Collagen triple helices are stabilized by 4-hydroxyproline residues. No function is known for the much less common 3-hydroxyproline (3Hyp), although genetic defects inhibiting its formation cause recessive osteogenesis imperfecta. To help understand the pathogenesis, we used mass spectrometry to identify the sites and local sequence motifs of 3Hyp residues in fibril-forming collagens from normal human and bovine tissues. The results confirm a single, essentially fully occupied 3Hyp site (A1) at Pro^986 in A-clade chains α1(I), α1(II), and α2(V). Two partially modified sites (A2 and A3) were found at Pro^944 in α1(II) and α2(V) and Pro^707 in α2(I) and α2(V), which differed from A1 in sequence motif. Significantly, the distance between sites 2 and 3, 237 residues, is close to the collagen D-period (234 residues). A search for additional D-periodic 3Hyp sites revealed a fourth site (A4) at Pro^479 in α2(V), 237 residues N-terminal to site 3. In contrast, human and bovine type III collagen contained no 3Hyp at any site, despite a candidate proline residue and recognizable A1 sequence motif. A conserved histidine in mammalian α1(III) at A1 may have prevented 3-hydroxylation because this site in chicken type III was fully hydroxylated, and tyrosine replaced histidine. All three B-clade type V/XI collagen chains revealed the same three sites of 3Hyp but at different loci and sequence contexts from those in A-clade collagen chains. Two of these B-clade sites were spaced apart by 231 residues. From these and other observations we propose a fundamental role for 3Hyp residues in the ordered self-assembly of collagen supramolecular structures.

Collagens are the most abundant and ubiquitous proteins in multi-cellular animals. It is well established that 4-hydroxyproline (4Hyp) residues stabilize the collagen triple helix through water-bridged intramolecular hydrogen bonding (1). However, the function of the much less abundant 3-hydroxyproline (3Hyp), although discovered 50 years ago, is unknown (2). Only 1–2 residues of 3Hyp occur per chain in collagen types I and II and 3–6 residues occur per chain in collagen types V and XI. The content is highest in type IV collagens of basement membranes in which 10% of the total hydroxyproline can be 3Hyp (3).

Specific prolyl 3-hydroxylases (P3Hs) are responsible for 3Hyp synthesis. Three different genes encoding P3H1, P3H2, and P3H3 are present in the human genome, which show tissue specificity in their expression (4, 5). Substrate proline residues occur in a prerequisite sequence -Pro-4Hyp-Gly. The α1(I) chain has only one established 3Hyp site at Pro^986 in a motif conserved across vertebrate species (human GLPGPIGPPGPR) a close variant of which also occurs in type II collagen (human GIPGPPIGPPGPR).

Renewed interest in 3Hyp was recently sparked by the discovery that a recessive form of osteogenesis imperfecta (OI) is caused by mutations in CRTAP. This gene encodes a protein (cartilage-associated protein) that is bound to P3H1 and cyclophilin B in the endoplasmic reticulum and is required for prolyl 3-hydroxylation at the Pro^986 site in collagen α1(I) and α1(II) chains (6, 7). Further studies showed that mutations in P3H1 itself also caused recessive, severe OI (8, 9). A key question is whether the brittle bone phenotype in OI is caused by the absence of 3Hyp in bone matrix collagen or an intracellular assembly and transport defect caused by the malfunctioning enzyme complex or both.

Because little is known about the distribution of 3Hyp in normal fibril-forming collagens beyond the single Pro^986 site in α1(I) and α1(II) chains, we used protein mass spectrometry to locate further sites in all A-clade and B-clade gene products used in vertebrate collagen fibril formation. Collagen type I fibrils are assembled on a filamentous template of collagen type V, and collagen type II fibrils are assembled on a template of collagen type XI (10). To provide a basis for understanding the overall post-translational effects of mutations in CRTAP, LEPRE1 (encodes P3H1), and other genes involved in collagen prolyl-3-hydroxylation, it is important to identify all of the sites of prolyl-3-hydroxylation in normal collagens from human and other vertebrate tissues.

Our results reveal several partially hydroxylated sites of 3Hyp in the various fibrillar collagen chains, in addition to the usually fully hydroxylated primary site at Pro^986 in α1(I), α1(II), and α2(V). All of the additional sites lack the distinctive sequence motif of Pro^986 but share common features with known 3Hyp-containing sequences in type IV collagen. One important finding is the D-periodic spacing between sites A2 and A3 and between sites A3 (α2(I) and α2(V)) and A4 (α2(V)) of A-clade chains (α1(II), α2(V), and α2(I)) and between sites B2 and B3 of B-clade chains (α1(V), α1(XI), and α2(XI)). In contrast, mammalian type III collagen lacks any 3Hyp despite having a recog-
nizable primary site motif at Pro\(^{986}\). From the conserved sites, sequence motifs, and spacing of 3Hyp sites along the collagen chains, we speculate a role for 3Hyp in mediating inter-triple-helical interactions and in aiding the supramolecular assembly of collagen.

**EXPERIMENTAL PROCEDURES**

**Source of Tissues**—Adult human bone, cartilage, and meniscus (20–40 years old) were purchased from the Northwest Tissue Center (Seattle, WA). Fetal human bone and cartilage were obtained from the Birth Defects Research Laboratory of the University of Washington with Internal Review Board (IRB) approval. Human intervertebral disc tissue was obtained from normally discarded surgical tissue with patient-informed consent and IRB approval. Chicken skin (12–14 weeks old) was dissected from chicken wings purchased at a local supermarket. Bovine vitreous was dissected from adult steer eyes (18 months old) obtained from a local abattoir.

**Preparation of Collagens**—Types I and V collagens were prepared from adult and fetal human bone. Powdered bone was defatted at 4 °C in methanol/chloroform (1/3 v/v) and demineralized at 4 °C in 0.5 mM EDTA, 0.05 mM Tris-HCl, pH 7.5. Type III collagen was prepared from defatted chicken skin and adult human meniscus. Type II collagen was prepared from human adult articular cartilage, fetal epiphyseal cartilage, and adult nucleus pulposus, and bovine meniscus and vitreous (18 months old). Type XI collagen was prepared from fetal human articular cartilage. Proteoglycans were removed from cartilaginous tissues with 4 M guanidine HCl, 0.05 M Tris-HCl, pH 7.5, with protease inhibitors (5 mM 1,10-phenanthroline and 2 mM phenylmethylsulfonyl fluoride) for 24 h at 4 °C, and the residue was washed thoroughly.

Collagens from all of the tissues were solubilized with pepsin (1:20 w/w, pepsin/dry tissue) in 3% acetic acid for 24 h at 4 °C (11). Serial precipitations of solubilized bone collagen with 0.7 and 1.8 M NaCl separated types I and V collagens, respectively. Serial precipitations of solubilized articular cartilage, nucleus pulposus, and vitreous collagens with 0.8 and 1.2 M NaCl separated type II and type XI collagens, respectively. Skin type III collagen was precipitated at 0.8 M NaCl. Meniscus collagens were serially precipitated with 0.7, 0.9, and 1.2 M NaCl to separate types I/III, type II, and types V/XI, respectively. Collagen type II is a minor component of the meniscus and is highly modified post-translationally, causing it to precipitate at 0.9 M NaCl, separated from the bulk type I collagen (12). Portions of demineralized bone and guanidine HCl-extracted cartilage residue were digested with CNBr in 70% formic acid at room temperature for 24 h (13), and the resulting CB peptides were freeze-dried. For microsequence analysis, \(\alpha_1(II)\) CB9,7 was prepared from bovine nucleus pulposus and digested with trypsin, and individual peptides were resolved by reverse phase HPLC.

**SDS-PAGE**—The method of Laemmli (14) was used with 6% gels for pepsinized collagen and 12.5% gels for CNBr peptides.

**Microsequence Analysis**—N-terminal sequence analysis was carried out by Edman chemistry on a Portion 2090E machine equipped with on-line HPLC analysis of cleaved phenylthiohydantoin amino acids.

**RESULTS**

**3Hyp in Gene A-clade Collagen Chains \(\alpha_1(I), \alpha_1(II), \alpha_2(I),\) and \(\alpha_2(V)\)**—Peptides from the known fully occupied site of 3Hyp at Pro\(^{986}\) in \(\alpha_1(I),\) \(\alpha_1(II),\) and \(\alpha_2(V)\) revealed close to 100% hydroxylation by tryptic peptide mass spectrometry (Fig. 1). Early studies had established by Edman sequencing that this proline residue in \(\alpha_1(I)\) was 3Hyp (18).

In addition, tryptic peptides prepared from individual CB peptides and whole \(\alpha\)-chains on in-gel trypsin digestion were surveyed for mass variants (+16 Da) indicating 3Hyp at other GPP sites. From \(\alpha_1(II),\) a second partially hydroxylated site was identified at Pro\(^{944}\) in CB peptide, CB9,7, from human and bovine articular cartilage (Fig. 2; results for bovine shown). Table 1 lists fragment y and b ions used to interpret the MS/MS spectra from Fig. 2. This will serve also as a guide in interpreting the results presented in all the spectra. In addition, 3Hyp at Pro\(^{944}\) from bovine articular \(\alpha_1(II)\) prompted a survey of other tissue sources of type II collagen. Consistent species- and tissue-dependent variations at this site were revealed. These analytical results are summarized in Fig. 2. The degree of Pro\(^{944}\) 3-hydroxylation estimated from the mass ratios (and site of the +16 addition established by MS/MS fragmentation profile) ranged consistently from more than 80% in bovine vitreous type II collagen to less than 20% in bovine articular type II collagen.
FIGURE 1. Tandem mass spectra of tryptic peptides containing Pro<sup>986</sup> (site A1), the single, fully occupied 3Hyp site in collagen chains α1(I), α1(II), and α2(IV). A, peptides were prepared by in-gel trypsin digestion after SDS-PAGE of CNBr digests of human bone collagen (lane 1), human articular cartilage collagen (lane 2), and pepsin-solubilized human bone collagen, 1.2 M NaCl precipitate (lane 3). B, full scan mass spectra from the tryptic peptide LC-MS profiles of α1(I)CB6 and α1(II)CB9,7 across the elution window of the post-translational variants of the tryptic peptide containing Pro<sup>986</sup>. Any unhydroxylated peptides would be included so this provides a measure of hydroxylation of Pro<sup>986</sup>. This site is fully 3-hydroxylated in type I collagen of bone and type II collagen of cartilage. C, full scan mass spectrum from the LC-MS profile of the α2(IV) chain over the elution window of the tryptic peptide containing Pro<sup>986</sup> (upper) and MS/MS analysis of ion 773.72 (lower). The y<sub>10</sub> fragment ion establishes the added 16 Da on Pro<sup>986</sup>. The 773.72<sup>+</sup> peptide ion lacks the 4-hydroxylation at Pro<sup>978</sup>, whereas the 782.72<sup>+</sup> ion has both 3Hyp at Pro<sup>986</sup> and 4Hyp at Pro<sup>978</sup>; α2(IV) is consistently under-hydroxylated at Pro<sup>978</sup>. (The 782.72<sup>+</sup> ion is not a contaminant from α1(I).) P<sup>+</sup>, 3Hyp; P*<sup>+</sup>, 4Hyp.

FIGURE 2. Tandem mass spectra of an α1(II) trypsin peptide revealing a secondary, variably hydroxylated 3Hyp site at Pro<sup>944</sup> (site A2). A, upper four panels are full scan mass spectra from LC-MS profiles of in-gel trypsin digests of α1(II)CB9,7 from cartilage (bovine), nucleus pulposus (human), meniscus (bovine), and vitreous (bovine) over the elution window of hydroxylated and prolyl versions of the Pro<sup>944</sup>-containing peptide. The relative abundance of the ions shown provides an index of the degree of hydroxylation at Pro<sup>944</sup>. B, the bottom two panels show MS/MS spectral analyses of the prolyl and suspected 3-hydroxyprolyl versions of the peptide, from which the y ion ladder establishes the position of the added 16 Da on Pro<sup>944</sup>. Hydroxylation ranged from 10% for hyaline cartilage to 87% for vitreous type II collagen with intermediate values for intervertebral disc and meniscus collagens. See Table 1 for a guide to how fragment ions establish the sequence and position of the 3Hyp residue. P<sup>+</sup>, 3Hyp; P*<sup>+</sup>, 4Hyp.
TABLE 1
Guide to interpretation of mass spectrometric peptide sequencing (data from Fig. 2)

| Fragment | Y6 | Y7 | Y8 | Y9 | Y10 | Y11 | Y12 | Y13 | Y14 | Y15 |
|----------|----|----|----|----|-----|-----|-----|-----|-----|-----|
| parent ions | 570.3 | 759.4 | 1138.5 | 1311.5 | 1664.6 | 1761.7 | 1931.7 |
| C-term ions | 570.3 | 759.4 | 1138.5 | 1311.5 | 1664.6 | 1761.7 | 1931.7 |
| N-term ions | 458.3 | 756.3 | 756.3 | 1567.7 | 1583.5 |

* Threonine often shows a neutral water loss upon ms/ms fragmentation. Ions in bold carry the extra hydroxyl (16Da).

The results show that 4Hyp occurs only in the Y position of the (GXY)ₙ repeat of collagens, so hydroxylated proline at X is strong evidence in itself for 3Hyp. To rule out 4Hyp (because mass spectrometric results alone cannot distinguish 3Hyp from 4Hyp), Edman microsequencing was applied to the isolated tryptic peptide containing hydroxylated site A2 from bovine α2(I). The results are shown in Fig. 3 for the fully 3-hydroxylated peptide isolated from calf nucleus pulposus α2(I). At cycle 11, the phenylthiohydantoin-derivative reverse phase HPLC profile is similar to that reported as characteristic of 3Hyp phenylthiohydantoin degradation products (19) and quite distinct from that given by 4Hyp (see cycle 12).

Manual evaluation of the mass spectra of all tryptic peptides from the α2(I) chain of bone type I collagen (Fig. 4) revealed a third site of 3Hyp at Pro⁷⁰⁷ (site A3). (No candidate -GPP- or sequence motif was recognizable where site A1 or A2 would be in α2(I) of human or other vertebrate species we examined (Ensemble entry: ENSG0000164692).) Residue Pro⁷⁰⁷ was 80% hydroxylated in human α2(I) (similar in bovine). Screening human α2(V) similarly, the Pro⁷⁰⁷ locus was also hydroxylated to a similar extent. Thus in α2(V) all three sites, A1, A2, and A3, were hydroxylated.

The near D-periodic spacing between Pro⁷⁰⁷ and Pro⁹⁴⁴ (237 residues versus D = 234) (20) prompted us to search by tandem mass spectrometry for any further 3Hyp site spaced by one or two D-periods more N-terminal. Fig. 5 shows the results of analysis of a tryptic peptide containing a candidate proline at Pro⁷⁰⁷ (site A4) from α2(V) of bone. The residue was indeed partially hydroxylated as confirmed by MS/MS fragment analysis of the 3Hyp and Pro versions of the peptide. Additional analyses showed evidence of variable levels of 3-hydroxylation of a proline in the equivalent tryptic peptide from the α1(I) chain, but only from cell culture so the biological significance for α1(I) at present is unclear (results not shown).

Collagen. The small pool of type II collagen from fibrocartilaginous meniscus (3–6% of total collagen; Ref. 12) was highly 3-hydroxylated (66%) at Pro⁹⁴⁴. Similarly, nucleus pulposus type II collagen (results for human shown in Fig. 2, but also bovine) was also heavily hydroxylated (~40%) at Pro⁹⁴⁴ (Fig. 2; see also Fig. 8). The Pro⁹⁴⁴ site in human α2(V) from bone was 60% 3-hydroxylated (mass spectrometry results not shown).

We know that 4Hyp occurs only in the Y position of the (GXY)ₙ repeat of collagens, so hydroxylated proline at X is strong evidence in itself for 3Hyp. To rule out 4Hyp (because mass spectrometric results alone cannot distinguish 3Hyp from 4Hyp), Edman microsequencing was applied to the isolated tryptic peptide containing hydroxylated site A2 from bovine α2(I). The results are shown in Fig. 3 for the fully 3-hydroxylated peptide isolated from calf nucleus pulposus α2(I). At cycle 11, the phenylthiohydantoin-derivative reverse phase HPLC profile is similar to that reported as characteristic of 3Hyp phenylthiohydantoin degradation products (19) and quite distinct from that given by 4Hyp (see cycle 12).

Manual evaluation of the mass spectra of all tryptic peptides from the α2(I) chain of bone type I collagen (Fig. 4) revealed a third site of 3Hyp at Pro⁷⁰⁷ (site A3). (No candidate -GPP- or sequence motif was recognizable where site A1 or A2 would be in α2(I) of human or other vertebrate species we examined (Ensemble entry: ENSG0000164692).) Residue Pro⁷⁰⁷ was 80% hydroxylated in human α2(I) (similar in bovine). Screening human α2(V) similarly, the Pro⁷⁰⁷ locus was also hydroxylated to a similar extent. Thus in α2(V) all three sites, A1, A2, and A3, were hydroxylated.

The near D-periodic spacing between Pro⁷⁰⁷ and Pro⁹⁴⁴ (237 residues versus D = 234) (20) prompted us to search by tandem mass spectrometry for any further 3Hyp site spaced by one or two D-periods more N-terminal. Fig. 5 shows the results of analysis of a tryptic peptide containing a candidate proline at Pro⁷⁰⁷ (site A4) from α2(V) of bone. The residue was indeed partially hydroxylated as confirmed by MS/MS fragment analysis of the 3Hyp and Pro versions of the peptide. Additional analyses showed evidence of variable levels of 3-hydroxylation of a proline in the equivalent tryptic peptide from the α1(I) chain, but only from cell culture so the biological significance for α1(I) at present is unclear (results not shown).

Lack of 3Hyp in Mammalian Type III Collagen—The protein sequence of human and other mammalian type III collagens (Ensemble entry: ENSG0000168542) shows a recognizable motif and GPP at the primary site Pro⁹⁸⁶. Mass spectrometry, however, showed peptides of the mass of the proline form but none for the 3Hyp form from bovine and human collagen III prepared from skin, aorta, and other tissues (Fig. 6 shows results from human meniscus α1(III)). In comparing the genomic data base (Ensemble) for all available COL3A1 sequencxs, all had GHx in place of GLP or GIP the triplet before GPIGPP, predicting a lack of substrate recognition by prolyl 3-hydroxylase (see Fig. 6 for sample sequences). On inspecting a broader range of vertebrate COL3A1 sequences, chicken stood out with GYP not GHP (Fig. 6). To see whether this enabled 3Hyp formation in the neighboring GPP in vivo, collagen III was purified from chicken skin and analyzed by mass spectrometry. As shown in Fig. 6, the candidate tryptic peptide from chicken α1(III) was 100% hydroxylated at the homologous locus Pro⁹⁸⁹. It appears, therefore, that a hydrophobic residue is required at residue 980 (Ile, Leu, or Tyr) or at least not a histidine, for the P3H complex to recognize the prolyl 3-hydroxylase (see Fig. 6 for sample sequences). On looking at the genomic data base (Ensemble) for all available COL3A1 sequences, all had GHx in place of GLP or GIP the triplet before GPIGPP, predicting a lack of substrate recognition by prolyl 3-hydroxylase (see Fig. 6 for sample sequences). On inspecting a broader range of vertebrate COL3A1 sequences, chicken stood out with GYP not GHP (Fig. 6). To see whether this enabled 3Hyp formation in the neighboring GPP in vivo, collagen III was purified from chicken skin and analyzed by mass spectrometry. As shown in Fig. 6, the candidate tryptic peptide from chicken α1(III) was 100% hydroxylated at the homologous locus Pro⁹⁸⁹. It appears, therefore, that a hydrophobic residue is required at residue 980 (Ile, Leu, or Tyr) or at least not a histidine, for the P3H complex to recognize the prolyl 3-hydroxylase (see Fig. 6 for sample sequences).
Pro434, Pro665, and Pro692 (Fig. 7) that we refer to here as sites B3, B2, and B1 (consistent with the right to left order used along clade A chains). Sites B1 and B2 are in the same tryptic peptide 27 residues apart (Fig. 7). The results are shown only for /H92511(V).

The MS/MS fragmentation patterns of the 3/H11001 parent ions established the specific locations of the added hydroxyl groups. The /H92511(XI) and /H92512(XI) equivalent peptides from a cartilage type XI collagen preparation showed a similar degree of hydroxylation at these two sites. In Fig. 6B, site B3 (Pro434) mass spectral results are shown for /H92511(XI) from cartilage, but again /H92511(V) and /H92512(XI) gave very similar levels of 3-hydroxylation at this site.

Fig. 8 summarizes the molecular locations and local sequence motifs for all of the 3Hyp sites identified in the /H92511(I), /H92512(I), /H92511(II), /H92511(III), /H92511(V), /H92512(V), /H92511(XI), and /H92512(XI) chains. It is possible that other GPP sites may be 3-hydroxylated, particularly in the type V/XI collagen B-clade chains, because not all GPP-containing sequences gave informative tryptic peptides. But the early literature reporting amino acid compositions of isolated chains and derived cyanogen bromide peptides is consistent with one residue/chain at the single site in /H92511(I) and /H92512(I) (21–23), one or two residues in /H92511(II) (24), no 3Hyp in /H92511(III) (25), three or four residues in /H92511(V) (26, 27), and two or three residues in /H92512(V) (26). The present results show some evidence for clustering, for example 3Hyp sites B1 and B2 spaced 27 residues apart. Also the underlined proline in the GP#P*GPP* sequence (where P# indicates 3Hyp and P* indicates 4Hyp) at site B3 (Fig. 8) showed significant hydroxylation (50%) on analysis of the /H92511(V) chain prepared from bovine meniscus but not from bone (data not shown). Notably from meniscus, the /H92511(II) chain consistently was more hydroxylated than /H92511(II) from articular cartilage at site A2 (Fig. 8).

**DISCUSSION**

Our findings establish several sites of prolyl 3-hydroxylation not previously identified in fibril-forming collagens. Most of the data on 3Hyp in collagen in the literature were gathered from amino acid analyses as the different chain types were discovered and their cyanogen bromide-derived peptides were characterized (21–27). The present results are consistent with these original quantitative measurements, which showed, for example, one residue of 3Hyp per α1(I) and α2(I) chain of type I collagen (22, 23). The primary site (site A1) in the α1(I) chain at Pro434 was originally estab-
motif is unlike that of the A1 site in $\alpha_1(I)$. The $\alpha_2(I)$ sequence has no recognizable A1 proline site. The $\alpha_1(II)$ chain as shown here and previously has an A1 site that is almost fully 3-hydroxylated in cartilage tissue (6). The lack of any 3-hydroxylation of the A1 site motif at Pro$^{992}$ in $\alpha_1(III)$ (Fig. 6), despite a candidate proline, is consistent with an earlier reported absence of 3Hyp from human and bovine type III collagen based on amino acid composition and Edman sequencing analyses (25).

**Lack of 3Hyp in Mammalian Type III Collagen**—The fully hydroxylated Pro$^{986}$ primary site in chicken $\alpha_1(III)$ but not in mammals (Fig. 6) most probably reflects the lack of a recognizable substrate sequence. This appears to be an evolutionary loss in mammals. Inspecting the COL3A1 sequences of the zebra finch, the only other bird in the genomic data base (Ensemble entry: ENSTGVG0000010955); the anole lizard, a reptile (Ensemble entry: ENSCAG00000015062); and Xenopus tropicalis, an amphibian (Ensemble entry: ENSXETG00000010783), all show the same GXXPGPPI(P)PGPR at site 1, which predictably from chicken versus mammalian sequences in Fig. 6 means that their Pro$^{986}$ site in tissue type III collagen will be 3-hydroxylated.

A key question, therefore, is whether the lack of 3Hyp in type III collagen of mammals has any consequences in terms of the functional behavior of type III collagen in mammalian extracellular matrices. Collagen III does not form thick fibrils in its own right but occurs copolymerized on the surface of type I collagen fibrils in skin and other tissues (28) and on type II collagen fibrils in mature articular cartilage (29, 30). The main roles for type III collagen appear to be in wound healing, matrix repair, and tissue development and as a structural component of mechanically pliable “soft” tissues, such as arterial walls. It always coexists, it seems, as a component of fibrils formed from more abundant type I and/or type II collagens, at least in mammals. Whether collagen III can function in a more independent fibrillar role in species in which its A1 site can be 3-hydroxylated is an interesting question. For example, perhaps it could polymerize independently on a template of collagen V/XI as do types I and II fibrils (10).

**Comparison of A-clade and B-clade 3Hyp Sites**—The A1 sequence motif is evident in $\alpha_1(I)$, $\alpha_1(II)$, $\alpha_1(III)$, and $\alpha_2(V)$, all A-clade collagen gene products. Their common motif is GXXPGPPI(P)PGPR. The other fibrillar collagen sites (Fig. 8, sites A2–A4 and B1–B3) lack this sequence but share some common features with each other and with known prolyl 3-hydroxylation sites in type IV collagen (19). Their most recognizable feature, beyond the -PP*G- requirement, is a phenylalanine residue nine residues or less N-terminal to the substrate proline. This can be seen at sites A2, A3, B2, and B3 in Fig. 8. Site B1 lacks such a phenylalanine but follows closely after B2 in the same tryptic peptide. Such placement of 3Hyp residues following phenylalanine is evident at two sites previously reported for the $\alpha_1(IV)$ chain in the homologous sequence -GFXXPGP*P-GP- (19). Whether phenylalanine is required for enzyme recognition or is simply a coincident feature of the recognized substrate sequence remains to be seen. Also relevant is the observed importance of phenylalanine in model triple-helical collagen peptides in the homologous sequence -GFXXPGP*P-GP- (19).
promoting higher order structures through interactions with Pro/Hyp in neighboring molecules (31).

Recent studies imply that the enzyme variant P3H2 is responsible for prolyl 3-hydroxylation of type IV collagen (5). It is possible therefore that the non-A1 sites in A-clade and B-clade collagen chains are not hydroxylated by P3H1. However, P3H1 does seem to be the main isoform expressed by cells that make fibrillar collagens, whereas P3H2 is most prominently expressed in basement membrane-rich tissues (5). Alternatively the non-A1 3Hyp sites may not require the same enzyme complex as site A1. The latter site is normally hydroxylated by a trimeric protein complex of P3H1, CRTAP protein, and cyclophilin B (32). Without CRTAP, P3H1 fails to hydroxylate site A1 in collagen type I (H9251) and type II (H9251) from studies on the crtap null mouse (6) and human CRTAP-null OI patients (9). The situation is less clear from analyses of P3H1 (LEPRE1)-null human cells even for site A1 where some residual hydroxylation was observed in cell culture (8, 9). Perhaps if expressed, P3H2 can act to some extent as a 3-hydroxylase for both A1 and non-A1 sites in A-clade and B-clade collagen chains as well as for type IV collagen, because it appears not to form a complex with CRTAP and cyclophilin B (5). Clearly, analyses of the differential effects of CRTAP and LEPRE1 mutations on the various prolyl 3-hydroxylation sites should be helpful in understanding the significance of 3Hyp formation for normal collagen biology and its defective formation in the pathogenesis of recessive OI.

Origin of the A1 Site as a Substrate—It is tempting to speculate that the A1 site in fibrillar collagen chains appeared quite late in eukaryote evolution just prior to the emergence of vertebrates. Although 3Hyp is present in invertebrate collagens as far back as porifera (sponges), the most primitive extant multicellular animals (33), 3-hydroxylation of a recognizable A1 site

Molecular Sites of 3-Hydroxyproline in Collagen

FIGURE 7. Tandem mass spectral identification of 3Hyp sites in α1(V) and α1(XI) collagen chains (sites B1, B2, and B3). The peptides were prepared by in-gel trypsin digestion after SDS-PAGE of collagen types V and XI prepared by pepsin digestion and salt precipitation from human bone and articular cartilage, respectively. A, full scan mass spectrum from the tryptic peptide LC-MS profile of α1(V) across the elution window of the peptide of sequence shown (upper spectrum). An MS/MS spectrum of the 1371.03 ion confirms the presence of two extra 16-Da units as hydroxyls on Pro665 and Pro692 (lower spectrum). The fragment ions resulting from neutral hexose losses (−162 and −324) are also indicated. Similarly, MS/MS spectra showed that 1366.3 had one extra 16 Da on Pro665 and Pro1360 3Hyp (not shown). B, full scan mass spectrum from the tryptic peptide LC-MS profile of α1(XI) across the elution window of the peptide of sequence shown (upper spectrum). An MS/MS spectrum of the 1152.3 ion establishes an extra 16 Da as a hydroxyl on Pro434 (lower spectrum). Similarly, the MS/MS spectrum of 1144.2 showed that it lacked the 16Da on Pro434 (not shown). P, 3Hyp; P*, 4Hyp; galglc, glucosylgalactosyl.
sequence motif makes its appearance in primitive vertebrates.

A single P3H gene is present in the ascidian *Ciona intestinalis* genome (a primitive chordate) and ancestrally at least as far back as *Cnidaria* (34). Because ancestors of basement membrane type IV collagen and fibril-forming collagens are recognizable in sponges (*Porifera*) (33), we speculate that the A1 sequence is in evolutionary terms a relatively new substrate for P3H that became recognizable perhaps when the P3H1/CRTAP/cyclophilin B complex (or its ancestral form) first appeared. Presumably the event that created hydroxylation activity at this site occurred before the series of whole or partial genomic duplications that led to the divergence of A-clade collagen genes (*α1(I)*, *α2(I)*, *α1(II)*, *α1(III)*, and *α2(V)*) in vertebrates (35) and perhaps also before or soon after the ancestral leprecan (P3H) gene was duplicated twice and eventually diverged into three copies (*4, 34, 36, 37*).

Because the sequence motif at site A1 differs from that at sites A2–A4 and B1–B3 (Fig. 8) and from the 3Hyp motifs in type IV collagen, a gain of function in P3H activity, perhaps through *P3H1* associating with CRTAP, seems more likely than simply a collagen sequence change alone. Such an explanation would also fit the differences evident among vertebrate A-clade collagen chains in their relative prolyl 3-hydroxylation levels at sites A2, A3, and A4 (Fig. 8), which by the logic of this concept are more ancient substrates than the A1 site. These findings are perhaps best explained by site-specific changes in A2, A3, and A4 substrate activities as their sequences in the five A-clade genes diverged.

The *α2(V)* chain shows the most complete pattern with 3-hydroxylation across all four sites, A1, A2, A3, and A4. It should be noted that *α2(V)* is an A-clade gene product, but it functions exclusively in heterotrimers in combination with two B-clade chains, for example, two *α1(V)* chains, one *α1(XI)*, or two *α1(XI)* chains dependent on the tissue (10). Because collagen V/XI acts as a template for collagen types I and II fibril polymerization and growth, it is tempting to suspect a role for the D-periodic spacing of 3Hyp in the A-clade chain of the V/XI oligomer in recruiting A-clade type I or II molecules to form a hybrid fibril.

**Genetic Defects Affecting Prolyl 3-Hydroxylation**—The importance of the Pro<sup>996</sup> (site A1) 3Hyp site for normal bone and cartilage development was revealed in studies on the CRTAP-null mouse (6). Tandem mass spectral analysis of the tryptic peptide containing this known prolyl 3-hydroxylation site showed a complete absence of 3Hyp. This was not a complete surprise because the CRTAP protein has strong sequence homology to the N-terminal half of P3H1 (but no active site and so no enzyme activity) and was known to be complexed with P3H1 and cyclophilin B protein in the endoplasmic reticulum (4). Further work showing that mutations in CRTAP and P3H1 caused recessive forms of human OI confirmed the association of disease expression with absent or diminished 3Hyp content at site A1 in collagen type I (6–9). Still not resolved, however, is whether a lack of 3Hyp in the extracellular tissue collagen of the mice or recessive OI cases is in itself responsible for defective...
tissue. For example, does this 3Hyp domain present a binding site for a fibril-associated protein that might be necessary for collagen to mineralize properly? Or is the pathogenesis due to a collagen chaperone defect in the endoplasmic reticulum that causes a secondary cellular dystrophy and consequent deficiency of adequately assembled matrix collagen? Mutating the A1 site at Pro^986 in α1(I) from proline to another amino acid in a transgenic mouse could test this.

**Speculative Function for 3-Hydroxyproline in Collagen**—In considering all that is known about 3Hyp in collagen biology,
Molecular Sites of 3-Hydroxyproline in Collagen

OI pathogenesis, and the function of 4Hyp in stabilizing the triple helix, we suspect a fundamental role for 3Hyp residues in supramolecular assembly by forming hydrogen bonds between adjacent collagen triple helices. The D-spacing between 3Hyp residues (Fig. 8, sites A2 to A3, A3 to A4, and B2 to B3) suggests such interactions between 3Hyp-containing domains could be involved in fine-tuning the D-periodic relationship and forming dimers in register through inter-triple-helical hydrogen bonds. There is good evidence that aggregates of aligned procollagen molecules exit the Golgi (38), and aggregates appear to be a better substrate for BMP-1, the procollagen C-propeptidase, than individual procollagen molecules (39). Initial studies with synthetic peptides suggested a possible destabilization of the triple helix by 3Hyp (40), but further work concluded a marginal added stability (41). The crystal structure of a synthetic peptide containing 3Hyp and a Gly-Xaa-Xaa repeat showed that the 3-OH on proline pointed out from the triple helix and so could mediate hydrogen bonding to other protein molecules (42). One logical binding partner would be another triple helix, perhaps through a water molecule, analogous to how 4Hyp stabilizes the triple helix itself through interchain hydrogen bonding. (It is notable that crystals of collagen-like peptides can form inter-triple-helical hydrogen bonds between 4Hyp hydroxyls (43–46).) If so, mutual interactions might be strongest between adjacent 3Hyp-containing domains of neighboring molecules staggered by D-periods or in register with each other.

Rather than driving the D-stagger itself, which at 234 residues (20) is three residues less than the observed A-clade 3Hyp interval (237 residues; Fig. 8), mutual 3Hyp hydrogen bonding between the hydroxyls and backbone carbonyls, directly or through water, could strengthen the relationship driven by electrostatic and hydrophobic forces. This hypothesis is particularly attractive because it could contribute forces that help fine-tune the assembly of molecules in a D-staggered array to form fibrils with an optimal placement of intermolecular cross-links. If registered dimers of procollagen molecules rather than monomers were the subunits for fibrillogenesis in the late Golgi and secretory vesicles, it would explain the efficient formation of mature intermolecular cross-links between the two nearest neighbor pairs of molecules in register and another staggered by 4D-periods (47) (Fig. 9). Such mature cross-linking is a hallmark of vertebrate skeletal tissues and particularly of bone and cartilage collagen (48).

These concepts are illustrated in Fig. 9. The packing arrangement of tetragonally packed dimers (Fig. 9B) with potential supercoiling of dimeric subunits was a model originally considered by Woodhead-Galloway (48) to be a better fit for the x-ray diffraction data and the measured protein density of collagen fibrils than are the more densely packed quasihexagonal arrays of monomers or pentafibrils that continue to be reference standards (49). A square packing arrangement of dimers is also attractive in considering how bone collagen fibrils can have the space to accommodate ordered internal plates of mineral crystallites, aligned between sheets of cross-linked collagen molecules, to form an intimate composite (50). Moreover, disruption of ordered molecular packing that specifically accommodates mineral crystallite deposition could be especially detrimental to bone properties as evidenced in osteogenesis imperfecta (51).

In summary, we conclude that the advent of the A1 3Hyp site in an A-clade collagen founder gene was superimposed on a background of more ancient 3Hyp sites. To speculate, this new feature impacted the mechanism of collagen assembly at the threshold of vertebrate evolution with subsequent influence on tissue-related diversifications in collagen fibril subunit composition and cross-linking properties (10).

REFERENCES
1. Berg, R. A., and Prockop, D. J. (1973) Biochem. Biophys. Res. Commun. 52, 115–120
2. Ogle, J. D., Arlinghaus, R. B., and Logan, M. A. (1962) J. Biol. Chem. 237, 3667–3673
3. Gryder, R. M., Lamon, M., and Adams, E. (1975) J. Biol. Chem. 250, 2470–2474
4. Vranka, J. A., Sakai, L. Y., and Baechinger, H. P. (2004) J. Biol. Chem. 279, 23615–23621
5. Tiainen, P., Pasanen, A., Sormunen, R., and Myllyharju, J. (2008) J. Biol. Chem. 283, 19432–19439
6. Morello, R., Bertin, T. K., Chen, Y., Hicks, J., Tonachini, L., Monticone, M., Castagnola, P., Rauch, F., Glorieux, F. H., Vranka, J., Bächinger, H. P., Pace, J. M., Schwarze, U., Byers, P. H., Weis, M., Fernandes, R. J., Eyre, D. R., Yao, Z., Boyce, F. B., and Lee, B. (2006) Cell 127, 291–304
7. Barnes, A. M., Chang, W., Morello, R., Cabral, W. A., Weis, M., Eyre, D. R., Leikin, S., Makareeva, E., Kuznetsova, N., Uveges, T. E., Ashok, A., Flor, A. W., Mulvihill, J. J., Wilson, P. L., Sundaram, U. T., Lee, B., and Marini, J. C. (2006) N Engl. J. Med. 355, 2757–2764
8. Cabral, W. A., Chang, W., Barnes, A. M., Weis, M., Scott, M. A., Leikin, S., Makareeva, E., Kuznetsova, N. V., Rosenbaum, K. N., Tiffit, C. J., Bulas, D. I., Kozma, C., Smith, P. A., Eyre, D. R., and Marini, J. C. (2007) Nat. Genet. 39, 359–365
9. Baldridge, D., Schwarze, U., Morello, R., Lennington, J., Bertin, T. K., Pace, J. M., Peprin, M. G., Weis, M., Eyre, D. R., Walsh, J., Lambert, D., Green, A., Robinson, H., Michelson, M., Houge, G., Lindman, C., Martin, J., Ward, J., Lemyre, E., Mitchell, J. J., Krakow, D., Raimo, D. L., Cohn, D. H., Byers, P. H., and Lee, B. (2008) Hum. Mutat. 29, 1435–1442
10. Wu, J. J., Weis, M. A., Kim, L. S., Carter, B. G., and Eyre, D. R. (2009) J. Biol. Chem. 284, 5539–5545
11. Miller, E. J. (1972) Biochemistry 11, 4903–4909
12. Eyre, D. R., and Wu, J. J. (1983) FEBS Lett. 158, 265–270
13. Eyre, D. R., and Muir, H. (1975) Biochem. J. 151, 595–602
14. Laemmli, U. K. (1970) Nature 227, 680–685
15. Eyre, D. R. (1987) Methods Enzymol. 144, 115–139
16. Hanna, S. L., Sherman, N. E., Kinter, M. T., and Goldberg, J. B. (2000) Microbiology 146, 2495–2508
17. Eyre, D. R., Weis, M. A., and Wu, J. J. (2008) Methods 45, 65–74
18. Fietzek, P. P., Reuxrodt, F. W., Wendt, P., Stark, M., and Kühn, K. (1972) Eur. J. Biochem. 130, 163–168
19. Schuppan, D., Glanville, R. W., and Timpl, R. (1982) Eur. J. Biochem. 123, 505–512
20. Doyle, B. B., Hulmes, D. J., Miller, A., Parry, D. A., Piez, K. A., and Woodhead-Galloway, I. (1974) Proc. R. Soc. Lond. B 187, 37–46
21. Piez, K. A., Eigner, E. A., and Lewis, M. S. (1963) Biochemistry 2, 58–66
22. Butler, W. T., Piez, K. A., and Bornstein, P. (1967) Biochemistry 6, 3771–3780
23. Click, E. M., and Bornstein, P. (1970) Biochemistry 9, 4699–4706
24. Miller, E. J., and Lunde, L. G. (1973) Biochemistry 12, 3153–3159
25. Seyer, J. M., and Kang, A. H. (1981) Biochemistry 20, 2621–2627
26. Burgeson, R. E., El Adli, F. A., Kaitila, I., and Hollister, D. W. (1976) Proc. Natl. Acad. Sci. 73, 2579–2583
27. Rhodes, R. K., and Miller, E. J. (1979) J. Biol. Chem. 254, 12084–12087
28. Fleischmajer, R., MacDonald, E. D., Perlish, J. S., Burgeson, R. E., and Fisher, L. W. (1990) J. Struct. Biol. 105, 162–169
29. Young, R. D., Lawrence, P. A., Duance, V. C., Aigner, T., and Monaghan, P.
Molecular Sites of 3-Hydroxyproline in Collagen

(2000) J. Histochem. Cytochem. 48, 423–432
30. Eyre, D. R., Weis, M. A., and Wu, J. J. (2006) Eur. Cell. Mater. 12, 57–63
31. Kar, K., Ibrar, S., Nanda, V., Getz, T. M., Kunapuli, S. P., and Brodsky, B. (2009) Biochemistry 48, 7959–7968
32. Ishikawa, Y., Wirz, J., Vranka, J. A., Nagata, K., and Bächinger, H. P. (2009) J. Biol. Chem. 284, 17641–17647
33. Garrone, R. (1985) Biology of Invertebrate and Lower Vertebrate Collagens (Bairati, A., and Garrone, R., eds) pp. 157–175, NATO ASI Series A, Vol. 93, Plenum, New York
34. Capellini, T. D., Dunn, M. P., Passamaneck, Y. J., Selleri, L., and Di Gregorio, A. (2008) Genesis 46, 683–696
35. Zhang, X., Boot-Handford, R. P., Huxley-Jones, J., Forse, L. N., Mould, A. P., Robertson, D. L., Lili Athiyal, M., and Sarras, M. P., Jr. (2007) J. Biol. Chem. 282, 6792–6802
36. Dunn, M. P., and Di Gregorio, A. (2009) Dev. Biol. 328, 561–574
37. Vranka, J., Stadler, H. S., and Bächinger, H. P. (2009) Cell Struct. Funct. 34, 97–104
38. Polishchuk, E. V., Di Pentima, A., Luini, A., and Polishchuk, R. S. (2003) Mol. Biol. Cell 14, 4470–4485
39. Hojima, Y., Behta, B., Romanic, A. M., and Prockop, D. J. (1994) Anal. Biochem. 223, 173–180
40. Jenkins, C. L., Bretschler, L. E., Guzei, I. A., and Raines, R. T. (2003) J. Am. Chem. Soc. 125, 6422–6427
41. Mizuno, K., Peyton, D. H., Hayashi, T., Engel, J., and Bächinger, H. P. (2008) FEBS J. 275, 5830–5840
42. Schumacher, M. A., Mizuno, K., and Bächinger, H. P. (2006) J. Biol. Chem. 281, 27566–27574
43. Kramer, R. Z., Bella, J., Brodsky, B., and Berman, H. M. (2001) J. Mol. Biol. 311, 131–147
44. Kar, K., Amin, P., Bryan, M. A., Persikov, A. V., Mobs, A., Wang, Y. H., and Brodsky, B. (2006) J. Biol. Chem. 281, 33283–33290
45. Berisio, R., De Simone, A., Ruggiero, A., Improta, R., and Vitagliano, L. (2009) J. Pept. Sci. 15, 131–140
46. Okuyama, K., Hongo, C., Wu, G., Mizuno, K., Noguchi, K., Ebisuzaki, S., Tanaka, Y., Nishino, N., and Bächinger, H. P. (2009) Biopolymers 91, 361–372
47. Eyre, D. R., Paz, M. A., and Gallop, P. M. (1984) Annu. Rev. Biochem. 53, 717–748
48. Woodhead-Galloway, J. (1980) Proc. R. Soc. Lond. B 209, 275–297
49. Orgel, J. P., Irving, T. C., Miller, A., and Wess, T. J. (2006) Proc. Natl. Acad. Sci. U.S.A. 103, 9001–9005
50. Burger, C., Zhou, H. W., Wang, H., Sics, I., Hsiao, B. S., Chu, B., Graham, L., and Glimcher, M. J. (2008) Biophys. J. 95, 1985–1992
51. Marini, J. C., Cabral, W. A., Barnes, A. M., and Chang, W. (2007) Cell Cycle 6, 1675–1681