Review

Journal of PeptideScience

Received: 1 July 2008 Revised: 23 September 2008 Accepted: 24 September 2008 Published online in Wiley Interscience: 3 December 2008

(www.interscience.com) DOI 10.1002/psc.1082

Role of side chains in collagen triple helix stabilization and partner recognition‡

Rita Berisio,[a] Alfonso De Simone,[a,b] Alessia Ruggiero,[a] Roberto Improta[a] and Luigi Vitagliano[a]*

Collagen is a widespread protein family involved in a variety of biological processes. The complexity of collagen and its fibrous nature prevent detailed investigations on the full-length protein. Reductionist approaches conducted by dissecting the protein complexity through the use of model peptides have proved to be quite effective. There are, however, several issues regarding structure–stability relationships, aggregation in higher-order assemblies, and partner recognition that are still extensively investigated. In this review, we discuss the role that side chains play in triple helix stabilization and in partner recognition. On the basis of recent literature data, we show that collagen triple helix stability is the result of the interplay of different factors. As a general trend, interactions established by amino/imino acid side chains within the triple helix scaffold effectively modulate the intrinsic residue propensity for this common structural motif. The use of peptide models has also highlighted the role that side chains play in collagen self-association and in its interactions with receptors. Valuable examples in these fields are illustrated. Finally, future actions required to obtain more detailed information on the structure and the function of this complex protein are also delineated.

Keywords: collagen; iminoacids; triple helix; protein stability; molecular recognition

Background

Collagen is the most abundant protein in vertebrates as it is an essential structural component of all connective tissues such as cartilage, bones, ligaments, and skin. In these species, it accounts for one-third of the total protein weight. Collagen-like molecules are also widespread in other organisms. Indeed, they have been found in lower eukaryotic, prokaryotic, and viral genomes [1,2].

The unique shape and properties of collagen molecules are due to their peculiar amino acid composition and sequence. Collagen sequences are characterized by the repetition of triplets of the type Gly-Xaa-Yaa. Although all types of amino acids may be located at positions X and Y of the triplets, they are frequently occupied by iminoacids [Pro and its post-translationally modified form 4*R*-hydroxy-2*S*-proline (4*R*Hyp)]. In vertebrate collagens, as a consequence of the specificity of the prolyl-hydroxylation process, Pro and 4*R*Hyp are almost exclusively located at the X and Y positions, respectively. From the structural point of view, collagen single molecules are composed of three polypeptide chains in polyproline II (PPII) conformation, each containing hundreds of amino acid triplets, wrapped around a common axis (triple-helix motif).

The complexity of the collagen molecule and its fibrous nature prevent detailed investigations on the full-length protein. To overcome this limitation, a number of diversified strategies have been adopted. Because of the repetitive nature of collagen sequence/structure, the use of peptide models embedding specific motifs has been successful [3–8]. Through this approach important milestones have been achieved. Structural characterizations of collagen-like peptide models have provided an atomic resolution picture of collagen triple helix [9–28] (Table 1). Crystallographic analyses have also shown the relationship between sequence and global/local triple helix structure. These investigations have paved the way for subsequent theoretical investigations. In this framework, X-ray models were used as starting structures in several MD simulations [29–34]. Furthermore, quantum mechanics studies have been performed to address specific unsolved issue related to collagen structure and stability [29,35–38].

In addition, the design and the characterization of host–guest peptide models embedding naturally encoded amino acids have led to the generation of a detailed propensity scale for collagen triple helix [4,39–41]. These data have been of fundamental importance for unveiling sequence–stability and structure–stability relationships. Finally, through the use of the model peptide approach, new compounds with a triple helical scaffold endowed with special properties (themostability, inhibition of metalloproteases, and ability to mimic and modulate collagen interactions with its partners) have been developed [42–44]. These peptides have shown an important potential for applications in diversified fields (medicine, biology, and bioengineering). Although investigations on peptide models have generated a wealth of information on collagen structure, stability, and function, there are several aspects of these fundamental issues that are highly debated.

Over the years, several excellent reviews have been reported on the use of triple-helical peptides for elucidating collagen structure and function [3,5,6,8,40,45–49]. Here we illustrate, on the basis of

* *Correspondence to: Luigi Vitagliano, Istituto di Biostrutture e Bioimmagini, CNR via Mezzocannone 16, I-80134 Napoli, Italy. E-mail: luigi.vitagliano@unina.it*

a *Istituto di Biostrutture e Bioimmagini, CNR via Mezzocannone 16, I-80134 Napoli, Italy*

b *Department of Chemistry, University of Cambridge, Lensfield Road, CB2 1EW Cambridge, UK*

‡ *11th Naples Workshop on Bioactive Peptides.*

Biography

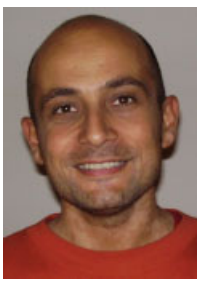
Dr Rita Berisio was born in 1972 in Naples, Italy. She graduated in chemistry at the University of Naples 'Federico II' (1995). From 1996 to 1998 she worked at the European Molecular Biology Lab in Hamburg and in 1999 she received her Ph.D. at the University of Naples. In 2000, Rita Berisio has been a post-doc of the University of Naples and in 2001 she became a permanent researcher at the Institute of Biostructures and Bioimaging of the Italian Research National Council, in Naples, Italy. In 2002–2004 she joined the group of Prof. Ada Yonath at the Max-Planck-Institute in Hamburg, Germany. Rita Berisio has been awarded the prize of the Italian Crystallography Association (AIC) in 2006 for her contribution to structural biology. The research activities of Rita Berisio are focussed on several aspects of the structure/function relationship in macromolecules of biological interest. These activities have been carried out by combining macromolecular crystallography with other physical–chemical techniques (CD spectroscopy, molecular modelling, molecular dynamics and statistical analyses).

Dr Alfonso De Simone earned his B.Sc. degree in chemistry at the University of Naples Federico II, Italy, in 2003. From 2004 to 2007 he worked in different European labs including the National Institute for Medical Research (London, UK) in the lab of Prof. Guy Dodson, and AMOLF (Amsterdam, NL) in the lab of Prof. Daan Frenkel. He received his Ph.D. at the University of Padua, Italy, in 2007. Currently, He is an European Molecular Biology Organization (EMBO) postdoctoral fellow at the University of Cambridge in the group of Prof. Chris Dobson. So far Dr De Simone has focused on the physical chemistry of proteins by carrying out researches in the field of protein hydration, protein misfolding, free energy of amyloid formation and statistical approaches to structural databases. His current research activity is mainly focused on the development of new approaches to integrate nuclear magnetic resonance (NMR) and simulations for an accurate determination of structure and dynamics of proteins in solution.

available literature data, the role played by collagen side chains in stabilizing the triple helix, in directing collagen association in large assemblies, and in modulating collagen interactions with its biological partners. A particular attention is devoted to the structural aspects of the topic. Collagen triple-helix symmetry and its dependence on local sequence have been extensively discussed in both original and review papers [12,26,50–53], and hence will not be considered here.

Biography

Dr Alessia Ruggiero graduated in Chemistry at the University of Naples 'Federico II', Italy, in 2001. She received her Ph.D. in Molecular Physiology and Structural Biology from the University of Padua, Italy, in 2004. Since 2004, she is a postdoc at the 'Istituto di Biostrutture e Bioimmagini' Napoli of the National Research Council (CNR). In 2005–2006, she spent 6 months as visiting postdoc at the European Molecular Biology Laboratory (EMBL) outstation of Hamburg. The research activities of Dr Ruggiero are focused on the structural characterization of biological macromolecules. Over the years, she has been working on different systems involved in a variety of biological processes (protein biosynthesis, redox regulation and bacterial infection). Dr Ruggiero's expertise includes molecular biology, protein biochemistry, X-ray crystallography, spectroscopy and molecular dynamics.



Dr Roberto Improta was born in Naples on 12 May 1970. He graduated and received a Ph.D. in Chemistry in 1998 from the University of Naples 'Federico II'. He spent his postdoctoral fellowships in Pisa, in Houston, and in Naples. Since 2001, he has been a researcher at the Italian National Research Council (CNR). Beyond the theoretical study of the spectroscopic properties of organic molecules in solution, his main research interest concerns the validation and application of theoretical models and computational methods for the study of large size molecules, with special attention to the structure and reactivity of excited states in the condensed phase.

Collagen Triple-Helix Stability: Role of Imino Acids

Unveiling sequence–stability and structure–stability relationships is a major goal of protein chemistry and structural biology. Despite the enormous efforts devoted, answers to these issues remain elusive [54]. In principle, if compared with globular proteins, collagen represents an ideal system for such investigations because of its simplified sequence and regular structure. However, the definition of the molecular basis of collagen triple-helix stability has so far proved to be a difficult task.

Proline and its derivatives have a prominent role in the modulation of collagen triple-helix stability. Simple considerations on the conformation adopted by the individual polypeptide chains within the triple helix (PPII) and the high propensity exhibited by proline residues for this specific secondary structure element provide a straightforward explanation for the major role played by this residues in collagen [55,56]. Nevertheless, the rationalization

Journal of PeptideScience

Biography

Dr Luigi Vitagliano is currently Senior Scientist at the 'Istituto di Biostrutture e Bioimmagini' Napoli of the National Research Council (CNR). He received his degree in Chemistry (1990) and his Ph.D. in Chemical Sciences (1994) at the University of Naples 'Federico II'. Dr Vitagliano has carried out research activities in several international institutions (Mount Sinai School of Medicine, USA; CNRS Strasburg, France; Rutgers University, USA; University of Aarhus, Denmark) as visiting Ph.D. student, Postdoc or scientist. He is (or has been) the project leader of projects funded by regional, national and European institutions. Dr Vitagliano is a member of ELETTRA Synchrotron Review Committee for beam-time allocation. The research activities of Dr Vitagliano have been focused in the large area of structural characterization of biological macromolecules. Indeed, most of his studies have been conducted on structure-function relationships of proteins and polypeptides. These analyses have been performed by using both computational and experimental techniques. In particular, the expertise are of Dr Vitagliano include X-ray crystallography, spectroscopy, molecular modeling, molecular dynamics and data mining in structural databases. Dr Vitagliano is coauthor of nearly 100 publications on international journals or books (85 in peer-revied journals) and one patent.

of the stabilizing/destabilizing effects produced by proline derivatives was unexpectedly complicated. Indeed, although proline and its derivatives share common structural features they have diversified effects on the triple helix [57,58].

The high abundance of the specific diastereoisomer 4*R*Hyp in collagen sequences has attracted the attention of the scientific community since the early 1950s [59]. Collagen sequence analyses and the use of peptide models (Table 2) have clearly demonstrated that 4Hyp residues strongly stabilize collagen triple helix when the residue is located in the Y position of the recurrent Gly-Xaa-Yaa motif [55,58]. Intriguingly, the same diastereoisomer has strong destabilizing effect when located at the X position. It is also worth mentioning that the diastereoisomer 4*S*-hydroxy-2*S*-proline (4*S*Hyp) always destabilizes the triple helix independently of the position assumed (Table 2) [57]. Although structural differences between proline and its hydroxylated derivatives are minimal (a simple OH group), no obvious explanation for this observation is available. Indeed, any structural mechanism should simultaneously answer the following questions: (a) why does proline hydroxylation produce contrasting effect depending on modification site (X or Y position) and (b) why do similar diastereoisomers, such as 4*R*Hyp and 4*S*Hyp, exhibit very different behaviors?

The initially proposed models predicted an essential role for water molecules. This hypothesis, originally proposed by Ramachandran [67], suggests that water molecules can mediate hydrogen bonding interactions between the OH group of Hyp residues and the rest of the protein. Although high-resolution structure determinations of collagen models have shown that these peptides are indeed highly hydrated [9–11], the inability of

Table 1. The presence of amino acids in collagen-like peptides whose three-dimensional structure has been determined by X-ray crystallography

Amino acid position	Peptide sequence	PDB code	Resolution (Å)
Ala			
X position	$(POG)_3ITGARGLAG(PPG)_4$	1BKV	2.00
Y position	$(POG)_3ITGARGLAG(PPG)_4$	1BKV	2.00
Arg			
X position	$(GPO)_4GPRGRT(GPO)_4$	2F6A	3.29
Y position	$(GPO)_4GPRGRT(GPO)_4$	2F6A	3.29
Y position	$(POG)_3ITGARGLAG(PPG)_4$	1BKV	2.00
Glu			
X position	$(GPO)_2GFOGER(GPO)_3$	1Q7D	1.80
X position	$(POG)_4EKG(POG)_5$	1QSU	1.75
Ile			
X position	$(POG)_3ITGARGLAG(PPG)_4$	1BKV	2.00
Leu			
X position	$(POG)_4LOG(POG)_5$	2DRT	1.60
X position	$(POG)_4(LOG)_2(POG)_5$	2DRX	1.40
X position	$(POG)_3ITGARGLAG(PPG)_4$	1BKV	2.00
Lys			
Y position	$(POG)_4EKG(POG)_5$	1QSU	1.75
Phe			
X position	$(GPO)_2GFOGER(GPO)_3$	1Q7D	1.80
Thr			
Y position	$(POG)_3ITGARGLAG(PPG)_4$	1BKV	2.00
Y position	$(GPO)_4GPRGAT(GPO)_4$	2F6A	3.29
Y position	$(PPG)_4OTG(PPG)_4$	Not available	

Glycine, proline, and its derivatives have not been considered. A single letter code has been used for peptide sequences. The letter O denotes 4*R*Hyp residues.

this explanation to account for stabilizing and destabilizing effects of 4*R*Hyp/4*S*Hyp as function of their position and for the data collected on fluoroproline derivatives [42,61] (Table 2) has shifted the interest toward alternative mechanisms. In this scenario, hypotheses relying on stereoelectronic effects and/or intrinsic iminoacid propensities have been proposed [16,61,68]. According to the hypothesis based on inductive effects, electron withdrawing groups such as F, OH in the 4*R* diastereoisomer should favor the *trans* peptide bond state, that is the one observed in folded triple helices. Although this suggestion is able to explain the stabilizing effects produced by 4*R*Hyp and 4*R*-fluoro-2*S*-proline (4*R*Flp) when located in the Y position, it cannot account for destabilization effects produced by 4*R*Hyp in the X position when the residue is inserted in a (Gly-4*R*Hyp-Pro)$_n$ or (Pro-Pro-Gly)$_n$ environment [22,57].

The near atomic resolution structure of the collagen-like peptide (Pro-Pro-Gly)$_{10}$ has shown a clear correlation between Pro conformation (puckering) (Figure 1) and the residue position along the sequence [16]. Indeed, proline residues adopt alternating conformations in the peptide. In particular, Pro in X and Y assumes down and up states, respectively. Since down and up conformers also exhibit differences at backbone level, the alternating down-up states in Pro-Pro-Gly structure are important for the correct triple helix assembly. Statistical surveys carried

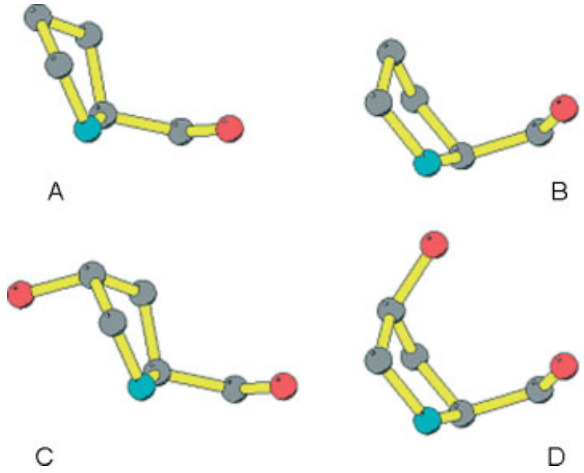


Figure 1. Proline and some of its hydroxylated derivatives. Up and down states of proline residues are shown in panels A and B, respectively. The up state of 4*R*Hyp and the down state of 4*S*Hyp are shown in panels C and D, respectively.

out using protein and peptide structure databases along with quantum mechanics calculations indicate that Pro can assume both up and down conformations whereas 4*R*Hyp (and 4*R*Flp) and 4*S*Hyp (4*S*Flp) adopt preferentially the up state and the down state, respectively [16,36,68] (Figure 1). Taking into account the specific conformational preferences of 4*R*Hyp and 4*S*Hyp, these residues are expected to destabilize the triple helix when located in the X and Y positions, respectively. On the other hand, the intrinsic conformational preferences of 4*R*Hyp and 4*S*Hyp should favor triple-helix formation when these residues are located in the Y and X positions, respectively. Indeed, their strong intrinsic tendency to adopt the conformation required in these positions likely reduces the entropy cost of the folding process. Apparently, a conflict occurs for 4*S*Hyp in X position. This should be stabilizing according to the model, but the experimental data demonstrate that it produces destabilizing effects (Table 2). However, this discrepancy could be solved by considering that 4*S*Hyp in X position generates severe steric clashes within the triple-helix scaffold [16].

These tendencies (and the resulting effects) are stronger in the fluoro-containing derivatives of proline, likely because of the larger electronegativity of the fluorine atom. Data collected on these peptides [42,63,69–71] show an almost perfect agreement with the propensity-based model predictions (Table 2). It is worth mentioning that, in contrast to (4*S*Hyp-Pro-Gly)$_{10}$, (4*S*Flp-Pro-Gly)$_{10}$ is able to form a triple helix. In this case, steric clashes between adjacent chains are likely attenuated by different local geometries of 4*S*Flp and 4*S*Hyp. In addition to the differences between C–F and C–OH bond distances, these residues may also display differences in the preferred dihedral angles that may have a significant impact on triple-helix formation. An indirect support to this idea comes from the observation that a single Gly-Pro-4*R*Flp triplet has destabilizing effects if embedded in a Gly-Pro-4*R*Hyp context [71]. It is also worth noticing that the replacement of Pro with 4*S*Flp in X position (T_m 34 to >58 °C) [63] has smaller stabilizing effects when compared with the replacement of Pro with 4*R*Flp in Y (T_m 34 to >80–90 °C) [61]. This indicates that also the insertion of 4*S*Flp in X may generate some steric clashes. A definitive answer to this puzzling issue will be likely identified when structural data on Flp-containing triple helical peptides will become available.

It this worth mentioning that the propensity-based model is also in line with the experimental stabilities of triple-helical peptide

Table 2. Melting temperatures (T_m) of some collagen-like peptides containing a hydroylated and fluorinated proline derivatives

Peptide	T_m (°C)	Prediction	References
(Pro-Pro-Gly)$_{10}$	34	Reference peptide	58,60
(Pro-4*R*Hyp-Gly)$_{10}$	62	Stabilization	58,60
(Pro-4*S*Hyp-Gly)$_{10}$	No helix	Destabilization	58
(4*R*Hyp-Pro-Gly)$_{10}$	No helix	Destabilization	58
(4*S*Hyp-Pro-Gly)$_{10}$	No helix	*Stabilization*[a]	58
(Pro-4*R*Flp-Gly)$_{10}$	91	Hyper-stabilization	61
(Pro-4*S*Flp-Gly)$_{10}$	No helix	Destabilization	62
(4*R*Flp-Pro-Gly)$_{10}$	No helix	Destabilization	63
(4*S*Flp-Pro-Gly)$_{10}$	54.5	*Hyper-stabilization*[b]	63
Host–guest peptides in a (Pro-Pro-Gly) context			
(Pro-Pro-Gly)$_4$-Pro-Pro-Gly-(Pro-Pro-Gly)$_4$	17.7	Reference peptide	22
(Pro-Pro-Gly)$_4$-Pro-4*R*Hyp-Gly-(Pro-Pro-Gly)$_4$	21.2	Stabilization	22
(Pro-Pro-Gly)$_4$-4*R*Hyp-Pro-Gly-(Pro-Pro-Gly)$_4$	15.5	Destabilization	22
(Pro-Pro-Gly)$_4$-Pro-4*S*Hyp-Gly-(Pro-Pro-Gly)$_4$	15.7	Destabilization	22
4*R*Hyp in the X position in other contexts			
(Gly-Pro-Thr)$_{10}$	No helix	Reference peptide	64
(Gly-4*R*Hyp-Thr)$_{10}$	18.0	*Destabilization*	64
(Gly-Pro-4*R*Hyp)$_{10}$	61	Reference peptide	58,60
(Gly-4*R*Hyp-4*R*Hyp)$_{10}$	65	*Destabilization*	60,65,66

The prediction of the propensity-based model is also reported (incorrect predictions are in italics).
[a] The destabilization of 4*S*Hyp in the X position has been attributed to repulsive effects of its OH group with the atoms of an adjacent polypeptide chain.
[b] The reduced stabilization of 4*S*-fluoro-2*S*-proline (4*S*Flp) in X position may be attributed to some repulsive effects of its F substituent with the atoms of an adjacent polypeptide chain.

Journal of PeptideScience

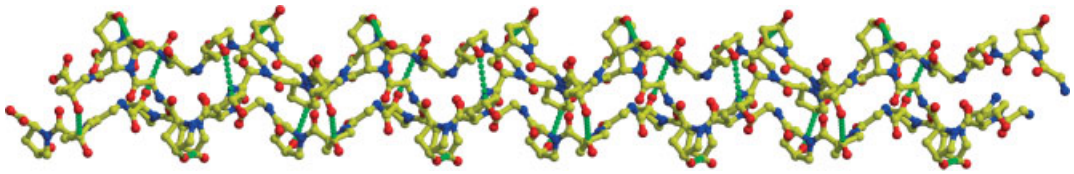


Figure 2. Recurrent dipole–dipole interactions (green) in (Gly-4*R*Hyp-4*R*Hyp)$_{10}$ triple helix established by 4*R*Hyp side chains. The figure has been generated by using the coordinates of the Protein Data Bank entry 1YM8.

containing 4-methyl proline residues [72]. Although in this case the intrinsic preferences of the amino acid are dictated by steric rather than stereoelectonic effects, 4*R*methyl-2*S*-proline that exhibits a strong preference for the down pucker stabilizes the triple helix when located in the X position. On the other hand, 4*S*-methyl-2*S*-proline that shows a strong preference for the up state stabilizes the triple helix when it is positioned in Y. Surprisingly, 4*S*-mercapto-2*S*-proline (4*S*-Mpc) produces destabilizing effects when inserted in the Y position in a Gly-Pro-Hyp context despite its preference for the up state [73]. Although steric effects have been invoked to explain these findings [73], these results could also be ascribed to differences between the up states of 4*S*-Mpc and 4*R*-Hyp. The analysis of triple helix formation by (Gly-Pro-4*S*-Mpc)$_n$ peptides could provide insights into this puzzling issue.

The validity of the propensity-based model has been recently supported by recent thermodynamics and structural investigations on the peptides: (Pro-Pro-Gly)$_4$-Pro-4*R*Hyp-Gly-(Pro-Pro-Gly)$_4$, (Pro-Pro-Gly)$_4$-4*R*Hyp-Pro-Gly-(Pro-Pro-Gly)$_4$, and (Pro-Pro-Gly)$_4$-Pro-4*S*Hyp-Gly-(Pro-Pro-Gly)$_4$ [22]. As shown in Table 2, there is a very good agreement between the predicted and the experimental stabilities of these peptides. Moreover, their crystallographic characterization shows that triple helix constraints dictate conformation of the iminoacids, independently of their intrinsic preferences [22]. Indeed, in the compound (Pro-Pro-Gly)$_4$-4*R*Hyp-Pro-Gly-(Pro-Pro-Gly)$_4$, the 4*R*Hyp adopts the down conformation, which intrinsically disfavored for this residue. Along this line, in (Pro-Pro-Gly)$_4$-Pro-4*S*Hyp-Gly-(Pro-Pro-Gly)$_4$ 4*S*Hyp adopts the up conformation, despite its propensity for the down state [23].

The concomitant presence of multiple functional groups in side chains, which may potentially interact, highly complicates this scenario. In these cases, stabilizing/destabilizing effects are often non-cumulative. As an example, the concomitant presence of 4*S*Flp and 4*R*Hyp in the X and Y positions, respectively, of the Gly-Xaa-Yaa triplet produces strong destabilizing effects [70]. Similarly, strands with 4*S*Flp in the X position and 4*R*Flp in the Y position are unable to form stable triple helices [74]. Particularly puzzling is the discovery that, in particular contexts, 4*R*Hyp may have stabilizing effects on the triple helix when located in the X position (Table 2). Indeed, in contrast to the propensity-based model prediction, the replacement of the X Pro with a Hyp residues in (Gly-Pro-Hyp) [60,65,66,71] and (Gly-Pro-Thr) [64] triplet has slight, but significant, stabilizing effects. To explain these data, hydrogen-bonding interactions, direct or water-mediated, established by the side chains of the residues located at the X and Y positions have been proposed [27,60,64,75]. According to molecular dynamics (MD) simulations the presence of 4*R*Hyp in the X position is important in the modulation of the strength of triple-helix interchain H-bonds [31]. It is interesting to note that the stabilizing role of the triplet Gly-4*R*Hyp-4*R*Hyp is controversial, despite the three independent high-resolution characterizations of model peptides embedding this motif reported in the last years [27,28](see also G. Wu, K. Noguchi, K. Okuyama, K. Mizuno, H.-P. Bächinger, Protein Data Bank (PDB) code 2D3H). Very recently, starting from available X-ray data, we carried out an extensive quantum chemistry analysis of the mutual interactions established by 4*R*Hyp residues located at the X and Y positions of the Gly-Xaa-Yaa motif [35]. Our data clearly indicate that C^{γ}–O^{δ} groups of opposing 4*R*Hyp rings establish significant dipole–dipole interactions that play an important role in triple-helix stabilization (Figure 2). This study has also demonstrated that Hyp C^{γ}–O^{δ} groups also establish relevant interactions with dipoles present on the main chain. Further investigations are required to check whether dipole–dipole interactions between side chains play also a role in the stabilizing effects exerted by Gly-4*R*Hyp-Thr triplets.

One of the most striking examples of the role played by iminoacid side chain–side chain interactions in the stabilization/destabilization of the triple helix motif is the possibility to obtain heterotrimeric triple helical association from peptides that are individually unable to form stable helices. Although (Pro-Pro-Gly)$_7$ and (4*S*Flp-4*R*Flp-Gly)$_7$ do not form stable homotrimeric species, mixtures of these two peptides form stable heterotrimeric helices containing one (Pro-Pro-Gly)$_7$ strand and two (4*S*Flp-4*R*Flp-Gly)$_7$ strands [74].

In conclusion, data on peptides containing a single proline derivative may be essentially explained on the basis of residue conformational propensities, which are dictated by the presence of specific substituents on the side chain. On the other hand, repulsive or attractive interactions between side chains must be considered to explain data on peptides containing more than one functional group [70]. Although the precise nature of these interactions has been identified in several cases, in many instances further structural studies are required.

Collagen Triple Helix and Amino Acids

Extensive studies carried out by the group of Brodsky and coworkers on host–guest peptide models have provided clear evidence of the sequence dependence of triple helix stability [4,40,41]. On analogy with studies conducted for the secondary structure elements found in globular proteins (α-helix, β-sheet, and PPII), these investigations have provided an accurate amino acid propensity scale for collagen triple helix (see [5] for further details). These analyses have unveiled that there is a significant correlation between the T_m values of host–guest peptides embedding a specific amino acid and its propensities for PPII motifs [4,76,77]. This indicates that the PPII propensity represents the basic contribution to the triple-helix stabilization by amino acids. The collagen propensity scale also correlates with the frequency of amino acids in collagens sequences. These analyses have also highlighted that the two positions X and Y of the Gly-Xaa-Yaa triplet motif are not equivalent. Indeed, several amino acids show clear preferences for one specific position. The correlation of triple helix propensity with PPII frequency is better for the X position

than the Y, likely because of its greater exposure to solvent. In more specific terms, the most stable residues in the Y position are Hyp and Arg, followed by Met. In the X position, the most stabilizing Pro residue is followed by charged residues, and then Ala and Gln. As most of the peptides investigated present an elevated content of iminoacids, crystallographic structural data on amino acid conformational preferences within the collagen triple-helix framework are rather limited (Table 1). Nevertheless, the combination of experimental and theoretical analyses has provided some interesting indications on the structural basis of amino acid propensities for collagen triple helix. Even in the absence of high-resolution collagen triple-helix structures, molecular mechanics analyses indicated the high stability of the Arg side chain conferred to the triple helix when located in the Y position is likely to be attributed to its hydrogen bonding to a C═O in a neighboring strand as well as to van der Waals interactions with the backbone [77]. These findings have been corroborated by later X-ray diffraction studies [14]. Computational modeling also showed similar favorable non-bonded interactions between the Met side chain and the triple-helix backbone [77]. Modeling also shows that residues with side chain branching at the C^{δ} can be more easily accommodated in the X position than in the Y position; as a consequence, all C^{δ}-branched residues show a greater propensity for the X position. Finally, the stabilizing role of Asp residues has been recently ascribed to solvation effects [78].

On analogy with iminoacids, interactions between amino acid side chains may strongly affect the stability of triple-helix model peptides. In this context, particularly evident is the role played by charged amino acids. Extensive thermodynamic and structural studies have highlighted that oppositely charged amino acid may strongly stabilize triple-helical structures. The role of these interactions has been recently corroborated by the analysis of the collagen-like domain of the *Streptococcus pyogenes* cell-surface protein Scl2 [79]. This protein forms a stable triple helix with a thermal stability close to that seen for mammalian collagens, despite the absence of hydroxylated proline residues in its sequence. The thermodynamic characterization of model peptides containing fragments of the sequence of this protein indicates that ion pairs play a major role in stabilizing the Scl2 triple helix [80]. This study has also suggested that the enthalpic stabilization of the triple helix likely involves interactions of polar groups with an ordered hydration network. This latter observation is in line with recent MD simulations that revealed a solvent-mediated stabilization of the triple helix in amino acid-rich regions [33,81] and that polar/charged side chains may modulate these effects [33].

The development of stable non-covalent heterotrimeric collagen-like helices by Gauba and Hartgerink represents one of the most brilliant applications of triple-helix stabilization through electrostatic interactions [82,83]. Such interstrand electrostatic interactions were also responsible for the quantitative oxidative disulfide-linked homotrimerization of a model peptide containing the *C*-terminal Cys–Cys sequence of collagen type III [84,85].

As many members of collagen family are formed by the assembly of distinct polypeptide chains, the search for stable heterotrimeric peptide models has been quite extensive over the last decade. Several intriguing approaches have been developed to efficiently link together different chains by covalent interactions [48,86–88]. The detection of stable non-covalent assemblies obtained by simply mixing different chains has recently received much attention. This approach, initially developed to characterize intermediate species [89,90], has been later used to generate stable heterotrimeric collagen-like assemblies from chains unable to form homotrimeric helices. Gauba and Hartgerink have shown, by means of appropriate modulations of the charges, that even residues or fragments with low/moderate propensities for the triple helix could be 'rescued' in heterotrimeric species. This approach was successfully used to synthesize collagen heterotrimers endowed with a very high stability (65 °C). Subsequent applications have been extended to the analysis of the alterations related to the disease *Osteogenesis Imperfecta* that are induced by replacements of Gly residues in the Gly-Xaa-Yaa repeat with a bulky aminoacid [91]. In contrast to previous studies employing peptide models, the authors could, for the first time [46], characterize collagen models with structures more relevant to native forms of this disease containing only one or two glycine substitutions by alanine. The general validity of this appealing approach paves the way for future applications aimed at elucidating unsolved aspects of collagen function.

Collagen Triple-Helix Self-Association

The function of many members of the collagen family is essentially structural. Lateral and axial association of collagen triple helices generates higher-order assemblies that are a fundamental constituent of the connective tissue. It has been shown that triple helix-tripe helix association is also important for conferring collagen additional stability [92].

The analysis of the crystal packing of collagen-like peptides represents a valuable source of information on triple-helix self-association modes. Although collagen-like peptides represent highly simplified models for collagen, packing of their triple helices in the crystal state often resembles the one adopted by real collagen molecules. Therefore, the analysis of the contacts between crystallographically related molecules in the crystal state provides clues on the interactions that potentially favor collagen self-assembly. A survey of the crystal packing of various peptides has evidenced that triple helix–triple helix association may be driven by different types of interactions. Since the early characterization of collagen-like peptides, it was demonstrated that exposed Hyp side chains located on adjacent molecules could establish rather strong H-bonds [13,15,93]. Interestingly, turbidity experiments indicate that the triple helical form of $(Pro\text{-}Hyp\text{-}Gly)_{10}$ self-associate, with a nucleation-growth mechanism [93]. A number of characteristics of these triple-helical assemblies resemble those obtained in collagen fibril formation. It is likely that the H-bonding interactions formed by Hyp side chains identified in the crystal state may contribute to the formation of the fibers. Electrostatic interactions formed by lysine and glutamic acid residues and by the charged termini of the peptides also play an important role in directing triple-helix association [13]. Finally, the recent structure determination of peptides containing Leu residues have also highlighted that triple-helix self-association may also be driven by hydrophobic interactions [24]. In this structure, Leu residues adopt rotameric states that allow the side chain to protrude along the radial direction of the rod-like molecule. The relevance of these findings for real collagen is demonstrated by the observation that Leu-Hyp-Gly sequence is one of the most frequently occurring triplets in type I collagen.

In recent years, the use of peptide models to generate large collagen-like fibers has been quite successful. In this field, the ability of short collagen fragments, in which the three strands are held in a staggered array by disulfide bonds, to generate fiber-like aggregates has been tested [94]. A combination of different

Journal of PeptideScience

techniques indicates that these 'sticky-ended' fragments self-assemble *via* intermolecular triple helix formation. The resulting fibrils resemble natural collagen, and some are longer (>400 nm) than any known collagen [94]. More recently, model peptides' activity was self-assembled into supramolecular fibrils exhibiting collagen-like biological properties by placing aromatic groups on the ends of a representative 30-mer (Gly-Pro-Hyp)$_{10}$, as with L-phenylalanine and L-pentafluorophenylalanine [95].

Collagen Triple-helix Recognition

Collagen is a promiscuous protein that interacts with a plethora of biomolecules, despite its limited sequence variability and the rigid rod-like shaped structure [96]. Collagen-interacting proteins are expressed by both eukaryotes and prokaryotes. Since collagen is involved in a variety of biological processes (survival signals, morphogenic processes, cell migration, and diseases), mammalian cells have developed several mechanisms to target this protein. Indeed, a number of transmembrane proteins, collectively designed as mammalian collagen–receptors, are able to interact with triple helical motifs. Collagen is also the target of adhesive proteins from Gram-positive bacteria, which do not share any sequence similarity with the mammalian receptors [97]. It has been shown that the ability to interact with collagen provides a general advantage to bacteria in pathogenesis.

From the structural point of view the interaction of collagen with receptors and adhesive proteins is still poorly understood. The analysis of structural features of the collagen-interacting proteins indicates that they can adopt a variety of different folds. In general collagen–receptor recognition occurs through two distinct mechanisms [96]. In the first one, the collagen partner complex recognizes a specific collagen sequence. In other cases, the receptor simply recognizes and binds the basic triple helix motif. Examples of the first class include the interaction with integrins and discoidin domain receptors and possibly the chaperone FKBP65, whereas the second class includes glycoprotein VI, CNA, YadA, saratin, and LAIR-1 [and references therein, [96,98–100].

So far, only two structures of a collagen-interacting proteins in complex with triple-helical models have been reported: (a) the structure of the integrin $\alpha_1\beta_1$ bound to a peptide containing the critical -Gly-Phe-Hyp-Gly-Glu-Arg- motif [25] and the structure of *Staphylococcus aureus* CNA complexed with the peptide (Gly-Pro-Hyp)$_4$-Gly-Pro-Arg-Gly-Arg-Thr-(Gly-Pro-Hyp)$_4$ peptide [97]. These two complexes have provided indications on sequence-specific and on triple-helix-based collagen recognition.

The binding of the peptide (Gly-Pro-Hyp)$_2$-Gly-Phe-Hyp-Gly-Glu-Arg-(Gly-Pro-Hyp)$_3$ to the integrin $\alpha_1\beta_1$ relies on different types of interactions (Figure 3). The most important one is represented by the participation of a Glu side chain of the peptide to the coordination of a metal ion of the receptor. The recognition is completed by the hydrophobic interaction involving a Phe residue of the peptide and an ion pair. The role of these residues in the binding affinities has been confirmed by mutagenesis analyses [25]. Modeling allowed also the understanding of the structural basis for the binding of a homotrimeric collagen peptide mimicking the integrin binding site of collagen type IV [86].

As for the integrin, the *S. aureus* CNA binds the triple helix in a trench present on the adhesin surface [97]. Each of the two (Gly-Pro-Hyp)$_4$ motifs present in the peptide binds a CNA molecule (Figure 4). The interactions that stabilize the complex are essentially hydrophobic along with few H-bonds. The limited resolution of the study (3.29 Å) does not provide any information on the solvent structure, which may play a relevant role in the binding of this intrinsically highly hydrated peptide.

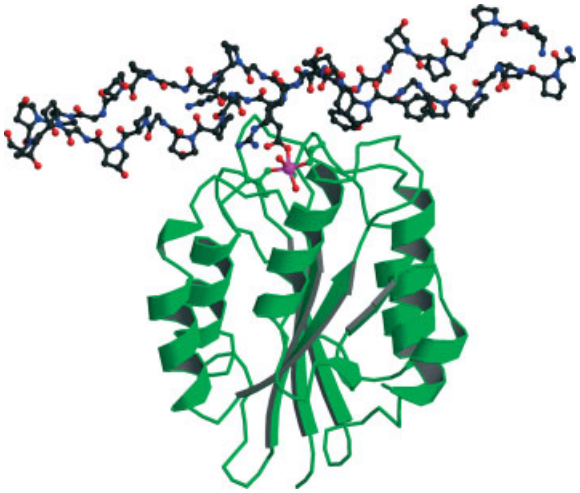


Figure 3. Three-dimensional structure of the complex between the integrin $\alpha_1\beta_1$ (green) and the triple helical peptide (Gly-Pro-4*R*Hyp)$_2$-Gly-Phe-4*R*Hyp-Gly-Glu-Arg-(Gly-Pro-4*R*Hyp)$_3$ (PDB entry 1DZI). The coordination of the integrin metal is also shown.

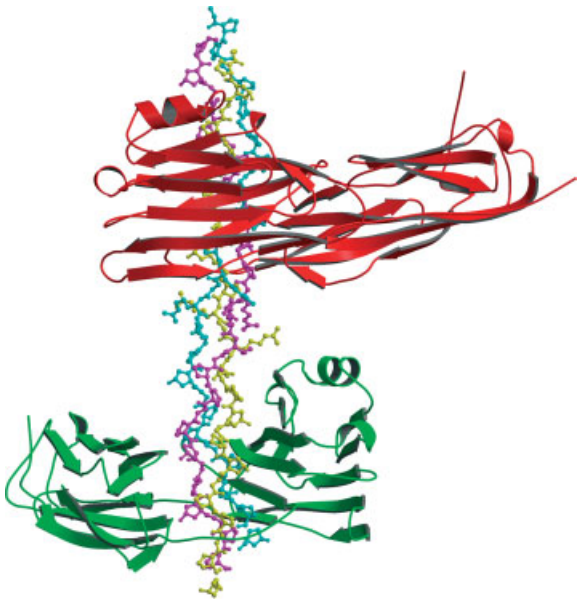


Figure 4. Three-dimensional structure of the complex between *S. aureus* CNA and the peptide (Gly-Pro-4*R*Hyp)$_4$-Gly-Pro-Arg-Gly-Arg-Thr-(Gly-Pro-4*R*Hyp)$_4$. The coordinates of the PDB entry 2F6A have been used to generate the picture.

Collagen interactions with other proteins may also generate potentially deleterious effects. Indeed, collagen plays an active role in the aggregation of β_2-microglobulin under physiopathological conditions (dialysis-related amyloidosis) [101]. The authors provide evidence that collagen positively charged side chains play a major role in inducing aggregation. Experiments carried out with standard peptides such as (Gly-Pro-Hyp)$_n$ and (Gly-Pro-Pro)$_n$ could provide information on whether the simple triple helix motif is also important in favoring the aggregation.

Conclusion and Perspectives

Complexity is one of the intrinsic features of the collagen molecule, despite some evident regularities in its sequence and structure. Not only the collagen characterization is important for understanding and modulating its many biological functions, but it also holds implications for the design and development of new biomaterials. Although, in principle, the collagen molecule size and its involvement in various biological processes require detailed characterizations of the entire molecule, some intrinsic features of this complex molecule have hitherto limited these investigations.

Reductionistic approaches, through the use of peptide models, have been widely applied to obtain information on the structure of this complex system (see papers of G. Fields and L. Moroder). The success of this approach has been illustrated in the previous sections. However, the necessity to mimic the collagen triple helix with smaller compounds has been often met by enriching these models with imino acids which are endowed with a high propensity for this secondary structure element. It should be noticed that the great majority of thermodynamic and structural studies and the ensuing stabilization models have been obtained using peptides with over-represented iminoacid content, if compared with the real imino acid/amino acid distribution in collagen. Indeed, the percentage of triplets with iminoacids in both X and Y is only 13% in real collagen. Furthermore, these triplets are infrequently consecutive in the collagen sequence, which, on the other hand, presents a large abundance of triplets with amino acids in both X and Y (41%) positions or with an imino acid in single (46%) either X or Y position. These simple considerations indicate that the currently available data and hypotheses, essentially derived using iminoacid-rich sequences, may be heavily biased. As a consequence, the extension of these results to real collagen may be somewhat limited by the specific composition of these models. In this context, it has been proposed that important general properties of collagen may be sequence dependent [12,33]. Focused investigations aimed at collecting information on models with low/moderate content of imino acids are particularly needed. In line with recent findings [44,48,102,103], it is likely that these models will be characterized by an enhanced triple-helix plasticity.

Available structural data on collagen peptide models are still rather limited. As shown in Table 1, there are several amino acids that have never been characterized within a triple helix context. It is not surprising that very few structures with biologically relevant sequences have been reported. This is likely because of the difficulties related to the growth of well-diffracting crystals of these peptides. More extensive applications of computational techniques may compensate for the lack of structural information. Indeed, as triple-helix scaffold limits the number of accessible states for the protein backbone, accurate conformational sampling of these systems may be achieved through MD simulations. These techniques can provide reliable descriptions of side chain mobility and peptide hydration. In addition, the repetitive nature of some structural features of collagen gives the opportunity to *ad hoc* design small-sized systems that could be investigated by quantum mechanics. Computational methods could be very useful to investigate the role played by long-range dipole–dipole and/or electrostatic interactions in modulating the triple-helix stability in a deeper detail. The characterization of non-local cooperative interactions, caused by the presence within the triple helix of repeating sequences of aligned dipoles, may also provide insightful information on the structural determinants of triple-helix stability. The continuous increase of available computational resources will favor the application of these methodologies to systems with a progressively augmented complexity. Finally, the definition of collagen–receptors recognition mechanism will benefit from the use of integrated molecular docking and MD techniques.

Acknowledgements

This work was supported by the Miur (FIRB – Contract number RBNE03PX83). A.D.S is supported by the European Molecular Biology Organization (EMBO), fellowship number: 584-2007.

References

1. Engel J. Versatile collagens in invertebrates. *Science* 1997; **277**: 1785–1786.
2. Smith MC, Burns N, Sayers JR, Sorrell JA, Casjens SR, Hendrix RW. Bacteriophage collagen. *Science* 1998; **279**: 1834.
3. Moroder L. Studies of protein folding and structure with model peptides. *J. Pept. Sci.* 2005; **11**: 258–261.
4. Persikov AV, Ramshaw JAM, Kirkpatrick A, Brodsky B. Amino acid propensities for the collagen triple-helix. *Biochemistry* 2000; **39**: 14960–14967.
5. Brodsky B, Persikov AV. Molecular structure of the collagen triple helix. *Adv. Protein Chem.* 2005; **70**: 301–339.
6. Jenkins CL, Raines RT. Insights on the conformational stability of collagen. *Nat. Prod. Rep.* 2002; **19**: 49–59.
7. Engel J, Bächinger H. Structure, stability and folding of the collagen triple helix. *Top. Curr. Chem.* 2005; **247**: 7–33.
8. Bächinger H, Engel J. The thermodynamics and kinetics of collagen folding. In *Protein Folding Handbook*, Vol. 2, Buchner J, Kiefhaber T (eds). Wiley-VCH: Weinheim, 2005; 1059–1111.
9. Bella J, Eaton M, Brodsky B, Berman HM. Crystal and molecular structure of a collagen-like peptide at 1.9 Å resolution. *Science* 1994; **266**: 75–81.
10. Bella J, Brodsky B, Berman HM. Hydration structure of a collagen peptide. *Structure* 1995; **3**: 893–906.
11. Kramer RZ, Vitagliano L, Bella J, Berisio R, Mazzarella L, Brodsky B, Zagari A, Berman HM. X-Ray crystallographic determination of a collagen-like peptide with the repeating sequence (Pro-Pro-Gly). *J. Mol. Biol.* 1998; **280**: 623–638.
12. Kramer RZ, Bella J, Mayville P, Brodsky B, Berman HM. Sequence dependent conformational variations of collagen triple-helical structure. *Nat. Struct. Biol.* 1999; **6**: 454–457.
13. Kramer RZ, Venugopal MG, Bella J, Mayville P, Brodsky B, Berman HM. Staggered molecular packing in crystals of a collagen-like peptide with a single charged pair. *J. Mol. Biol.* 2000; **301**: 1191–1205.
14. Kramer RZ, Bella J, Brodsky B, Berman HM. The crystal and molecular structure of a collagen-like peptide with a biologically relevant sequence. *J. Mol. Biol.* 2001; **311**: 131–147.
15. Berisio R, Vitagliano L, Mazzarella L, Zagari A. Crystal structure of a collagen-like polypeptide with repeating sequence pro-hyp-gly at 1.4 Å resolution: Implications for collagen hydration. *Biopolymers* 2000; **56**: 8–13.
16. Vitagliano L, Berisio R, Mazzarella L, Zagari A. Structural bases of collagen stabilization induced by proline hydroxylation. *Biopolymers* 2001; **58**: 459–464.
17. Berisio R, Vitagliano L, Mazzarella L, Zagari A. Crystal structure of the collagen triple helix model $[(\text{Pro-Pro-Gly})_{10}]_3$. *Protein Sci.* 2002; **11**: 262–270.
18. Okuyama K, Arnott S, Takayanagi M, Kakudo M. Crystal and molecular structure of a collagen-like polypeptide $(\text{Pro-Pro-Gly})_{10}$. *J. Mol. Biol.* 1981; **152**: 427–443.
19. Nagarajan V, Kamitori S, Okuyama K. Crystal structure analysis of collagen model peptide $(\text{Pro-Pro-Gly})_{10}$. *J. Biochem. (Tokyo)* 1998; **124**: 1117–1123.
20. Nagarajan V, Kamitori S, Okuyama K. Structure analysis of a collagen-model peptide with a (Pro-Hyp-Gly) sequence repeat. *J. Biochem. (Tokyo)* 1999; **125**: 310–318.

21. Okuyama K, Hongo C, Fukushima R, Wu G, Narita H, Noguchi K, Tanaka Y, Nishino N. Crystal structures of collagen model peptides with Pro-Hyp-Gly repeating sequence at 1.26 Å resolution: Implications for proline ring puckering. *Biopolymers* 2004; **76**: 367–377.
22. Jiravanichanun N, Hongo C, Wu G, Noguchi K, Okuyama K, Nishino N, Silva T. Unexpected puckering of hydroxyproline in the guest triplets, Hyp-Pro-Gly and Pro-*allo*-Hyp-Gly sandwiched between Pro-Pro-Gly sequence. *ChemBioChem* 2005; **6**: 1184–1187.
23. Jiravanichanun N, Nishino N, Okuyama K. Conformation of *allo*-Hyp in the Y position in the host-guest peptide with the Pro-Pro-Gly sequence: Implication of the destabilization of (Pro-*allo*-Hyp-Gly)$_{10}$. *Biopolymers* 2006; **81**: 225–233.
24. Okuyama K, Narita H, Kawaguchi T, Noguchi K, Tanaka Y, Nishino N. Unique side chain conformation of a Leu residue in a triple-helical structure. *Biopolymers* 2007; **86**: 212–221.
25. Emsley J, Knight CG, Farndale RW, Barnes MJ, Liddington RC. Structural basis of collagen recognition by integrin $\alpha_1\beta_2$. *Cell* 2000; **101**: 47–56.
26. Schumacher MA, Mizuno K, Bächinger HP. The crystal structure of a collagen-like polypeptide with 3(*S*)-Hydroxyproline residues in the Xaa position forms a standard 7/2 collagen triple helix. *J. Biol. Chem.* 2006; **281**: 27566–27574.
27. Schumacher M, Mizuno K, Bächinger HP. The crystal structure of the collagen-like polypeptide (glycyl-4(*R*)-hydroxyprolyl-4(*R*)-hydroxyprolyl)$_9$ at 1.55 Å resolution shows up-puckering of the proline ring in the Xaa position. *J. Biol. Chem.* 2005; **280**: 20397–20403.
28. Kawahara K, Nishi Y, Nakamura S, Uchiyama S, Nishiuchi Y, Nakazawa T, Ohkubo T, Kobayashi Y. Effect of hydration on the stability of the collagen-like triple-helical structure of [4(*R*)-hydroxyprolyl-4(*R*)-hydroxyprolylglycine]$_{10}$. *Biochemistry* 2005; **44**: 15812–15822.
29. Mooney SD, Klein TE. Structural models of Osteogenesis imperfecta-associated variants in the col1α1 gene. *Mol. Cell Proteomics* 2002; **1**: 868–875.
30. Park S, Klein TE, Pande VS. Folding and misfolding of the collagen triple helix: Markov analysis of molecular dynamics simulations. *Biophys. J.* 2007; **93**: 4108–4115.
31. Radmer RJ, Klein TE. Triple helical structure and stabilization of collagen-like molecules with 4(*R*)-hydroxyproline in the Xaa position. *Biophys. J.* 2006; **90**: 578–588.
32. Radmer RJ, Klein TE. Severity of osteogenesis imperfecta and structure of a collagen-like peptide modeling a lethal mutation site. *Biochemistry* 2004; **43**: 5314–5323.
33. De Simone A, Vitagliano L, Berisio R. Role of hydration in collagen triple helix stabilization. *Biochem. Biophys. Res. Commun.* 2008; **372**: 121–125.
34. Mogilner IG, Ruderman G, Grigera JR. Collagen stability, hydration and native state. *J. Mol. Graph. Model.* 2002; **21**: 209–213.
35. Improta R, Berisio R, Vitagliano L. Contribution of dipole-dipole interactions to the stability of the collagen triple helix. *Protein Sci.* 2008; **17**: 955–961.
36. Improta R, Benzi C, Barone V. Understanding the role of stereoelectronic effects in determining collagen stability. 1. A quantum mechanical study of proline, hydroxyproline, and fluoroproline dipeptide analogues in aqueous solution. *J. Am. Chem. Soc.* 2001; **123**: 12568–12577.
37. Improta R, Mele F, Crescenzi O, Benzi C, Barone V. Understanding the role of stereoelectronic effects in determining collagen stability. 2. A quantum mechanical/molecular mechanical study of (proline-proline-glycine)$_n$ polypeptides. *J. Am. Chem. Soc.* 2002; **124**: 7857–7865.
38. Palfi VK, Perczel A. How stable is a collagen triple helix? An ab initio study on various collagen and β-sheet forming sequences. *J. Comput. Chem.* 2008; **29**: 1374–1386.
39. Shah NK, Ramshaw JA, Kirkpatrick A, Shah C, Brodsky B. A host-guest set of triple-helical peptides: Stability of Gly-X-Y triplets containing common nonpolar residues. *Biochemistry* 1996; **35**: 10262–10268.
40. Persikov AV, Ramshaw JA, Brodsky B. Collagen model peptides: Sequence dependence of triple-helix stability. *Biopolymers* 2000; **55**: 436–450.
41. Persikov AV, Ramshaw JA, Brodsky B. Prediction of collagen stability from amino acid sequence. *J. Biol. Chem.* 2005; **280**: 19343–19349.
42. Holmgren SK, Taylor KM, Bretscher LE, Raines RT. Code for collagen's stability deciphered. *Nature* 1998; **392**: 666–667.
43. Lauer-Fields J, Brew K, Whitehead JK, Li S, Hammer RP, Fields GB. Triple-helical transition state analogues: A new class of selective matrix metalloproteinase inhibitors. *J. Am. Chem. Soc.* 2007; **129**: 10408–10417.
44. Ottl J, Gabriel D, Murphy G, Knauper V, Tominaga Y, Nagase H, Kroger M, Tschesche H, Bode W, Moroder L. Recognition and catabolism of synthetic heterotrimeric collagen peptides by matrix metalloproteinases. *Chem. Biol.* 2000; **7**: 119–132.
45. Raines RT. 2005 Emil Thomas Kaiser award. *Protein Sci.* 2006; **15**: 1219–1225.
46. Brodsky B, Thiagarajan G, Madhan B, Kar K. Triple-helical peptides: An approach to collagen conformation, stability, and self-association. *Biopolymers* 2008; **89**: 345–353.
47. Berisio R, Vitagliano L, Mazzarella L, Zagari A. Recent progress on collagen triple helix structure, stability and assembly. *Protein Pept. Lett.* 2002; **9**: 107–116.
48. Renner C, Saccà B, Moroder L. Synthetic heterotrimeric collagen peptides as mimics of cell adhesion sites of the basement membrane. *Biopolymers* 2004; **76**: 34–47.
49. Okuyama K. Structural study of collagen based on single crystal analyses of model peptides. *J. Pept. Sci.* 2001; **37**: 263–264.
50. Okuyama K, Okuyama K, Arnott S, Takayanagi M, Kakudo M. Crystal and molecular structure of a collagen-like polypeptide (Pro-Pro-Gly)$_{10}$. *J. Mol. Biol.* 1981; **152**: 427–443.
51. Okuyama K, Wu G, Jiravanichanun N, Hongo C, Noguchi K. Helical twists of collagen model peptides. *Biopolymers* 2006; **84**: 421–432.
52. Okuyama K, Xu X, Iguchi M, Noguchi K. Revision of collagen molecular structure. *Biopolymers* 2006; **84**: 181–191.
53. Okuyama K. Revisiting the molecular structure of collagen. *Connect. Tissue Res.* 2008; **49**: 299–310.
54. Petsko GA. Structural basis of thermostability in hyperthermophilic proteins, or "there's more than one way to skin a cat". *Methods Enzymol.* 2001; **334**: 469–478.
55. Privalov PL. Stability of proteins. Proteins which do not present a single cooperative system. *Adv. Protein Chem.* 1982; **35**: 1–104.
56. Burjanadze TV. New analysis of the phylogenetic change of collagen thermostability. *Biopolymers* 2000; **53**: 523–528.
57. Inouye K, Sakakibara S, Prockop DJ. 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* 1976; **420**: 133–141.
58. Inouye K, Kobayashi Y, Kyogoku Y, Kishida Y, Sakakibara S, Prockop DJ. 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.* 1982; **219**: 198–203.
59. Gustavson KH. The function of hydroxyproline in collagens. *Nature* 1955; **175**: 70–74.
60. Berisio R, Granata V, Vitagliano L, Zagari A. Imino acids and collagen triple helix stability: Characterization of collagen-like polypeptides containing Hyp-Hyp-Gly sequence repeats. *J. Am. Chem. Soc.* 2004; **126**: 11402–11403.
61. Holmgren SK, Bretscher LE, Taylor KM, Raines RT. A hyperstable collagen mimic. *Chem. Biol.* 1999; **6**: 63–70.
62. Bretscher LE, Jenkins CL, Taylor KM, DeRider ML, Raines RT. Conformational stability of collagen relies on a stereoelectronic effect. *J. Am. Chem. Soc.* 2001; **123**: 777–778.
63. Doi M, Nishi Y, Uchiyama S, Nishiuchi Y, Nakazawa T, Ohkubo T, Kobayashi Y. 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.* 2003; **125**: 9922–9923.
64. Bann JG, Bachinger HP. Glycosylation/hydroxylation-induced stabilization of the collagen triple helix. 4-*trans*-hydroxyproline in the Xaa position can stabilize the triple helix. *J. Biol. Chem.* 2000; **275**: 24466–24469.
65. Mizuno K, Hayashi T, Peyton DH, Bachinger HP. Hydroxylation-induced stabilization of the collagen triple helix. Acetyl-(glycyl-4(*R*)-hydroxyprolyl-4(*R*)-hydroxyprolyl)$_{10}$-NH_2 forms a highly stable triple helix. *J. Biol. Chem.* 2004; **279**: 38072–38078.
66. Doi M, Nishi Y, Uchiyama S, Nishiuchi Y, Nishio H, Nakazawa T, Ohkubo T, Kobayashi Y. Collagen-like triple helix formation of synthetic (Pro-Pro-Gly)$_{10}$ analogues: (4(*S*)-hydroxyprolyl-4

(*R*)-hydroxyprolyl-Gly)$_{10}$, (4(*R*)-hydroxyprolyl-4(*R*)-hydroxyprolyl-Gly)$_{10}$ and (4(*S*)-fluoroprolyl-4(*R*)-fluoroprolyl-Gly)$_{10}$. *J. Pept. Sci.* 2005; **11**: 609–616.
67. Ramachandran GN, Bansal M, Bhatnagar RS. A hypothesis on the role of hydroxyproline in stabilizing collagen structure. *Biochim. Biophys. Acta* 1973; **322**: 166–171.
68. Vitagliano L, Berisio R, Mastrangelo A, Mazzarella L, Zagari A. Preferred proline puckerings in *cis* and *trans* peptide groups: Implications for collagen stability. *Protein Sci.* 2001; **10**: 2627–2632.
69. Hodges JA, Raines RT. Stereoelectronic effects on collagen stability: The dichotomy of 4-fluoroproline diastereomers. *J. Am. Chem. Soc.* 2003; **125**: 9262–9263.
70. Barth D, Milbradt AG, Renner C, Moroder L. A (4*R*)- or a (4*S*)-fluoroproline residue in position Xaa of the (Xaa-Yaa-Gly) collagen repeat severely affects triple-helix formation. *ChemBioChem* 2004; **5**: 79–86.
71. Persikov AV, Ramshaw JA, Kirkpatrick A, Brodsky B. Triple-helix propensity of hydroxyproline and fluoroproline: Comparison of host-guest and repeating tripeptide collagen models. *J. Am. Chem. Soc.* 2003; **125**: 11500–11501.
72. Shoulders MD, Hodges JA, Raines RT. Reciprocity of steric and stereoelectronic effects in the collagen triple helix. *J. Am. Chem. Soc.* 2006; **128**: 8112–8113.
73. Cadamuro SA, Reichold R, Kusebauch U, Musiol HJ, Renner C, Tavan P, Moroder L. Conformational properties of 4-mercaptoproline and related derivatives. *Angew. Chem., Int. Ed. Engl.* 2008; **47**: 2143–2146.
74. Hodges JA, Raines RT. Stereoelectronic and steric effects in the collagen triple helix: Toward a code for strand association. *J. Am. Chem. Soc.* 2005; **127**: 15923–15932.
75. Nishi Y, Uchiyama S, Doi M, Nishiuchi Y, Nakazawa T, Ohkubo T, Kobayashi Y. Different effects of 4-hydroxyproline and 4-fluoroproline on the stability of collagen triple helix. *Biochemistry* 2005; **44**: 6034–6042.
76. Berisio R, Loguercio S, De Simone A, Zagari A, Vitagliano L. Polyproline helices in protein structures: A statistical survey. *Protein Pept. Lett.* 2006; **13**: 847–854.
77. Vitagliano L, Nemethy G, Zagari A, Scheraga HA. Stabilization of the triple-helical structure of natural collagen by side-chain interactions. *Biochemistry* 1993; **32**: 7354–7359.
78. Raman SS, Parthasarathi R, Subramanian V, Ramasami T. Role of aspartic acid in collagen structure and stability: A molecular dynamics investigation. *J. Phys. Chem. B* 2006; **110**: 20678–20685.
79. Xu Y, Keene DR, Bujnicki JM, Hook M, Lukomski S. Streptococcal scl1 and scl2 proteins form collagen-like triple helices. *J. Biol. Chem.* 2002; **277**: 27312–27318.
80. Mohs A, Silva T, Yoshida T, Amin R, Lukomski S, Inouye M, Brodsky B. Mechanism of stabilization of a bacterial collagen triple helix in the absence of hydroxyproline. *J. Biol. Chem.* 2007; **282**: 29757–29765.
81. Ravikumar KM, Hwang W. Region-specific role of water in collagen unwinding and assembly. *Proteins* 2008; **72**: 1320–1332.
82. Gauba V, Hartgerink JD. Surprisingly high stability of collagen abc heterotrimer: Evaluation of side chain charge pairs. *J. Am. Chem. Soc.* 2007; **129**: 15034–15041.
83. Gauba V, Hartgerink JD. Self-assembled heterotrimeric collagen triple helices directed through electrostatic interactions. *J. Am. Chem. Soc.* 2007; **129**: 2683–2690.
84. Boulegue C, Musiol HJ, Gotz MG, Renner C, Moroder L. Natural and artificial cystine knots for assembly of homo- and heterotrimeric collagen models. *Antioxid. Redox Signal.* 2008; **10**: 113–125.
85. Barth L, Sinner E, Cadamuro S, Renner C, Oesterhelt D, Moroder L. Homotrimeric collagen peptides as model systems for cell adhesion studies. In *Peptides for Youth*, Escher E, Lubell W, Del Valle S (eds). American Peptide Society: In press.
86. Saccà B, Sinner EK, Kaiser J, Lübken C, Eble JA, Moroder L. Binding and docking of synthetic heterotrimeric collagen type IV peptides with $\alpha_1\beta_1$ integrin. *ChemBioChem* 2002; **3**: 904–907.
87. Moroder L, Musiol HJ, Gotz M, Renner C. Synthesis of single- and multiple-stranded cystine-rich peptides. *Biopolymers* 2005; **80**: 85–97.
88. Slatter DA, Foley LA, Peachey AR, Nietlispach D, Farndale RW. Rapid synthesis of a register-specific heterotrimeric type I collagen helix encompassing the integrin $\alpha_2\beta_1$ binding site. *J. Mol. Biol.* 2006; **359**: 289–298.
89. Berisio R, Granata V, Vitagliano L, Zagari A. Characterization of collagen-like heterotrimers: Implications for triple-helix stability. *Biopolymers* 2004; **73**: 682–688.
90. Slatter DA, Miles CA, Bailey AJ. Asymmetry in the triple helix of collagen-like heterotrimers confirms that external bonds stabilize collagen structure. *J. Mol. Biol.* 2003; **329**: 175–183.
91. Gauba V, Hartgerink JD. Synthetic collagen heterotrimers: Structural mimics of wild-type and mutant collagen type I. *J. Am. Chem. Soc.* 2008; **130**: 7509–7515.
92. Leikina E, Mertts MV, Kuznetsova N, Leikin S. Type I collagen is thermally unstable at body temperature. *Proc. Natl. Acad. Sci. U.S.A.* 2002; **99**: 1314–1318.
93. Kar K, Amin P, Bryan MA, Persikov AV, Mohs A, Wang YH, Brodsky B. Self-association of collagen triple helic peptides into higher order structures. *J. Biol. Chem.* 2006; **281**: 33283–33290.
94. Kotch FW, Raines RT. Self-assembly of synthetic collagen triple helices. *Proc. Natl. Acad. Sci. U.S.A.* 2006; **103**: 3028–3033.
95. Cejas MA, Kinney WA, Chen C, Vinter JG, Almond HR Jr, Balss KM, Maryanoff CA, Schmidt U, Breslav M, Mahan A, Lacy E, Maryanoff BE. Thrombogenic collagen-mimetic peptides: Self-assembly of triple helix-based fibrils driven by hydrophobic interactions. *Proc. Natl. Acad. Sci. U.S.A.* 2008; **105**: 8513–8518.
96. Leitinger B, Hohenester E. Mammalian collagen receptors. *Matrix Biol.* 2007; **26**: 146–155.
97. Zong Y, Xu Y, Liang X, Keene DR, Hook A, Gurusiddappa S, Hook M, Narayana SV. A 'collagen hug' model for *Staphylococcus aureus* cna binding to collagen. *EMBO J.* 2005; **24**: 4224–4236.
98. Gronwald W, Bomke J, Maurer T, Domogalla B, Huber F, Schumann F, Kremer W, Fink F, Rysiok T, Frech M, Kalbitzer HR. Structure of the leech protein saratin and characterization of its binding to collagen. *J. Mol. Biol.* 2008; **381**: 913–927.
99. Ishikawa Y, Vranka JA, Wirz JA, Nagata K, Bachinger HP. The rough endoplasmic reticulum-resident fk506-binding protein fkbp65 is a molecular chaperone that interacts with collagens. *J. Biol. Chem.* 2008 (in press). DOI: 10.1074/jbc.M802535200.
100. Leo JC, Elovaara H, Brodsky B, Skurnik M, Goldman A. The yersinia adhesin yada binds to a collagenous triple-helical conformation but without sequence specificity. *Protein Eng. Des. Sel.* 2008; **21**: 475–484.
101. Relini A, Canale C, De Stefano S, Rolandi R, Giorgetti S, Stoppini M, Rossi A, Fogolari F, Corazza A, Esposito G, Gliozzi A, Bellotti V. Collagen plays an active role in the aggregation of β2-microglobulin under physiopathological conditions of dialysis-related amyloidosis. *J. Biol. Chem.* 2006; **281**: 16521–16529.
102. Saccà B, Fiori S, Moroder L. Studies of the local conformational properties of the cell-adhesion domain of collagen type IV in synthetic heterotrimeric peptides. *Biochemistry* 2003; **42**: 3429–3436.
103. Saccà B, Renner C, Moroder L. The chain register in heterotrimeric collagen peptides affects triple helix stability and folding kinetics. *J. Mol. Biol.* 2002; **324**: 309–318.

