydroxylation is an Xaa–Pro–Gly tripeptide, and Pro is the preferred residue in the Xaa position, though hydroxylation can occur at lower rates with a variety of residues at this position (Kivirikko et al., 1972; Rapaka et al., 1978).

P4Hα interacts with substrates in two sites, the peptide-substrate-binding domain and the active site. The peptide-substrate-binding domain binds to polyproline II-type structures. Polyproline itself is not hydroxylated by P4H, though it does bind to the enzyme and is a competitive inhibitor of enzymatic activity (Prockop and Kivirikko, 1969). The affinity of P4H for peptide substrates increases with increasing peptide length (Lambert et al., 1995; Myllyharju and Kivirikko, 1997). The three-dimensional structure of the peptide-substrate-binding domain (Phe144–Ser244) was determined by X-ray crystallography (Pekkala et al., 2004). The largely α-helical structure forms a concave, "bowl-like" surface containing a number of hydrophobic amino acids that likely compose the peptide-substrate-binding site. The three-dimensional structure of the entire P4H tetramer is unknown. The structure of yeast PDI is known (Tian et al., 2006), but provides little insight as to how mammalian PDI might associate with P4Hα to form a tetramer.

P4Hα also contains the catalytic active site. The iron is bound in the active site by two histidine residues and an aspartate residue. The spatial orientation of these three residues around the iron is not known in P4H, though that orientation is critical for enzymatic activity (Gorres et al., 2009). This 2-His–1-carboxylate motif is common to the α-ketoglutarate-dependent, iron(II) dioxygenases. Structural, spectral and computational analyses of Pro–Gly sequences in substrate peptides and proteins suggested that adoption of a β-turn conformation is required for their recognition by P4H (Rapaka et al., 1978; Brahmachari and Ananthanarayan, 1979; Chopra and Ananthanarayan, 1982; Atreya and Ananthanarayan, 1991). The β-turn structure forms the structural requirement for binding and catalysis in the active site, and longer substrates having a polyproline II-type helical structure add to the binding interaction by making contacts with the peptide-substrate-binding domain. Hydroxylation of the proline residues then results in a "straightening" of that turn, which allows the collagen triple helix to form. In previous studies, the residues surrounding the Pro–Gly sequence were varied to influence substrate conformation, and it was assumed that the peptide bond was in the trans conformation. More recent work describes P4H recognition of the conformation of the proline ring itself, and perhaps the cis conformation of the peptide bond. Peptide substrates containing proline derivatives that vary in ring pucker preference were used to reveal that P4H recognizes the C-endo ring pucker (Gorres et al., 2008). Proline derivatives that are P4H substrates form peptide bonds with a low trans:cis ratio. Upon hydroxylation, the switch to the Cʼ-exo ring pucker and trans peptide bond could provide a mechanism for P4H to avoid product inhibition.

Given that collagen is a polymeric substrate, the question arises as to whether P4H acts in a processive or distributive manner. To date, no evidence has been presented in support of processive catalysis by P4H. In peptide fragments derived from collagen, proline residues in the Yaa position are hydroxylated incompletely (Bornstein, 1967a; 1967b). In (Pro–Pro–Gly)n collagen-related peptides, the proline residues in the Yaa position of the third and fourth triplets are hydroxylated preferentially (Kivirikko et al., 1971). P4H tetramers contain two α subunits, each containing an active site, and a substrate of sufficient length could interact with both binding sites (de Waal and de Jong, 1988; Pekkala et al., 2004). This mode of action is, however, distinct from processive catalysis (de Jong et al., 1991).

Collagenous domains as substrates

The (Xaa–Yaa–Gly)n amino-acid sequence that is characteristic of collagen is found in other proteins as well. These sequences form triple helices, though the triple-helical domains are usually much shorter than those in collagen. The triple-helical domains can act as a spacer between globular domains, as oligomerization domains, or as a binding site for interacting partners.

The asymmetric form of acetylcholinesterase is composed of catalytic subunits attached to a collagenous, triple-helical domain (Rosenbloom and Cywinski, 1976b). Acetylcholinesterase hydrolyzes the neurotransmitter acetylcholine, which ends signal transmission at neuromuscular junctions. As in collagen, the collagenous domain consists of Xaa–Yaa–Gly triplets that contain Hyp. This collagen-like tail domain is responsible for attaching acetylcholinesterase to the basal lamina via heparin sulfate proteoglycans (Deprez et al., 2003). Gene mutations that prevent formation of the triple helix result in an absence of acetylcholinesterase at the neuromuscular junction and lead to dysfunction (Aldunate et al., 2004).

The complement protein C1q also contains a collagenous domain. C1q is a subunit of C1 that operates in the complement pathway of the innate immune response. The globular head domains of C1q bind antigen-bound immunoglobulins (Igs), as well as ligands.
associated with pathogens. The globular domains are held together by an oligomerization domain, which is composed of collagen-like triple helices made from an (Xaa–Yaa–Gly)₃ amino-acid sequence containing Hyp (Porter and Reid, 1978). Six trimers of C1q are linked by a collagen microfibril that produces an overall “bunch-of-tulips” structure for C1q. Individual C1q globular domains bind weakly to IgG and IgM, but oligomerization increases the strength of the interaction with clusters of IgG. The hydroxylation of proline residues is critical, as C1q secretion and function is decreased in the presence of either the iron chelator α,α-dipyridyl or 3,4-dehydroproline, which inhibit P4H (Muller et al., 1978; Mocharla et al., 1987).

Collectins are a class of proteins that contain a lectin domain, in addition to collagenous domains (van de Wetering et al., 2004). Also involved in the innate immune response, the lectin domains bind carbohydrates on the surface of pathogen cells. The collagen-like triple-helical domains perform a number of functions. Oligomerization of the lectin domains is accomplished by the collagenous domains. Binding of a single lectin domain to its carbohydrate ligand is weak without multivalency. Inhibition of prolyl hydroxylation, and thus triple-helix formation, in the collectin mannan-binding lectin (MBL) prevents proper oligomerization (Ma et al., 1997). The collagenous domain also dictates the shape and spacing of the lectin domains. The triple helix acts as a spacer in lung surfactant protein D (SP-D), whereas SP-A and MBL have interruptions in the Xaa–Yaa–Gly triplets that cause the collagenous domain to bend (Figure 4) (Voss et al., 1988).

The collagenous domains also play a role in the effector function of the collectins and C1q. When a lectin domain binds a ligand, the proteases that bind the collagenous region are activated and lead to initiation of the complement pathway. The collagenous domain can also bind to cell-surface receptors that then elicit many responses, including phagocytosis, chemotaxis, coagulation, and regulation of the adaptive immune response (Kishore et al., 2006).

Collagens, the collagen-domain-containing proteins discussed thus far, and the hibernation proteins HP-20, -25, and -7 that also contain collagenous domains (Takamatsu et al., 1993) are all secreted proteins. There are, in addition, integral membrane proteins with collagenous domains (Franzke et al., 2005). The macrophage scavenger receptors were the first known collagenous membrane proteins. Their collagen-like domain contains positively charged residues that bind a wide range of negatively charged ligands, including oxidized low-density lipoprotein. Ligand binding can lead to endocytosis or phagocytosis, or mediate adhesion.

Elastin as a substrate

Elastin is a structural protein that provides elasticity in connective tissues. Elasticity is especially important for blood vessels and lung tissues, which have an expectedly high elastin content. The amino-acid composition of elastin is rich in proline and glycine, like that of collagen, but elastin does not have glycine as every third residue, nor does it have a triple-helical structure. Instead, elastin is rich in alanine and valine. A prototypical elastin sequence is Val–Pro–Gly–Val–Gly, and peptides composed of (Val–Pro–Gly–Val–Gly)₃ repeats are substrates for P4H (Bhatnagar et al., 1978). The creation of Hyp in elastin is catalyzed by the collagen P4H, but there is less Hyp in elastin than in collagen (Rosenbloom and Cywinski, 1976b) and Hyp is not required for elastin biosynthesis or secretion (Rosenbloom and Cywinski, 1976a). The accumulation of elastin is, however, affected by levels of ascorbic acid. Cell cultures grown in the presence of ascorbate produce over-hydroxylated elastin that is less cross-linked and more soluble. Apparently, Hyp levels affect the formation of elastin fibrils (Dunn and Franzblau, 1982). Replacing Hyp with Flp or flp has dichotomous effects on the self-assembly of elastin peptides in vitro, indicative of a stereoelectronic effect analogous to that in collagen (Kim et al., 2005).

Prion protein as a substrate

The conversion of the cellular prion protein (PrP⁰) to a partially protease-resistant, aggregated scrapie
form (PrPSc) leads to neurodegenerative disorders. The C-terminal portion of PrP is mostly α-helical in PrPC and changes to all β-sheet in PrPSc. The physiological function(s) of PrP remains unknown, although roles have been proposed in antiapoptosis; antioxidation; sensing and transport of copper or other metals; neuronal development, differentiation, and maintenance; and even in the immune system (Marc et al., 2007). The proposed functions are based on interactions between PrP and metals, other proteins, or nucleic acids. The majority of these interactions occur within the N-terminus of PrP. The N-terminal domain of PrP is unstructured, but contains distinct regions of nonapeptide repeats and octapeptide repeats. A portion of the N-terminus also has a polyproline II-type helical structure. This region contains a Pro–Gly sequence that is hydroxylated in PrP produced in mammalian cell culture and from the brains of scrapie-infected mice (Gill et al., 2000). A peptide derived from residues 37–53 is hydroxylated in vitro by purified human P4H (K.L. Gorres, R.T. Raines, and E.S. Eberhardt, unpublished results). The biological consequence of this modification is unknown. It is possible that hydroxylation results in structural changes within PrP, or alters the metal–protein or protein–protein interactions required for the normal function of PrPC or the conversion and transmission of PrPSc.

Conotoxins as a substrate

Cone snails (genus Conus) produce venomous peptides that often target ion channels in the nervous system. These peptide toxins, known as conotoxins, are translated by the ribosome, and are highly cross-linked by disulfide bonds. Conotoxins also contain a large number of post-translational modifications, including prolyl hydroxylation (Buczek et al., 2005). Hyp has been identified in several conotoxins, and the hydroxylation seems to be sequence-specific because some peptides contain both Pro and Hyp. Hyp is found in amino-acid sequences that are distinct from the Pro–Gly sequence hydroxylated in collagen, and there is no obvious consensus sequence among hydroxylated conotoxins. No prolyl hydroxylase from Conus has been characterized, and the Hyp could be produced by either a specific prolyl 4-hydroxylase or one with broad specificity that also produces 4-hydroxyvaline (Pisarewicz et al., 2005).

Hyp in conotoxins affects folding, structure, and biological activity. An NMR structure of the O10P variant of conomarphin (where O = Hyp) revealed structural differences compared to the native peptide (Huang and Du, 2009). A study of peptides from each of the μ-, α-, and ω-conotoxin families revealed a variable effect of prolyl hydroxylation (Lopez-Vera et al., 2008). Removal of all three Hyp hydroxyl groups in the μ-GIIIA conotoxin slightly increases its folding rate, but greatly decreases its biological activity. In these and other peptides, Hyp mediates the conotoxin peptide–protein interaction. The α-conotoxins do not contain Hyp, though replacing Pro with Hyp in α-1ml or α-GI increases the rate of folding and decreases bioactivity. Hyp in the α-conotoxins interrupts peptide–protein interactions. Hyp has no effect on the biological activity of ω-MVIIA conotoxin, but does improve the yield of folded peptide and rate of folding. Accordingly, the role of Hyp in conotoxins could be to stabilize structure, enable molecular recognition, or encourage other post-translational modifications.

Argonaute 2 as a substrate

RNA interference (RNAi) is enacted by RNA-induced silencing complexes (RISCs). RISCs are composed of small interfering RNAs (siRNAs) and proteins from the Argonaute family. Argonaute 2 (Ago2) cleaves target mRNAs (Liu et al., 2004). Ago2 interacts with collagen P4H, and Hyp has been identified as residue 700 in Ago2 (Qi et al., 2008). Yet, hydroxylation is not required for the catalytic activity of Ago2 or for siRNA binding. Pro700 is located within the Pro–Gly dipeptide sequence that is hydroxylated in collagen. Other proline residues within Pro–Gly sequences of Ago2 are not hydroxylated, suggesting specificity. Although Ago2 is located largely in the cytosol, there is evidence for some Ago2 in the ER. Cytosolic prolyl hydroxylase domain proteins (vide infra) do not hydroxylate Ago2 in vitro.

The hydroxylation of Ago2 at Pro700 increases the physiological stability of Ago2 (Qi et al., 2008). The P700A variant of Ago2 has less conformational stability than does the wild-type enzyme, and the cellular half-life of Ago2 is diminished upon P4H inhibition. The mechanism by which Hyp700 stabilizes Ago2 is unknown. The degradation of Ago2 appears to be proteasome-mediated, but what is the role of Hyp? Does Hyp stabilize the structure of Ago2, as it does for collagen? Does Hyp promote the binding of another protein that stabilizes Ago2, or does the absence of Hyp allow recognition of Ago2 and its direction to the proteasome? A key will be to learn whether prolyl hydroxylation affects other post-translational modifications of Ago2, such as ubiquitination.

Prolyl 3-hydroxylase

Collagen also contains (2S,3S)-3-hydroxyproline (3-Hyp), though 3-Hyp is much less abundant than 4-Hyp (Rhodes and Miller, 1978). 3-Hyp is more prevalent in the Type IV collagen of basement membranes, which contain 10–15 3-Hyp residues, than in Type I
Prolyl hydroxylase domain protein (PHD)

Hypoxia inducible factor α as a substrate

In animals, molecular oxygen is detected and its homeostasis is maintained through the hypoxia-inducible transcription factors (HIFs) (Kaelin and Ratcliffe, 2008; Chowdhury et al., 2008). HIFs direct the transcription of >100 genes through regulatory hypoxia response elements (HRE) (Ke and Costa, 2006). HIF-regulated genes are involved in cell proliferation, angiogenesis, erythropoiesis, and metabolism. The principal HIF, HIF-1, is composed of two subunits, HIF-1α and HIF-1β, both of which are produced constitutively. The level of HIF-1α, however, is regulated by the availability of molecular oxygen. Under normal oxygen levels, HIF-1α is polyubiquitinated and degraded rapidly by the proteasome (Figure 5). During hypoxia, HIF-1α is not degraded, but translocates to the nucleus and dimerizes with HIF-1β to form the active transcription factor.

The concentration of cytosolic oxygen is sensed by prolyl hydroxylase domain proteins (PHDs) that act on HIF-1α. Under normal oxygen conditions (normoxia), PHDs hydroxylate two highly conserved proline residues (Pro402 and Pro564) located within the oxygen-dependent degradation domain (ODDD). The presence of Hyp within the ODDD of HIF-1α is recognized by the von Hippel-Lindau tumor suppressor protein (pVHL), which is a component of a ubiquitin–protein E3 ligase complex, along with elonginB, elonginC, cul2, and rbx1. Upon hydroxylation, HIF-1α is recognized by the ubiquitin E3 ligase, polyubiquitinated, and directed to the proteasome for degradation (Figure 5). Under hypoxic conditions PHD activity is decreased due to its need for molecular oxygen as a cosubstrate.

The interaction between HIF-1α and the pVHL–elonginC–elonginB (VCB) complex is controlled by prolyl hydroxylation. A 20-residue peptide derived from HIF-1α that encompasses Pro564 can be hydroxylated by a PHD and then recognized by pVHL. The three-dimensional structure of the VCB complex co-crystallized with a HIF-1α peptide containing Hyp is known. The HIF-1α peptide and the pyrrolidine ring of Hyp form contacts with hydrophobic areas of pVHL. The hydroxyl group of Hyp564 in the HIF-1α peptide forms hydrogen bonds with the hydroxyl group of Ser111 and the imidazolyl group of His115 in pVHL (Figure 6) (Hon et al., 2002; Min et al., 2002). The presence of Hyp in a peptide fragment of HIF-1α increases its affinity for the VCB complex by 10^4-fold (Hon et al., 2002).

HIF prolyl hydroxylases

A prolyl hydroxylase domain protein (PHD) that acts on HIF-1α is known (Bruick and McKnight, 2001). There are three isoforms of PHDs: PHD1–3, which are also known as HIF-P4Hs (HPHs) 3–1 or EGLNs 2, 1 and 3. The PHDs are like collagen P4H in that they require molecular oxygen, α-ketoglutarate, and iron(II) for catalytic activity, and the PHDs have the 2-His–1-Asp iron-binding motif (Bruick and McKnight, 2001). PHDs likely utilize a mechanism similar to those as P4Hs. PHDs are, however, distinct from the P4H involved in collagen biosynthesis in being cytosolic enzymes. The apparent K_m value of PHDs for O_2 is higher than that for collagen P4H and is greater than the concentration of molecular oxygen in tissues, which allows the enzymatic activity to report on O_2 concentrations throughout the physiological range (Hirslä et al., 2003; Ehrismann et al., 2007).

The two proline residues in HIF-1α that are hydroxylated by PHDs are located in LXXLAP motifs. The preferences for the N-terminal oxygen-dependent degradation
domain (ODDD) versus the C-terminal ODDD vary among the HIFα and PHD isoforms. Collagen P4H cannot hydroxylate the LXXLAP motif in HIF-1α (Jaakkola et al., 2001). Recognition of the sequence by PHDs is, however, quite flexible, with the presence of the alanine residue being the strictest requirement (Li et al., 2004). The minimum length for a peptide substrate is eight residues, but peptides of 19–20 residues are hydroxylated much more efficiently. There is no evidence for secondary structural requirements within the HIF-1α peptide for PHD recognition.

The proline residue within the HIF-1α peptide is required for binding to PHD (Li et al., 2004). When proline is replaced with the analogs 3,4-dehydroproline or l-azetidine-2-carboxylic acid, the rate of uncoupled α-ketoglutarate decarboxylation increases. (2S,4S)-4-Hydroxyproline (hyp) and flp are substrates for PHD when incorporated into peptides derived from HIF-1α (Loenarz et al., 2009). The structure of a HIF-1α-derived peptide bound to PHD2 revealed the substrate proline residue to have a Cγ-endo ring pucker (Figure 7) (Chowdhury et al., 2009). These results suggest that PHDs recognize the prolyl ring pucker in a manner similar to P4H (Gorres et al., 2008).

Large subunit of RNA polymerase II as a substrate

The RNA polymerase II complex, responsible for transcribing DNA into mRNA, transitions from transcription initiation to elongation through phosphorylation of the C-terminal domain of the large subunit Rpb1. In response to ultraviolet irradiation or oxidative stress, hyperphosphorylated Rpb1 is bound by pVHL and decorated with ubiquitin. The ubiquitination of Rpb1 does not lead to its degradation. Binding of pVHL is dependent on the hyperphosphorylation of the C-terminal domain of Rpb1 and the hydroxylation of Pro1465 (Kuznetsova et al., 2003).

Rpb1 shares some amino-acid sequence similarity with HIF-1α, including an LXXLAP motif, suggesting the involvement of a PHD rather than a collagen P4H. PHD1 was found to be the major catalyst of Rpb1 prolyl hydroxylation (Mikhaylova et al., 2008). Surprisingly,
PHD2 inhibited hydroxylation of Pro1465 and phosphorylation. The role of prolyl hydroxylation in Rpb1, as in HIF-1α, is to recruit pVHL. Rpb1 is translocated from the soluble fraction to the chromatin-engaged fraction by pVHL under conditions of oxidative stress. The PHDs were also found in the chromatin fraction. The consequences of Rpb1 hydroxylation and pVHL binding within a cell are unknown. The regulation of Rpb1, and thus RNA polymerase, by pVHL could be involved in transcription elongation that alters gene expression during stresses that result in DNA damage.

**iκB kinase-β as a substrate**

NFκB is a transcription factor involved in fundamental aspects of the innate immune response and inflammation, and is important for tumor development. Hypoxia has been shown to activate NFκB. The link between oxygen sensing and NFκB appears to be prolyl hydroxylation by the same PHD that is crucial for oxygen sensing and the HIF response. Inhibition of PHD, particularly PHD1, by either small-molecule inhibitors or siRNA results in NFκB activation (Cummins et al., 2006). Conversely, overexpression of PHD1 under normal oxygen conditions causes a decrease in NFκB activity. PHD does not, however, act directly on NFκB.

NFκB is controlled by a cascade of inhibitory proteins. NFκB is sequestered in the cytosol by its interaction with inhibitory xB (IxB). Phosphorylation of IxB by IκB kinase-β (IKKβ) leads to the ubiquitination and degradation of IxB, exposing the nuclear localization signal of NFκxB. IKKβ contains a conserved LXXLAP motif, which is the same sequence that is required for hydroxylation in HIFα. When the proline residue in the LXXLAP motif in IKKβ is replaced (as in the P191A variant), NFκxB is no longer induced by hypoxia (Cummins et al., 2006). Hydroxylation of these substrates remains to be confirmed by mass spectrometry.

The hydroxylation of Pro191 could change the conformation of the activation loop, making the kinase inactive. Alternatively, hydroxylation might disrupt the binding of the substrate. Hydroxylation could also induce the binding of another protein, possibly pVHL, which would block the phosphorylation and activation of IKKβ.

**Activating transcription factor 4 as a substrate**

There is also evidence for prolyl hydroxylation-dependent degradation of activating transcription factor 4 (ATF-4) (Koditz et al., 2007). ATF-4 was found to interact with PHD3, but not PHD1 or PHD2. Like HIF-1α, ATF-4 is stabilized by PHD inhibitors, hypoxia, and proteasome inhibitors. The interaction was mapped to a portion of the zipper II domain, which contains five proline residues, though none are within an LXXLAP motif. ATF-4 variants lacking this region or all five proline residues are more stable than wild-type ATF-4. Replacing individual proline residues does not, however, elicit the same effect. The combination of Hyp residues required for protein stabilization is unknown. ATF-4, incubated under appropriate conditions for prolyl hydroxylation, did not interact with pVHL. Rather, degradation of ATF-4 was found to be dependent on the SCFTrCP ubiquitin ligase (Lassot et al., 2001). It remains to be determined whether prolyl hydroxylation is required for this interaction or one with another ubiquitin ligase or adaptor protein, or whether hydroxylation has an important structural consequence.

**β2-Adrenergic receptor as a substrate**

The β-adrenergic receptors, members of the G protein-coupled receptor family, are stimulated by the catecholamines norepinephrine and epinephrine, and regulate cardiovascular and pulmonary functions. Signaling through this pathway is modulated by the number of receptors on the cell surface. For example, receptors are down-regulated by continuous agonist stimulation. Hypoxia, though, results in an increase in the β2-subtype adrenergic receptor (β2AR). This response to molecular oxygen occurs via prolyl hydroxylation (Xie et al., 2009). Like HIFα, hydroxylation of proline residues in β2AR promotes the binding of pVHL-E3 ligase, which ubiquitinates
the \( \beta_{\text{AR}} \), marking it for proteasomal degradation. Hyp was found at Pro382 and Pro395, though neither proline is located in an LXXLAP motif. \( \beta_{\text{AR}} \) is insensitive to oxygen when both of these proline residues are replaced with alanine. \( \beta_{\text{AR}} \) interacts with EGLN3 (PHD3), but not EGLN1 or EGLN2, and depletion of EGLN3 leads to an increase in \( \beta_{\text{AR}} \) under normoxic conditions. The regulation of \( \beta_{\text{AR}} \) by ELGN3 evidences a HIF-independent oxygen-sensing role for prolyl hydroxylation that could have implications in cardiovascular pathogenesis.

### Transmembrane prolyl 4-hydroxylase

A known prolyl 4-hydroxylase, P4H-TM or PH-4, contains a transmembrane domain near its N-terminus (Oehme et al., 2002; Koivunen et al., 2007). P4H-TM is associated with the membrane of the ER. By comparison of amino-acid sequences, P4H-TM is related more closely to the catalytic C-terminal region of collagen P4H than to the PHDs, though P4H-TM does not show any sequence similarity to the N-terminal peptide-substrate-binding domain of P4H. P4H-TM, however, decreases transcriptional activation by HIF-1\( \alpha \). In vitro, P4H-TM hydroxylates HIF-1\( \alpha \) but does not hydroxylate collagen, even though in cellulo, its active site resides in the ER lumen. P4H-TM expression is induced under hypoxic conditions in cell culture, although it cellular location does not change. How the active site of P4H-TM inside the ER can act upon a (typically) cytosolic protein and the role of the cellular localization of P4H-TM are not known. It is possible that P4H-TM has a specialized function in regulating HIF-1\( \alpha \). Alternatively, HIF-1\( \alpha \) might not be the primary substrate, and P4H-TM could be active in other pathways.

### Plant and algal prolyl 4-hydroxylases

Prolyl hydroxylation occurs in a number of proteins in plants and algae. Peptides containing Hyp are part of systemin defense mechanisms (Ryan and Pearce, 2003; Pearce et al., 2009), and Hyp is found in some secreted and vacuolar proteins (Shimizu et al., 2005). Hyp is abundant in a large class of proteins, termed hydroxyproline-rich glycoproteins (HPRGs), in which 15–25% of the residues are Hyp. HPRGs are the major proteinaceous components of the cell walls in higher plants and green algae. In addition to functioning in cell-wall assembly and rigidity, HPRGs play roles in plant growth, development, cell–cell interactions, and cellular communication (Wu et al., 2001). The HPRGs are subgrouped by the type of residues in characteristic repetitive sequences. The extensins typically contain a Ser–Hyp motif, the repetitive proline-rich proteins have variations of pentapeptide repeats containing much Hyp and some Ser, the arabinogalactan proteins contain Hyp alternating with other residues, and other HPRGs have contiguous Hyp residues (Kieliszewski and Lamport, 1994; Kieliszewski and Shpak, 2001).

Some Hyp residues in plants and algae are modified further by the addition of oligoarabinose or arabinogalactan. The extent and type of O-glycosylation can be predicted by the Hyp contiguity hypothesis, in which glycosylation correlates with the location and context of Hyp residues (Kieliszewski, 2001). Where Hyp residues are contiguous in the amino-acid sequence, arabinosylation is predominant, whereas arabinogalactan polysaccharides are added to clustered, non-contiguous Hyp residues. Glycosylation of Hyp has not been found in animals.

DNA encoding plant prolyl 4-hydroxylases has been cloned from Arabidopsis thaliana (Hieta and Myllyharju, 2002; Tiainen et al., 2005), Nicotiana tabacum (Yuasa et al., 2005), and the green alga Chlamydomonas reinhardtii (Keskiaho et al., 2007). Prolyl 4-hydroxylases in plants, like those in animals, utilize molecular oxygen, \( \alpha \)-ketoglutarate, iron(II), and ascorbate. In general, plant P4Hs are smaller in size, \( \sim 30–60 \text{kDa} \), compared to collagen P4H. Plant and algal P4Hs are soluble monomers, and the three-dimensional structure of the C. reinhardtii P4H (Cr-P4H-1) has been determined by X-ray crystallography (Koski et al., 2007). The 2-His–1-carboxylate iron-binding residues and overall structure are consistent with what is known about P4H and PHD. Cr-P4H-1 seems, however, to be more similar to P4H in that it contains a polyproline-binding domain. An N-terminal transmembrane domain was identified in a P4H from N. tabacum, and is predicted by sequence analysis to exist in other plant P4Hs. This membrane-bound P4H localizes to the ER and Golgi (Yuasa et al., 2005).

The plant prolyl 4-hydroxylases differ substantially from the animal enzymes in their substrate specificity. P4Hs isolated from plants can hydroxylate polyproline, which is a competitive inhibitor of the collagen P4Hs. The product of this reaction, poly(4-hydroxyproline), has an even greater tendency than polyproline to adopt a polyproline II-type conformation (Horng and Raines, 2006). Peptides that mimic collagen, \((Xaa-Pro-Gly)_{n}\), are hydroxylated by some plant P4Hs, though generally inefficiently (Tanaka et al., 1981; Kaska et al., 1987). The A. thaliana At-P4H-1 enzyme does hydroxylate collagen-like peptides, as well as a peptide derived from HIF-1\( \alpha \) that has only one proline residue. The At-P4H-2 enzyme does not, however, hydroxylate efficiently either the collagen-like peptide or the HIF-1\( \alpha \) peptide. Both At-P4H-1 and At-P4H-2 hydroxylate peptides representing the plant proline-rich proteins, arabinogalactan protein and extensin.

Despite differences in the amino-acid sequence of native substrates for plant and animal prolyl 4-hydroxylases, the recognition of the proline residue
through its ring pucker seems to be a commonality. P4H and PHD prefer substrates containing proline derivatives that favor the C-endo ring pucker, and do not bind Hyp-containing peptides that favor the C-exo ring pucker (Gorres et al., 2008; Loenarz et al., 2009). Similarly, the structure of the algal P4H, Cr-P4H-1, complexed with a (Pro–Ser)5 peptide substrate, revealed the Pro in the active site to have a C-endo ring pucker (Figure 7) (Koski et al., 2009). Tyr140 in the Cr-P4H-1 active site could prevent Hyp from binding.

**Prolyl 4-hydroxylases in microorganisms**

**Protozoan prolyl 4-hydroxylases**

Skp1 is a eukaryotic protein that is a subunit in several multi-subunit complexes, but is well studied as an adapter in the SCF (Skp1-cullin-F box protein) E3 ubiquitin ligase complex. In the amoeba Dictyostelium discoideum, commonly referred to as slime mold, Pro143 of Skp1 is glycosylated after hydroxylation (West et al., 2004). The ensuing pentasaccharide is added by five glycosyltransferases. Although glycosylation of hydroxyproline is common in secreted plant cell wall proteins, Skp1 is a cytosolic and nuclear protein.

A gene encoding P4H from D. discoideum, phyA, has been cloned and characterized (van der Wel et al., 2005). The activity of recombinant D. discoideum P4H1, DdP4H1, requires molecular oxygen, α-ketoglutarate, and ascorbate; and activity decreases in the presence of iron chelators. Recombinant Skp1 is a substrate, but a peptide derived from Skp1 is not. DdP4H1 was found to be a soluble cytosolic protein. The phyA gene for DdP4H1 encodes the conserved 2-His–1-Asp iron-binding residues and is related more closely to the gene of the PHDs than of the P4Hs. The hydroxylated proline in Skp1 is not, however, within an LXXLAP motif. Like PHDs, DdP4H1 appears to sense molecular oxygen and regulate D. discoideum development (West et al., 2007).

DNA sequences that encode proteins resembling P4H have been discovered in the genomes of other eukaryotic microorganisms, such as the diatom Thalassiosira pseudonana and the oomycete Phytophthora sojae. Interestingly, these genes are predicted to be bifunctional, encoding the first glycosyltransferase in the pathway in addition to a P4H. The P4H/glycosyltransferase pathway might also exist in Toxoplasma gondii, the causative agent of toxoplasmosis (West et al., 2006).

**Bacterial prolyl 4-hydroxylases**

Hyp is also found in bacterial antibiotic peptides. These peptides are synthesized by enzymatic pathways rather than by the ribosome. These non-ribosomal peptides often contain a high percentage of non-natural and modified amino acids, including Hyp. As in animals and plants, bacterial Hyp is formed by stereospecific hydroxylation at the 4R position (Baldwin et al., 1993), and the hydroxyl oxygen is derived from O2 (Diegelmann et al., 1969). In addition to Hyp, other isomers of hydroxyproline and other proline modifications occur in bacteria. Hyp and (2S)-4-ketoproline (Kep) are found in actinomycins produced by Streptomyces antibioticus (Katz et al., 1962), both diastereomers of (2S)-3-hydroxyproline are found in telomycin (Sheehan et al., 1968), (2R,4R)-4-hydroxyproline is found in etamycin (Katz et al., 1979), and hyp is found in microcolin A (Koehn et al., 1992). Pro is the precursor to all the different forms of hydroxyproline in bacteria. A major difference from all other organisms, however, is that bacterial Hyp is produced from free proline instead of peptidyl proline (Adefarati et al., 1991).

The enzymes that catalyze the hydroxylation of free proline are identified as the proline hydroxylases, and are distinct from the prolyl hydroxylases that hydroxylate peptidyl proline. A proline 4-hydroxylase and a proline 3-hydroxylase have been purified from Streptomyces. Proline 4-hydroxylase forms Hyp in the production of etamycin (Lawrence et al., 1996), and proline 3-hydroxylase catalyzes the formation of (2S,3S)-hydroxyproline (Mori et al., 1997). A proline 4-hydroxylase converting Pro to hyp is also known (Hara and Kino, 2009). Like P4H and PHD, these proline hydroxylases are thought to be members of the non-heme iron(II) dioxygenase family. They also require molecular oxygen, α-ketoglutarate, and iron(II). The proline hydroxylases seem to show less substrate specificity than does P4H in that the disparate analogs (2S)-3,4-dehydroproline and L-pipecolic acid are substrates (Baldwin et al., 1994). A three-dimensional structure of proline 3-hydroxylase reveals the canonical 2-His–1-Asp iron-binding residues in the active site (Clifton et al., 2001). The structure also implicates a number of charged residues that could bind the amino and carboxyl groups of the proline substrate.

Although prolyl hydroxylation in bacteria occurs mainly on free proline, a bacterial peptidyl-prolyl hydroxylase is known. This Bacillus anthracis enzyme, designated anthrax-P4H, is homodimeric and dependent on molecular oxygen, α-ketoglutarate, and iron(II) (Miller et al., 2008). Unlike other bacterial hydroxylases that hydroxylate free proline, anthrax-P4H binds the collagen-like peptide (Gly–Pro–Pro)4 with an affinity similar to that of P4H. The three-dimensional structure of anthrax-P4H reveals an overall fold and a 2-His–1-Asp active site characteristic of α-ketoglutarate-dependent iron(II) dioxygenases (Culpepper et al., 2010). The
physiological substrate and role of anthrax-P4H is unknown.

**Viral prolyl 4-hydroxylases**

An enzyme catalyzing prolyl hydroxylation has also been identified in the eukaryotic algal virus *Paramecium bursaria* chlorella virus-1 (PBCV-1) (Eriksson *et al*., 1999). The PBCV-1 prolyl 4-hydroxylase sequence shows similarity to the C-terminal region of the catalytic subunit of P4H. PBCV-1 P4H is a monomer, and can hydroxylate a collagen-like peptide, as well as polyproline, the typical P4H. PBCV-1 P4H is a monomer, and can hydroxylate a collagen-like peptide, as well as polyproline, the typical plant P4H substrate. The viral genome contains open characteristic (XPG)$_n$ collagen sequence, but proline residues preceding glycine (PG) in non-collagenous proteins also undergo hydroxylation. The hydroxylated prolines in conotoxins, though, do not seem to be in any consensus sequence. The LXXLAP motif is a common substrate for PHDs, as in HIF$\alpha$, RNA polymerase II Rpb1, and IKK$\beta$. Still, proteins without this motif, such as ATF-4 and $\beta$AR, are also PHD substrates. Plant P4Hs hydroxylate sequences rich in proline residues with a variety of repetitive motifs, and bacteria are unique in that they hydroxylate proline as a free amino acid.

The function of Hyp also varies greatly. In collagen, Hyp plays a structural role preorganizing the collagen strands to stabilize the triple-helical structure. Hyp can also act as a recognition motif for protein–protein interactions that can lead to a variety of consequences. Hyp allows pVHL recognition of HIF$\alpha$ that leads to protein degradation, conotoxin binding to target ion channels, and bacterial nonribosomal peptide antibacterial activity. Prolyl hydroxylation inhibits the enzymatic activity of IKK$\beta$, which may be caused by a change in protein conformation or another protein binding. In plants and algae, Hyp is abundant and provides a substrate for the addition of sugars that have many functions on the cell surface. Intriguingly, the biological consequences of the presence of Hyp in place of Pro are yet to be revealed in many proteins (Table 2). Moreover, it is likely that evidence for the action of prolyl 4-hydroxylases will continue to be discovered in additional proteins and host organisms.

**Protein structure**

All prolyl and proline hydroxylases are members of a family of enzymes that utilize molecular oxygen, $\alpha$-ketoglutarate, and iron(II), and most show increased activity in the presence of ascorbate. Studies on $\alpha$-ketoglutarate-dependent iron(II) dioxygenases have revealed a common iron-binding motif that includes two His residues and one Asp/Glu residue (Schöfied and Zhang, 1999). An exception to the 2-His–1-carboxylate motif is found in the active site of halogenases, which catalyze the addition of a halo group instead of a hydroxyl moiety (Blasiak *et al*., 2006; Wong *et al*., 2009). In the halogenases, the carboxylate (Asp or Glu) is replaced by an alanine residue and a halide ion. Simply replacing the active-site Asp of human P4H with an alanine residue does not, however, endow the enzyme with halogenase activity (Gorres *et al*., 2009).

Overall, the $\alpha$-ketoglutarate-dependent dioxygenases show low sequence identity, but do share a common three-dimensional structural fold. Their 2-His–1-carboxylate motifs occupy a similar position within the $\beta$-barrel jelly roll motif (Figure 8). Outside that motif, the enzymic structures vary to accommodate disparate substrates.

**Conclusions**

Prolyl and proline hydroxylases, their substrates, and their biological functions are summarized in Table 2. The amino-acid sequences of the substrates are quite diverse, with Pro itself being the only commonality. The P4Hs involved in collagen biosynthesis recognize the characteristic (XPG)$_n$ collagen sequence, but proline residues preceding glycine (PG) in non-collagenous proteins also undergo hydroxylation. The hydroxylated prolines in conotoxins, though, do not seem to be in any consensus sequence. The LXXLAP motif is a common substrate for PHDs, as in HIF$\alpha$, RNA polymerase II Rpb1, and IKK$\beta$. Still, proteins without this motif, such as ATF-4 and $\beta$AR, are also PHD substrates. Plant P4Hs hydroxylate sequences rich in proline residues with a variety of repetitive motifs, and bacteria are unique in that they hydroxylate proline as a free amino acid.

The function of Hyp also varies greatly. In collagen, Hyp plays a structural role preorganizing the collagen strands to stabilize the triple-helical structure. Hyp can also act as a recognition motif for protein–protein interactions that can lead to a variety of consequences. Hyp allows pVHL recognition of HIF$\alpha$ that leads to protein degradation, conotoxin binding to target ion channels, and bacterial nonribosomal peptide antibacterial activity. Prolyl hydroxylation inhibits the enzymatic activity of IKK$\beta$, which may be caused by a change in protein conformation or another protein binding. In plants and algae, Hyp is abundant and provides a substrate for the addition of sugars that have many functions on the cell surface. Intriguingly, the biological consequences of the presence of Hyp in place of Pro are yet to be revealed in many proteins (Table 2). Moreover, it is likely that evidence for the action of prolyl 4-hydroxylases will continue to be discovered in additional proteins and host organisms.
Table 2. Prolyl and proline hydroxylases.

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Substrate</th>
<th>Sequence</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>P4H</td>
<td>Collagen</td>
<td>(XPG)$_n$</td>
<td>Conformational stability</td>
</tr>
<tr>
<td>P4H</td>
<td>Collagen-domain proteins</td>
<td>(XPG)$_n$</td>
<td>Conformational stability</td>
</tr>
<tr>
<td>P4H</td>
<td>Elastin</td>
<td>PG</td>
<td>?</td>
</tr>
<tr>
<td>P4H</td>
<td>Prion protein</td>
<td>PG</td>
<td>?</td>
</tr>
<tr>
<td>P4H</td>
<td>Conotoxin</td>
<td>No consensus</td>
<td>Conformational stability and activity</td>
</tr>
<tr>
<td>P4H</td>
<td>Ago2</td>
<td>PG</td>
<td>Conformational stability</td>
</tr>
<tr>
<td>P3H</td>
<td>Collagen</td>
<td>(POG)$_n$</td>
<td>?</td>
</tr>
<tr>
<td>PHD</td>
<td>HIFα</td>
<td>LXXLAP</td>
<td>Protein–protein interaction</td>
</tr>
<tr>
<td>PHD</td>
<td>RNA polymerase II Rpb1</td>
<td>LXXLAP</td>
<td>Protein–protein interaction</td>
</tr>
<tr>
<td>PHD</td>
<td>1xB kinase-βI</td>
<td>LXXLAP</td>
<td>Enzymatic activity</td>
</tr>
<tr>
<td>PHD</td>
<td>ATF-4α</td>
<td>Not LXXLAP</td>
<td>Conformational stability</td>
</tr>
<tr>
<td>PHD</td>
<td>β₂AR</td>
<td>Not LXXLAP</td>
<td>Protein–protein interaction</td>
</tr>
<tr>
<td>P4H-TM</td>
<td>HIF-1α</td>
<td>LXXLAP</td>
<td>?</td>
</tr>
<tr>
<td>Plant and algal P4H</td>
<td>HPRGs</td>
<td>Polypeptide and proline-rich sequences</td>
<td>Glycosylation</td>
</tr>
<tr>
<td>DdP4H1</td>
<td>Skp1</td>
<td>KNDFTPEEQIRK</td>
<td>Glycosylation</td>
</tr>
<tr>
<td>Proline hydroxylase</td>
<td>Peptide antibiotics</td>
<td>ProOH</td>
<td>Antibacterial activity</td>
</tr>
<tr>
<td>Anthrax-P4H</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>PBCV P4H</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>

* Hyp not yet identified directly (e.g. by mass spectrometry).

Acknowledgements

K.L.G. was supported by Chemistry–Biology Interface Training Grant T32 BM008505 (NIH). Work in our laboratory on prolyl 4-hydroxylases is supported by grant R01 AR044276 (NIH).

Declaration of interest

The authors report no conflicts of interest. The authors alone are responsible for the content of this paper.

References


Shoulders MD, Satyshur KA, Forest KT and Raines RT. 2010. Stereoelectronic and steric effects in side chains preor-


