MOLECULAR STRUCTURE OF THE COLLAGEN TRIPLE HELIX

By BARBARA BRODSKY AND ANTON V. PERSIKOV

Department of Biochemistry, University of Medicine and Dentistry of New Jersey, Robert Wood Johnson Medical School, Piscataway, New Jersey 08854

Abstract

The molecular conformation of the collagen triple helix confers strict amino acid sequence constraints, requiring a $(Gly\text{-}X\text{-}Y)_n$ repeating pattern and a high content of imino acids. The increasing family of collagens and proteins with collagenous domains shows the collagen triple helix to be a basic motif adaptable to a range of proteins and functions. Its rodlike domain has the potential for various modes of self-association and the capacity to bind receptors, other proteins, GAGs, and nucleic acids. High-resolution crystal structures obtained for collagen model peptides confirm the supercoiled triple helix conformation, and provide new information on hydrogen bonding patterns, hydration, sidechain interactions, and ligand binding. For several peptides, the helix twist was found to be sequence dependent, and such variation in helix twist may serve as recognition features or to orient the triple helix for binding. Mutations in the collagen triple-helix domain lead to a variety of human disorders. The most common mutations are single-base substitutions that lead to the

ADVANCES IN PROTEIN CHEMISTRY, Vol. 70
DOI: 10.1016/S0065-3233(04)70009-1

Copyright 2005, Elsevier Inc.
All rights reserved.

0065-3233/05 $35.00

replacement of one Gly residue, breaking the Gly-X-Y repeating pattern. A single Gly substitution destabilizes the triple helix through a local disruption in hydrogen bonding and produces a discontinuity in the register of the helix. Molecular information about the collagen triple helix and the effect of mutations will lead to a better understanding of function and pathology.

I. Introduction

Collagen has always been considered to be an important protein because of its abundance in the human body and its commercial applications. Fifty years ago, the molecular conformation for collagen was proposed to be a novel triple-helix structure. Fiber diffraction analysis and model building, together with early amino acid composition and sequence data, led to the concept of three chains, each in a polyproline-II-like conformation, supercoiled about a common axis (Ramachandran, 1967; Ramachandran and Kartha, 1955; Rich and Crick, 1955, 1961). The close packing of the three chains near the common axis places steric constraints on every third position, such that only glycine can be accommodated without chain distortion. This generates the $(\text{Gly-X-Y})_n$ repeating sequence, recognized as the signature of a collagen. Residues in the X and Y positions are largely solvent accessible and can accommodate any side-chain. In fact, imino acids are particularly favorable because their fixed ψ angle and restricted φ angle are close to those found in the triple helix. When Pro is incorporated into the Y position in a collagen chain, it becomes posttranslationally modified to hydroxyproline (Hyp), which has a highly stabilizing effect on the triple helix.

Early on, it was realized that peptides could be useful models for the features of collagen. Polyglycine and polyproline played an important role in elucidating the triple-helix structure (Cowan *et al.*, 1955; Rich and Crick, 1955), and polytripeptides were designed to adopt the collagen conformation. Initially, the peptides were heterogeneous polymers, but with the advent of solid state peptide synthesis, peptides of defined length and sequence could be synthesized to clarify principles of triple-helix stability and mimic biological activity (Fields and Prockop, 1996; Goodman *et al.*, 1998; Heidemann and Roth, 1982; Jenkins and Raines, 2002). Most peptides are studied as monomers that self-associate into trimers, while, in some cases, crosslinked trimers are synthesized to facilitate trimer formation, stabilize the triple helix, or create heterotrimers. Triple-helical peptides are proving more amenable to structural characterization than the longer native collagen molecules, and peptide investigations have led to clarification of the molecular details of collagen structure by

X-ray crystallography, nuclear magnetic resonance (NMR), and other biophysical techniques.

Because of its strict structural constraints, triple-helix domains can be identified on the basis of amino acid sequence. With many amino acid sequences of proteins available through molecular biology, there is clearly a growing family of collagen proteins with a common triple-helix motif present in diverse tissue structures and an expanding collection of collagen-like domains in various other proteins. The collagen triple helix can form a straight or kinked rod, has the capacity for self-association into various supramolecular structures, and has the potential to bind ligands and receptors. These structural features of collagen have adapted to a wide range of biological roles, and now the triple-helix motif is more appropriately viewed in a general context as a basic and versatile protein motif. Additionally, a growing number of mutations in collagens and collagen-like domains have been associated with specific diseases. The need to understand the effects of these mutations has lent a growing impetus to characterization of the normal and variant forms of molecular conformation and higher order structure.

This review will focus on recent advances in the molecular structure of the collagen triple helix, including information from high-resolution structures, the effect of amino acid sequence on stability, ligand binding, and disease. Important articles on the molecular properties of collagen were published in earlier volumes of this series (Harrington and von Hippel, 1961; Privalov, 1982; Traub and Piez, 1971). There are excellent reviews of earlier work on the molecular structure of collagen (Fraser and MacRae, 1973; Fraser *et al.*, 1987; Ramachandran, 1967) and several comprehensive overviews of collagens, their structures, assembly, and biochemistry (Kielty and Grant, 2002; Myllyharju and Kivirikko, 2004).

II. The Biological Role and Occurrence of the Collagen Triple-Helix Motif in Proteins

Collagens are defined as structural molecules in the extracellular matrix that contain a triple-helix domain; at this time, 27 distinct types of human collagens have been identified (Fig. 1; Table I; Kielty and Grant, 2002; Myllyharju and Kivirikko, 2004). These different collagen types carry out specialized functions in diverse tissues and have distinctive modes of supramolecular organization. Some molecules are homotrimers, while others are heterotrimers, with two or three distinguishable chain types. The most abundant of these are found in characteristic collagen fibrils (major types I, II, III, and minor types V and XI), which form the structural basis of skin, tendon, bone, cartilage, and other tissues. Other

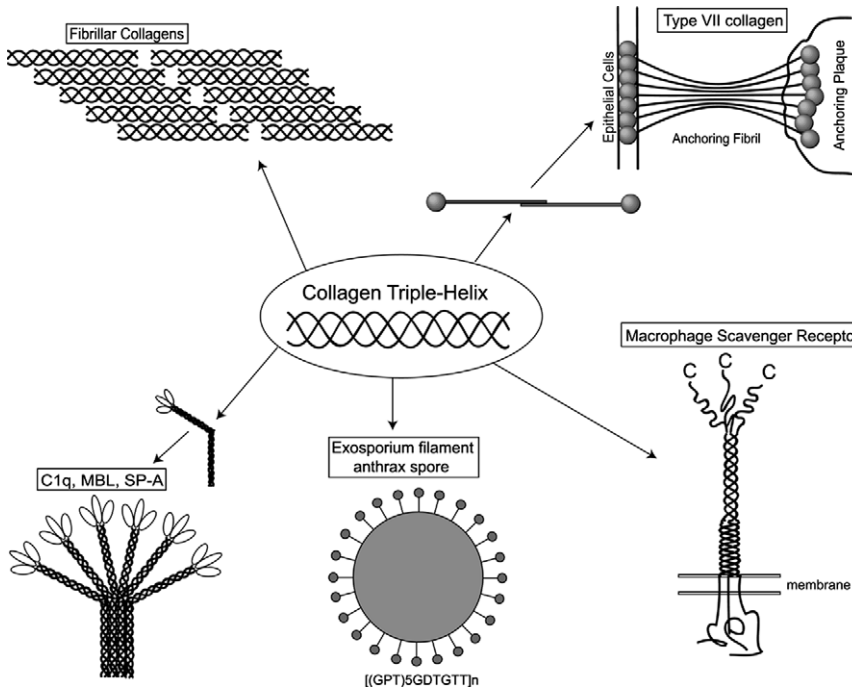

FIG. 1. Illustration of some of the biological forms of the collagen triple-helix domains.

collagens are found on the surface of fibrils (FACIT types IX, XII, XIV, XVI, XIX, XX, XXI, XXII, XVI); within basement membrane networks (type IV collagen); in hexagonal networks (types VIII and X); as beaded filaments (type VI); in the anchoring fibrils of skin (type VII); or as membrane proteins (types XIII, XVII, XXIII, XXV). Type XVIII collagen is a basement membrane heparan sulfate proteoglycan, and contains a C-terminal noncollagenous fragment, endostatin, which inhibits angiogenesis and tumor growth (Kawashima *et al.*, 2003). A thorough discussion of the various collagen types, their structures, and their functions is included in several excellent reviews (Kielty and Grant, 2002; Myllyharju and Kivirikko, 2004).

In addition to being the defining feature of collagens, the collagen triple helix is present as a motif in a variety of proteins, many of which are involved in host-defense functions (Fig. 1; Table I; Kielty and Grant, 2002; Lu *et al.*, 2002; Matsuzawa *et al.*, 2004; Myllyharju and Kivirikko, 2004;

TABLE I
Occurrence of Proteins with Collagen Triple-Helix Domains

Proteins	Supramol	Tissue	Number of $(GXY)_n$	Breaks
Vertebrates				
Type I collagen	D-periodic fibril	Tendon, bone, skin	338	0
Type II collagen	D-periodic fibril	Cartilage, vitreous	338	0
Type IV collagen	Network	Basement membrane	437	21
Type VII collagen	Antiparallel	Anchoring fibrils	472	20
Type IX collagen	Surface of fibrils	Ubiquitous	85	4
Type XII collagen	(FACIT)		197	5
C1q	Hexamer	Serum, complement	24	1
MBP	Oligomer	Serum	19	1
SP-A	Hexamer	Lung surfactant	23	1
SP-D	Tetramer	Lung surfactant	59	0
MSR	Membrane	Macrophage	24	0
Invertebrates				
C. elegans cuticle collagen		Exoskeleton	~9, ~40	1–4
Drosophila type IV collagen	Network	Basement membrane	444	17
Hydra minicollagen	Disulfide linked polymers	Nematocyst inner wall	12–16	0
Bacteria and Viruses (~120 types)				
Scl1	Cell surface	*Streptococcus* group A	50	0
BclA	Exosporium filament	*Bacillus anthracis*	70	0
Lymphocytis disease virus	Unknown	Unknown	55	1
Shrimp white spot virus	Unknown	Unknown	389	1

Shirai *et al.*, 1999). The kinked triple-helix domain in C1q binds to the serine proteases C1r and C1s, and mediates the self-association of six trimer molecules that generates the bouquet-type structure important for activity. The family of collectins, which includes the mannose-binding lectin SP-A and SP-D, all contain a collagenous domain, a coiled coil α-helical domain, and a terminal carbohydrate recognition domain. Ficolins have a triple-helix domain with a terminal carbohydrate binding fibringen domain. Both collectins and ficolins are host-defense molecules, which bind to the carbohydrate groups of microbes, leading to complement activation and phagocytosis (Lu *et al.*, 2002). The triple-helix domain of the macrophage scavenger receptor is responsible for ligand recognition (Doi *et al.*, 1993; Shirai *et al.*, 1999), while the collagenous tail of the asymmetric form of acetylcholinesterase binds to heparan sulfate, localizing the enzyme to the neuromuscular junction (DePrez *et al.*, 2000). In addition to these examples, numerous other proteins have been observed to contain collagen-like domains, some of which are listed in Table I.

Orthologues of fibril-forming collagens and of type IV collagen are found in a number of invertebrates, in addition to specialized collagens such as the cuticle collagens of *C. elegans* and the hydra nematocyst minicollagens (Exposito *et al.*, 2002). It was thought that collagen was a defining feature of multicellular animals, but recent observations show triple-helix domains present in bacteria and viruses. Collagen-like domains Scl1 and Scl2 were observed in proteins expressed on the cell surface of group A streptococcus, and were shown to adopt a triple-helix conformation (Xu *et al.*, 2002). A highly repetitive collagen-like domain was also identified in the filament protein of the exosporium of anthrax spores, and the length of this collagenous domain appears to determine the filament length (Sylvestre *et al.*, 2003). A search of the genomes of various bacteria and viruses indicated the presence of at least 100 novel proteins containing collagen-related structural motifs (Rasmussen *et al.*, 2003). These findings confirm that the collagen triple-helix motif can be part of many different kinds of proteins and can fill a wider than expected set of biological niches (Table I).

The occurrence of the collagen triple helix illustrates its role in protein function (Fig. 1). In all cases, the triple helix forms a rodlike structure, but this rod can have a kink, as in C1q and MBL, or have flexible interruptions, as in type IV collagen. In most cases, the collagen triple helix self-associates to form a higher-order structure. Many possible modes of association are seen, including staggered parallel (fibrils), parallel unstaggered (stalk of MBP, SP-A and C1q), antiparallel (type VII and type VI collagens) and other more complex relationships in networks and fibers. A number of collagens and proteins with collagenous domains contain

transmembrane domains (MSR, and collagen types XIII, XVII, XXIII, and XXV). Fibrillar collagen structure and network-forming collagen structures are the focus of Chapters 10 and 11. The triple helix is an important ligand binding domain in most proteins, binding to receptors, proteases, extracellular matrix proteins, GAGs, nucleic acids, and lipids in different cases (see section VII below). In contrast to early thinking, it is increasingly clear that the collagen triple helix is not a domain responsible for trimerization. The triple helix folds very slowly, and is usually dependent on a neighboring coiled coil or globular domain for nucleation (McAlinden *et al.*, 2003). But once trimerized, its rodlike structure has the potential for a wide range of modes of self-association and the ability to bind diverse ligands, as seen in the growing collection of proteins with triple-helix domains.

III. Molecular Structure: Collagen and Collagen Model Peptides

The advances in fiber diffraction analysis, amino acid composition data, molecular modeling, and polypeptide structures all converged during the mid-1950s to clarify the unique structure of collagen. The supercoiled triple-helix conformation was proposed for collagen in 1955 independently by Ramachandran, who proposed a similar model without supercoiling in 1954, (Ramachandran and Kartha, 1954); by Rich and Crick (1955); and by Cowan, McGavin, and North (Cowan *et al.*, 1955). In this conformation, three polypeptide chains, each in an extended left-handed polyproline II-helix conformation, are supercoiled in a right-handed manner around a common axis, with a stagger of one residue between adjacent chains. The most accurate model available for collagen is based on linked-atom least-squares refinement of the Rich and Crick II model using the excellent fiber diffraction data from highly stretched partially dehydrated kangaroo tail tendon, as reported by Fraser *et al.* (1979). The basic conformation was confirmed and complemented by a range of spectroscopic and hydrodynamic studies on collagens from various species. The historical developments and collagen investigations are reviewed in Fraser and MacRae (1973).

Although there were early reports of crystallization of a cyanogen bromide fragment of collagen (Yonath and Traub, 1975) and important NMR studies were carried out on isotopically labeled collagen in tissues (Batchelder *et al.*, 1982; Sarkar *et al.*, 1983, 1987), the collagen molecule itself has not proved amenable to investigations at the molecular level. The path to the molecular details of the collagen triple helix has been through collagen model peptides, which have yielded high-resolution X-ray structures and allowed NMR characterization of dynamic and conformational features. At

this time, the crystal structures of nine different triple-helical peptides have been reported, one with a bound integrin domain (Table II; Fig. 2). The fiber diffraction models of collagen left unresolved controversies about the nature of interchain hydrogen bonding, including the possibility of hydrogen bonds involving alpha carbon atoms, and the precise geometrical parameters of both the basic helix and the supercoil. The high-resolution crystal structures of peptides confirmed the conformation derived from fiber diffraction data of collagen, and resolved a number of long-standing controversies about hydrogen bonding and hydration. The molecular details also raise new issues for consideration, including variability in helix twist and the mechanism of hydroxyproline stabilization.

A. Sequence Dependence of Triple Helix Twist

Fiber diffraction patterns of collagen in tail tendon are indexed with a 10/3 symmetry (10 units in 3 turns) (Fraser *et al.*, 1979; Rich and Crick, 1961). Therefore, it was surprising when Okuyama *et al.* (1981) reported that the first crystal structure of a collagen-like peptide, $(Pro\text{-}Pro\text{-}Gly)_{10}$, had a 7/2 symmetry (7 units in 2 turns). The difference in triple-helix symmetry between the $(Pro\text{-}Pro\text{-}Gly)_{10}$ crystal and collagen fibers could have arisen from crystal packing effects, the unusually high imino acid content in the peptide compared with collagen, or from the absence of Hyp in the peptide. Recent crystallographic structures have confirmed the 7/2 symmetry is present in $(Gly\text{-}Pro\text{-}Hyp)_{10}$ and the G→A peptide, as well as $(Pro\text{-}Pro\text{-}Gly)_{10}$ (Table II). The EKG peptide, which is homologous to $(Pro\text{-}Hyp\text{-}Gly)_{10}$ but with one Glu-Lys-Gly triplet in the center, also shows strict 7/2 symmetry (Kramer *et al.*, 2000).

However, a nonuniform helical twist has been observed in the crystal structures of peptides T3-785 and IBP (Emsley *et al.*, 2000, 2004; Kramer *et al.*, 1999, 2001). These two peptides each can be viewed as having three zones: N-terminal Gly-Pro-Hyp repeats; a central collagen sequence; and C-terminal Gly-Pro-Hyp repeats (Fig. 3). In T3-785, the two terminal Gly-Pro-Hyp regions show 7/2 symmetry, while the central Gly-Ile-Thr-Gly-Ala-Arg-Gly-Leu-Ala region is closer to 10/3 symmetry. In IBP, the terminal Gly-Pro-Hyp repeats have 7/2 symmetry, and the central Gly-Phe-Hyp-Gly-Glu-Arg sequence is intermediate between 7/2 and 10/3. In each case, these three zones are slightly bent and twisted with respect to each other. It appears that the 7/2 symmetry is generated by the steric restrictions of repeating Gly-imino acid-imino acid units and is maintained when only one Gly-X-Y triplet is introduced (e.g., EKG peptide). The presence of 2 or 3 tripeptide units, where X and Y are not imino acids, starts to change the helix twist towards 10/3 symmetry.

TABLE II
Collagen-Like Peptides with High-Resolution Crystal Structures Entered in the Protein Data Bank

Peptide	Sequence	Number of residues	PDB ID	Resolution (Å)	Helical twist	Reference
G->A	$(POG)_4$ POA $(POG)_5$	30	1CAG	1.85	7/2	Bella *et al.*, 1994
T3-785	$(POG)_3$ITGARGLAG$(POG)_4$	30	1BKV	2.00	Variable	Kramer *et al.*, 1999
EKG	$(POG)_4$ EKG $(POG)_5$	30	1QSU	1.75	7/2	Kramer *et al.*, 2000
Hyp-	$(POG)_4$ PG $(POG)_5$	29	1EI8	2.00	7/2	Liu, 2000
PPG10	$(PPG)_{10}$	30	1K6F	1.30	7/2	Berisio *et al.*, 2002
GPP-foldon	$(GPP)_9$-foldon	27	1NAY	2.60	7/2	Stetefeld *et al.*, 2003
PPG9	$(PPG)_9$	27	1ITT	1.00	7/2	Hongo *et al.*, unpub.
POG10	$(POG)_{10}$	30	1V4F	1.26	7/2	Okuyama *et al.*, unpub.
Integrin binding peptide	$(GPO)_2$ GFOGER $(GPO)_3$	21	1Q7D	2.10	Variable	Emsley *et al.*, 2004
	In complex with integrin	21	1DZI	1.80	Variable	Emsley *et al.* 2000

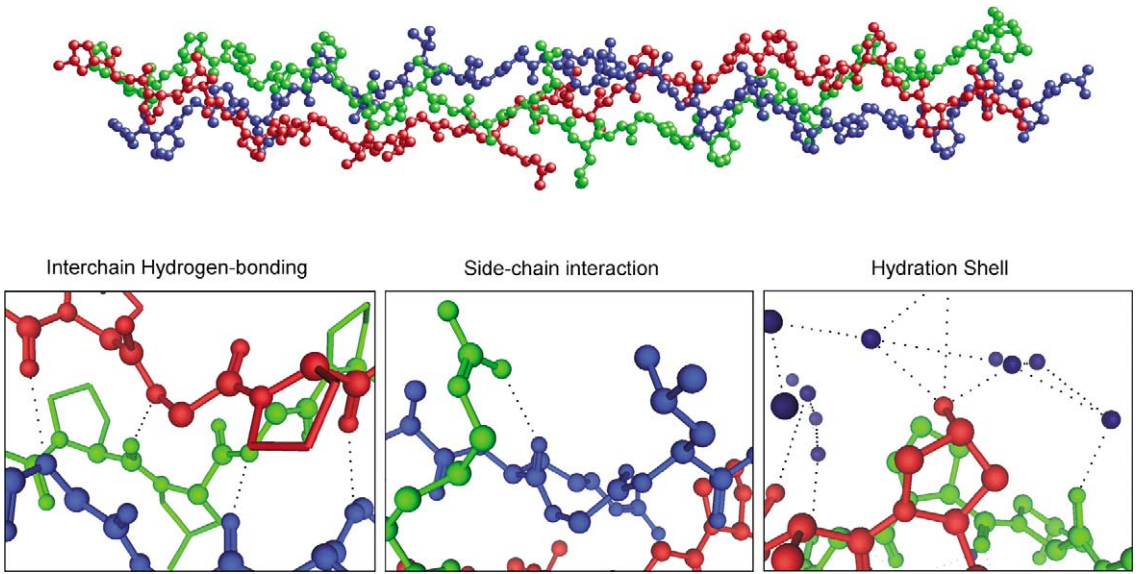

FIG. 2. Molecular structures of peptide T3-785 showing hydrogen bonding, sidechain orientation, and hydration.

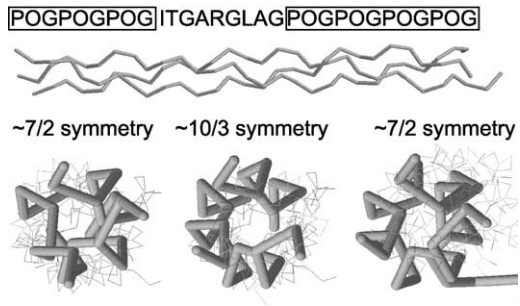

FIG. 3. Variation in helix twist seen in the three domains of peptide T3-785.

These peptide results suggest the collagen molecule, with its varied Gly-X-Y sequence, is likely to have a nonuniform helical twist along its length. It is likely sequences poor in imino acids will have a symmetry close to 10/3, while stretches of Gly-Pro-Hyp units may have 7/2 symmetry. In type I collagen, the longest sequence of repeating Gly-Pro-Hyp triplets is found at the C-terminus, and 7/2 symmetry could play a role in its functioning as a nucleation domain (Xu *et al.*, 2003). It is not clear at this time whether collagen chains have a continuous variation in superhelix twist, or domains of different symmetry connected with kinks and bends. The X-ray pattern of tail tendons does show a predominance of the 10/3 symmetry, but coherent regions may be overrepresented in fiber diffraction patterns. The conformational differences between the 7-fold and 10-fold symmetric triple-helices are subtle, but the small difference in units per turn (3.50 vs. 3.33) would translate into a significant difference in the helix repeat over a long distance and could affect recognition.

B. Hydrogen Bonding

Hydrogen bonding is a critical part of triple-helix stabilization. The very favorable enthalpy reported for collagen compared to other proteins is consistent with the importance of hydrogen bonding (Privalov, 1982). The triple helix has repetitive backbone hydrogen bonding networks, but differs from beta sheets or alpha helices in that the repeating tripeptide unit consists of three nonequivalent peptide groups, and not all backbone peptide groups participate in hydrogen bonding. It was clear from earliest models that one strong interchain peptide NH...OC bond could be formed per Gly-X-Y tripeptide unit (Ramachandran, 1967; Rich and Crick,

1961). Ramachandran originally argued that a second NH...OC bond was possible when the residue in the X position was not Pro (Ramachandran and Kartha, 1955), but because of distortion and steric problems, this view was modified to suggest an interaction mediated by water (Ramachandran and Chandrasekharan, 1968).

All crystal structures show a hydrogen bond between the NH of Gly in one chain and the C═O of the residue in the X position of the neighboring chain, as predicted (Fig. 4). In addition, peptides with sequences where the X position is occupied by a residue other than Pro, show a second interchain hydrogen bond between the amide group of the X position residue and the C═O of the Gly residue, which is mediated by one water molecule (Emsley *et al.*, 2000, 2004; Kramer *et al.*, 1999, 2000). This second set of hydrogen bonds connects chains in a direction opposite to that of the first set, as proposed by Ramachandran and Chandrasekharan (1968). In some cases, the water molecules involved in the NH (X position)...CO (Gly) hydrogen bond make additional hydrogen bonds with Hyp or side chains. When the Y position is occupied by an amino acid, rather than an imino acid, its amide group is hydrated by water molecules directed into the solvent, which is not likely to contribute to stability.

The hydrogen bonding patterns seen in peptide crystals are consistent with hydrogen exchange studies and thermodynamic analyses (Fan *et al.*, 1993; Privalov, 1982; Yee *et al.*, 1974). NMR hydrogen exchange studies on 15N-labeled Gly-Leu-Ala residues in the central zone of peptide T3-785 demonstrated that Gly NH exchanged the slowest, Leu NH exchanged almost as slowly, and Ala NH showed an exchange rate similar to peptides exposed to bulk solvent (Fan *et al.*, 1993). The water-mediated hydrogen bond of Leu, involving the X position amide, slows the hydrogen exchange process almost as much as the direct Gly interchain hydrogen bonds. Hydrogen exchange studies on collagen also show two distinct slowly exchanging sets of amide hydrogens (Privalov, 1982; Yee *et al.*, 1974). Therefore, the second set of water-mediated hydrogen bonds is likely to contribute to stability and may serve to reinforce the triple helix in regions lacking Pro.

The possibility of $C^{\alpha}H$...CO hydrogen bonds in polyglycine II and collagen was raised by Ramachandran (1967) and Krimm and Kuroiwa (1968). A detailed analysis of the high resolution structure of the G→A peptide supports the existence of two kinds of such bonds: an H-bond between the Gly C^{α} in one chain and the Gly and Pro C═O groups from the other two chains; and an H-bond connecting Hyp C^{α} in one chain with a Pro C═O group on the neighboring chain (Bella and Berman, 1996). These $C^{\alpha}H$...CO bonds are suggested to define a network of weak interactions that may add additional stability to the strong NH...CO bonds.

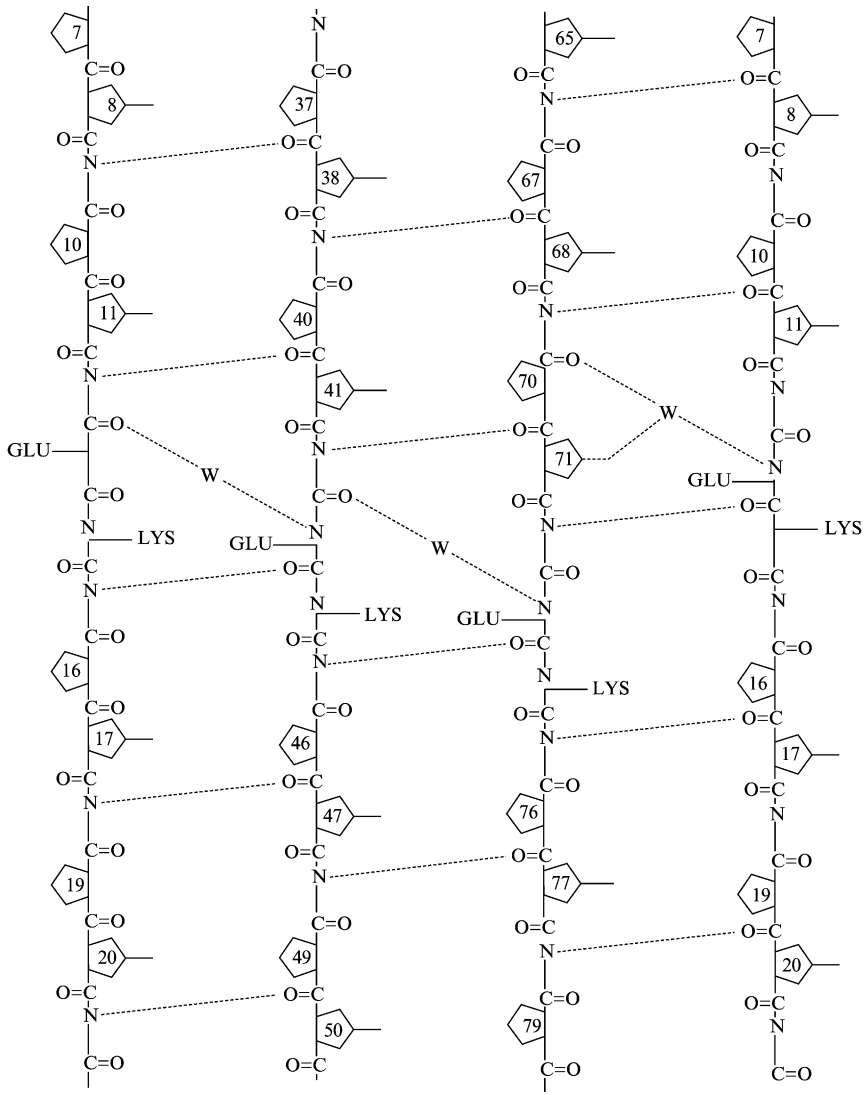

FIG. 4. Schematic of the hydrogen bonding patterns seen in the high-resolution structure of EKG peptide.

C. *Hydration Networks*

Water is an integral part of the collagen molecule (Fraser and MacRae, 1973; Harrington and Von Hippel, 1961; Traub and Piez, 1971). The collagen triple helix has tightly bound water, and experimental evidence supports it being a highly ordered hydration network (Berendsen and Migchelsen, 1965; Grigera and Berendsen, 1979; Suzuki *et al.*, 1980). The water was proposed to bind to available backbone carbonyls; since two C═O groups per tripeptide do not participate direct NH. . .CO bonds, and original concepts of single water binding were expanded to suggest waters bonded to each other and then to several main chain atoms (Ramachandran and Chandrasekharan, 1968). The function of hydroxyproline has also been related to this hydration network. On the basis of an elegant analysis of thermodynamics and hydrogen exchange data, Privalov found that Hyp stabilization of collagen is correlated with increased enthalpy (Privalov, 1982). Since Hyp cannot directly bond with the backbone of the molecule, enthaplic stabilization supports the involvement of an ordered water hydrogen-bonding network.

The determination of the high-resolution structure of the G→A peptide provided the first visualization of the elaborate water network that surrounds collagen molecules (Fig. 2; Bella *et al.*, 1994, 1995). Water molecules are seen to bridge C═O groups within and between molecules and to link C═O and Hyp hydroxyl groups within and between molecules. Often four or five water molecules participate in these bridges, with pentagonal, clathrate-like geometry. Strikingly repetitive networks of these water patterns are seen along the chain. Water tends to order near hydrophobic surfaces, and Bella *et al.* (1995) suggest the abundance of pentagonal water clusters in the triple-helical cylinder of hydration may result from the solvent exposure of nonpolar Pro and Hyp residues, together with the availability of backbone CO groups and Hyp OH groups to anchor the ordered water network to the peptide. Initially, questions were raised about whether these intricate water networks were real (Engel and Prockop, 1998), but the increasing number of high-resolution structures have confirmed that extended water networks are an inherent feature of all collagen triple-helix peptide crystal structures (Berisio *et al.*, 2001, 2002; Kramer *et al.*, 2000, 2001). NMR studies have indicated the kinetically labile nature of this collagen hydration shell (Melacini *et al.*, 2000).

The nature and order of the water network appears to depend on the molecular packing in the crystal and the specific sequence present. In peptide T3-785, less ordered water was seen in the terminal Gly-Pro-Hyp regions that were not well packed, than in the central Gly-Ile-Thr-Gly-Ala-Arg-Gly-Leu-Ala region with close molecular packing (Kramer *et al.*, 2001).

For the EKG peptide, there was some disruption of the regular hydration pattern in the central charged region, where the packing is also less dense (Kramer *et al.*, 2000). In addition to interactions involving backbone CO groups and Hyp, water is also seen to interact with sidechains, mediating hydrogen bonds between ionizable sidechains and between sidechains and the backbone (Kramer *et al*, 2000, 2001). In the central region of peptide T3-785, where there are no Hyp residues present, the waters involved in the hydrogen bonds between NH(X)...CO(Gly) also interact with Thr and Arg in the Y positions rather than Hyp (Kramer *et al.*, 2001).

The high-resolution structures of triple-helical peptides allow visualization of the ordered water network that was expected from studies on collagen, and has elucidated the specific bonds involved in this network and their repetitive nature. There is strong evidence supporting the role of this ordered water in maintaining molecular packing in fibrils (Leikin *et al.*, 1994, 1995, 1997). The biological and physical significance of this water network with respect to molecular stability has been a subject of much debate (Jenkins and Raines, 2002). The Hyp hydroxyl groups are clearly key players in repetitive hydration networks in the high-resolution structures, but one cannot assess from a crystal structure alone whether the entropic cost of localizing the waters will be less or greater than the highly favorable enthalpy from hydrogen bond formation in the collagen molecule (see Section IV below).

D. Side Chain Interactions

The availability of high-resolution structures of peptides EKG, T3-785, and IBP, which include residues other than Pro and Hyp in the X and Y positions, offers the opportunity to investigate the conformation and interactions of sidechains from residues typically found within the collagen triple helix. In the peptide with an EKG tripeptide sequence, the Lys and Glu residues did not form direct intermolecular or intramolecular ion pairs, even though such pairs are sterically feasible (Kramer *et al.*, 2000). Instead, the Lys side chains bond to Y position carbonyl groups of an adjacent chain, while one Glu directly interacts with a Hyp hydroxyl group. There was also a range of water-mediated interactions involving the polar sidechains. In peptide T3-785, with the central region Gly-Ile-Thr-Gly-Ala-Arg-Gly-Leu-Ala, the Arg side chains make direct contacts with backbone carbonyl groups on an adjacent chain, confirming predictions of this interaction and clarifying the high stability of Arg in the Y position (Fig. 2; Kramer *et al.*, 2001; Vitagliano *et al.*, 1993). In T3-785, the Arg sidechains also make hydrophobic interactions with Leu and Ile sidechains

from the same or neighboring molecules, forming nonpolar clusters that minimize exposure to solvent. The Thr sidechains are involved in bonding with the water, which mediates hydrogen bonds between the amide groups from the X position and Gly C═O groups. The participation of Thr in the Y position in the water network, much as Hyp does when it is in the Y position, suggests Thr could play a similar stabilizing role in invertebrates and bacteria. In IBP, with the central Gly-Phe-Hyp-Gly-Glu-Arg sequence, one Glu is involved in an intrahelix water mediated interaction with Arg, while Glu (as well as the Arg sidechains) are involved in direct interactions with backbone peptide groups of neighboring molecules.

All sidechains in X and Y positions of the triple helix are exposed to solvent and appear to have multiple options of interacting through solvent and with available backbone carbonyl groups, in addition to sidechain-sidechain interactions. NMR studies on collagen fibrils show there are two interconverting conformations of Leu in fibrils, and considerable reorientation around the helix axis (Batchelder *et al.*, 1982; Sarkar *et al.*, 1983). This suggests there may be switching between alternative interaction sets in solution and even in fibrils. Thus, the sidechain orientations and interactions seen in the crystal structure may represent one of a number of possibilities, rather than a uniquely determined interaction.

E. Molecular Packing

One surprise in the X-ray diffraction studies was the resemblance of the molecular packing of peptides in the crystal structure to that seen by fiber diffraction for collagen molecules in tendon fibrils. The packing of molecules in the crystals of G→A, EKG, and POG is quasi-hexagonal with intermolecular spacings near 14Å, an arrangement and distance very similar to collagen molecular packing (Fig. 5; Fraser *et al.*, 1983). The 14Å spacing between triple-helices is spanned by highly ordered water molecules that connect neighboring molecules. This type of network is consistent with the attractive hydration forces that have been suggested to function in collagen assemblies (Leikin *et al.*, 1994, 1995, 1997). The strong similarity between peptide and collagen organization suggests that the 14Å lateral spacing and quasi-hexagonal packing observed in collagen assemblies are sequence-independent and determined by the effective diameters of the cylinder of hydration coating the triple helix.

The most common intermolecular interaction—observed in GEK, GPO, and T3-785 crystals—was Hyp-Hyp hydrogen bonds between neighboring triple-helices (Fig. 5; Berisio *et al.*, 2001; Kramer *et al.*, 2000, 2001). The observation of Hyp-Hyp intermolecular interactions in the peptide crystals and the report that recombinant unhydroxylated collagen will not form

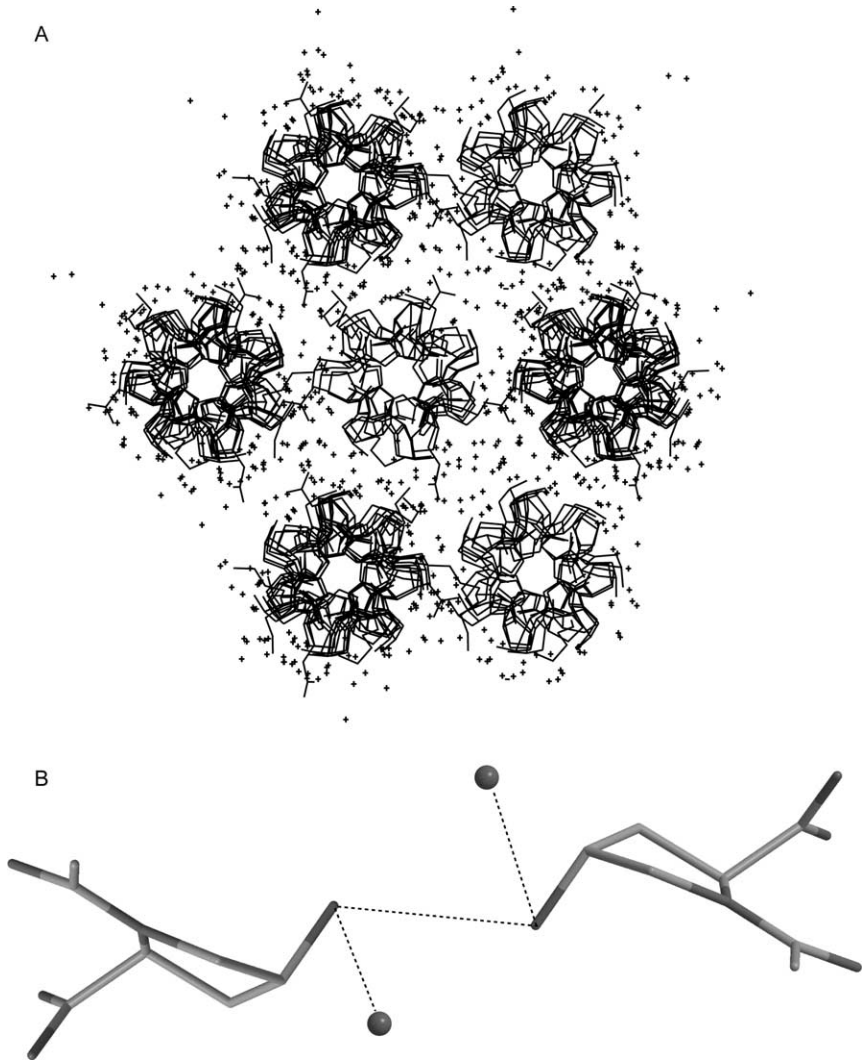

FIG. 5. (A) Cross-section of the molecular packing of the EKG peptide in the crystal. (B) The Hyp-Hyp interactions seen between neighboring molecules in the crystal.

fibrils (Perret *et al.*, 2001) support the original hypothesis of Gustavson (1955) that Hyp plays an important role in molecular association and fibril formation. There are Hyp residues all along triple helices, and direct interactions between two Hyp groups in adjacent molecules could be important in stabilization or orientation of molecular packing.

While the peptide lateral packing in crystals resembles that in collagen fibrils, the axial relationships, such as the D staggering (234 residues or 670Å), are likely to reflect interactions between charged and/or hydrophobic side chains. It was assumed that such interactions would be direct, between side chains of neighboring molecules, but in the crystal structures, an Arg or Lys group of one molecule usually binds to the backbone of a neighboring molecule (directly or through water) rather than to another side chain (Emsley *et al.*, 2004; Kramer *et al.*, 2001). The EKG peptide shows a staggering of neighboring molecules with charged Lys and Glu residues of one peptide aligned with the charged N and C-termini of other molecules (Kramer *et al.*, 2000), supporting the importance of electrostatic interactions in determining axial relationships between molecules.

F. Molecular Dynamics

Even though the collagen triple helix is considered a rigid rod, NMR, fluorescence, and molecular dynamics studies indicate there are sequence-dependent motions of the backbone and side chains. It appears that Gly-Pro-Hyp is the most rigid, as well as the most stabilizing, sequence. In model peptides, all backbone amides show high-order parameters typical of a rigid structure, but the Gly amide shows a slower hydrogen exchange rate when in a Gly-Pro-Hyp environment (Gly-Pro-Hyp-**Gly**-Pro-Hyp–Gly-Pro-Hyp) than in the imino acid poor environment (Gly-Ala-Arg-**Gly**-Leu-Ala-Gly-Pro-Hyp) (Fan *et al.*, 1993). Side chains of residues other than Pro and Hyp have considerable mobility in the collagen molecule, usually adopting more than one preferred orientation (Batchelder *et al.* 1982). Fluorescence studies of host-guest peptides showed that Trp in the Y position has a more restricted motion than Trp in the X position, which is consistent with the greater solvent accessibility of the X position (Simon-Lukasik *et al.*, 2003). Solid-state NMR studies of collagen with isotopically labeled residues indicate that the imino acid rings have angular fluctuations and that amino acid side chains are dynamic, reorienting around at least two side chain bonds (Sarkar *et al.*, 1983, 1987). Significant azimuthal motions around the helix axis were also observed (Sarkar *et al.*, 1983).

The sequence-dependent mobility of the collagen triple helix is important for its biological function. The unique collagenase cleavage site of types I and III collagen is at a boundary with a stable N-terminal region and a highly unstable C-terminal sequence (Fields, 1991). A molecular dynamics study carried out on the collagenase cleavage site of type III collagen, based on the crystal structure of peptide T3-785, suggested an alternative partially unfolded conformation that could expose the triple-helix backbone to cleavage (Stultz, 2002). Recently, experimental studies

indicate that collagenase binds and locally unwinds the triple-helix conformation in type I collagen prior to hydrolysis of its peptide bonds (Chung *et al.*, 2004). Locally unstable regions, which would be expected to be more dynamic, have also been implicated in collagen binding of various ligands as discussed below in Section VII.

Computational studies starting with collagen triple-helix structures now available in the Protein Data Bank are becoming increasingly important for collagen. As described above, molecular dynamics simulations can suggest models that are amenable to experimental verification. Computational analyses and molecular dynamics studies have also been carried out on Gly substitutions in the triple helix (see Section VIII below).

IV. Mechanism of Hydroxyproline Stabilization

Typically, the imino acid content of collagen is about 20%, with at least half in the form of hydroxyproline. Since collagens are the only animal proteins with a high content of hydroxyproline, an important role is indicated for this posttranslationally modified residue. Early proposals by Gustavson supported a role for Hyp in the stabilization of collagen fibrils, while later studies focused on a role in stabilization of the triple-helical molecule (Fraser and MacRae, 1973; Gustavson, 1955; Ramachandran, 1967). Evidence that Hyp contributes to triple-helix stability came from the greater stability of $(\text{Pro-Hyp-Gly})_{10}$ (T_m=60 °C) compared with $(\text{Pro-Pro-Gly})_{10}$ (T_m=30 °C), and from the decreased stability of unhydroxylated collagen synthesized in the presence of prolyl hydroxylase inhibitors (Rosenbloom *et al.*, 1973; Sakikabara *et al.*, 1973). Recent studies on recombinant collagen expressed in tobacco plants confirm the decrease in collagen stability when hydroxylation of proline is absent (Perret *et al.*, 2001).

The nature and the basis of the stabilizing effect of hydroxyproline have been actively investigated through studies on repeating polytripeptides and host-guest peptides (Table III). The predominant hydroxyproline found in collagen domains is 4-Hyp (i.e., on the gamma carbon of the imide ring). 4-Hyp is exclusively in the R diastereoisomer form and the importance of this stereospecificity is illustrated by the inability of $(\text{Pro-4SHyp-Gly})_{10}$ to form a triple helix (Inouye *et al.*, 1976). The location of (4R)-Hyp in the Y position of the Gly-X-Y repeating sequence is also critical since $(\text{4RHyp-Pro-Gly})_{10}$, with Hyp in the X position rather than the Y, does not adopt a triple-helical conformation (Inouye *et al.*, 1982). Interestingly, triple-helices can be formed with (4R)-Hyp in the X position if (4R)-Hyp or Thr is in the Y position (Table III; Bann and Bachinger, 2000; Berisio *et al.*, 2004). A very small amount of (3S)-Hyp is present in most

TABLE III
Thermal Stability of Repeating Polytripeptides $(Gly\text{-}X\text{-}Y)_{10}$ and Host-Guest Peptides, $(GPO)_3\text{-}GXY\text{-}(GPO)_4$, Containing Hydroxyproline and Fluoroproline*

Triplet	T_m (°C)		
	Repeating	Host-guest	References
Pro-Pro-Gly	33 (Persikov *et al.*, 2003)	45.5	(Persikov *et al.*, 2003)
Pro-(4R)Hyp-Gly	60 (Persikov *et al.*, 2003)	47.3	(Persikov *et al.*, 2003)
(4R)Hyp-Pro-Gly	<4	43.0	(Inouye *et al.*, 1982; Persikov *et al.*, 2003)
Pro-(4S)Hyp-Gly	<4		Inouye *et al.*, 1982
(4S)Hyp-Pro-Gly	<4		Inouye *et al.*, 1982
(4R)Hyp-(4R)Hyp-Gly	65	47.3	(Berisio *et al.*, 2004; Persikov *et al.*, 2003)
Gly-Pro-(3S)Hyp	<4	37.5	(Mizuno *et al.*, 2004)
Gly-(3S)Hyp-(4R)Hyp	<4	49.6	(Mizuno *et al.*, 2004)
Pro-(4R)Flp-Gly	87	43.7	(Holmgren *et al.*, 1999; Persikov *et al.*, 2003)
(4R)Flp-Pro-Gly	<4		(Doi *et al.*, 2003)
Pro-(4S)Flp-Gly	<4		(Bretscher *et al.*, 2001)
(4S)Flp-Pro-Gly	58		(Doi *et al.*, 2003)
Gly-Pro-Thr	<4		(Bann and Bachinger, 2000)
Gly-(4R)Hyp-Thr	19		(Bann and Bachinger, 2000)

* Hodges and Raines (2003) carried out experiments for (4R)Flp and (4S)Flp in the X position using a $(X\text{-}Y\text{-}Gly)_7$ design, and reached similar conclusions to those seen for the ten repeating units.

collagens in the X position, but model polytripeptides, including this residue in the X or Y position, were not triple-helical and it had a destabilizing influence in host-guest peptides (Jenkins *et al.*, 2003; Mizuno *et al.*, 2004).

The mechanism of stabilization by Hyp has generated considerable controversy (Fig. 6). Since the hydroxyl group of Hyp is directed outside and is unable to participate in any direct hydrogen bonds within the triple helix, the possibility of hydrogen bonding through water was suggested (Ramachandran *et al.*, 1973; Suzuki *et al.*, 1980). The strongest evidence for Hyp stabilization through water-mediated H bonding comes from Privalov's analysis showing that, in different collagens, increased Hyp content is correlated with increased enthalpic stabilization (Privalov,

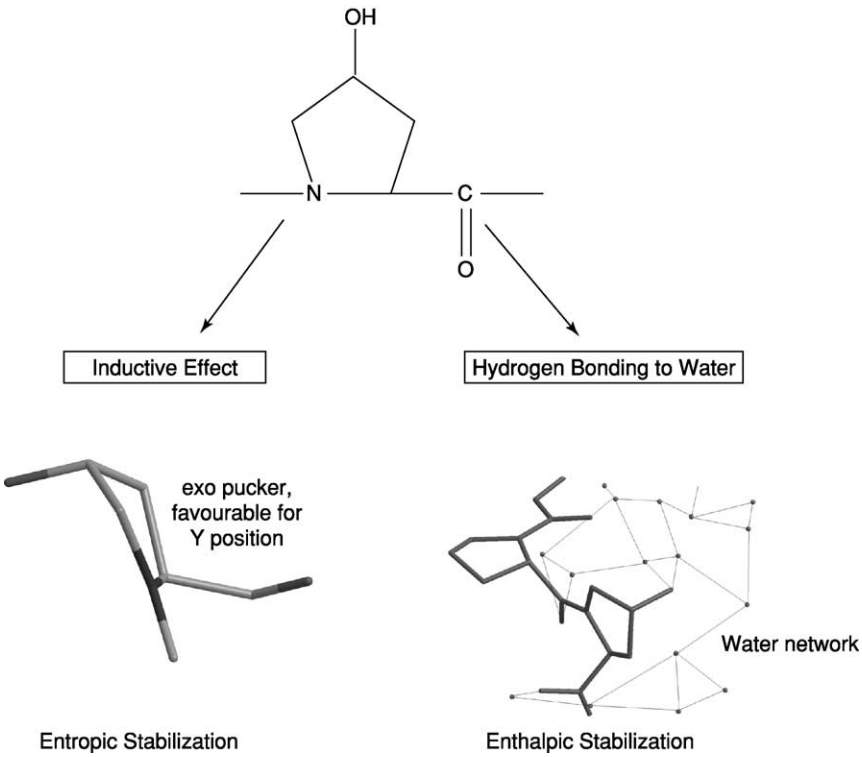

FIG. 6. Two proposed mechanisms for hydroxyproline stabilization.

1982). When the first high-resolution crystal structures were solved, hydroxyl groups of Hyp were seen to act as anchoring points for both intramolecular and intermolecular multispan water bridges (Bella *et al.*, 1994, 1995). It was suggested that the upward pucker of Hyp rings could be important because of Hyp's favorable orientation for binding to water networks. Arguments were raised about the existence of this hydration network in the crystal structures (Engel and Prockop, 1998), but such organized water was later confirmed in various laboratories for all triple-helical peptides. A highly ordered network was also seen in (Pro-Pro-Gly)$_{10}$ crystals, in the absence of Hyp, and questions were raised about whether the increased stability of (Pro-Hyp-Gly)$_{10}$ peptide originates from the hydrogen bonding of hydroxyprolines with the hydration network (Holmgren *et al.*, 1999).

In the late 1990s, Ronald Raines's group stimulated this area by taking a new approach, incorporating fluoroproline (Flp) residues in the design of collagen-like peptides (Jenkins and Raines, 2002). Fluorine has a greater electron-withdrawing effect on the imide ring than a hydroxyl group, and has a low tendency to form hydrogen bonds. The observation that (Pro-Flp-Gly)$_{10}$ (T_m~90 °C) was much more stable than (Pro-Hyp-Gly)$_{10}$ (T_m~60 °C) led to investigations about the basis of Flp stabilization and a reexamination of the nature of Hyp stabilization. Raines proposed that electron-withdrawing effects in (4R)-Flp would lead to a strong preference for the puckering of the imino ring in an up (or exo) conformation, and to the favoring of the trans vs. cis imide peptide bond (De Rider *et al.*, 2002). The puckering of imidic acids in the triple helix seen in the crystal structures largely confirmed the presence of down or endo for imino acids in the X position and up or exo conformation for imino acids in the Y position, as previously concluded from fiber diffraction models (Fraser *et al.*, 1979) and NMR data (Fan *et al.* 1993). Vitagliano *et al.* (2001) suggested that the stabilization of the triple helix by (4R)-Hyp in the Y position is due to its preference for the exo puckered conformation. Observations on the stabilizing effect of 4SFlp, which strongly prefers the down pucker in the X position, is consistent with this hypothesis (Table III; Doi *et al.*, 2003; Hodges and Raines, 2003). Hydrogen bonding through a water network is not a possibility for Flp, so its stabilization of the triple helix, which is even greater than that of Hyp, must come only from ring pucker and trans/cis preferences, mediated by its electronegative, inductive effect.

Thus, recent experiments involving Flp as well as Hyp, in different stereoisomeric forms and in X and Y positions, have made major contributions to our understanding of triple-helix stabilization (Berisio *et al.*, 2002, 2004; Persikov *et al.*, 2003). However, it is not clear that Flp and Hyp stabilize the collagen triple helix by the same mechanism. The experimental results support a predominant role for the inductive effect and its subsequent effect on the exo ring pucker in the stabilization of the triple helix by Flp. However, the large enthalpic stabilization seen for (Gly-Pro-Hyp)$_{10}$ and for collagens suggests that hydrogen bonding of Hyp through hydration networks is likely to play an important role in hydroxyproline stabilization, in addition to its favoring of the exo pucker.

The unique Hyp residue may be important both at the molecular and higher-order structure levels in collagen. Bacteria and viruses lack prolyl hydroxylase and have no hydroxyproline in their collagen-like domains. The significant differences in their amino acid composition and sequence compared to animal collagens suggest the use of alternative stabilization strategies for the triple helix (Rasmussen *et al.*, 2003).

V. Breaks in the Gly-X-Y Repeating Pattern

The basic amino acid features of the collagen triple helix can be reconsidered in light of the many sequences now available for this motif. Early studies on fibril-forming type I collagen indicated that Gly should be every third residue throughout a triple-helix domain. But it is clear that the sequences of most collagens and collagen-like domains include sporadic interruptions in the repeating pattern, and that the fibril-forming collagens, with their $(Gly\text{-}X\text{-}Y)_{338\text{-}341}$ structure, are unusual in having such a strict requirement. Type IV and VII collagen each contain more than 20 breaks in their very long triple-helical sequences. A number of collagens have between 2–10 interruptions (e.g., types VIII, X, VI), while some domains have a single break (e.g., mannose binding protein, SP-A, C1q) (Table I). The structural and functional implications of these breaks in the Gly-X-Y repeat may vary in different molecules. Early studies on C1q indicated that the site of the single interruption in the triple helix led to a rigid kink (Kilchherr *et al.*, 1985), while some breaks in type IV collagen and type XVI have been associated with flexible sites (Kassner *et al.*, 2004; Siebold *et al.*, 1987). As seen in the G→A peptide structure (where one Gly is replaced by an Ala) or the Hyp–peptide structure (where one Hyp is absent), one consequence of a break is to induce a discontinuity in registration between the domains on either side, serving to mark different regions (see Section VIII below; Bella *et al.*, 1994; Liu, 2000). A strict requirement for Gly as every third residue must be met in a domain of at least 6–7 tripeptide units in order to adopt a triple-helical conformation, but this conformation may be extended through breaks that do not meet this requirement.

VI. Amino Acid Sequence and Stability

All collagens have a high content of stabilizing imino acids, including the posttranslationally formed hydroxyproline, but there is a range of these important residues in collagens and collagen-like domains. In using translated sequences of DNA, it is assumed, based on experimental studies of collagens C1q and MBL, that all Pro residues in the Y position will be hydroxylated to Hyp, since prolyl hydroxylase specifically recognizes this position. Type I collagen has an equal amount of Pro and Hyp, while most collagens have more Hyp than Pro residues. The thermal stability of collagens from different animals correlates with their upper environmental temperature and with their imidic acid and Hyp content (Burjanadze, 2000; Rigby and Robinson, 1975).

Sequences of the form Gly-Pro-Hyp confer maximal stability to the collagen triple helix, while variations in the identities of the residues in the X and Y positions determine global thermal stability and modulate local stability and energetics that are required for self-association, recognition, and binding. Recent studies on recombinant collagen have shown that there are domains of varying stability along the collagen chain (Steplewski *et al.*, 2004). As seen in Section IV, the use of repeating polytripeptides have been extremely important in defining conformational features, but repeats of most tripeptide sequences in collagen will not lead to a stable triple helix. One approach to analysis of more varied collagen sequences has been host-guest peptides, where one or two tripeptide units are introduced into a constant, stabilizing Gly-Pro-Hyp framework. This approach has been used to determine the propensity of all 20 amino acids for the X position in a Gly-X-Hyp triplet, and for all 20 residues in the Y position in a Gly-Pro-Y triplet (Persikov *et al.*, 2000) The most stabilizing residues for the X position are Pro, Glu, Ala, Lys, Arg, Gln, and Asp, while the most stabilizing residues for the Y position are Hyp, Arg, Met, Ile, Gln, and Ala. The least stabilizing residues for both positions are the aromatic residues and Gly. These data provide a scale for the propensities of a given residue in the X or Y position in a homotrimer, measured in terms of the destabilizing influence of each residue compared to Gly-Pro-Hyp.

The propensities of individual residues for the triple-helix conformation may be affected by intramolecular interactions. For a given Gly-X1-Y1-Gly-X2-Y2 sequence, interchain interactions are sterically possible. Because the three chains (A, B, and C) are staggered by one residue, there are close contacts between sidechain X1 of chain A and Y1 of chain B, and between Y1 of chain C and X2 of chain A. In addition, intrachain X1–X2 and Y1–Y2 interactions are possible. Measurements of thermal stability of host-guest peptides with varying residues in the guest positions show that a subset of all sterically possible ion pairs lead to favorable interactions, and some favorable hydrophobic interactions as well (Persikov *et al.*, 2002, 2005). Helix stability is increased by interchain interactions such as Gly-Arg-Asp and Gly-Lys-Asp sequences, but the most energetically favorable interactions occur when sequences of the form Lys-Gly-Glu and Lys-Gly-Asp are present. The thermal stability of the 400 possible Gly-X-Y sequences are presented in Table IV, with the experimentally determined values shown in bold; those values predicted on the basis of the additivity of individual residues in the X and Y positions are shown in italics. A significant variation in thermal stability can be seen varying from the most stabilizing Gly-Pro-Hyp unit ($T_m = 47\,^{\circ}C$) to the low stability Gly-Gly-Phe ($T_m = 20\,^{\circ}C$).

TABLE IV
Predicted (*italics*) and Experimentally Observed (***bold***) T_m Values (°C) for all Possible Gly-X-Y Tripeptide Units in a Triple-Helix, Based on Host-Guest Peptide Studies*

X\Y	O	R	M	I	Q	A	V	E	T	C	K	H	S	D	G	L	N	Y	F	W
P	**47**	**47**	**43**	**42**	**41**	**41**	**40**	**40**	**40**	**38**	**37**	**36**	**35**	**34**	**33**	**32**	**30**	**30**	**28**	**26**
E	**43**	**40**	*38*	*37*	**38**	**35**	**35**	*35*	**36**	*33*	**35**	*31*	*31*	**30**	*29*	*28*	**30**	*26*	*24*	*22*
A	**42**	**38**	*37*	*36*	*36*	**33**	*34*	*34*	*34*	*32*	**31**	*30*	**33**	**33**	*27*	**28**	*26*	*25*	**22**	*21*
K	**42**	**39**	*37*	*36*	**39**	*35*	*34*	**35**	*34*	*32*	*31*	*30*	*29*	**36**	*27*	*27*	**32**	*24*	*23*	*20*
R	**41**	*41*	*36*	*35*	*35*	*34*	*33*	**34**	*33*	*31*	**30**	*29*	**31**	**35**	*26*	*26*	*25*	*24*	*22*	*19*
Q	**40**	**40**	*36*	*35*	*34*	*34*	*33*	*33*	*33*	*31*	**33**	*29*	*28*	*27*	*26*	*26*	*25*	*23*	*22*	*19*
D	**40**	**37**	*35*	*34*	*34*	**32**	*33*	*33*	*33*	*31*	**31**	*29*	*28*	*27*	*26*	*26*	*25*	*23*	*21*	*19*
L	**39**	*39*	*34*	*33*	**36**	**31**	*32*	*31*	*31*	*29*	**31**	*27*	*27*	*26*	*25*	**27**	*23*	*22*	*20*	*18*
V	**39**	*39*	*34*	*33*	*33*	*33*	*32*	*31*	*31*	*29*	**33**	*27*	*27*	*26*	*25*	*24*	*23*	*22*	*20*	*18*
M	**39**	*39*	*34*	*33*	*33*	*32*	*31*	*31*	*31*	*29*	**32**	*27*	*26*	*25*	*24*	*24*	*23*	*22*	*20*	*17*
I	**38**	*38*	*34*	*33*	*32*	**34**	*31*	*31*	*31*	*29*	*28*	*27*	*26*	*25*	*24*	*24*	*23*	*21*	*20*	*17*
N	**38**	*38*	*34*	*33*	*32*	*32*	*31*	*31*	*31*	*29*	*28*	*27*	*26*	*25*	*24*	*24*	*23*	*21*	*19*	*17*
S	**38**	*38*	*33*	*32*	*32*	*32*	*31*	*30*	*30*	*28*	*28*	*26*	*26*	*25*	*24*	*23*	*22*	*21*	*19*	*17*
H	**37**	*36*	*32*	*31*	*31*	*30*	*29*	*29*	*29*	*27*	*26*	*25*	*24*	*23*	*22*	*22*	*21*	*19*	*18*	*15*
T	**36**	*36*	*32*	*30*	*30*	*30*	*29*	*29*	*29*	*27*	*26*	*25*	*24*	*23*	*22*	*22*	*21*	*19*	*17*	*15*
C	**36**	*36*	*31*	*30*	*30*	*30*	*29*	*29*	*29*	*27*	*26*	*25*	*24*	*23*	*22*	*22*	*21*	*19*	*17*	*15*
Y	**34**	*34*	*30*	*29*	*28*	*28*	*27*	*27*	*27*	*25*	*24*	*23*	*22*	*21*	*20*	*20*	*19*	*17*	*15*	*13*
F	**34**	*33*	*29*	*28*	*28*	**24**	*26*	*26*	*26*	*24*	*23*	*22*	*21*	*20*	*19*	*19*	*18*	*16*	*15*	*12*
G	**33**	*33*	*29*	*27*	*27*	**26**	*26*	*26*	*26*	*24*	**27**	*22*	*21*	*20*	*19*	**25**	*18*	*16*	**20**	*12*
W	**32**	*32*	*27*	*26*	*26*	*26*	*25*	*24*	*24*	*22*	*21*	*20*	*20*	*19*	*18*	*17*	*16*	*15*	*13*	*11*

*The rows of amino acids are listed in order of their X position propensity for triple-helix formation, while the amino acids in columns are listed in order of their Y position propensity. The predicted T_m values are based on simple additivity of the stability of the residue in the X position and that in the Y position (see Persikov *et al.*, 2002).

Quantification of the intrinsic propensities and interactions for the most common sequences can now be used as a basis for predicting stability of peptides and local stability in collagens. Comparison of the predicted stability of more than 20 peptides with observed T_m values shows there is good agreement in the majority of the cases (Persikov *et al.*, 2002, 2005). This is an important asset in designing peptides that will form stable triple-helices, which are being used increasingly to map binding sites (see Section VII below) as well as for investigating interactions. It had been proposed some years ago that it would be possible to predict the local stability along collagen on the basis of individual tripeptide propensities (Bachinger and Davis, 1991), and these host-guest peptide data provide a realistic basis for such predictions.

VII. Ligand Binding

The binding of various ligands to the collagen triple helix is critical to biological function (Deprez *et al.*, 2000; Di Lullo *et al.*, 2002; Kadler, 1994; Knight *et al.*, 2000). Collagen binding to integrins and other cellular receptors mediates cell adhesion, while interaction of collagen with proteoglycans and other matrix molecules organizes the extracellular matrix, giving each tissue its distinctive mechanical properties. The binding of microbial surface components recognizing adhesive matrix molecules (MSCRAMMs) to collagen mediates bacterial adhesion to the integrins of the host cell (Patti *et al.*, 1995). The turnover of collagen involves recognition of one specific site on the triple helix by matrix metalloproteinases (collagenase) and the biosynthesis of collagen involves binding of the chaperone Hsp47. Binding of ligands to collagen-like domains in host-defense proteins plays an important role in their physiological activity. The complement serine proteinases C1r and C1s bind to the collagenous domain of C1q during activation, while the MASP serine proteases bind to mannose binding lectin (MBL). Polyanionic ligands, including oxidized LDL, interact with the triple-helix domain of the macrophage scavenger receptor (MSR); the collagenous domain of asymmetric AChE binds to heparan sulfate in the basal lamina of neuromuscular junctions. As the list of proteins with collagen domains grows, so does the number of observed and proposed interactions. It is not clear how interactions mediated by a collagen triple helix compare to those found with globular proteins in terms of the nature of the interactions, degree of specificity, and strength of binding.

Rotary shadowing and transmission electron microscopy of collagen with bound ligand or antibodies have been used to roughly locate the binding site, and binding assays using cyanogen bromide fragments of

collagen have often narrowed down the region of interest (Di Lullo *et al*, 2002). The precise definition of the sequence involved has been determined using triple-helical peptides that cover the region of interest (Glattauer *et al.*, 1997; Knight *et al.*, 2000), or mutational analysis for smaller collagen-like domains that can be studied by recombinant DNA methodology (Doi *et al.*, 1993; Wallis *et al*, 2004). The known binding sites along type I collagen, both at a low and high resolution level, have been mapped by the group of San Antonio, in order to envision their relationships with each other and the D-periodic fibril organization (DiLullo *et al.*, 2002).

Although more than 50 molecules are listed as interacting with type I collagen (Di Lullo *et al.*, 2002) and ligands are known to bind to collagenous domains in host defense protein, it has proved difficult to define $(\text{Gly-X-Y})_n$ sequence binding sites within the triple helix. At this time, the specific sequence in the triple helix responsible for binding has been defined in a small number of cases: for integrin binding to type I collagen; the binding of a monoclonal antibody to type III collagen; heparin sulfate binding to the collagenous tail of acetylcholinesterase; ligand binding to MSR; Hsp47 to type I collagen; decorin binding to type I collagen; and MASP binding to MBL (Table V). In addition, the location of the unique

TABLE V
List of Known Binding Sequences in Collagens and Collagen-Like Proteins

Binding of	Binds to	Sequence	Reference
Type I collagen	Integrin $\alpha1\beta1$, $\alpha2\beta1$	GFOGER	Knight *et al.*, 2000
	HSP-47	Gly-X-Arg	Koide *et al.*, 2002
	Heparin	KGHRGF	Sweeney *et al.*, 1998
	Collagenase (cleavage site)	GLA or GIA	Fields, 1991
	Crosslinking site	Gly-X-Hyl-Gly-His-Arg-Gly	Kadler, 1994
Type III collagen	Monoclonal antibody	GLAGAOGLR	Glattauer *et al.*, 1997
Type IV collagen	Integrin $\alpha1\beta1$	$\alpha1$(IV)Asp461, $\alpha2$(IV)Arg461	Golbik *et al.*, 2000
Type V collagen	Heparin	GKPGPRGQRGPTGPRGER	Delacoux *et al.*, 2000
Collagenous tail of AchE	Heparin	GROGRKGRO, GROGKRGKQGQK	Deprez *et al.*, 2000
MBL collagen domain	MASP	GLRGLQGPOGKLGPOG	Wallis *et al.*, 2004

collagenase cleavage site and the two crosslinking sites have been well established, defining the collagen sequences that are recognized (Fields, 1991; Kadler, 1994).

In some cases there is evidence that a local decrease in stability or looseness promotes binding. The type III collagen epitope for a monoclonal antibody is flanked by destabilizing Gly-Gly-Y triplets (Shah *et al.*, 1997), and characterization of model peptides for the type IV cell adhesion has suggested conformational heterogeneity (Sacca *et al.*, 2003). Triple-helical peptide models of the collagenous tail of asymmetric acetylcholinesterase (AChE) show more effective binding to heparin when their stability is lower (Deprez *et al.*, 2000). The recognition sequence for serine proteinases (MASPs) on MBL appears to be just C-terminal to the destabilizing Gly-Gln-Gly interruption in the triple helix (Wallis *et al.*, 2004), and the collagenase cleavage site has a C-terminal locally unstable region (Fields, 1991).

A major advance in understanding collagen-ligand interactions came with the high-resolution structure of cocrystals in the I domain of $\alpha 2\beta 1$ integrin with a triple-helical peptide $(Gly\text{-}Pro\text{-}Hyp)_2$-Gly-Phe-Hyp-Gly-Glu-Arg-$(Gly\text{-}Pro\text{-}Hyp)_3$ containing the known type I collagen binding sequence (designated the IBP peptide) (Fig. 7; Emsley *et al.*, 2000). The

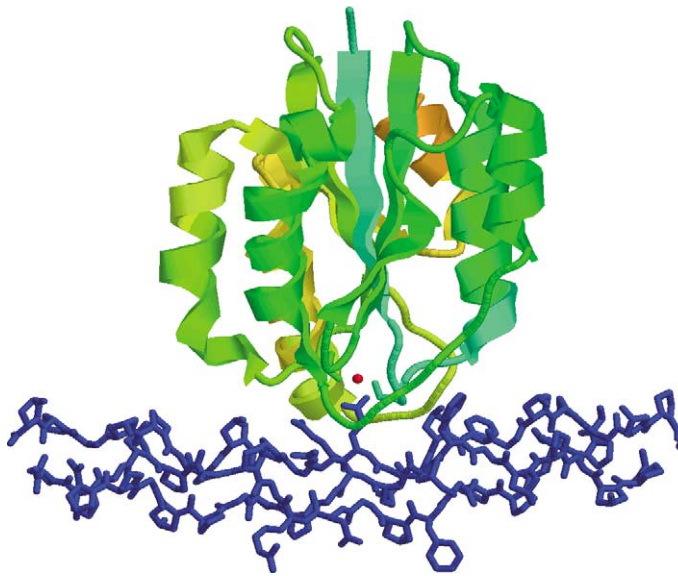

FIG. 7. Structure of the integrin I domain bound to the IBP triple-helix peptide.

three strands of this homotrimer—designated leading, middle, and trailing—have unique environments. Interactions with the I domain are largely mediated by the middle strand. The middle strand Glu completes the coordination sphere of a metal formed by the I domain MIDAS (metal ion-dependent adhesion site) motif, while the middle strand Arg forms a salt bridge to the I domain. The Phe residues of the middle and trailing strand make contact with the surface of the I domain. Peptide studies showed the Glu is essential and the Arg is required for high-affinity binding (Knight *et al.*, 2000). As discussed above in Section III.A, the IBP peptide has three distinct regions in terms of helix twist, and the kinking and bending at the junctions of these regions appears to play a critical role in optimizing interactions with the I domain (Emsley *et al*, 2004). The dominance of one strand in interactions with the integrin may have important implications for heterotrimeric collagen domains, and suggests the importance of recent strategies for synthesizing heterotrimeric collagen model peptides (e.g., see Sacca *et al.*, 2003).

VIII. Mutations and Disease

More than 1000 different mutations leading to human disorders have been observed in various types of collagens, and the clinical manifestations of these diseases reflect the specific tissue distribution of the particular collagen type that contains the mutation. Several excellent reviews of collagen diseases have been published (Byers and Cole, 2002; Myllyharju and Kivirikko, 2001, 2004). Well-characterized diseases include osteogenesis imperfecta (OI), where bone fragility results from mutations in type I collagen; Ehlers Danlos syndrome type IV, with aortic rupture as a result of mutations in type III collagen; Alport syndrome, with progressive renal failure resulting from basement membrane Type IV collagen mutations; and the dystrophic form of epidermolysis bullosa, with scarring and blistering of skin due to mutations in the skin anchoring fibrils Type VII collagen (Myllyharju and Kivirikko, 2001, 2004). Collagen mutations are generally dominant, because of their presence in multisubunit molecules and higher-order assemblies.

The first collagen genetic disease to be characterized was OI, a clinically heterogeneous disorder characterized by varying degrees of bone fragility (Byers and Cole, 2002). Cases are classified in four major categories, ranging from mild to perinatal lethal. OI results from mutations in the genes that code for the $\alpha 1$(I) or $\alpha 2$(I) chains of type I collagen, the major structural protein in bone. The majority of mutations involve single base substitutions in Gly codons in the triple helix, which lead to the replacement of a single Gly within the $(\text{Gly-X-Y})_{338}$ sequence by one of eight

bulkier amino acids (Ala, Arg, Asp, Cys, Glu, Ser, Trp, Val). More than 150 distinct mutations have been reported for different cases, and the mutation sites are located at varying sites along the length of both α1(I) and α2(I) collagen chains. The defective mineralization of bone collagen may be related to delayed collagen folding, a small decrease in helix stability, excess posttranslational modification, increased intracellular breakdown, reduced secretion, and abnormal fibril assembly (Byers and Cole, 2002). It is not known why different Gly substitution mutations lead to differing degrees of OI clinical severity, but it has been suggested to be related to the identity of the residue replacing Gly, the location of the mutation site with respect to the C-terminus, and the sequence immediately surrounding the mutation (Byers, 2001; Marini *et al.*, 1993). Disorders caused by single base substitutions in Gly codons within the triple helix are also common in other collagens, even in cases such as type IV collagen, where breaks in the Gly-X-Y repeating pattern are found normally (Hudson *et al.*, 2003).

Gly substitution mutations in the triple-helix domains of mannose-binding protein and C1q have also been associated with clinical disorders (Petry *et al.*, 1997; Turner, 2003). A Gly→Asp mutation in the fifth codon of the triple-helix domain of MBL is a common variant, reaching as high as 30% in some populations. This defect leads to MBL deficiency in the serum, resulting in susceptibility to infection, particularly in young children. Additional MBL mutations, Gly→Glu in the sixth codon of the triple helix and Arg→Cys in the Y position of the fourth codon are also observed. The Gly45Asp variant has been associated with defects in the oligomerization of MBL and also with binding to its associated MASPs, so that its function as host-defense is compromised, leading to increased infections (Wallis *et al.*, 2004).

Since Gly replacements are such a common feature in collagen diseases, it is important to understand the consequences of such a break in the Gly-X-Y repeating pattern for the collagen triple-helix structure, and to see if any molecular features correlate with clinical severity in OI. Peptides have been used to investigate the structural and energetic consequences of a Gly replacement on the collagen triple helix. A single replacement of a Gly by an Ala in $(\text{Pro-Hyp-Gly})_{10}$ was very destabilizing (Long *et al.*, 1993), and the crystal structure of this homotrimer G→A peptide showed only a small bulge at the Ala site, with a local disruption of NH...CO bonds, which are replaced by water-mediated hydrogen bonds (Bella *et al.*, 1994). The triple-helices in the Gly-Pro-Hyp sequences at both ends are well-ordered 7/2 helices, but the registration between the two ends is lost as a result of a slight untwisting at the replacement site. This loss of spatial coherence could lead to alterations in fibril formation and failure to

mineralize. More realistic peptide models, including a sequence of collagen with an OI site capped by C-terminal Gly-Pro-Hyp triplets, have been characterized by calorimetry, circular dichroism, and NMR spectroscopy (Baum and Brodsky, 1999). The introduction of a Gly to Ser or Ala replacement interrupts the C-to N-terminal folding of this peptide, suggesting that some renucleation event is needed to continue through Gly substitution sites.

The X-ray results on the Gly→Ala peptide have been complemented by computational studies to understand the effect of Gly replacements on the triple helix, largely carried out in Teri Klein's group (Klein and Huang, 1999; Mooney *et al.*, 2001; Radmer and Klein, 2004). Molecular dynamics and free energy calculations were done on the G→A peptide. All peptide models of mutations are limited by the use of homotrimers, while computational methods were used to evaluate the consequences of a more realistic model of having a Gly replacement in just one or two chains. More recently, molecular dynamic simulations were done on a sequence modeling the site of a lethal OI mutation, examining the effects of neighboring residues on the hydrogen bonding networks.

All Gly substitutions dramatically destabilize the triple helix in peptide models. Studies on host-guest peptides show the degree of destabilization depends on the identity of the residue replacing Gly. The residues, in order from least to most destabilizing, are: Ala < Ser < Cys < Arg <Val < Glu, Asp < Trp (Beck *et al.*, 2000). This order of destabilization correlates with the clinical severity of OI, demonstrated when two cases with different Gly replacements occur at the same site in α1(I) chains. For instance, Gly883Ser gives rise to a mild OI phenotype, while Gly883Asp gives rise to a lethal phenotype (Byers and Cole, 2002). The spectrum of amino acids replacing Gly in OI cases is significantly different from that predicted on the basis of nucleotide mutations rates and codon usage in type I collagen (Persikov *et al.*, 2004). This difference is most striking for the nonlethal cases, where the least and most destabilizing residues are underrepresented. This suggests that not every Gly replacement leads to a clinically detectable disorder, and supports the hypothesis that there is underrepresentation of the most destabilizing cases because they are not viable, and underrepresentation of the least destabilizing Gly→X substitutions because they are too mild to be clinically classified as OI. A schematic illustration of the number of different replacements observed in lethal and nonlethal cases of OI is shown in Fig. 8. The nonlethal and lethal categories represent observed data, which were scaled by the total number of Ser (the most frequent replacement in OI) and used to estimate the number that were undetectable for being too mild or too severe (Fig. 8).

more severe ---------- less severe	less destabilizing ---------- more destabilizing						
Gly→	Ala	Ser	Cys	Arg	Val	Glu	Asp
Not observed Normal	[17] 65%	[0] 0%	[1] 5%	[1] 3%	[1] 4%	[0] 0%	[0] 0%
Observed Non-lethal	6 23%	30 51%	14 70%	15 41%	2 9%	1 9%	1 2%
Observed Lethal	3 12%	29 49%	5 25%	13 35%	10 44%	1 9%	8 18%
Not observed Not viable	[0] 0%	[0] 0%	[0] 0%	[8] 21%	[10] 43%	[9] 82%	[36] 80%

FIG. 8. Schematic showing the relationship between the degree of destabilization caused by a Gly replacement by a given residue and the degree of clinical severity seen in OI. The nonlethal and lethal phenotypes are observed data, while the normal and nonviable are inferred by scaling the total number of each possible replacement to the number of observed for Ser, as described in Persikov *et al.* (2004).

IX. CONCLUSIONS

The discovery of collagen triple-helix domains in a whole range of proteins, including bacteria and viruses, illustrates the adaptability of this motif to a range of biological functions. As more human disorders are found to be related to mutations in the collagen triple helix, the functional importance of some collagen types is clarified. These mutations provide a motivation for basic research on sequence-dependent alterations of triple-helix properties, and for investigating how such alterations may lead to a disease phenotype. This past decade has been marked by the determination of high-resolution structures of collagen triple-helix peptides and the availability of NMR data on specifically labeled residues. These studies have been complemented by dynamic, thermodynamic, and computational analyses. In the coming years, crystal structures of heterotrimers and of other ligands bound to collagen domains will be needed to further clarify the molecular basis of function.

X. ABBREVIATIONS AND NOTATION

The standard one-letter and three-letter amino acid notations are used, with hydroxyproline designated as Hyp in the three-letter code and O in the one-letter code. Fluoroproline is designated as Flp. Abbreviations used for different proteins include MBL, mannose binding lectin, also known as mannose binding protein; SP-A and SP-D, surfactant apoproteins A and D; MSR, macrophage scavenger receptor; IBP, integrin binding peptide; Scl1 and Scl2,

Streptococcus collagen-like protein 1 and 2; BclA, Bacillus collagen-like protein in anthrax. Peptides are indicated by the notation used in Table II, or by the single letter code for the residues other than Gly-Pro-Hyp triplets.

The individual collagen types are denoted by roman numerals, with individual chain types indicating alpha chains. For example, $\alpha 1$(V) indicates the $\alpha 1$ chain of type V collagen.

The symmetry of the triple helix is designated here as 10/3 (twist angle of 36 degrees, 10-fold symmetry, which is also designated as 10_7 in crystallographic screw symmetry notation) and 7/2 (twist angle of 51.4 degrees, sevenfold symmetry, which is also designated as 7_5 in crystallographic screw symmetry notation).

Acknowledgments

This work is supported by National Institutes of Health (NIH) grant GM60048 (to B.B.), by the Michael Geisman OI Research Fellowship (to A.P.), and a Ruth Kirschstein NRSA NIH award (to B.B.).

References

Bachinger, H. P., and Davis, J. M. (1991). Sequence specific thermal stability of the collagen triple-helix. *Int. J. Biol. Macromol.* **13,** 152–156.

Bann, J. G., and Bachinger, H. P. (2000). Glycosylation/hydroxylation-induced stabilization of the collagen triple helix. 4-trans-hydroxyproline in the Xaa position can stabilize the triple helix. *J. Biol. Chem.* **275,** 24466–24469.

Batchelder, L. S., Sullivan, C. E., Jelinski, L. W., and Torchia, D. A. (1982). Characterization of leucine side-chain reorientation in collagen fibrils by solid state 2H NMR. *Proc. Natl. Acad. Sci.* **79,** 386–389.

Baum, J., and Brodsky, B. (1999). Folding of peptide models of collagen and misfolding in disease. *Curr. Opin. Struct. Biol.* **9,** 122–128.

Beck, K., Chan, V. C., Shenoy, N., Kirkpatrick, A., Ramshaw, J. A., and Brodsky, B. (2000). Destabilization of osteogenesis imperfecta collagen-like model peptides correlates with the identity of the residue replacing glycine. *Proc. Natl. Acad. Sci.* **97,** 4273–4278.

Bella, J., and Berman, H. M. (1996). Crystallographic evidence for C α-H$\cdots$O═C hydrogen bonds in a collagen triple helix. *J. Mol. Biol.* **264,** 734–742.

Bella, J., Eaton, M., Brodsky, B., and Berman, H. M. (1994). Crystal and molecular structure of a collagen-like peptide at 1.9 Å resolution. *Science* **266,** 75–81.

Bella, J., Brodsky, B., and Berman, H. M. (1995). Hydration structure of a collagen peptide. *Structure* **3,** 893–906.

Berendsen, H. J. C., and Migchelsen, C. (1965). Hydration structure of fibrous macromolecules. *Ann. N.Y. Acad. Sci.* **125,** 365–379.

Berisio, R., Vitagliano, L., Mazzarella, L., and Zagari, A. (2001). Crystal structure of a collagen-like polypeptide with repeating sequence Pro-Hyp-Gly at 1.4Å resolution: Implications for collagen hydration. *Biopolymers* **56,** 8–13.

Berisio, R., Vitagliano, L., Mazzarella, L., and Zagari, A. (2002). Recent progress on collagen triple helix structure, stability and assembly. *Prot. Pept. Lett.* **9,** 107–116.

Berisio, R., Granata, V., Vitagliano, L., and Zagari, A. (2004). Imino acids and collagen triple helix stability: Characterization of collagen-like peptides containing Hyp-Hyp-Gly sequence repeats. *J. Am. Chem. Soc.* **126,** 11402–11403.

Bretscher, L. E., Jenkins, C. L., Taylor, K. M., DeRider, M. L., and Raines, R. T. (2001). Conformational stability of collagen relies on a stereoelectronic effect. *J. Am. Chem. Soc.* **123,** 777–778.

Burjanadze, T. V. (2000). New analysis of the phylogenetic change of collagen thermostability. *Biopolymers* **53,** 523–528.

Byers, P. H. (2001). Folding defects in fibrillar collagens. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **356,** 151–158.

Byers, P. H., and Cole, W. G. (2002). Osteogenesis Imperfecta. *In* ''Connective tissue and its hereditable disorders,'' (P. M. Royce and B. Steinmann, Eds.), pp. 385–430. Wiley-Liss., New York.

Chung, L., Dinakarpandian, D., Yoshida, N., Lauer-Fields, J. L., Fields, G. B., Visse, R., and Nagase, H. (2004). Collagenase unwinds triple-helical collagen prior to peptide bond hydrolysis. *EMBO J.* **23,** 3020–3030.

Cowan, P. M., McGavin, S., and North, A. C. T. (1955). The polypeptide configuration of collagen. *Nature* **176,** 1062–1064.

De Rider, M. L., Wilkens, S. J., Waddell, M. J., Bretcher, L. E., Weinhold, F., Raines, R. T., and Markely, J. L. (2002). Collagen stability: Insights from NMR spectroscopic and hybrid density functional computational investigations of the effect of electronegative substituents on prolyl ring conformations. *J. Am. Chem. Soc.* **124,** 2497–2505.

Delacoux, F., Fichard, A., Cogne, S., Garrone, R., and Ruggiero, F. (2000). Unraveling the amino acid sequence crucial for heparin binding to collagen V. *J. Biol. Chem.* **275,** 29377–29382.

Deprez, P., Doss-Pepe, E., Brodsky, B., and Inestrosa, N. C. (2000). Interaction of the collagen-like tail of asymmetric acetylcholinesterase with heparin depends on triple-helical conformation, sequence and stability. *Biochem. J.* **350,** 283–290.

Di Lullo, G. A., Sweeney, S. M., Korkko, J., Ala-Kokko, L., and San Antonio, J. D. (2002). Mapping the ligand-binding sites and disease-associated mutations on the most abundant protein in the human, type I collagen. *J. Biol. Chem.* **277,** 4223–4231.

Doi, T., Higashino, K., Kurihara, Y., Wada, Y., Miyazaki, T., Nakamura, H., Uesugi, S., Imanishi, T., Kawabe, Y., Itakura, H., Yazaki, Y., Matsumoto, A., and Kodama, T. (1993). Charged collagen structure mediates the recognition of negatively charged macromolecules by macrophage scavenger receptors. *J. Biol. Chem.* **268,** 2126–2133.

Doi, M., Nishi, Y., Uchiyama, S., Nishiuchi, Y., Nakazawa, T., Ohkubo, T., and Kobayashi, Y. (2003). Characterization of collagen model peptides containing 4-fluoroproline; $(4(S)$-fluoroproline-Pro-Gly$)_{10}$ forms a triple helix, but $(4(R)$-fluoroproline-Pro-Gly$)_{10}$ does not. *J. Am. Chem. Soc.* **125,** 9922–9923.

Emsley, J., Knight, C. G., Farndale, R. W., Barnes, M. J., and Liddington, R. C. (2000). Structural basis of collagen recognition by integrin $\alpha 2\beta 1$. *Cell* **101,** 47–56.

Emsley, J., Knight, C. G., Farndale, R. W., and Barnes, M. J. (2004). Structure of the integrin alpha2beta1-binding collagen peptide. *J. Mol. Biol.* **335,** 1019–1028.

Engel, J., and Prockop, D. J. (1998). Does bound water contribute to the stability of collagen? *Matrix Biol.* **17,** 679–680.

Exposito, J. Y., Cluzel, C., Garrone, R., and Lethias, C. (2002). Evolution of collagens. *Anat. Rec.* **268,** 302–316.

Fan, P., Li, M., Brodsky, B., and Baum, J. (1993). Backbone dynamics of (Pro-Hyp-Gly)10 and a designed collagen-like triple-helical peptide by 15N NMR relaxation and hydrogen-exchange measurements. *Biochemistry* **32,** 13299–13309.

Fields, GB. (1991). A model for interstitial collagen catabolism by mammalian collagenases. *J. Theor. Biol.* **153,** 585–602.

Fields, G. B., and Prockop, D. J. (1996). Perspectives on the synthesis and application of triple-helical, collagen-model peptides. *Biopolymers* **40,** 345–357.

Fraser, R. D. B., and MacRae, T. P. (1973). ''Conformation of Fibrous Proteins.'' Academic Press, New York.

Fraser, R. D. B., MacRae, T. P., and Suzuki, E. (1979). Chain conformation in the collagen molecule. *J. Mol. Biol.* **129,** 463–481.

Fraser, R. D. B., MacRae, T. P., Miller, A., and Suzuki, E. (1983). Molecular conformation and packing in collagen fibrils. *J. Mol. Biol.* **167,** 497–521.

Fraser, R. D. B., MacRae, T. P., Chew, M. W. K, and Squire, J. M. (1987). Collagen and Elastin. *In* ''Fibrous Protein Structure,'' (J. M. Squire and P. J. Vibert, Eds.), pp. 173–191. Academic Press, London.

Glattauer, V., Werkmeister, J. A., Kirkpatrick, A., and Ramshaw, J. A. M. (1997). Identification of the epitope for a monoclonal antibody that blocks platelet aggregation by type III collagen. *Biochem. J.* **323,** 45–49.

Golbik, R., Eble, J. A., Ries, A., and Kuhn, K. (2000). The spatial orientation of the essential amino acid residues arginine and aspartate within the alpha1beta1 integrin recognition site of collagen IV has been resolved using fluorescence resonance energy transfer. *J. Mol. Biol.* **297,** 501–509.

Goodman, M., Bhumralkar, M., Jeffersohn, E. A., Kwak, J., and Locardi, E. (1998). Collagen mimetics. *Biopolymers* **47,** 127–142.

Grigera, J. R., and Berendsen, H. J. C. (1979). The molecular details of collagen hydration. *Biopolymers* **18,** 47–57.

Gustavson, K. H. (1955). The function of hydroxyproline in collagen. *Nature* **175,** 70–74.

Harrington, W. F., and von Hippel, P. F. (1961). The Structure of Collagen and Gelatin. *In* ''Advances in Protein Chemistry,'' (C. B. Anfinsen, Ed.), Vol. 16, pp. 1–138. Academic Press, New York.

Heidemann, E., and Roth, W. (1982). Synthesis and investigation of collagen model peptides. *Adv. Polym. Sci.* **43,** 142–203.

Hodges, J. A., and Raines, R. T. (2003). Stereoelectronic effects on collagen stability: The dichotomy of 4-fluoroproline diastereomers. *J. Am. Chem. Soc.* **125,** 9262–9263.

Holmgren, S. K., Bretscher, L. E., Taylor, K. M., and Raines, R. T. (1999). A hyperstable collagen mimic. *Chem. Biol.* **6,** 63–70.

Hudson, B. G., Tryggvason, K., Sundarmoorthy, M., and Neilson, E. G. (2003). Alport's syndrome, Goodpasture's syndrome, and type IV collagen. *N. Engl. J. Med.* **348,** 2543–2556.

Inouye, K., Sakakibara, S., and Prockop, D. J. (1976). Effects of the stereo-configuration of the hydroxyl group in 4-hydroxyproline on the triple-helical structures formed by homogenous peptides resembling collagen. *Biochim. Biophys. Acta* **420,** 133–141.

Inouye, K., Kobayashi, Y., Kyogoku, Y., Kishida, Y., Sakakibara, S., and Prockop, D. J. (1982). Synthesis and physical properties of (Hydroxyproline-Proline-Glycine)$_{10}$: Hydroxyproline in the X-position decreases the melting temperature of the collagen triple helix. *Arch. Biochem. Biophys.* **219,** 198–203.

Jenkins, C. L., and Raines, R. T. (2002). Insights on the conformational stability of collagen. *Nat. Prod. Rep.* **19,** 49–59.

Jenkins, C. L., Bretscher, L. E., Guzei, I. A., and Raines, R. T. (2003). Effect of 3-hydroxyproline residues on collagen stability. *J. Am. Chem. Soc.* **125,** 6422–6427.

Kadler, K. (1994). Extracellular matrix fibril forming collagen. *Protein Profile* **1,** 519–638.

Kassner, A., Tiedemann, K., Notbohm, H., Ludwig, T., Morgelin, M., Reinhardt, D. P., Chu, M.-L., Bruckner, P., and Grassel, S. (2004). Molecular structure and interaction of recombinant human type XVI collagen. *J. Mol. Biol.* **339,** 835–853.

Kawashima, H., Watanabe, N., Hirose, M., Sun, X., Atarashi, K., Kimura, T., Shikata, K., Matsuda, M., Ogawa, D., Heljasvaara, R., Rehn, M., Pihlajaniemi, T., and Miyasaka, M. (2003). Collagen XVIII, a basement membrane heparan sulfate proteoglycan, interacts with L-selectin and monocyte chemoattractant protein-1. *J. Biol. Chem.* **278,** 13069–13076.

Kielty, C. M., and Grant, M. E. (2002). The Collagen Family: Structure Assembly and Organization in the Extracellular Matrix. *In* ''Connective Tissue and its Heritable Disorders,'' (P. M. Royce and B. Steinmann, Eds.), pp. 159–222. Wiley-Liss, New York.

Kilchherr, E., Hofmann, H., Steigemann, W., and Engel, J. (1985). Structural model of the collagen-like region of C1q comprising the kink region and the fibre-like packing of the six triple helices. *J. Mol. Biol.* **186,** 403–415.

Klein, T. E., and Huang, C. C. (1999). Computational investigations of structural changes resulting from point mutations in a collagen-like peptide. *Biopolymers* **49,** 167–183.

Knight, C. G., Morton, L. F., Peachey, A. R., Tuckwell, D. S., Farndale, R. W., and Barnes, M. J. (2000). The collagen-binding A-domains of integrins $\alpha1\beta1$ and $\alpha2\beta1$ recognize the same specific amino acid sequence, GFOGER, in native (triple-helical) collagens. *J. Biol. Chem.* **275,** 35–40.

Koide, T., Takahara, Y., Asada, S., and Nagata, K. (2002). Xaa-Arg-Gly triplets in the collagen triple helix are dominant binding sites for the molecular chaperone HSP47. *J. Biol. Chem.* **277,** 6178–6182.

Kramer, R. Z., Bella, J., Mayville, P., Brodsky, B., and Berman, H. M. (1999). Sequence dependent conformational variations of collagen triple-helical structure. *Nat. Struct. Biol.* **6,** 454–457.

Kramer, R. Z., Venugopal, M. G., Bella, J., Mayville, P., Brodsky, B., and Berman, H. M. (2000). Staggered molecular packing in crystals of a collagen-like peptide with a single charged pair. *J. Mol. Biol.* **301,** 1191–1205.

Kramer, R. Z., Bella, J., Brodsky, B., and Berman, H. M. (2001). The crystal and molecular structure of a collagen-like peptide with a biologically relevant sequence. *J. Mol. Biol.* **311,** 131–147.

Krimm, S., and Kuroiwa, K. (1968). Low temperature infrared spectra of polyglycines and CH. . .O═C hydrogen bonding in polyglycine II. *Biopolymers* **6,** 401–407.

Leikin, S., Rau, D. C., and Parsegian, V. A. (1994). Direct measurement of forces between self-assembled proteins. Temperature-dependent exponential forces between collagen triple-helices. *Proc. Natl. Acad. Sci.* **91,** 276–280.

Leikin, S. Rau, D. C., and Parsegian, V. A. (1995). Temperature-favoured assembly of collagen is driven by hydrophilic not hydrophobic interactions. *Nat. Struct. Biol.* **2,** 205–210.

Leikin, S., Parsegian, V. A., Yang, W.-H., and Walrafen, G. E. (1997). Raman spectral evidence for hydration forces between collagen triple-helices. *Proc. Natl. Acad. Sci.* **94,** 11312–11317.

Liu, J. (2000). X-ray crystallographic studies of two nucleic-acid binding protein/complex systems and one collagen-like peptide. *Ph.D. thesis, Rutgers University.*

Long, C. G., Braswell, E., Zhu, D., Apigo, J., Baum, J., and Brodsky, B. (1993). Characterization of collagen-like peptides containing interruptions in the repeating GLy-X-Y sequence. *Biochemistry* **32,** 11688–11695.

Lu, J., the, C., Kishore, U., and Reid, K. B. M. (2002). Collectins and ficolins: Sugar pattern recognition molecules of the mammalian innate immune system. *Biochim. Biophys. Acta* **1572,** 387–400.

Marini, J. C., Lewis, M. B., Wang, Q., Chen, K. J., and Orrison, B. M. (1993). Serine for glycine substitutions in type I collagen in two cases of type IV osteogenesis imperfecta (OI). Additional evidence for a regional model of OI pathophysiology. *J. Biol. Chem.* **268,** 2667–2673.

Matsuzawa, Y., Funahashi, T., Kihara, S., and Shimomura, I. (2004). Adiponectin and metabolic syndrome. *Arterioscler. Thromb. Vasc. Biol.* **24,** 29–33.

McAlinden, A., Smith, T. A., Sandell, L. J., Ficheux, D., Parry, D. A. D., and Hulmes, D. J. S. (2003). Alpha-helical coiled coil oligomerization domains are almost ubiquitous in the collagen superfamily. *J. Biol. Chem.* **278,** 42200–42207.

Melacini, G., Bonvin, A. M., Goodman, M., Boelens, R., and Kaptein, R. (2000). Hydration dynamics of the collagen triple helix by NMR. *J. Mol. Biol.* **300,** 1041–1049.

Mizuno, K., Hayashi, T., Peyton, D. H., and Bachinger, H. P. (2004). The peptides acetyl-(Gly-3(S)Hyp-4(R)Hyp)$_{10}$-NH2 and acetyl-(Gly-Pro-3(S)Hyp)$_{10}$-NH2 do not form a collagen triple helix. *J. Biol. Chem.* **279,** 282–287.

Mooney, S. D., Huang, C. C., Kollman, P. A., and Klein, T. E. (2001). Computed free energy differences between point mutations in a collagen-like peptide. *Biopolymers* **58,** 347–353.

Myllyharju, J., and Kivirikko, K. I. (2001). Collagen and collagen-related diseases. *Ann. Med.* **33,** 7–21.

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

Okuyama, K., Arnott, S., Takayanagi, M., and Kakudo, M. (1981). Crystal and molecular structure of a collagen-like polypeptide (Pro-Pro-Gly)$_{10}$. *J. Mol. Biol.* **152,** 427–443.

Patti, J. M., House-Pompeo, K., Boles, J. O., Garza, N., Gurusiddappa, S., and Hook, M. (1995). Critical residues in the ligand-binding site of the Staphylococcus aureus collagen-binding adhesion (MSCRAMM). *J. Biol. Chem.* **270,** 12005–12011.

Perret, S., Merle, C., Bernocco, S., Berland, P., Garrone, R., Hulmes, D. J., Theisen, M., and Ruggiero, F. (2001). Unhydroxylated triple helical collagen I produced in transgenic plants provides new clues on the role of hydroxyproline in collagen folding and fibril formation. *J. Biol. Chem.* **276,** 43693–43698.

Persikov, A. V., Ramshaw, J. A., Kirkpatrick, A., and Brodsky, B. (2000). Amino acid propensities for the collagen triple-helix. *Biochemistry* **39,** 14960–14967.

Persikov, A. V., Ramshaw, J. A., Kirkpatrick, A., and Brodsky, B. (2002). Peptide investigations of pairwise interactions in the collagen triple-helix. *J. Mol. Biol.* **316,** 385–394.

Persikov, A. V., Ramshaw, J. A. M., Kirkpatrick, A., and Brodsky, B. (2005). Electrostatic interactions involving lysine make major contributions to collagen triple-helix stability. *Biochemistry* **44,** 1414–1422.

Persikov, A. V., Ramshaw, J. A., Kirkpatrick, A., and Brodsky, B. (2003). Triple-helix propensity of hydroxyproline and fluoroproline: Comparison of host-guest and repeating tripeptide collagen models. *J. Am. Chem. Soc.* **125,** 11500–11501.

Persikov, A. V., Pillitteri, R. J., Amin, P., Schwarze, U., Byers, P. H., and Brodsky, B. (2004). Stability related bias in residues replacing glycines within the collagen triple helix (Gly-Xaa-Yaa) in inherited connective tissue disorders. *Hum. Mut.* **24,** 330–337.

Petry, F., Hauptmann, G., Goetz, J., Grosshans, E., and Loos, M. (1997). Molecular basis of a new type of C1q deficiency associated with a non-functional low molecular weight (LMW) C1q: Parallels and differences to other known genetic C1q defects. *Immunopharmacology* **38,** 189–201.

Privalov, P. L. (1982). Stability of proteins: Proteins which do not present a single cooperative system. *Adv. Prot. Chem.* **35,** 1–104.

Radmer, R. J., and Klein, T. E. (2004). Severity of osteogenesis imperfecta and structure of a collagen-like peptide modeling a lethal mutation site. *Biochemistry* **43,** 5314–5323.

Ramachandran, G. N. (1967). Structure of Collagen at the Molecular Level. *In* ''Treatise on Collagen,'' (G. N. Ramachandran, Ed.), pp. 103–183. Academic Press, New York.

Ramachandran, G. N., and Kartha, G. (1954). Structure of collagen. *Nature* **174,** 269–270.

Ramachandran, G. N., and Kartha, G. (1955). Structure of collagen. *Nature* **176,** 593–595.

Ramachandran, G. N., and Chandrasekharan, R. (1968). Interchain hydrogen bonds via bound water molecules in the collagen triple-helix. *Biopolymers* **6,** 1649–1658.

Ramachandran, G. N., Bansal, M., and Bhatnagar, R. S. (1973). A hypothesis on the role of hydroxyproline in stabilizing collagen structure. *Biochim. Biophys. Acta* **322,** 166–171.

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

Rich, A., and Crick, F. H. C. (1955). The structure of collagen. *Nature* **176,** 915–916.

Rich, A., and Crick, F. H. C. (1961). The molecular structure of collagen. *J. Mol. Biol.* **3,** 483–506.

Rigby, B. J., and Robinson, M. S. (1975). Thermal transitions in collagen and the preferred temperature range of animals. *Nature* **253,** 277–279.

Rosenbloom, J., Harsh, M., and Jimenez, S. (1973). Hydroxyproline content determines the denaturation temperature of chick tendon collagen. *Arch. Biochem. Biophys.* **158,** 478–484.

Sacca, B., Fiori, S., and Moroder, L. (2003). Studies of the local conformational properties of the cell-adhesion domain of collagen type IV in synthetic heterotrimeric peptides. *Biochemistry* **42,** 3429–3436.

Sakikabara, S., Inouye, K., Shudo, K., Kishida, Y., Kobayashi, Y., and Prockop, D. J. (1973). Synthesis of $(Pro\text{-}Hyp\text{-}Gly)_n$ of defined molecular weights. Evidence for stabilization of collagen triple-helix by hydroxyproline. *Biochim. Biphys. Acta* **303,** 198–202.

Sarkar, S. K., Sullivan, C. E., and Torchia, D. A. (1983). Solid state 13C NMR study of collagen molecular dynamics in hard and soft tissues. *J. Biol. Chem.* **258,** 9762–9767.

Sarkar, S. K., Hiyama, Y., Niu, C. H., Young, P. E., Gerig, J. T., and Torchia, D. A. (1987). Molecular dynamics of collagen side chains in hard and soft tissues. A multinuclear magnetic resonance study. *Biochemistry* **26,** 6793–6800.

Shah, N. K., Sharma, M., Kirkpatrick, A., Ramshaw, J. A. M., and Brodsky, B. (1997). Gly-Gly containing triplets of low stability adjacent to a type III collagen epitope. *Biochemistry* **36,** 5878–5883.

Shirai, H., Murakami, T., Yamada, Y., Doi, T., Hamakubo, T., and Kodama, T. (1999). Structure and function of type I and II macrophage scavenger receptors. *Mech. Ageing Dev.* **111,** 107–121.

Siebold, B., Qian, R. A., Glanville, R. W., Hofmann, H., Deutzmann, R., and Kuhn, K. (1987). Identification of the cleavage sites by a hemorrhagic metalloproteinase in type IV collagen. Construction of a model for the aggregation and cross-linking region (7S domain) of type IV collagen based upon an evaluation of the primary structure of the alpha 1 and alpha 2 chains in this region. *Eur. J. Biochem.* **168,** 569–575.

Simon-Lukasik, K. V., Persikov, A. V., Brodsky, B., Ramshaw, J. A., Laws, W. R., Ross, A., and Ludescher, R. D. (2003). Fluorescence determination of tryptophan side-chain accessibility and dynamics in triple-helical collagen-like peptides. *Biophys. J.* **84,** 501–508.

Steplewski, A., Majsterek, I., McAdams, E., Rucker, E., Brittingham, R. J., Ito, H., Kazuya, H., Adachi, E., Jimenez, S. A., and Fertala, A. (2004). Thermostability gradient in the collagen triple-helix reveals its multi-domain structure. *J. Mol. Biol.* **338,** 989–998.

Stetefeld, J., Frank, S., Jenny, M., Schulthess, T., Kammerer, R. A., Boudko, S., Landwehr, R., Okuyama, K., and Engel, J. (2003). Collagen stabilization at atomic level: Crystal structure of designed $(GlyProPro)_{10}$-foldon. *Structure* **11,** 339–346.

Stultz, C. M. (2002). Localized unfolding of collagen explains collagenase cleavage near imino-poor sites. *J. Mol. Biol.* **319,** 997–1003.

Suzuki, E., Fraser, R. D. B., and MacRae, T. P. (1980). Role of hydroxyproline in the stabilization of the collagen molecule via water molecules. *Int. J. Biol. Macrom.* **2,** 54–56.

Sweeney, S. M., Guy, C. A., Fields, G. B., and San Antonio, J. D. (1998). Defining the domains of type I collagen involved in heparin-binding and endothelial tube formation. *Proc. Natl. Acad. Sci.* **95,** 7275–7280.

Sylvestre, P., Couture-Tose, E., and Mock, M. (2003). Polymorphism in the collagen-like region of the Bacillus anthracis BclA protein leads to variation in exosporium filament length. *J. Bacteriol.* **185,** 1555–1563.

Traub, W., and Piez, K. A. (1971). The chemistry and structure of collagen. *Adv. Prot. Chem.* **25,** 243–352.

Turner, M. W. (2003). The role of mannose binding lectin in health and disease. *Mol. Immunol.* **40,** 423–429.

Vitagliano, L., Nemethy, G., Zagari, A., and Scheraga, H. A. (1993). Stabilization of the triple-helical structure of natural collagen by side chain interactions. *Biochemistry* **32,** 7354–7359.

Vitagliano, L., Berisio, R., Mazzarella, L., and Zagari, A. (2001). Structural bases of collagen stabilization induced by proline hydroxylation. *Biopolymers* **58,** 459–464.

Wallis, R., Shaw, J. M., Uitdehaag, J., Chen, C. B., Torgersen, D., and Drickamer, K. (2004). Location of the serine protease-binding sites in the collagen-like domain of mannose-binding protein. *J. Biol. Chem.* **279,** 14065–14073.

Xu, Y., Keene, D. R., Bujnicki, J. M., Hook, M., and Lukomski, S. (2002). Streptococcal Scl1 and Scl2 proteins form collagen-like triple-helices. *J. Biol. Chem.* **277,** 27312–27318.

Xu, Y., Hyde, T., Wang, X., Bhate, M., and Brodsky, B. (2003). NMR and CD spectroscopy show that imino acid restriction of the unfolded state leads to efficient folding. *Biochemistry* **42,** 8696–8703.

Yonath, A., and Traub, W. (1975). Crystallization of cyanogen bromide peptides from chick cartilage collagen. *FEBS Lett.* **57,** 93–95.

Yee, R. W., Englander, S. W., and Von Hippel, P. H. (1974). Native collagen has a two-bonded structure. *J. Mol. Biol.* **83,** 1–16.