Collagen Prolyl 4-Hydroxylase as a Therapeutic Target

Miniperspective

James D. Vasta† and Ronald T. Raines‡,†,‡,§

†Department of Biochemistry and ‡Department of Chemistry, University of Wisconsin—Madison, Madison, Wisconsin 53706, United States
§Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, United States

ABSTRACT: Collagen is the dominant protein of the extracellular matrix. Its distinguishing feature is a three-stranded helix of great tensile strength. (2S,4R)-4-Hydroxyproline residues are essential for the stability of this triple helix. These residues arise from the post-translational modification of (2S)-proline residues by collagen prolyl 4-hydroxylases (CP4Hs), which are members of the Fe(II)- and α-ketoglutarate (AKG)-dependent dioxygenase family. Here, we provide a framework for the inhibition of CP4Hs as the basis for treating fibrotic diseases and cancer metastasis. We begin with a summary of the structure and enzymatic reaction mechanism of CP4Hs. Then, we review the metal ions, metal chelators, mimetics of AKG and collagen strands, and natural products that are known to inhibit CP4Hs. Our focus is on inhibitors with potential utility in the clinic. We conclude with a prospectus for more effective inhibitors.

INTRODUCTION

Comprising a third of the proteins in the human body and three-quarters of the dry weight of human skin, collagen is the most abundant protein in humans and other animals. Different types of collagen play myriad roles in biology. Collagen is, however, predominantly a structural protein, serving as the principle component of connective tissue, bone, and the extracellular matrix (ECM).

The distinguishing feature of collagen is its triple helix (Figure 1), an elegant structural motif in which three polypeptide chains are interwoven to form a right-handed superhelical structure of great tensile strength (Figure 1A,B). The individual polypeptide chains that form the triple helix are composed of numerous triplet repeats of the sequence Xaa-Yaa-Gly, where Xaa is often (2S)-proline (Pro) and Yaa is often (2S,4R)-4-hydroxyproline (Hyp), two amino acids discovered by Emil Fischer. The presence of Hyp in collagen is due to the post-translational hydroxylation of Pro residues in the Yaa position by collagen prolyl 4-hydroxylases (CP4Hs). This hydroxylation enhances greatly the conformational stability of the triple helix by imposing stereoelectronic effects that preorganize individual strands (Figure 1C). Thus, the activity of CP4H is critical to the structure of the most prevalent protein in the human body.

COLLAGEN BIOSYNTHESIS

As collagens are secretory proteins, their biosynthesis begins in the rough endoplasmic reticulum (ER) (Figure 2). Mature mRNAs are translated into individual procollagen polypeptides (also called preprocollagens) that undergo processing in the ER lumen. In addition to the excision of the signal peptide, the processing includes extensive post-translational modifications, many of which are essential for various aspects of triple-helix or supramolecular fibril formation. Modified procollagen strands fold into triple helices in a process facilitated by globular N- and C-terminal propeptide domains present in each procollagen strand. Typically, the C-terminal domains from three procollagen strands form a complex that is stabilized by interdomain interactions that enforce a particular homotrimeric or heterotrimeric composition. This complex also establishes the register for the pendant (Xaa-Yaa-Gly)n domains, which “zip up” from C-terminus to N-terminus to form a procollagen triple helix. Triple-helix assembly appears to benefit from general chaperone proteins in the secretory pathway as well as from more specific proteins. To date, 48 proteins have been identified as participants in collagen proteostasis, including Hsp47, cyclophilin B, FKBP10, and enzymes involved in post-translational modifications, including CP4Hs. During assembly, the N- and C-terminal propeptide domains serve to enhance solubility and prevent premature aggregation. Procollagen triple helices, which are substantially longer (0.3 μm) than typical secretory proteins, are transported from the endoplasmic reticulum to the Golgi in enlarged COPII-coated vesicles. Subsequent steps in processing and secretion await further clarification and could vary depending on collagen type. Ultimately, the N- and C-terminal propeptide domains are excised by specific proteases, and the resulting collagen triple helices aggregate into fibrils that associate with growing fibers in the ECM.

Received: May 23, 2018
Published: July 9, 2018
Abnormalities in the structure and metabolism of collagen are associated with a variety of heritable diseases. For example, the rare genetic disorders osteogenesis imperfecta (OI) and Ehlers–Danlos syndrome (EDS) involve mutations that decrease the production or structural stability of fibrillar collagens in a tissue. Diseases within the OI and EDS spectra are most commonly associated with single-base substitutions that lead to the substitution of an obligate glycine with another residue.

As catalysts of a key step in collagen biosynthesis, CP4H has been a drug target since the discovery of the enzyme itself. Of primary interest are diseases in which normal collagen is overproduced or produced in a manner that exacerbates a pathological process. The disruption of collagen biosynthesis could be especially beneficial in the treatment of such fibrotic diseases. Recent links between CP4Hs and cancer metastasis suggest new opportunities. For example, the remodeling of collagen in the ECM near tumors has been linked to cancer metastasis. Moreover, high levels of two CP4H isoforms correlate well with cancer patient mortality, and an inhibitor of CP4H activity decreases cancer metastasis in mice. The mechanism-of-action could entail normalization of the vasculature and decompression of blood vessels.

CP4Hs are susceptible to inhibition by a variety of chemical strategies. Here, we begin by reviewing the structure and enzymatic reaction mechanism of CP4H. We then discuss the metal ions, metal chelators, mimetics of the cosubstrate AKG, mimetics of collagen strands, and natural products that are...
known to inhibit CP4H, focusing on inhibition strategies of therapeutic value and those that have been tested in live cells. Lastly, we discuss the state-of-the-art for inhibition of CP4Hs in vivo.

### CP4H STRUCTURE

In mammals, CP4Hs exist as $\alpha_2\beta_2$ tetramers. The $\alpha$-subunit is the catalytic subunit and the $\beta$-subunit is protein disulfide isomerase, which is a multifunctional chaperone protein that maintains the $\alpha$-subunit in a soluble and active conformation. Three isoforms of the $\alpha$-subunit, $\alpha(I)$, $\alpha(II)$, and $\alpha(III)$, have been identified in humans. All $\alpha$-subunit isoforms form tetramers with the $\beta$-subunit, which we refer to herein as the CP4H1, CP4H2, and CP4H3 holoenzymes. The structure of the tetrameric complex is unknown, although low-resolution small-angle X-ray scattering data together with models derived from homologous structures (like those in Figure 3) and bioinformatic analyses suggest an elongated assembly with a $\beta\alpha\alpha\beta$ architecture.

### CP4H MECHANISM

As in other members of the Fe(II)- and AKG-dependent dioxygenase (FAKGD) superfamily, CP4H couples the oxidative decarboxylation of AKG to the hydroxylation of a hydrocarbon substrate via a radical mechanism involving a highly reactive Fe(IV)=O species (ferryl ion). The ferryl ion abstracts a hydrogen atom from a hydrocarbon substrate and effects hydroxylation via a radical rebound process (“Coupled Reaction” in Figure 4). Although human CP4Hs have been the object of insightful analyses, much of what we...
understand about CP4Hs comes from studies of the homologue from chicken and of other FAKGD enzymes as well as theoretical considerations. Classical kinetic studies of chicken CP4H1 have suggested that catalysis proceeds by an ordered mechanism in which AKG binds to the CP4H·Fe(II) complex, followed by a pro-collagen strand and O₂. The peptide bond of the Pro residue that is bound in the enzymic active site is likely in the cis (that is, E) conformation. After formation of the ferryl ion and subsequent hydroxylation of the Pro residue, the enzyme releases the nascent Hyp-containing strand and succinate.

The ferryl ion appears to decay via an unknown mechanism to an inactive Fe(III) state. This aberrant pathway is linked to scurvy. Reduced vitamin C (L-ascorbic acid) can reactivate the enzyme. During catalysis by the typical mechanism, there is no requirement for vitamin C, as the radical rebound process returns the iron center to its Fe(II) state. For example, chicken CP4H catalyzes prolyl hydroxylation at a maximal rate for at least a few cycles in the complete absence of vitamin C. Yet, even at saturating concentrations of the peptide substrate, CP4Hs (and other FAKGDs) have been observed to decarboxylate AKG oxidatively in a manner that does not result in the hydroxylation of the peptide substrate. Some ligands for the iron center enhance the rate of vitamin C-dependent reactivation.

### ALTERNATIVE METAL IONS

Most FAKGD enzymes have only modest affinity for Fe(II). Accordingly, other transition metal ions, including Mn(II), Co(II), Ni(II), Cu(II), and Zn(II), are modest inhibitors of these enzymes. CP4Hs have weak affinity for Fe(II) (e.g., apparent Kᵢ = 4 μM for human CP4H1 and 6 μM for human CP4H2) and are particularly susceptible to inhibition in this manner (e.g., Kᵢ = 0.6 μM for inhibition by Zn(II), which has an ionic radius similar to that of Fe(II)). Transition metal ions are, however, unlikely to be useful as CP4H inhibitors in a therapeutic context.

### METAL CHELATORS

As with metal ions, virtually all FAKGD enzymes are inhibited by metal chelators (Figure 5A). These inhibitors include simple chelators such as ethylenediaminetetraacetic acid (EDTA), deferroxamine, 2,2′-bipyridine (bipy), 1,10-phenanthroline, and catechol. Simple metal chelators bind to and sequester free iron and typically display Kᵢ values in the micromolar range depending on the amount of free iron added to assays in vitro. Indeed, bipy itself was used early on to disrupt collagen biosynthesis in a biological context, providing initial evidence that iron was required for the process. Although metal affinity is an important component of some CP4H inhibitors (vide infra), the simple metal chelators used clinically to treat conditions such as iron overload and heavy metal toxicity are too nonspecific to be of therapeutic value as inhibitors of CP4Hs. Nevertheless, functionalization or modification of chelators can generate more potent inhibitors of the enzyme. For example, 2,2′-bipyridine-5,5′-
bicarboxylic acid (bipy55′DC) is one of the most potent known inhibitors of a CP4H ($K_i = 0.14 \mu M$). Yet, the high residual iron affinity of bipy55′DC is a concern and complicates the characterization of the inhibitory mechanism of this and other bipy derivatives.40,45,46

**AKG MIMETICS**

The putative mechanism of catalysis suggests that analogues of AKG that cannot undergo decarboxylation could inhibit catalysis by CP4Hs.30 Moreover, the mechanism explains why AKG mimics that inhibit CP4Hs do not inhibit the essential enzyme AKG dehydrogenase, which does not require a metal for its catalytic activity.30 Soon after the enzymatic reaction mechanism was put forth,30 a variety of AKG-mimetic inhibitors of chicken CP4H were reported, including N-oxalylglycine (NOG), pyridine-2,4-dicarboxylic acid (24PDC), pyridine-2,5-dicarboxylic acid (25PDC),39 and 3,4-dihydroxybenzoic acid (DHB) (Figure 5B).48 Like AKG ($K_M = 20 \mu M$), these compounds possess a metal-chelating moiety and thus resemble the structure of AKG in the context of the enzymatic reaction mechanism. All inhibit human CP4Hs in vitro.17 Whereas 24PDC, 25PDC, and NOG are simple competitive inhibitors with respect to AKG,48 DHB is competitive with both AKG and vitamin C.49

**Pyridinedicarboxylic Acids.** Unlike 24PDC, which inhibits many members of the FAKGD with similar low micromolar potency,24 25PDC is selective for CP4Hs. Thus, considerable effort has been made to explore 25PDC derivatives. In particular, a variety of substitutions have been made at both the 2-position50 and the 5-position51 of the pyridine ring (Figure 5B). Regrettably, all of the substitutions for the 2-carboxyl group (i.e., N-acyl sulfonamides, tetrazole,
hydroxamic acids, 2-imidazolyl, 2-pyrrolyl, and others) were made in the absence of the $S$-carboxyl group, thereby obfuscating comparisons. In general, though, modification of the 2-position in the absence of the $S$-position carboxyl group did not yield potent inhibitors. For example, many inhibitors were more potent than the parent pyridine-2-carboxylic acid, but most still exhibited IC$_{50}$ values in the midmicromolar range and have not been developed further.

In comparison to 2-position modifications, those at the $S$-position have been more fruitful. Substitution of the $S$-carboxyl group in $N$-acyl sulfonamide derivatives yielded 25PDC analogues with similar or improved potency compared to the 25PDC parent. Most of these compounds inhibited with IC$_{50}$ values in the low micromolar range, with phenyl and 4-methoxyphenyl sulfonamides (Figure 5B) displaying the highest potency (IC$_{50}$ values of 1.1 and 1.0 μM, respectively, compared to 5.5 μM for 25PDC). Yet, considering the binding mode of AKG, it was highly surprising that substitution of the $S$-carboxyl group with neutral derivatives was tolerable in inhibitors. This finding raised the question of whether or not 25PDC and these analogues bound to the active site in the same manner as did AKG itself. No clinical inhibitors based upon these sulfonamide derivatives of 25PDC have been reported to date.

**N-Oxalyl Glycine and Its Derivatives.** NOG does not display the selectivity properties of 25PDC and is known to inhibit AKG dehydrogenase and other enzymes that utilize AKG. Nevertheless, the NOG scaffold has provided a means to evaluate stereochirality in the context of the AKG binding pocket. In contrast to the 25PDC scaffold, substitutions to the $\omega$-carboxyl group of the glycine moiety abolished inhibition, suggesting that NOG likely binds in the conventional AKG binding mode. Still, the $\alpha$ carbon of the glycine moiety has a tetrahedral geometry, so substitutions were made to evaluate the nature of the binding pocket on either side of the AKG substrate.

Upon substitution of the glycine moiety with a variety of other amino acids (e.g., alanine, phenylalanine, leucine, methionine, proline, and sarcosine), the only tolerated substitution was that of alanine, which diminished potency (IC$_{50}$ = 90.7 μM for racemic N-oxalyl alanine (NOA), IC$_{50}$ = 2.89 μM for NOG). A striking difference was observed between the $2R$ and $2S$ enantiomers, with ($S$)-NOA (IC$_{50}$ = 38.2 μM) (Figure 5B) being substantially more potent than ($R$)-NOA (IC$_{50}$ = 621 μM).

Attempts have been made to fuse elements of the N-oxalylglycine and pyridinedicarboxylic acid scaffolds. For example, amides of glycine with 4-hydroxycinnamyl 3-carboxylic acid or 4-hydroxyisouquinoline 3-carboxylic acid were reported to be potent inhibitors of chicken CP4H in vitro (IC$_{50}$ = 0.02 μM and IC$_{50}$ = 0.19 μM, respectively). As with NOA, the $2S$ enantiomer of alanine was preferred in place of glycine, which is consistent with the binding of these alkaloid-like compounds in the enzymic active site. A 7-bromo derivative of the isouquinolylglycine was reported to be especially effective at inhibiting the enzyme in chicken cells and live rats, perhaps because of protection against metabolic inactivation afforded by a 7-bromo group (Figure 5B). This isouquinolylglycine has, however, not been developed further, perhaps because of intrinsic cytotoxicity.

**COLLAGEN MIMETIC PEPTIDES**

CP4H inhibitors based upon the structure of collagen strands could offer enhanced selectivity compared to that of AKG mimetics. Whereas many features of the reaction catalyzed by CP4Hs are well characterized, a detailed understanding of the manner in which the protocollagen substrate is recognized by these CP4Hs remains incomplete, partially due to the limited structural information available for mammalian CP4Hs.

Although there is still not a clear understanding of the peptide substrate, a few individual examples of peptide-based inhibitors have been reported. For example, CP4Hs are known to be inhibited by polyproline strands in the PP-II conformation, which have been observed to bind in the PSB domain of human CP4H-1. Moreover, some simple analogues of the collagen-repeat sequence in which the composition is modified or rearranged inhibit the enzyme from chicken. For example, peptides of the sequence (Pro-Gly)$_n$ (Gly-Pro-Gly)$_n$ and (Pro-Pro-Gly-Gly)$_n$ have all been observed to inhibit chicken CP4H, although (Pro-Pro-Gly-Gly)$_n$ has also been observed to be a poor substrate. Moreover, short collagen-like polymers with triplet repeats containing Pro derivatives with methyl or fluoro groups at the 3- or 4-position serve as inhibitors of chicken CP4H. A small tetrapeptide containing the sequence Gly-Pro-flp-Gly (where flp refers to a (2S,4S)-4-fluoroproline residue) was found to be a substrate for human CP4H-1, furnishing a product containing (2S)-4-ketoproline. This product results from the decomposition of the fluorohydron that is generated by enzymatic catalysis. Thus, a Gly-Pro-flp-Gly substrate can serve as the basis for a continuous assay of enzymatic activity based on the detection of fluoride ion with a selective electrode. Moreover, the orthogonal reactivity of the ensuing keto group could serve as the basis for a high-throughput assay.

**IRREVERSIBLE INHIBITORS**

Most known inhibitors of CP4Hs are reversible. Nonetheless, a few inhibitors are time-dependent and irreversible. These molecules (Figure 5C) include certain ancistracyclines, peptides containing (2S)-5-oxaproline (OPr), and electrophilic analogues of AKG.

Among ancistracyclines, doxorubicin and daunorubicin are irreversible inhibitors of chicken CP4H. They appeared to function via a mechanism that requires Fe(II) in solution but that is competitive with vitamin C, with possible alkylation of the enzymic active site via a postulated radical intermediate. Dxorubicin is used clinically as a chemotherapeutic agent for the treatment of a variety of types of cancer. One of the common side effects of doxorubicin treatment is poor wound healing, which is consistent with the notion that a pleiotropic effect of doxorubicin is inactivation of CP4Hs (as has been observed in human skin fibroblasts).

Tetrapeptides containing OPr (Figure 5C) are irreversible inhibitors of chicken CP4H. The chemical mechanism of this inactivation is not known but requires Fe(II) and AKG and is diminished by high concentrations of (Pro-Pro-Gly)$_{10}$ or vitamin C. Peptides containing OPr were even found to inhibit CP4H activity in cultured fibroblasts, although much less effectively than the inhibition observed in vitro.

Finally, two electrophilic analogues of AKG inactivate CP4H. One is coumalic acid (Figure 5C), which resembles AKG in its chain length and placement of a pendant carboxyl group but appears to alkylate the enzyme irreversibly by a postulated nucleophilic attack on the coumarin ring. AKG protects against the inactivation, suggesting that the alkylation likely occurred in or near the enzymic active site. The second is racemic 4-oxo-5,6-epoxyhexanoic acid, in which a carboxy...
group of AKG is replaced with an epoxy group that could be activated for alkylation upon chelation to the active-site Fe(II). A nematode, *Caenorhabditis elegans*, exposed to the esterified compound displays the phenotype of animals lacking CP4H.

## INHIBITION OF CP4HS IN VIVO

In addition to some compounds described above, a few other CP4H inhibitors have been tested in vivo. Few, however, have proven to be sufficiently safe or efficacious to move into clinical studies. HOE-077 (Figure 5D), which is a prodrug of 24PDC, has been explored in preclinical studies in rats and dogs and was found to inhibit collagen synthesis in the liver. HOE-077 was, however, found to yield many metabolites and was found to disrupt collagen synthesis by inactivating cells involved in producing collagen rather than by direct inhibition of CP4H. FG-0041 (Figure 5D), which is a derivative of 1,10-phenanthroline, was shown to aid in the recovery of left ventricular function in rats that have suffered post myocardial infarction. This compound likely functions, however, by inhibiting hypoxia inducible factor prolyl hydroxylases, rather than CP4Hs.

Arguably, the state-of-the-art compound for inhibiting CP4Hs in vivo is ethyl 3,4-dihydroxybenzoate (EDHB) (Figure 5D). EDHB is a prodrug of DHB and is commonly used in cell culture experiments. Moreover, EDHB has even been shown to be safe enough to use in mouse models of cancer metastasis. A dose of 40 mg/kg per day in mice was found to decrease primary tumor growth, reduce the collagen content of the primary tumor, and significantly reduce metastasis to the lung, all without affecting overall body weight. Like other catechols, however, EDHB perturbs iron metabolism and affects redox homeostasis. Thus, EDHB is unlikely to be sufficiently selective for the therapeutic applications of interest here.

A recent screen of biheteroaryl compounds revealed that 2-(5-carboxythiazol-2-yl)pyridine-5-carboxylic acid (pythiDC), in which one pyridyl group of bipy55′ (5-carboxythiazol-2-yl)pyridine-5-carboxylic acid (pythiDC), was found to aid in the recovery of left ventricular function in rats that have suffered post myocardial infarction. This compound likely functions, however, by inhibiting hypoxia inducible factor prolyl hydroxylases, rather than CP4Hs.

## ACKNOWLEDGMENTS

We are grateful to M. D. Shoulders and I. C. Tanrikulu for comments on the manuscript. J.D.V. was supported by Molecular Biosciences Training Grant T32 GM007215 [National Institutes of Health (NIH)] and a fellowship from the Department of Biochemistry at the University of Wisconsin—Madison. Work on collagen prolyl 4-hydroxylase in the Raines laboratory is supported by grant R01 AR044276 (NIH).

## ABBREVIATIONS USED

AKG, α-ketoglutarate; CP4H, collagen prolyl 4-hydroxylase; FAKGD, Fe(II)- and AKG-dependent dioxygenase

## REFERENCES

for the catalytic mechanism of prolyl hydroxylase: Applicability to oxygenases and related enzymes. 


(20) Vasta, J. D.; Raines, R. T. Human collagen prolyl 4-hydroxylase is activated by ligands for its iron center. Biochemistry 2016, 55, 3224−3233.


(33) Wall, K. D.; Choudhary, A.; Jensen, K. H.; McGrath, N. A.; Rains, R. T. Prolyl 4-hydroxylase: Substrate isosteres in which an (E)- or (Z)-alkene replaces the prolyl peptide bond. Biochemistry 2017, 56, 219−227.


