doi:10.1016/j.jmb.2008.07.009 J. Mol. Biol. (2008) 382, 246–256

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Differential Unfolding of α1 and α2 Chains in Type I Collagen and Collagenolysis

Paul S. Nerenberg[1] and Collin M. Stultz[2]*

[1]*Department of Physics, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*

[2]*Harvard-MIT Division of Health Sciences and Technology, Department of Electrical Engineering and Computer Science, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*

Received 15 April 2008; received in revised form 17 June 2008; accepted 3 July 2008
Available online 11 July 2008

Collagenolysis plays a central role in many disease processes and a detailed understanding of the mechanism of collagen degradation is of immense interest. While a considerable body of information about collagenolysis exists, the details of the underlying molecular mechanism are unclear. Therefore, to further our understanding of the precise mechanism of collagen degradation, we used molecular dynamics simulations to explore the structure of human type I collagen in the vicinity of the collagenase cleavage site. Since post-translational proline hydroxylation is an important step in the synthesis of collagen chains, we used the DNA sequence for the α1 and α2 chains of human type I collagen, and the known amino acid sequences for bovine and chicken type I collagen, to infer which prolines are hydroxylated in the vicinity of the collagenase cleavage site. Simulations of type I collagen in this region suggest that partial unfolding of the α2 chain is energetically preferred relative to unfolding of α1 chains. Localized unfolding of the α2 chain leads to the formation of a structure that has disrupted hydrogen bonds N-terminal to the collagenase cleavage site. Our data suggest that this disruption in hydrogen bonding pattern leads to increased chain flexibility, thereby enabling the α2 chain to sample different partially unfolded states. Surprisingly, our data also imply that α2 chain unfolding is mediated by the non-hydroxylation of a proline residue that is N-terminal to the cleavage site in α1 chains. These results suggest that hydroxylation on one chain (α1) can affect the structure of another chain (α2), and point to a critical role for the non-hydroxylation of proline residues near the collagenase cleavage site.

Keywords: collagen; collagenolysis; collagen degradation; protein unfolding; free energy simulations

Edited by D. Case

Introduction

Collagens comprise the most prevalent family of proteins in the human body, and the extracellular matrix in particular.[1] While collagenolysis, or collagen degradation, is essential to normal physiological processes such as angiogenesis and wound healing, it also has a pivotal role in the progression of diseases such as cancer, atherosclerosis, and rheumatoid arthritis.[2–5] Approximately 90% of collagen found in the body is fibril-forming collagen – types I, II, III, V and XI – that crosslink to form fibrillar structures.[1] Of the fibril-forming (or fibrillar) collagens, type I collagen is the most common. Type I collagen is a major structural constituent of bone, tendon, skin, and a myriad of other tissues.[1] Therefore understanding the mechanism of degradation for the fibrillar collagens, specifically type I collagen, may lead to the design of novel therapies that slow the progression of human diseases that involve abnormal collagen catabolism.

Fibrillar collagens are degraded by a family of Zn^{2+}-dependent proteases known as matrix metalloproteinases (MMPs); e.g. primarily the collagenases MMP-1, MMP-8, and MMP-13.[6,7] MMPs are multidomain proteins that generally contain a catalytic domain connected *via* a linker region to a hemopexin-like domain.[6] Some MMPs contain additional domains, e.g. the fibronectin type II-like collagen-binding domains of the gelatinases MMP-2 and

*Corresponding author. E-mail address: cmstultz@mit.edu.

Abbreviations used: MMP, matrix metalloproteinase; pmf, potential of mean force; rms, root-mean-square; H-bonds, hydrogen bonds.

MMP-9, and the cytoplasmic domains of the membrane-type MMPs.[8] Despite these differences in the structure of various MMPs, a common paradox arises from analyses of different MMP structures and the structure of collagen. Crystallographic studies of collagen-like model peptides suggest that the scissile bond that is cleaved by collagenases is not solvent-accessible, and that the triple-helical collagen structure cannot fit into the MMP active site.[8–10] Consequently, hydrolysis of scissile peptide bonds in collagen must involve some form of collagen unfolding that leads to the formation of a solvent-exposed scissile bond that can fit into the MMP active site.[8,11,12] Early experimental studies were interpreted to show that collagen does not unfold spontaneously in solution, and that unwinding of the triple helix is due to the simultaneous binding and subsequent coordinated motion of collagenase domains in a process referred to as molecular tectonics.[8] However, a growing body of experimental and theoretical results suggests that the collagen triple helix may be more flexible than originally thought, and that thermal fluctuations in the structure of collagen may lead to partial unfolding in the region of the cleavage site.[13–18] In particular, theoretical calculations on collagen-like model peptides suggest that regions near the collagenase cleavage site can adopt both well-folded (native) and partially unfolded (vulnerable) states.[13] In this formalism, collagenolysis occurs when collagenases bind to and subsequently cleave preformed vulnerable states of collagen.[18]

A number of experimental and computational works have focused on models of type III collagen, a homotrimeric fibrillar collagen, in large part due to the ease of synthesizing self-assembling homotrimeric model peptides.[19,20] For example, the model peptide T3-785, which contains a portion of the type III collagen sequence immediately C-terminal to the collagenase cleavage site, has been studied using X-ray crystallography, NMR, and detailed molecular dynamics simulations.[10,13,14,21] Type I collagen, conversely, is a heterotrimer composed of two identical α1(I) chains and one α2(I) chain.[1] Synthesis of heterotrimeric model peptides is therefore more difficult and often requires the construction of covalent linkages to join chains with distinct sequences.[15,19,22,23]

Experiments with full-length type I collagen have been performed that provide some insight into the mechanism of collagen degradation. In particular, it has been argued that MMP-1 cleaves α1 and α2 chains in type I collagen with different efficiencies.[11] In earlier experiments, type I collagen was incubated with catalytically inactive MMP-1 at 25 °C followed by the introduction of other active proteases (e.g., catalytic domains of various MMPs and human leukocyte elastase). The result was that α1 chains were cleaved more rapidly than the α2 chains.[11] On the basis of these data it was suggested that inactive MMP-1 preferentially bound to the α2 chain, leaving the corresponding sites on α1 chains amenable to cleavage by noncollagenolytic proteases. That is, since inactive MMP-1 was unable to cleave the scissile bond on α2 chains, it likely remained bound to the α2 chain for a relatively long period of time, thereby protecting the scissile bond from being cleaved by other proteases. Therefore, to explore the physical basis underlying any potential preference for MMP association with either the α1 or α2 chains, we conducted unfolding simulations of each chain in type I collagen.

Results

The sequence of type I collagen inferred from sequence homologs

Sequences for human α1 and α2 chains were obtained from the UniProtKB/Swiss-Prot sequence database (accession numbers P02452 and P08123, respectively).[24] Since these amino acid sequences are deduced from nucleotide sequences, and hydroxylation is a post-translational modification, it is impossible to determine which prolines are hydroxylated from these data alone. Fortunately, the entire amino acid sequence of the homologous bovine type I collagen and a large fraction of the amino acid sequence of the homologous chicken type I collagen have been determined using Edman degradation.[25–28] As expected, when an imino acid appears in the X2 position of G-X1-X2 collagen triplets, it is most often hydroxyproline.[1,20,25] In fact, in the entire sequence of both bovine and chicken α1 chains there are only four locations where hydroxylation at the X2 position does not occur.[27,28] The most notable exception is a triplet that is N-terminal to the triplet containing the scissile bond (yellow residues in Fig. 1). Given the significant sequence homology between bovine/chicken and human type I collagen, particularly in the α1 chain, we modeled proline at this position instead of hydroxyproline (Fig. 1).

Conformational thermodynamics and flexibility of type I collagen near the cleavage site

Non-hydroxylation of proline in the X2 position of G-X1-X2 triplets is curious because of its relative rarity, and because the stability of collagen-like sequences is a function of their hydroxyproline content.[20,29,30] Therefore, we postulated that non-hydroxylation of proline in the vicinity of the collagenase cleavage site would affect the conformational thermodynamics of collagen in this region.

We computed conformational free energy profiles for unfolding all three chains (two α1 and one α2) in human type I collagen using the sequence shown in Fig. 1. As we were interested in the conformational thermodynamics of the region immediately surrounding the cleavage site, a stochastic boundary method similar to that outlined earlier was used.[13] Simulations were conducted at a 300 K, a temperature comparable to previous degradation experiments, which were conducted at 298 K (25 °C).[11]

```
α1 chain
Bovine: GEKGAOGADGPAGAOGTPGPQGIAGQRGVVGLOGQRGERGFOGLO
Chick:  GEKGSOGADGPIGAOGTPGPQGIAGQRGVVGLOGQRGERGFOGLO
Human:  GEKGSOGADGPAGAOGTPGPQGIAGQRGVVGLOGQRGERGFOGLO

α2 chain
Bovine: GEKGPSGEOGTAGPOGTOGPQGLLGAOGFLGLOGSRGERGLOGVA
Chick:  GEKGPSGEAGAAGPOGTOGPQGILGAOGILGLOGSRGERGLOGIA
Human:  GEKGPSGEAGTAGPOGTOGPQGLLGAOGILGLOGSRGERGLOGVA
```

Fig. 1. Alignments of sequences of type I collagen from bovine and chicken collagen in the region of the collagenase cleavage site. The sequence of human collagen is shown where the hydroxylation pattern is deduced from the sequence alignment. The triplets containing the scissile bonds are colored magenta. Residues that vary across species are underlined. The GTP triplet of the α1 chain and the GTO triplet of the α2 chain are indicated in yellow.

Folded triple-helical collagen conformations (Fig. 2a) have interchain hydrogen bonds (H-bonds) between the inward-facing amide hydrogen of glycine residues and the carbonyl oxygen of X1 residues in G-X1-X2 triplets (Fig. 2b). As nitrogen (of G) and oxygen (of residue X1) are the heavy atoms involved in these H-bonds, unfolding simulations were conducted using a reaction coordinate corresponding to the average interchain backbone nitrogen-to-oxygen (NO) distance. Distances of less than 3.6 Å are consistent with the formation of a H-bond between the amide hydrogen and carbonyl oxygen; i.e. H-bonds present in the folded state. A free energy profile (also known as potential of mean force or pmf) for unfolding each of the three chains in type I collagen was obtained by varying the average NO distance between one chain and the other two chains, thereby allowing us to monitor directly the separation of one chain from its adjacent chains.

Despite the fact that each α1 chain is in a distinct chemical environment (Fig. 2b), the pmfs for both α1 chains are very similar (Fig. 3a and b). Both pmfs have a single deep energy minimum at an average NO distance of 3.1 Å and 3.2 Å (Fig. 3a and b). Hence, on average the interchain H-bonds are preserved in structures corresponding to these minima. In contrast, the pmf for unfolding the α2 chain has two distinct local energy minima (Fig. 3c). In addition to the global energy minimum at an average NO distance of 3.0 Å, there is a minimum at 3.8 Å that has an energy within 0.9 kcal/mol of the global energy minimum (Fig. 3c). Taken together, these data suggest that unfolding either α1 chain is energetically unfavorable, while unfolding the α2 chain occurs more readily and leads to the formation of a state that has, on average, disrupted interchain H-bonds.

Representative structures from the α2 chain global energy minimum (i.e., the native state), which occurs at an average NO distance of 3.6 Å, resemble the familiar triple-helical structure of collagen (Fig. 4a). Two NO distances, in the vicinity of the triplet that contains the scissile bond (i.e., the scissile bond triplet), are slightly above 3.6 Å, suggesting that even in the native state the α2 chain can sample states where NO distances are stretched beyond that expected on the basis of the H-bonding pattern of an idealized triple helix. Nevertheless, most H-bonds in the vicinity of the scissile bond have NO distances below 3.3 Å (Fig. 4b). However, representative structures corresponding to the energy minimum at 3.8 Å are partially unfolded in the region N-terminal to the collagenase cleavage site (Fig. 4c). In particular, the α2 scissile bond triplet no longer donates a H-bond to an α1 chain (Fig. 4d). Moreover, the relatively large NO distances of the GPQ triplet, which is immediately N-terminal to the cleavage site, suggest that the α2 chain dissociates from the other two chains in the region (Fig. 4c and d). In particular, this triplet neither donates nor accepts H-bonds from other chains (Fig. 4d). We therefore refer to structures corresponding to this minimum as belonging to the vulnerable state (Fig. 3c). This classification is based on our previous formalism that defined vulnerable conformations as states that have partially unfolded

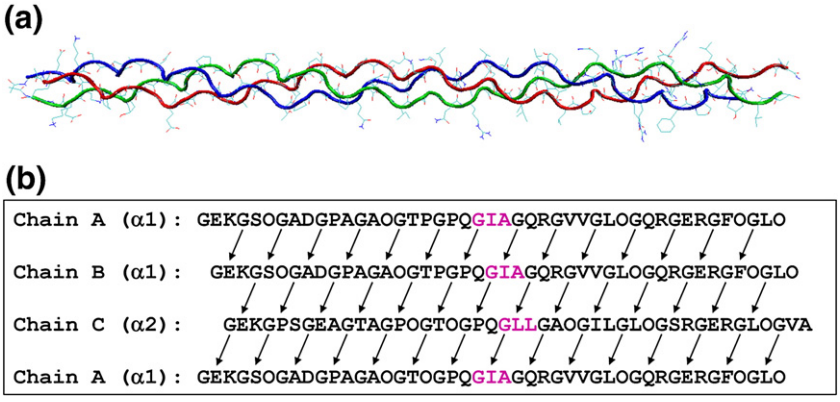

Fig. 2. (a) The prototypical triple-helical structure of type I collagen in the region of the collagenase cleavage site (generated with the Triple Helix Builder[36]). Chain A is colored blue, chain B is colored green, and chain C is colored red. (b) H-bonding pattern in the triple helical conformation. The triplets containing the scissile bonds are colored magenta. Arrows point from H-bond donors to H-bond acceptors.

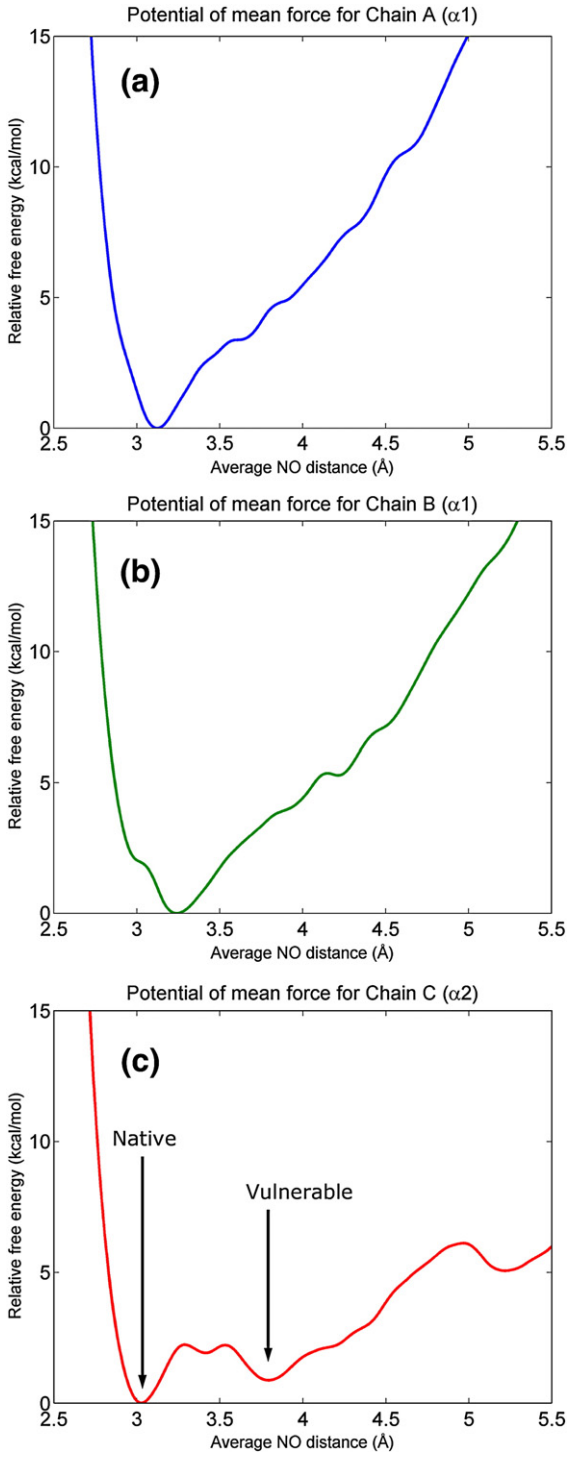

Fig. 3. Type I collagen. Potential of mean force for (a) chain A (α1); (b) chain B (α1); and (c) chain C (α2).

structures in the vicinity of the collagenase scissile bond.[13,21]

To characterize the effect that this disruption in H-bonding pattern has on the dynamics of the structure, we calculated the root-mean-square (rms) fluctuations of backbone C^{α} atoms in both the native and the vulnerable states for the α2 chain (Fig. 5). In the native state, the GPQ triplet has mildly elevated fluctuations relative to other residues — a finding consistent with the observation that one of the average NO distances of this triplet is 3.91 Å (Fig. 4b). However, in the vulnerable state, the NO distances of the GPQ triplet are significantly larger than those in the native state (Fig. 5). Therefore, in the vulnerable state, this triplet, which does not H-bond to any residue in adjacent α1 chains, has the largest rms fluctuations and consequently the greatest flexibility (Fig. 5).

The effect of hydroxylation near the cleavage site

Our data suggest that unfolding of the α2 chain in the vicinity of the collagenase cleavage site is energetically favored relative to unfolding of α1 chains in type I collagen. This is somewhat surprising, given that the α2 chain contains more hydroxyproline residues than the α1 chain near the scissile bond triplet. In particular, we model the X2 position of a G-X1-X2 triplet, which is N-terminal to the scissile bond in α1 chains, as containing proline instead of hydroxyproline, while the corresponding position in the α2 chain contains hydroxyproline (yellow residues in Fig. 1). Since collagen stability typically increases as the hydroxyproline content increases, this finding is at first counterintuitive.[20,29,30] Therefore, to further explore the role that hydroxylation at this site has on the conformational thermodynamics of collagen, we conducted additional unfolding simulations with this proline in the α1 chains replaced with hydroxyproline. We henceforth refer to the heterotrimer that contains these mutant α1 chains as the GTO mutant.

The resulting pmfs of all three chains in the GTO mutant have global energy minima with an average NO distance of less than 3.6 Å (Fig. 6a–c). In particular, the vulnerable state that was observed in unfolding simulations of the α2 chain is absent from α2 unfolding simulations with the GTO mutant (Fig. 6c). Native state structures from this minimum are triple-helical with preserved H-bonds in the vicinity of the collagenase cleavage site. Hence, α2 chain unfolding does not occur in the mutant structure. This is an interesting finding in light of the fact that we added hydroxyproline to the α1 chains but not the α2 chain.

While the pmf for α2 chain unfolding now has one distinct minimum, the pmfs for the α1 chains each have two minima. One α1 chain has a shelf-like broad minimum in the range of 3.4–4 Å (Fig. 6b), which occurs at a relative free energy of 4.2 kcal/mol; i.e. approximately 7 *RT* at 300 K. Consequently, it is unlikely that this state is significantly sampled at room temperature. The other α1 chain has minima at 3.7 Å and 4.6 Å (Fig. 6a). The minimum at 4. 6 Å has a relative free energy of 4.0 kcal/mol, which again suggest that this state is rarely sampled at room temperature. The minimum at 3.7 Å has a relative free energy of 1.3 kcal/mol (~2 *RT*), suggesting that this metastable state may correspond to an accessible state at 300K (Fig. 6a). Therefore, to explore the

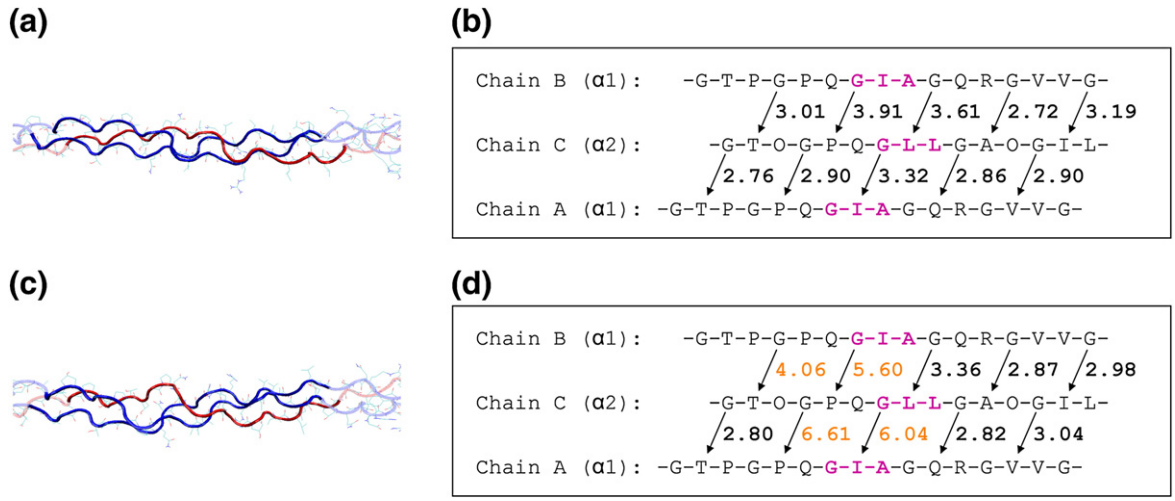

Fig. 4. (a) Representative structure from the native state of type I collagen. The α2 chain is colored red. (b) Interchain NO distances between the α2 chain and the α1 chains of the native state of type I collagen. The triplets containing the scissile bonds are colored magenta. (c) Representative structure of the vulnerable state of type I collagen. (d) Interchain NO distances between the α2 chain and the α1 chains of the vulnerable state. NO distances greater than 4 Å are highlighted in orange.

structure and dynamic properties of these states we analyzed representative structures from these minima.

Structures corresponding to the lowest energy state in Fig. 6a have, on average, preserved interchain H-bonds near the collagenase cleavage site (Fig. 7a). Additionally, triplets N-terminal to the scissile bond triplet all have NO distances of less than 3.1 Å (Fig. 7a). One H-bond, however, which is C-terminal to the scissile bond triplet, is broken as the corresponding average NO distance is ~5 Å (Fig. 7a). Representative structures of the metastable state in Fig. 6a have broken H-bonds both N-terminal and C-terminal to the scissile bond (Fig. 7b). Nevertheless, in both the native and metastable states, the central scissile bond triplet maintains both H-bonds. In addition, no triplet is dissociated from both neighboring chains in either the native or metastable state.

To examine the effect of this H-bonding pattern on chain flexibility, we computed rms fluctuations of the α1 chain C^{α} atoms for the native and metastable states of the GTO mutant. In the native state, the rms fluctuations of the GVV triplet, which is near the C terminus, are increased (Fig. 8) — an expected finding given that this triplet no longer receives a H-bond from one adjacent chain (Fig. 7a). In the metastable state, however, this triplet has

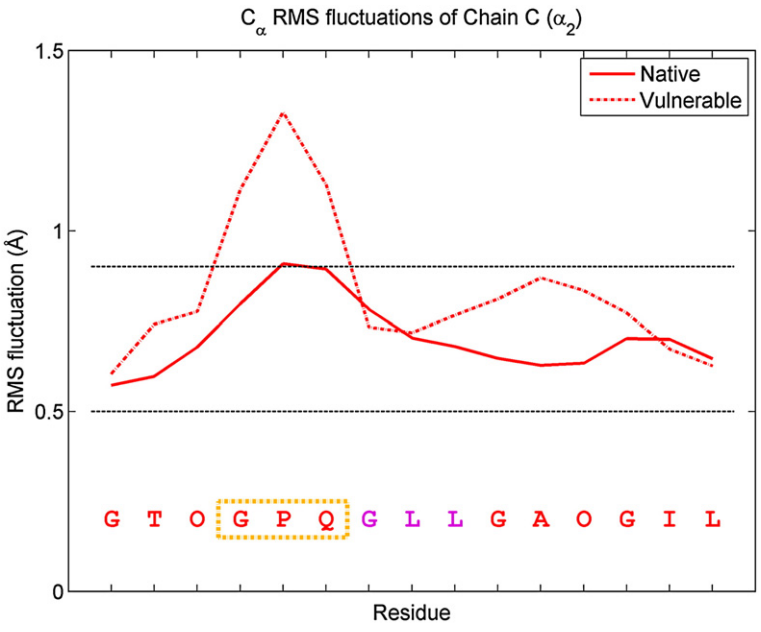

Fig. 5. The rms fluctuations of the α2 chain (chain C) backbone C^{α} atoms in the native and vulnerable states of type I collagen. The triplet containing the scissile bond is colored magenta. The triplet boxed in orange is not H-bonded to any residues in adjacent α1 chains in the vulnerable state. Black broken lines indicate the range of fluctuations within two standard deviations of the mean fluctuation in the native state. Fluctuations that fall outside this range are therefore greater than would be expected for native state residues.

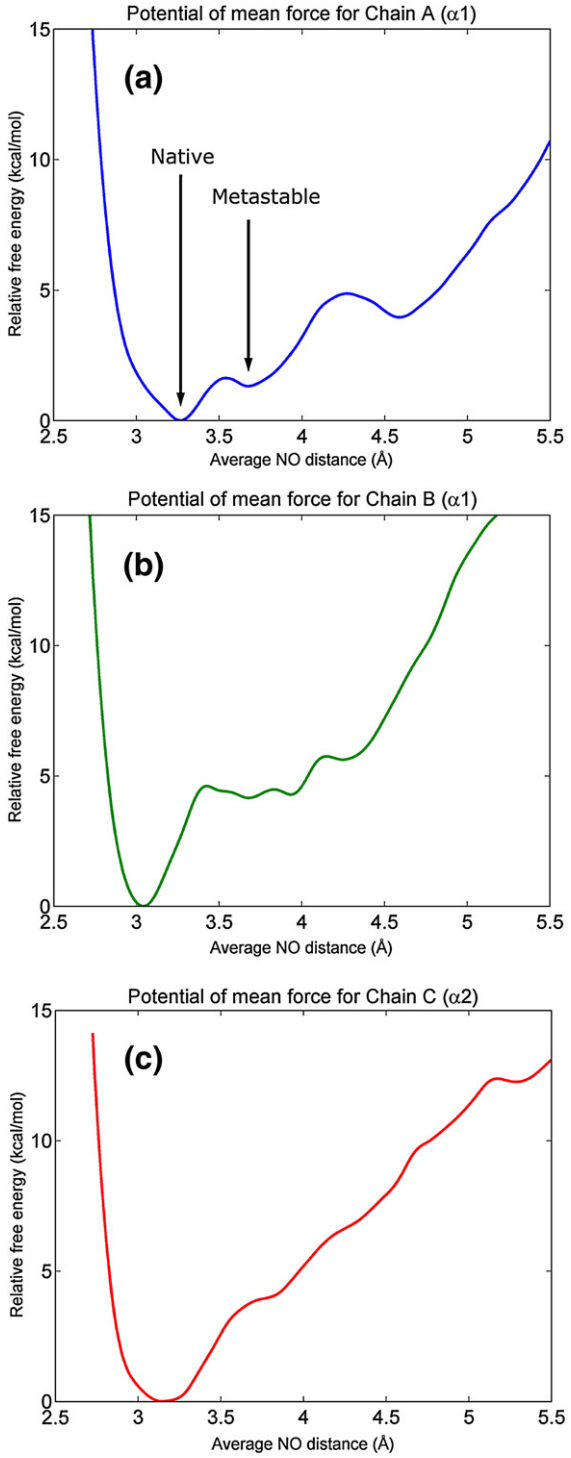

Fig. 6. GTO mutant type I collagen. Potential of mean force for: (a) chain A (α1); (b) chain B (α1); and (c) chain C (α2).

reduced rms fluctuations (Fig. 8). Although this GVV triplet does not receive a H-bond from one chain, it still donates a H-bond to another chain and the corresponding NO distance is smaller in the metastable state (Fig. 7b). Furthermore, in the metastable state, C^{α} rms fluctuations are elevated in the N-terminal region relative to the C-terminal region. This is consistent with the observation that two H-bonds are broken N-terminal to the central scissile bond in this state (Fig. 7b). Overall, both the native and metastable states have a similar range of rms fluctuations (Fig. 8).

Discussion and Conclusions

The molecular mechanism of collagenolysis is an enigma. Studies aimed at deciphering the structural changes needed to hydrolyze scissile bonds in collagen may help to elucidate the complex series of events that must occur to enable collagen degradation. To deduce structural changes in type I collagen that facilitate collagenase recognition and cleavage of scissile bonds in collagen, we explored the conformational free energy landscape of sequences that model regions near the collagenase cleavage site in collagen.

Although the DNA sequences of human α1 and α2 collagen chains are known, it is difficult to deduce the precise amino acid sequence of collagen chains from these data because collagen undergoes extensive post-translational modifications.[1] Most notably, a *sine qua non* of collagen post-processing is the hydroxylation of a significant number of proline residues in the X2 position of G-X1-X2 triplets. Since deciphering the precise hydroxylation pattern from the DNA sequence alone is not possible, the exact amino acid sequences of human α1 and α2 chains have yet to be determined. However, as the hydroxylation patterns of proline residues in bovine and chicken collagen are known, we used the significant amino acid sequence homology between bovine and chicken sequences to infer the hydroxylation pattern of the corresponding human sequences.[25,26,27,28]

Interestingly, the alignment of bovine and chicken sequences suggests that a proline residue located in the X2 position of an upstream triplet in the α1 chains is not hydroxylated. Collagen unfolding simulations using an amino acid sequence having proline in this position suggest that unfolding of the α2 chain near the collagenase cleavage site is energetically favored relative to unfolding α1 chains. More precisely, we find that the α2 chain can adopt a partially unfolded state having an energy similar to that of the native triple-helical conformation.

This partially unfolded, vulnerable state has a number of interesting structural properties. In an idealized triple helix, every G-X1-X2 triplet in the α2 sequence of type I collagen donates a H-bond to one α1 chain and accepts a H-bond from the other α1 chain.[20,30] However, in the vulnerable state, the α2 triplet that contains the scissile bond does not donate a H-bond to a neighboring α1 chain and consequently only forms one H-bond with another α1 chain. In addition, the triplet that is immediately N-terminal to the scissile bond does not form any H-bonds with neighboring α1 chains; i.e. this triplet dissociates from the other two chains. The loss of both H-bonds is associated with increased chain flexibility, as demonstrated by an increase in the rms

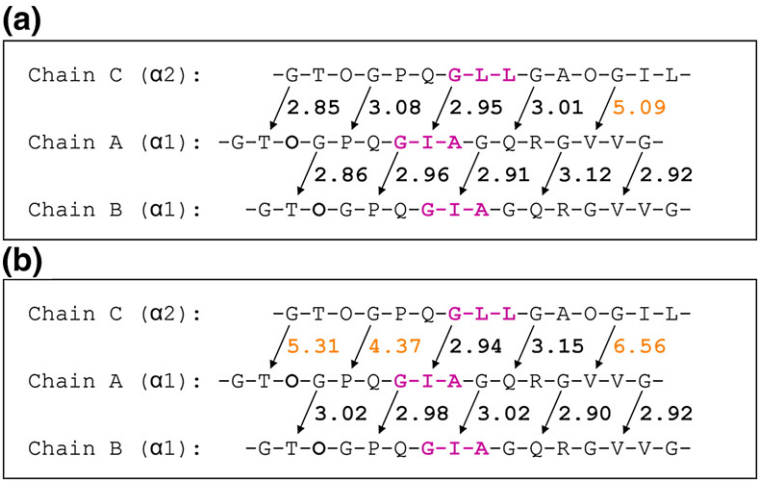

Fig. 7. Interchain NO distances between the α2 chain and the α1 chains of the native (a) and the metastable states (b) of GTO mutant type I collagen. The triplets containing the scissile bonds are colored magenta. NO distances greater than 4 Å are highlighted in orange. In both states, no triplet has broken donor and acceptor bonds, and the scissile bond triplet remains H-bonded to both neighboring chains.

fluctuations of C^{α} atoms in residues belonging to this triplet. An increase in backbone fluctuations suggests that the formation of interchain H-bonds helps to restrict the motion of collagen chains, and that local disruption of interchain H-bonding may constitute a mechanism that imparts structurally lability to regions of the collagen structure. Overall, an analysis of the rms fluctuations in stable states of both the presumed wild-type and GTO mutant collagen sequences support the hypothesis that H-bonding between adjacent chains has a direct effect on chain flexibility.

Our results were obtained using an amino acid sequence where a proline residue in the X2 position of a triplet upstream from the cleavage site is not hydroxylated. To explore the effect that this assumption has on our results, we conducted additional unfolding simulations where this proline residue was replaced with hydroxyproline. Interestingly, we find that the hydroxylation of this proline residue, which is found in α1 chains, leads to a dramatic change in the conformational free energy profile of the α2 chain. The vulnerable state of the α2 chain disappears when this proline is hydroxylated. These data therefore suggest that hydroxylation of residues on the α1 chains can affect the conformational thermodynamics on the α2 chain, and more specifically, that hydroxyproline residues can have stabilizing effects on adjacent chains.

It has been suggested that the stabilizing effect of hydroxyproline on collagen is mainly inductive.[33] Specifically, the hydroxyl group of hydroxyproline stabilizes the pyrrolidine ring in the $C_{\bar{a}}$-*exo* pucker, which in turn stabilizes both the ψ and ω (peptide bond) angles in a conformation that is compatible with the triple-helical structure.[20,31] Given this, it may be that hydroxyproline-mediated stabilization of α1 chains allows other chains to form stabilizing contacts and H-bonds to these restricted α1 chains. Along these lines we note that even though our GTO mutant is able to adopt a metastable state, all of the triplets in the α2 chain form at least one H-bond to a

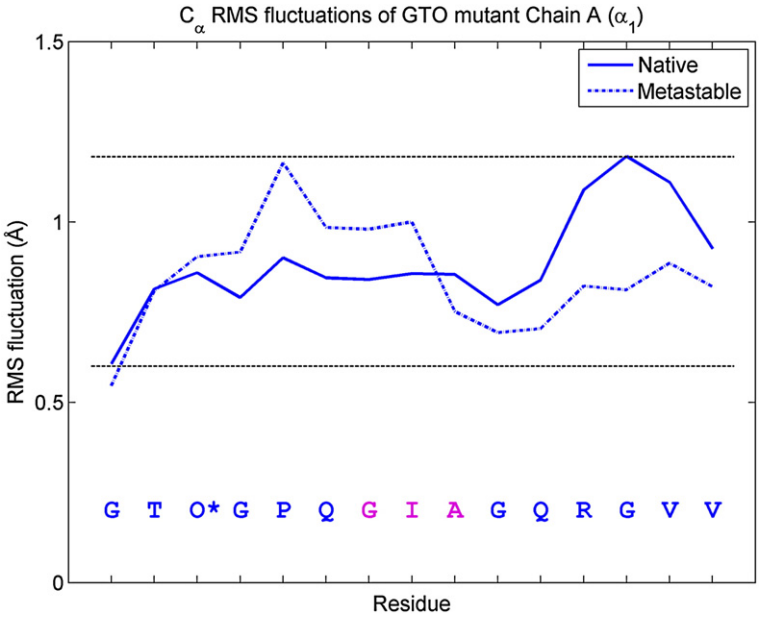

Fig. 8. The rms fluctuations of the α1 chain (chain A) backbone C^{α} atoms in the native and in the metastable states of GTO mutant type I collagen. The hydroxyproline that has been substituted for proline is denoted O*. The triplet containing the scissile bond is colored magenta. Black broken lines indicate the range of fluctuations within two standard deviations of the mean fluctuation in the native state.

neighboring chain. In addition, recent *ab initio* and classical electrostatic calculations suggest that hydroxyproline residues can form stabilizing dipole–dipole interactions with other hydroxyprolines on adjacent chains in collagen.[32] Overall, our data are consistent with these notions and suggest that non-hydroxylation of a proline residue may play a crucial role in determining the conformational flexibility and dynamics in the vicinity of the collagenase cleavage site.

This study was motivated by the results of earlier experiments, performed at 25 °C, that demonstrated that inactive MMPs can protect the α2 chain in type I collagen from peptide bond hydrolysis by other proteases that hydrolyze the MMP-specific scissile bond. On the basis of these observations, it was suggested that MMPs can bind α2 chains in the vicinity of the collagenase cleavage site.[11] Our data suggest that at room temperature localized unfolding and increased flexibility of the triplet upstream from the collagenase cleavage site may provide a mechanism that enables MMPs to recognize regions near the cleavage site. As such, our data help to explain and clarify these experimental observations.

Another more recent study utilized data obtained from X-ray fiber diffraction to deduce structural properties of type I collagen *in situ*.[33,34] Using the electron density obtained from experiments on fibrillar collagen at room temperature, a relaxed, energy-minimized, model of collagen was constructed and compared to an idealized triple-helical structure.[33] An analysis of these structures suggested that the structure of fibrillar collagen most differs from an idealized triple-helical conformation in a region of the α2 chain that is in the vicinity of the collagenase-scissile bond; i.e. in this region, the α2 chain is more dissociated from the center of the triple helix than α1 chains.[33] Additionally, residues immediately upstream from the scissile bond in the α2 chain are more separated from the center of the triple helix than residues immediately downstream from the scissile bond. These data are consistent with our results that were not obtained with a fibrillar model of collagen. As a result, our findings suggest that preferential unfolding of the α2 chain in the vicinity of the collagenase cleavage site is an inherent property of the local sequence of collagen and not a function of additional contacts that may be present in the collagen fibril. Moreover, the X-ray diffraction data suggest that the α2 chain *in situ* is oriented in a manner such that it is the most solvent-exposed chain and therefore likely to first come into contact with circulating MMPs.[33,34] Our data imply that the α2 chain has an inherent propensity to sample partially unfolded states in the vicinity of the collagenase cleavage site, even at temperatures below 37 °C. Taken together, these observations suggest that collagen fibrils have evolved a precise mechanism to ensure that the α2 chain is first to be cleaved.

Recent experiments performed at 37 °C using bovine type I collagen in solution and MMP-8 suggest similar catalytic rates for the α1 and α2 chains, implying that collagenases do not strongly prefer one sequence over the other.[35] However, as solutions of type I collagen are known to be thermally unstable at 37 °C and particularly flexible in the region of the cleavage site,[16,17] it is likely that all three chains exist in a partially unfolded state with disrupted H-bonds at this higher temperature. Although CD measurements indicate that a significant fraction of triple-helical structure exists in type I collagen at 37 °C,[35] CD spectroscopy can rarely distinguish small localized changes in structure. Furthermore, data from NMR experiments and computational simulations of collagen-like model peptides suggest that regions near the collagenase cleavage site are relatively unstable and may be the first to unfold.[10,13,15,21] If all three chains are equally unfolded at 37 °C, then the inability of MMP-8 to distinguish between the different chains in type I collagen is expected.

Similar experiments with solutions of bovine type I collagen and MMP-2, again at 37 °C, imply that gelatinases may prefer binding to the α1 chains, and cleave the scissile bond of the α2 chain.[8] When evaluating the results of experiments utilizing MMP-2, however, it is worth noting that MMP-2 contains a fibronectin type II-like collagen-binding domain in addition to the catalytic and hemopexin-like domains found in the fibrillar collagenases.[8] In addition, there are experimental data implying that the collagen-binding domain, and not the hemopexin-like domain, is primarily responsible for MMP-2 binding to collagen.[12] Hence, the preference of MMP-2 for particular exposed collagen chains will be dictated by inherent preferences that this domain may have for a given collagen chain.

Overall, our data are consistent with the notion that α2 chain unfolding is energetically preferred relative to α1 chain unfolding. Indeed, our data suggest that the α2 chain is likely to sample partially unfolded states in the vicinity of the scissile bond during its biological lifetime. Consequently, this process may provide a mechanism that enables collagenases to recognize and bind to regions near the collagenase scissile bond. Our data suggest also that non-hydroxylation of proline residues in α1 chains can increase the conformational flexibility of the α2 chain and this enables α2 chains to sample unfolded states. Our results offer a detailed picture into the mechanism of collagen degradation and highlight the importance of subtle sequence variations in the vicinity of the collagenase cleavage site.

Computational Methods

Construction of the initial model

The sequences of the α1 and α2 chains of type I collagen used in the pmf calculations included 45 residues in total — the triplet containing the scissile bond and the seven triplets adjacent to that triplet in both the N-and C-terminal directions. The initial

structure, which contained only heavy atoms, was first generated using Triple Helix Builder.[36] A polar hydrogen model of collagen was then constructed with CHARMM[37] version 33b2 using the CHARMM19 extended-atom force field.[38] Parameters for hydroxyproline were obtained from previous computational studies.[13,21] Different models of collagen were constructed, each having a different arrangement of chain stagger (e.g., the α2 chain can be chain A, B, or C in the triple helix). Initial simulations were run with each model; however, only the model where the α2 chain was assigned to chain C had stable trajectories with triple-helical structures. This structure was then minimized with 100 steps of steepest descent minimization *in vacuo* to relieve bad contacts.

Solvation and equilibration of model structure

The minimized structure was overlaid with a 30 Å sphere of equilibrated TIP3P water molecules about the center of the protein.[39] Water molecules overlapping with the protein structure were deleted; the remaining water molecules were subjected to 5 ps of molecular dynamics at 300 K, with all atoms of the protein remaining fixed. This solvent overlay procedure was repeated twice (always deleting any overlapping water molecules) to ensure adequate solvation and yielded a total of 3403 water molecules in the solvent sphere. Taking into account the volume excluded by the collagen triple helix, this number of water molecules is consistent with a density of ~1000 kg m^{-3}.

The solvated system was then minimized using 200 steps of steepest descent, followed by 200 steps of conjugate gradient minimization. The minimized system was heated linearly over 20 ps to 300 K and equilibrated for an additional 80 ps at 300 K with a Nosé-Hoover thermostat.[40] To allow for side chain relaxation during minimization, heating, and equilibration without significant distortion of the triple helix, the backbone atoms of the protein were harmonically restrained with a force constant of 100 kcal mol^{-1} $Å^{-2}$. Using atom-based cutoffs, electrostatic interactions were switched to zero between 8 Å and 12 Å, and van der Waals interactions were shifted to zero at 12 Å. The nonbonded list was truncated at 13 Å. Bond lengths involving hydrogen atoms were fixed with SHAKE/Roll and RATTLE/Roll, and a time-step of 1 fs was used with a velocity Verlet integrator.[41]

Umbrella sampling and pmf calculations

The potential of mean force for each chain was constructed using umbrella sampling with the average NO distance within the (solvated) reaction region serving as the reaction coordinate. This reaction region encompassed the seven central triplets of the sequence surrounded by the solvent sphere, with the scissile bond contained in the middle triplet. Simulations for a given chain constrained NO distances between that chain and the other two, e.g. a simulation for chain A would constrain the distances between chains A and B as well as chains A and C, hence modeling a chain separating from its two partners.

Simulation windows were centered on 1.79–6.64 Å in 0.14 Å intervals using a harmonic biasing potential with a force constant of 196 kcal mol^{-1} $Å^{-2}$. In practice, total NO distance (the number of NO pairs multiplied by average NO distance) is used for the simulations, which yields an integer, rather than fractional, increment between windows. Before beginning the umbrella sampling protocol, the system underwent an additional 100 ps of equilibration at 300 K with the H-bond lengths constrained to the starting structure. The starting window for all three chains was 2.93 Å, as this was the average interchain NO distance after side chain equilibration. Each window was simulated for 100 ps, including 40 ps of equilibration. The value of the reaction coordinate was saved every 50 fs, yielding 1200 data points per window. In some windows the equilibration time was observed to be longer, thus necessitating longer simulations. In these cases the simulation length was 200 ps, with 80 ps of equilibration, yielding 2400 data points per window. Parameters for these molecular dynamics simulations were the same as those described in the previous section.

A pmf for each chain was constructed from the biased probability distribution of each window using the weighted histogram analysis method (WHAM).[42,43] The resulting pmfs were smoothed using a robust linear fit method with a data span of 5%.

Calculation of average NO distances and C^{α} rms fluctuations

Average NO distances were obtained using the simulation window centered closest to the energy minimum of interest. Relevant NO distances were recorded every 50 fs using CHARMM, summed, and then divided by the number of data points in that window (a simple average). C^{α} rms fluctuations were obtained using the same simulation windows. An average structure for each window, along with the associated rms fluctuations, was calculated using CHARMM.

References

1. Gelse, K., Pöschl, E. & Aigner, T. (2003). Collagens – structure, function, and biosynthesis. *Adv. Drug Deliv. Rev.* **55**, 1531–1546.
2. McDonnell, S., Morgan, M. & Lynch, C. (1999). Role of matrix metalloproteinases in normal and disease processes. *Biochem. Soc. Trans.* **27**, 734–740.
3. Brinckerhoff, C. E. & Matrisian, L. M. (2002). Matrix metalloproteinases: a tail of a frog that became a prince. *Nature Rev. Mol. Cell. Biol.* **3**, 207–214.
4. Ray, J. M. & Stetler-Stevenson, W. G. (1994). The role of matrix metalloproteases and their inhibitors in tumour invasion, metastasis and angiogenesis. *Eur. Respir. J.* **7**, 2062–2072.

5. Chambers, A. F. & Matrisian, L. M. (1997). Changing views of the role of matrix metalloproteinases in metastasis. *J. Nat. Cancer Inst.* **89**, 1260–1270.
6. Bode, W., Fernandez-Catalan, C., Tschesche, H., Grams, F., Nagase, H. & Maskos, K. (1999). Structural properties of matrix metalloproteinases. *Cell Mol. Life Sci.* **55**, 639–652.
7. Jones, C. B., Sane, D. C. & Herrington, D. M. (2003). Matrix metalloproteinases: a review of their structure and role in acute coronary syndrome. *Cardiovasc. Res.* **59**, 812–823.
8. Overall, C. M. (2002). Molecular determinants of metalloproteinase substrate specificity. *Mol. Biotechnol.* **22**, 51–86.
9. Fields, G. B. (1991). A model for interstitial collagen catabolism by mammalian collagenases. *J. Theor. Biol.* **153**, 585–602.
10. Kramer, R. Z., Bella, J., Brodsky, B. & 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.
11. Chung, L., Dinakarpandian, D., Yoshida, N., Lauer-Fields, J. L., Fields, G. B., Visse, R. & Nagase, H. (2004). Collagenase unwinds triple-helical collagen prior to peptide bond hydrolysis. *EMBO J.* **23**, 3020–3030.
12. Tam, E. M., Moore, T. R., Butler, G. S. & Overall, C. M. (2004). Characterization of the distinct collagen binding, helicase and cleavage mechanisms of matrix metalloproteinase 2 and 14 (gelatinase A and MT1-MMP): the differential roles of the MMP hemopexin c domains and the MMP-2 fibronectin type II modules in collagen triple helicase activities. *J. Biol. Chem.* **279**, 43336–43344.
13. Stultz, C. M. (2002). Localized unfolding of collagen explains collagenase cleavage at imino-poor sites. *J. Mol. Biol.* **319**, 997–1003.
14. Fan, P., Li, M., Brodsky, B. & Baum, J. (1993). Backbone dynamics of $(Pro\text{-}Hyp\text{-}Gly)_{10}$ and designed collagen-like triple-helical peptide by ^{15}N NMR relaxation and hydrogen-exchange measurements. *Biochemistry*, **32**, 13299–13309.
15. Fiori, S., Saccá, B. & Moroder, L. (2002). Structural properties of a collagenous heterotrimer that mimics the collagenase cleavage site of collagen type I. *J. Mol. Biol.* **319**, 1235–1242.
16. Leikina, E., Mertts, M. V., Kuznetsova, N. & Leikin, S. (2002). Type I collagen is thermally unstable at body temperature. *Proc. Natl Acad. Sci. USA*, **99**, 1314–1318.
17. Makareeva, E., Mertz, E. L., Kuznetsova, N. V., Sutter, M. B., DeRidder, A. M. & Cabral, W. A. (2008). Structural heterogenity of type I collagen triple helix and its role in osteogenesis imperfecta. *J. Biol. Chem.* **283**, 4787–4798.
18. Nerenberg, P. S., Salsas-Escat, R. & Stultz, C. M. (2008). Do collagenases unwind triple-helical collagen before peptide bond hydrolysis? Reinterpreting experimental observations with mathematical models. *Proteins: Struct. Funct. Genet.* **70**, 1154–1161.
19. Fields, G. B. & Prockop, D. J. (1996). Perspectives on the synthesis and application of triple-helical, collagen-model peptides. *Biopolymers*, **40**, 345–357.
20. Brodsky, B. & Persikov, A. V. (2005). Molecular structure of the collagen triple helix. *Adv. Protein Chem.* **70**, 301–339.
21. Stultz, C. M. & Edelman, E. R. (2003). A structural model that explains the effects of hyperglycemia on collagenolysis. *Biophys. J.* **85**, 2198–2204.
22. Ottl, J., Battistuta, R., Pieper, M., Tschesche, H., Bode, W., Kühn, K. & Moroder, L. (1996). Design and synthesis of heterotrimeric collagen peptides with a built-in cystine-knot. Models for collagen catabolism by matrix-metalloproteases. *FEBS Lett.* **398**, 31–36.
23. Fields, C. G., Grab, B., Lauer, J. L., Miles, A. J., Yu, Y. C. & Fields, G. B. (1996). Solid-phase synthesis of triple-helical collagen-model peptides. *Lett. Peptide Sci.* **3**, 3–16.
24. The UniProt Consortium. (2008). The universal protein resource (UniProt). *Nucleic Acids Res.* **36**, D190–D195.
25. Bornstein, P. & Traub, W. (1979). The chemistry and biology of collagen. In *The Proteins* (Neurath, H. & Hill, R. L., eds), 3rd edit., vol. 4, pp. 411–632. Academic Press, New York.
26. Dixit, S. N., Mainardi, C. L., Seyer, J. M. & Kang, A. H. (1979). Covalent structure of collagen: amino acid sequence of α2-CB5 of chick skin collagen containing the animal collagenase cleavage site. *Biochemistry*, **18**, 5416–5422.
27. Glanville, R. W., Breitkreutz, D., Meitinger, M. & Fietzek, P. P. (1983). Completion of the amino acid sequence of the alpha 1 chain from type I calf skin collagen. Amino acid sequence of α1(I)B8. *Biochem. J.* **215**, 183–189.
28. Highberger, J. H., Corbett, C., Dixit, S. N., Yu, W., Seyer, J. M., Kang, A. H. & Gross, J. (1982). Amino acid sequence of chick skin collagen α1(I)-CB8 and the complete primary structure of the helical portion of the chick skin collagen α1(I) chain. *Biochemistry*, **21**, 2048–2055.
29. Jenkins, C. L. & Raines, R. T. (2002). Insights on the conformational stability of collagen. *Nature Prod. Rep.* **19**, 49–59.
30. Engel, J. & Bächinger, H. P. (2005). Structure, stability and folding of the collagen triple helix. *Top. Curr. Chem.* **247**, 7–33.
31. Holmgren, S. K., Bretscher, L. E., Taylor, K. M. & Raines, R. T. (1999). A hyperstable collagen mimic. *Chem. Biol.* **6**, 63–70.
32. Improta, R., Berisio, R. & Vitagliano, L. (2008). Contributions of dipole-dipole interactions to the stability of the collagen triple helix. *Protein Sci.* **17**, 955–961.
33. Perumal, S., Antipova, O. & Orgel, J. P. (2008). Collagen fibril architecture, domain organization, and triple-helical conformation govern its proteolysis. *Proc. Natl Acad. Sci. USA*, **105**, 2824–2829.
34. Orgel, J. P., Irving, T. C., Miller, A. & Weiss, T. J. (2006). Microfibrillar structure of type I collagen in situ. *Proc. Natl Acad. Sci. USA*, **103**, 9001–9005.
35. Gioia, M., Monaco, S., Fasciglione, G. F., Coletti, A., Modesti, A., Marini, S. & Coletta, M. (2007). Characterization of the mechanisms by which gelatinase A, neutrophil collagenase, and membrane-type metalloproteinase MMP-14 recognize collagen I and enzymatically process the two α-chains. *J. Mol. Biol.* **368**, 1101–1113.
36. Rainey, J. K. & Goh, M. C. (2004). An interactive triple-helical collagen builder. *Bioinformatics*, **20**, 2458–2459.
37. Brooks, B. R., Bruccoleri, R., Olafson, B., States, D., Swaninathan, S. & Karplus, M. (1983). CHARMM: a program for macromolecular energy, minimization and dynamics calculations. *J. Comput. Phys.* **4**, 187–217.
38. Neria, E., Fischer, S. & Karplus, M. (1996). Simulation of activation free energies in molecular systems. *J. Chem. Phys.* **105**, 1902–1921.
39. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. (1983). Comparison of

simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935.
40. Hoover, W. G. (1985). Canonical dynamics: equilibrium phase-space distributions. *Phys. Rev. A*, **31**, 1695–1697.
41. Lamoureux, J. & Roux, B. (2003). Modeling induced polarization with classical Drude oscillators: Theory and molecular dynamics simulation algorithm. *J. Chem. Phys.* **119**, 3025–3039.
42. Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H. & Kollman, P. A. (1992). The weighted histogram analysis method for free-energy calculations on biomolecules. *J. Comput. Chem.* **13**, 1011–1021.
43. Souaille, M. & Roux, B. (2001). Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations. *Comput. Phys. Commun.* **135**, 40–57.