jmb Article

Insight into the Degradation of Type-I Collagen Fibrils by MMP-8

Kuojung G. Lu and Collin M. Stultz

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

Correspondence to Collin M. Stultz: *The Institute for Medical Engineering and Science, Massachusetts Institute of Technology, Building 36-796, 77 Massachusetts Avenue, Cambridge, MA 02139, USA. cmstultz@mit.edu*
http://dx.doi.org/10.1016/j.jmb.2013.02.002
Edited by D. Case

Abstract

Although a number of studies have shed light on the mechanism of collagen degradation in solution, the precise mechanism of collagenolysis in the native fibrillar state remains unclear. To gain insight into the mechanism of fibrillar degradation, we calculated the conformational free-energy landscape for unfolding regions of the α2 chain of type-I collagen within the context of the microfibril. Our data suggest that, relatively, imino-rich sequences maintain the canonical triple-helical structure at body temperature. By contrast, the unique MMP (*m*atrix *m*etallo*p*roteinase) cleavage site adopts conformations where the α2 chain is dissociated from the rest of the fibril—behavior that is similar to what was observed in unfolding simulations of isolated collagen-like model peptides in solution. However, the dissociated cleavage site does not fit within the catalytic site of MMP-8, a representative fibrillar collagenase. Additional free-energy simulations suggest that the presence of the catalytic domain leads to a reorientation of the α2 chain such that it adopts a pose that is complementary to the enzyme's active site. These observations argue that, in the fibrillar state, there is a synergy between the normal thermal fluctuations of the substrate when the enzyme is absent and the fluctuations of the substrate when the enzyme is present. More precisely, our findings suggest that thermal fluctuations serve as the driving force for a degradative process that requires both an unfolded cleavage site and the presence of the enzyme.

Introduction

Collagen is the most prevalent biomolecule in mammals. Although a structural molecule that is the source of tensile strength in many tissues, collagen is dynamically regulated in that the net amount of collagen in any tissue is a function of both collagen synthesis and catabolism. Indeed, collagen turnover is crucial for many normal physiologic processes such as tissue repair and wound healing. However, excessive collagen degradation has been implicated in several diseases such as atherosclerosis, tumor metastasis, and arthritis.[1–3] While several studies have shed light on the mechanism of collagenolysis in solution, the precise mechanism of collagen degradation in the native fibrillar state remains mysterious. Consequently, the motivation of this work was to gain insight into how type-I collagen fibrils are degraded by MMPs (*m*atrix *m*etallo*p*roteinases), that is, proteases that play a role in the initial degradation of collagen.

There are at least 29 types of collagen in the human body, with type-I collagen being the most abundant.[4–6] All fibrillar collagens, including type-I collagen, is composed of smaller building blocks called tropocollagens or, equivalently, triple-helical units. In type-I collagen, each triple-helical unit is composed of two α1 chains and one α2 chain. MMP-mediated collagen degradation, or collagenolysis, begins at a unique site within each chain of each triple-helical unit. A number of studies suggest that this cleavage event preferentially occurs at the site

on the α2 chain, breaking the chain into three-quarter and one-quarter length fragments.[7–10] Subsequent unfolding of the triple helix followed by catabolism of the rest of the molecule then proceeds with the help of other proteases.[11]

Crystallographic studies on MMPs and collagen-like triple-helical peptides in solution argue that the scissile bond of collagen is not solvent accessible and does not fit into the active site of MMPs.[9,12–16] Hence, if the structural data are representative of the true *in vivo* state of collagen and their associated proteases (such as MMP-1 and MMP-8), then collagen degradation should not occur—clearly an erroneous conclusion. Several theories have been proposed to explain the physical basis of collagenolysis in order to resolve this dilemma.

In the first, the enzyme unwinds (or unfolds) triple-helical collagen leading to exposure of the scissile bond.[9] We note that MMPs are generally multi-domain proteases that contain both a catalytic domain and a C-terminal hemopexin-like domain.[17,18] In this formalism, the catalytic and hemopexin-like domains coordinate to unwind collagen, thereby exposing one of the triple-helical chains; that is, MMPs can also act as helicases.[19] An analysis of degradation data on MMP-mediated collagenolysis suggests that the helicase activity associated with MMP action entails a significant energetic cost.[20] However, collagen degradation is an energy-independent process. Recent structural data on MMP-1 suggest that the enzyme can alternate between open and closed states,[21] and it has been proposed that the existence of these states provides a mechanism that enables MMPs to unwind the cleavage site in an energy-independent manner.[22] Nevertheless, it is not clear if the energetic changes associated with moving between these two states provide enough energy to overcome the estimated energetic cost associated with MMP-mediated unwinding.[20]

Another mechanism proposes that the enzyme's involvement is to aid in what has been termed "stochastic unfolding" of the collagen triple helix. One interpretation of stochastic unfolding is that the mere presence of the enzyme increases the probability that chain unfolding occurs.[23] Hence, in this formalism, the energy for unfolding arises solely from the thermal energy. In yet another theory, it has been argued that the degradative process is driven by thermal energy alone in that thermal fluctuations cause collagen to sample partially unfolded states that are complementary to the enzyme active site.[16,20,24,25]

All of these theories were developed using data obtained from experiments with triple-helical collagen molecules in solution; therefore, it is not clear whether these ideas are applicable to the fibrillar state. While a number of studies on the action of MMPs on fibrillar collagen have been published, they are of insufficient resolution to unambiguously determine the precise mechanism whereby collagen is recognized and cleaved.[8] While a high-resolution structure of an inactive MMP-1 mutant bound to a collagen-like triple-helical peptide has recently been reported,[26] high-resolution structures of the fibrillar state have been particularly hard to obtain. Lower-resolution X-ray diffraction studies on the fibrillar state have provided important insights into structural features of the collagen microfibril. Early studies proposed an "end-overlapping" model with close packing of individual collagen molecules.[27] Additional studies using fiber diffraction clarified the packing, symmetry, and telopeptide orientation within the microfibrillar structure.[28–30] These studies make it clear that there is tight packing within the microfibril. Hence, it may be that the results obtained from analyses of collagen degradation in solution are not fully applicable to the process of collagen degradation in the fibrillar state.

Recent X-ray diffraction data on type-I collagen fibrils from rat tails led to a limited resolution (5.16 Å axial and 11.1 Å equatorial) electron density map of the microfibrillar structure.[31,32] In order to obtain atomic coordinates for the microfibril, were docked high-resolution structures of collagen-like peptides into the electron density, yielding a model where the placement of the C^{α} atoms could be estimated.[31,32] Hence, while the precise coordinates of all atoms within the microfibril could not be determined with high accuracy, the model represents an attempt to develop an atomistic model of the microfibril. To gain insight into the precise conformational fluctuations that may be present in the microfibrillar state, we constructed a low-energy all-heavy-atom model for the surface of a type-I collagen fibril using these structural data as a guide. We then conducted detailed free-energy calculations to explore the conformational thermodynamics of different regions of the collagen chain. We find that despite the tight packing of collagen molecules, there is still considerable flexibility in distinct regions of the microfibrillar structure. In particular, for the unique collagenase cleavage site, the dominant state is dissociated from the rest of the chain. Interestingly, however, the amount of unfolding is insufficient to lead to productive binding to the catalytic site of a representative fibrillar collagenase, MMP-8, a member of the fibrillar collagenase family that includes MMP-1 and MMP-13.[17] Additional unfolding simulations with the catalytic domain of MMP-8 present leads to complexed structures that have the α2 chain in an orientation that is conducive to scissile bond cleavage. In essence, when the enzyme is present, the local environment about the cleavage site is different (relative to the state when the enzyme is absent) and, in this new environment, the cleavage site samples, with relatively high probability, states that are complementary to the catalytic

site. These data are consistent with a model where collagenolysis requires both a dissociated chain and the presence of an enzyme that facilitates chain reorientation.

Results

An all-atom model for the collagen microfibril

A recent X-ray diffraction structure of the collagen microfibril served as the starting point for this study. While the relatively poor resolution of the structure itself only allowed for the determination of C^{α} atom locations in a collagen microfibril subunit,[31] the model itself serves as a useful initial starting point from which a complete coordinate set could be built.

Starting from the C^{α} coordinates [Protein Data Bank (PDB) ID 3HQV],[21,33] we generated an all-heavy-atom model of the microfibril using SABBAC[34]—a tool that facilitates the construction of atomic coordinates given a C^{α} backbone trace. In this model, most cleavage sites are covered by telopeptides arising from the N-terminal triple-helical unit (Fig. 1a). Cleavage sites can potentially become exposed when the N-terminal telopeptide is cleaved or when the N-terminal triple-helical unit is degraded.[32] With the use of the proposed quasi-hexagonal packing of microfibrils, the structure of the initial collagen model could be duplicated and translated to build a model of the microfibrillar structure and, hence, a model of the fibrillar surface[35] (Fig. 1b). This model corresponds to either (1) the region of the fibril that is located at the "N-terminus" or (2) the region of the fibril that is downstream from a triple-helical unit that has been cleaved.

Cleavage sites on triple-helical units that are located at the N-terminal region of the fibril, T_E, are the most solvent exposed (Fig. 1b and c), while cleavage sites on triple-helical units located closer to the C-terminus, T_C, are covered by the corresponding C-terminal telopeptide from the preceding N-terminal collagen molecule (Fig. 1d). This observation suggests that collagenolysis at the fibrillar surface occurs from the N-terminus to the C-terminus because the N-terminal sites are the most accessible.[32] Site "I" (Fig. 1e) corresponds to a relatively solvent exposed region that contains an imino-rich sequence that is not recognized by MMPs.

To explore the conformational thermodynamics at each of these sites, we computed the free energy for unfolding each uncovered site, that is, sites T_E and I (site T_C was excluded since it is buried; Fig. 1d). Computing the relative free-energy profiles for unfolding any molecule requires a suitable choice for the reaction coordinate, which quantifies the extent of unfolding. As in our prior work, we choose the average distance between hydrogen-bond pairs on adjacent amino acid chains.[10] Since a defining characteristic of the collagen chain is that the three chains all have a specific network of inter-residue hydrogen bonds, this serves as a reasonable metric to distinguish between both the folded state and the unfolded state. More precisely, in the prototypical folded state, for each G-X-Y triplet, the G amide hydrogen is hydrogen bonded to the X carbonyl oxygen on another chain, while in the completely dissociated state, all of these hydrogen bonds are broken. As prior work suggests that the α2 chain preferentially unfolds and is recognized first by MMPs, such as MMP-1, we focus on the free energy associated with unfolding the α2 chain.[9,10,32]

For each site, the relative free-energy profiles, or pmfs (*p*otentials of *m*ean *f*orce), were calculated by gradually causing the α2 chain to unfold in the region of interest, using an umbrella potential, and then subsequently calculating the associated free energy of the partially unfolded intermediates. We note that this formalism has been widely applied in the literature and yields the rigorous free energy of unfolding up to an additive constant.[16,36] That is, while the absolute value of the free energy of any conformer is not accurate, the relative free energies of different conformers are (i.e., they include all enthalpic and entropic contributions).[36] In all of the free-energy profiles shown, the lowest energy for the plot is set to 0. In this sense, the energies for a given plot are all relevant to the ground state energy.

The calculated pmfs are shown in Fig. 2. The *x*-axis indicates the average N–O distance, while the *y*-axis indicates the relative free energy. Site "I" has a deep minimum at approximately 3.3 Å, suggesting that, on average, the hydrogen bonds in the imino-rich site are preserved. While there is a local energy minimum at site "I" at approximately 7 Å, this state has an energy that is 7 kT higher than the ground state and therefore is not thermodynamically relevant. By contrast, site "T_E" has a minimum at 5.9 Å. Hence, the true cleavage site is more dissociated.

Representative structures from each of the minima along with the corresponding average N–O distances for each minimum are shown in Fig. 3. N–O distances smaller than 3.5 Å are generally considered to be indicative of hydrogen bonding. The imino-rich site, "I", largely maintains its canonical structure and has its hydrogen-bonding pattern preserved, especially in the position that is analogous to that of the scissile bond in the true cleavage site (Fig. 3a). However, the site corresponding to the true cleavage site, T_E, is more dissociated from the rest of the chain in that the majority of the hydrogen bonds about the scissile bond are disrupted (Fig. 3b).

MMP binding to the true cleavage site on the fibrillar surface

The ability of the MMP catalytic domain to localize to the true cleavage site is likely facilitated by non-

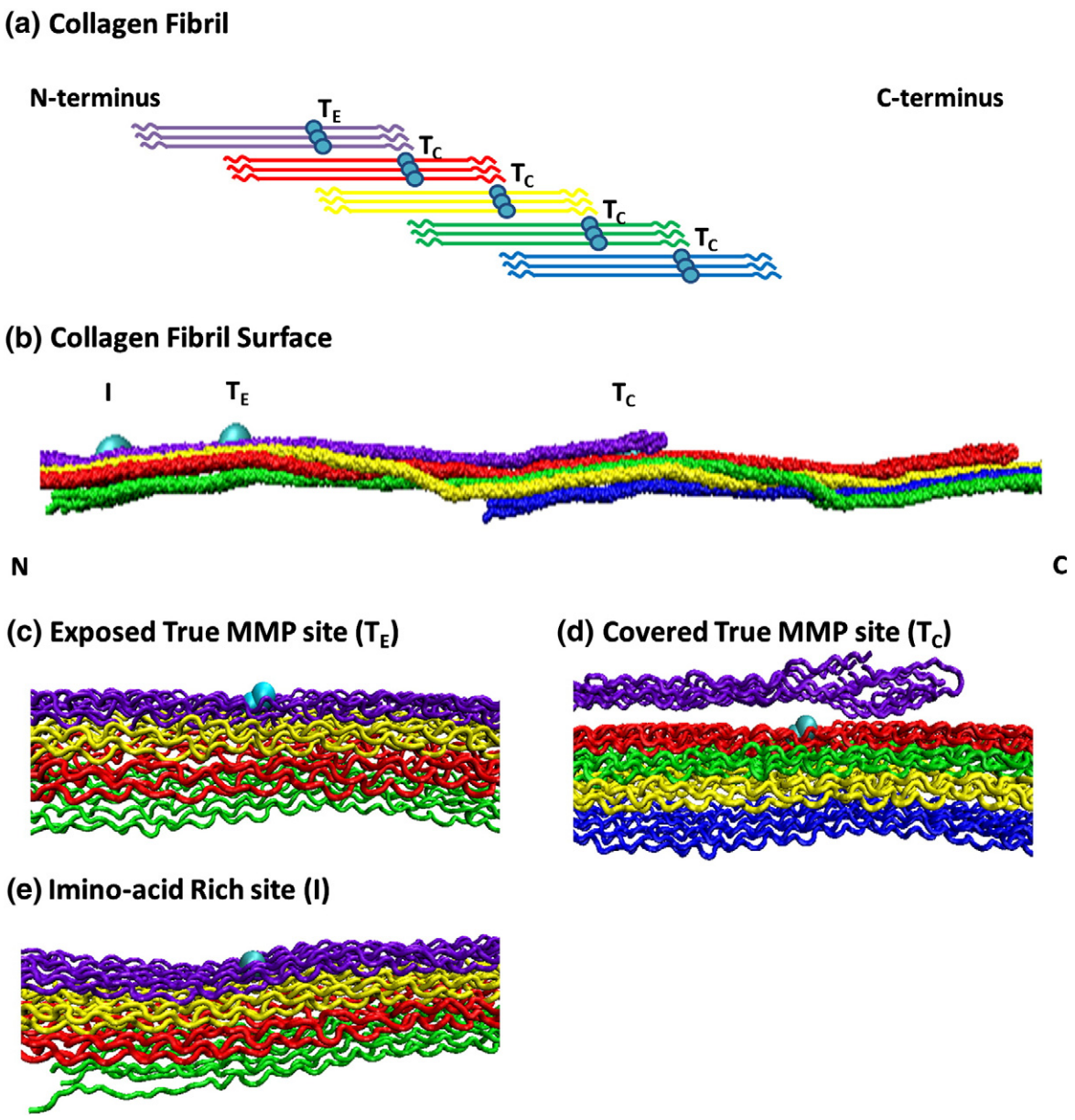

Fig. 1. Schematic of collagen fibrillar surface. (a) Schematic showing the arrangement of triple-helical units in the microfibril. (b) A collagen microfibril, representing the outermost microfibril on the fibril surface. Each microfibril consists of repeating triple-helical collagen units, each in a different color. Cyan spheres indicate atoms that comprise the scissile bond at site T_E (the exposed N-terminal true MMP cleavage sites) and site T_C (the scissile bond in a downstream MMP cleavage site). For site I, the cyan spheres denote the position of a Gly-Ile sequence in a region that has a relatively high imino acid content. (c–e) Close-up of the three locations on the fibrillar surface.

catalytic domains of MMPs, which may bind to specific sequences near the true collagenase cleavage site.[31,37] This process enables the catalytic domain to be "concentrated" in the vicinity of the true cleavage site. Once the catalytic domain is near the true cleavage site, we wished to determine whether the extent of chain dissociation at site T_E is sufficient to allow the catalytic domain to bind and cleave the corresponding scissile bond. Therefore, we docked the catalytic domain of a representative fibrillar collagenase, MMP-8, onto structures representing the ground state on the free-energy landscape of T_E.

Docking simulations were performed with 3000 structures chosen from the umbrella sampling window corresponding to the lowest-energy state for the T_E site. Rigid-body docking of the enzyme to each collagen structure was performed in a manner

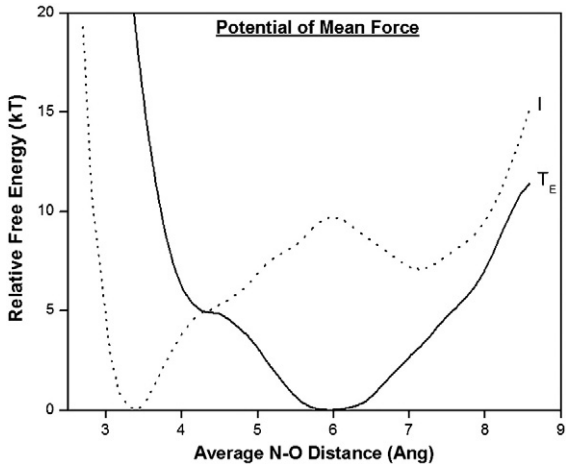

Fig. 2. pmf for unfolding the α2 chain in the type-I collagen microfibril. The true collagenase cleavage site (T_E) is the continuous line, and the imino-acid-rich site (I) is the broken line.

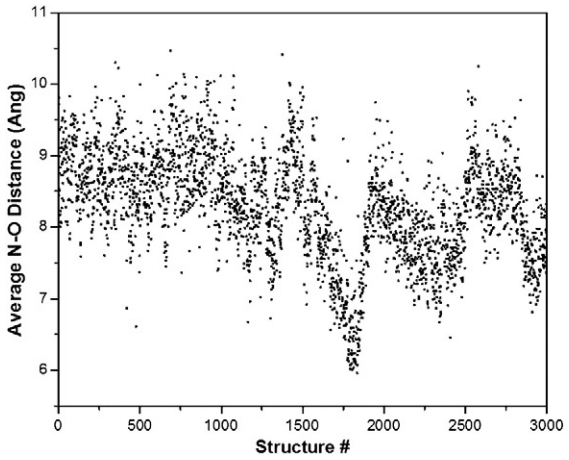

Fig. 4. Results of rigid-body docking of 3000 structures from the window corresponding to the lowest-energy state for the T_E site, into the catalytic site of MMP-8. The average N–O distance is calculated between the backbone atoms on the fibril and residues in the MMP-8 that are known to form background hydrogen bonds with the substrate.

reproduce the hydrogen-bonding pattern observed between MMP-8 and the backbone of a peptide in the active site of the MMP-8 co-crystal structure.[21] However, none of the 3000 structures docked well into the active site of MMP-8. The best-docked structures had fibril backbone atoms that were greater than 6 Å away from the enzyme backbone atoms on average (Fig. 4). By contrast, in the MMP-8 co-crystal structure, the docked peptide atoms were approximately 3 Å, on average, from the enzyme

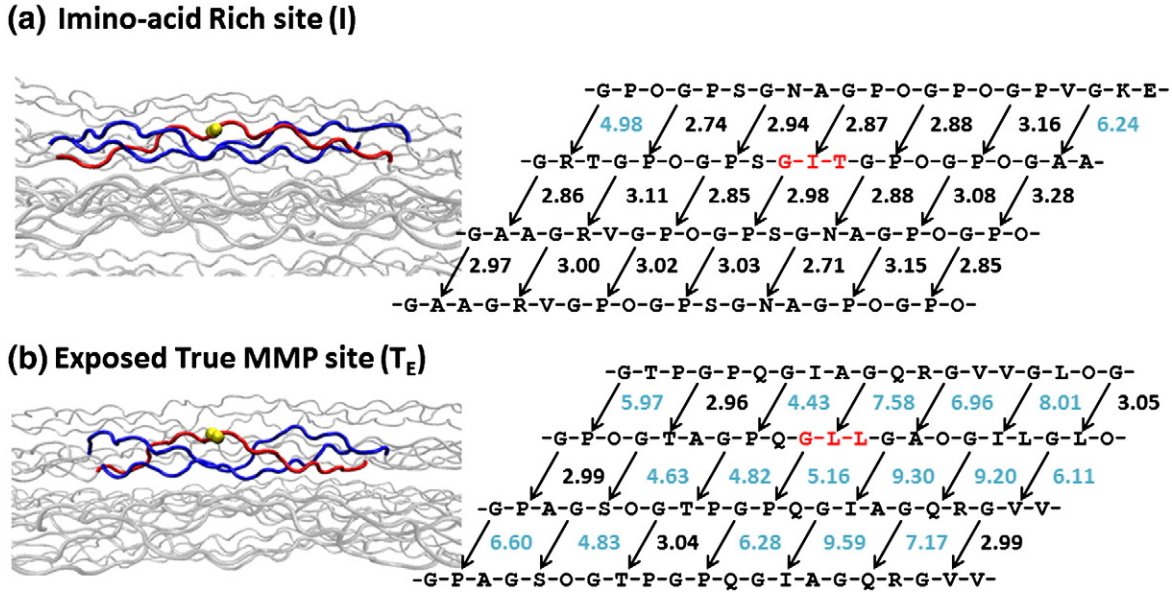

Fig. 3. Representative structures from the preferred dissociation distance for (a) imino-acid-rich sequence and (b) the true cleavage site. The solvated and unconstrained region (that undergoes full molecular dynamics) for each structure is shown in color, while the remainder of the fibril is shown in gray. The dark-blue chains are the α1 chains, and the red chain is the α2 chain. The two atoms making up the scissile peptide bond are highlighted in yellow. The right side of each figure indicates the inter-chain N–O distance of each of the three regions at the preferred dissociation distance, in angstroms. The triplet containing the scissile bond, or the central triplet for the non-cleaved sequence, is colored red. The arrows indicate the associated N–O inter-chain distances. N–O distances that are less than 3.5 Å are colored black (suggesting the presence of a hydrogen bond); distances that are greater than 3.5 are colored cyan.

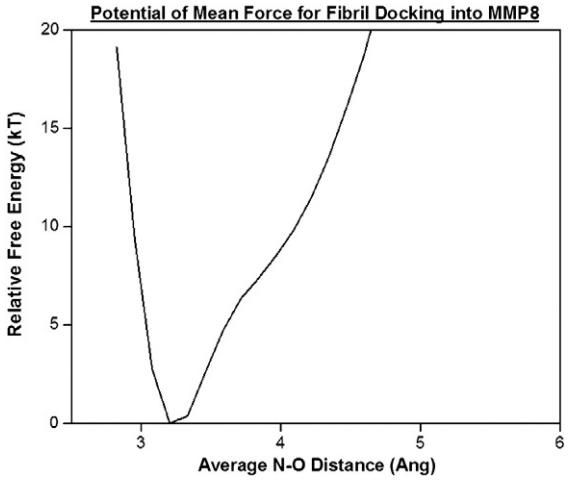

Fig. 5. pmf for binding of the α2 chain to the MMP-8 catalytic site.

backbone atoms. In addition, in many of the docked structures, the Leu785 in position P_1' was pointed away from the hydrophobic pocket S_1' on the catalytic domain, whereas pocket S1′ normally houses the side chain of a hydrophobic residue.[21]

To explore the effect that MMP-8 has on the thermally accessible states of the α2 chain, we calculated the free-energy profile associated with T_E binding to the catalytic site of MMP-8. For the starting structure, we picked one of the 3000 initial docked structures that had both a low Zn^{2+}-to-scissile bond distance and a Leu785 that was close to the hydrophobic pocket. We use the average N–O distance between the catalytic domain and the backbone of the α2 chain as the reaction coordinate together with an additional constraint that ensures that Leu785 would sample states where it bound the S_1' pocket. The resulting pmf is shown in Fig. 5. The *x*-axis represents the average N–O distances between the fibril and the catalytic domain, and the *y*-axis represents the relative free energy. The lowest-energy state has a deep minimum at 3.2 Å, suggesting that the α2 chain is well situated in the catalytic site, making, on average, favorable hydrogen bonds with the enzyme.

Figure 6a shows a representative docked structure from the lowest-energy window in Fig. 5, with a

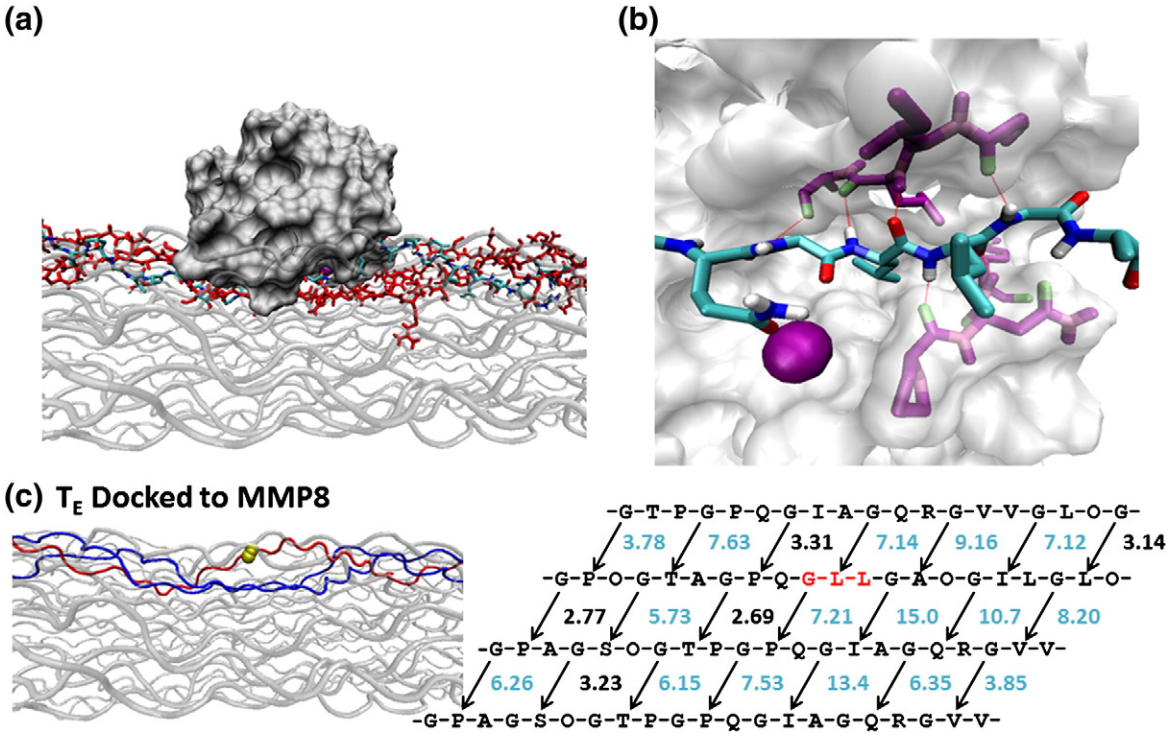

Fig. 6. A representative docked structure from the window corresponding to the lowest-energy state. (a) Side view of the structure docked into the catalytic site of MMP-8. The central triple helix of the fibril is shown in stick representation, and the rest of the fibril surface is shown as gray transparent tubes. The α1 chains are colored red, and the α2 chain is colored cyan. The enzyme is shown as a surface representation, in gray. (b) The view of the catalytic site, with the docked α2 chain. The Zn^{2+} ion in the catalytic site is shown in purple. The hydrogen-bonding pattern of the α2 chain backbone to the catalytic domain of MMP-8 is shown in red lines and closely follows that of the one shown in the crystal structure of MMP-8 with a bound peptide (PDB 2OY4). (c) Representative structure and average hydrogen-bonding distances of the T_E site, after it has been docked to MMP-8. The coloring of the hydrogen bonds is the same as in Fig. 3.

hydrogen-bond distance between the enzyme and the fibril that only differs on average 0.15 Å from that of the corresponding crystallographic structure. The active-site Zn^{2+} is in close proximity to the scissile bond (G784–L785), and the distance from Zn^{2+} to the glycine 784 carbonyl oxygen is 3.0 Å—a value similar to that observed in the crystallographic structure crystallographic structure (Fig. 6b). In this docked structure, the hydrogen bonds between the three chains of the triple helix around the true cleavage site are again mostly disrupted (Fig. 6c). However, the pattern and degree of disruption, shown by the average hydrogen bonds, are different than that of the true cleavage site without docked MMP-8 (Fig. 3b).

These results demonstrate that while unfolding of the α2 chain alone is insufficient, by itself, to explain binding of to the MMP catalytic site, the presence of the enzyme leads to a reorientation of the α2, such that it adopts a pose that is conducive to binding the enzymatic active site.

Discussion

While the mechanism of collagenolysis in solution has been extensively studied, these observations might not be applicable to the fibrillar state because collagen fibrils are composed of tightly packed triple-helical collagen units.[4,35] Experimental studies that provide sufficient resolution to directly address this question are quite difficult to perform; hence, in this study, we have taken a physics-based approach to understand the thermally accessible states of the substrate, collagen. Physics-based computations such as the present study allow us to ask detailed questions at an atomic level of detail and can provide insights that would be difficult to garner from experiment alone.

Our model for the microfibril was obtained from a previous fiber diffraction study on type-1 collagen.[32] In this model, the cleavage site on a given triple-helical unit is covered by the telopeptide from the N-terminal triple-helical unit. The model used in this study corresponds to either (1) the region of the fibril that is located at the "N-terminus" or (2) the region of the fibril that is downstream from a triple-helical unit that has been cleaved. Using this model as a starting point for unfolding simulations, we find that triple-helical units that are not covered by N-terminal telopeptide sample dissociated states at body temperature. Indeed, at the true collagenase cleavage site, essentially all inter-chain hydrogen bonds are broken. This behavior is to be distinguished from a prototypical imino-rich region that retains the canonical triple-helical structure at body temperature. These findings are consistent with and help to clarify prior observations on the mechanism of collagenolysis. In the original fiber diffraction study of Perumal *et al.*, chain dissociation was estimated using the relative C^{α} distance from the helix center.[32] While chain dissociation of the α2 chain was indeed significant at the true cleavage site, it was unclear whether dissociation at this site was significantly more than it was at many other positions along the fibrillar structure. Given the relatively low resolution of the fiber diffraction data, it is not clear how reliable the calculated C^{α} distances from the helix center are. Our studies suggest that the α2 chain dissociates from the microfibril at the true cleavage site. In this sense, we also find that the extent of chain dissociation in the fibrillar state is insufficient to ensure adequate binding to the catalytic site.

Docking of partially unfolded structures of collagen into the catalytic site of MMP-8 does not result in structures that place the scissile bond in an orientation that is suitable for catalysis. In the fibrillar state, the tightly packed triple-helical units prevent the enzyme from coming into close proximity to the atoms that form the scissile bond. However, although the enzyme's movement is restricted by the presence of other tightly packed collagen molecules, the presence of the catalytic domain changes the local environment at the cleavage site. The result is that the α2 chain samples states that are distinct from those that are observed when the enzyme is absent. Overall, the free-energy profile in the presence of the enzyme favors conformations of the α2 chain that are complementary to the active site. Representative structures of the bound state are similar to that observed in the crystal structure of MMP-8 with a bound peptide (PDB 2OY4).[21,33] In particular, in these structures, a Zn^{2+}-coordinated water molecule is positioned to act as a nucleophile that attacks the scissile bond.[38]

Our results suggest that catalysis in the fibrillar state may be driven by two physical phenomena: (1) the thermal energy of the substrate (collagen) and (2) the presence of the catalytic domain, where the latter effect is responsible for reorienting the α2 chain. In a recent study, Sarkar *et al.* used single molecule fluorescence to track individual MMP-1 molecules as they degraded type-I collagen fibrils.[39] Interestingly, MMP-1 spent ~90% of the time in a paused state. Two types of pauses were noted. Class I pauses occurred randomly along the fibril while class II pauses occurred at "hot spots" located at periodic intervals. Moreover, class II pauses were followed by biased motion indicative of collagen cleavage by the enzyme. In addition, class II pauses were approximately twice as likely at 37 °C relative to 25 °C. To explain the temperature dependence of class II pauses, the authors hypothesized that these events correspond to melting or rearrangement of collagen at specific sites leading to exposure of the cleavage site. Our data are consistent with the notion that cleavage site rearrangement may indeed play a

role in the etiology of class II pauses. Although our calculations suggest that the cleavage site is partially unfolded, unfolding alone is not enough to ensure that the cleavage site samples states that are complementary to the active site of the enzyme. When the enzyme is present, the probability that the cleavage site samples states that are compatible with the enzyme is increased. Hence, one potential explanation for the associated pauses is that once the enzyme is guided to a region near the cleavage site (a process facilitated by the hemopexin-like domain), it must "wait" until the cleavage site reorients to a conformation that is appropriate for binding.[40]

One limitation of our results is that the simulations with the enzyme only include the catalytic domain of MMP-8 instead of the full enzyme, which contains both a catalytic domain and a hemopexin-like domain. Hence, we cannot exclude the fact that a more complex mechanism may be at play when the full-length enzyme is present. In this regard, we note that it has been proposed that non-catalytic domains, such as the hemopexin-like domain, may help to localize the catalytic domain in the region of the true collagenase cleavage site.[31,37] In this sense, non-catalytic domains serve to increase the effective concentration of the catalytic domain in the vicinity of the true cleavage site.[20] In addition, non-catalytic domains may further facilitate collagenolysis by enabling the protein to preferentially bind preexisting partially unfolded states with high affinity.[20,24] Such effects are important because the cleavage rate of MMP-8 against type-I collagen in solution is very slow when the hemopexin-like domain is removed.[24]

Recently, Bertini *et al.* studied the collagenolytic mechanism of MMP-1 in solution.[22] Using a series of NMR experiments, the authors observed two distinct conformations for full-length MMP-1: a closed form and an open form. Starting from the premise that collagen does not adopt partially unfolded conformations in solution, they postulated that closed and open forms could work together to essentially dislodge one of the collagen chains from the triple helix. The authors then propose a mechanism for collagenolysis where the transition from open to closed states of the enzyme when it is bound to a collagen-like peptide helps to dislodge the α2 chain, thereby exposing the scissile bond. The HADDOCK[41] energy function was then used to evaluate the relative stability of the structures along this hypothetical pathway. The calculated energies of the structures were quite favorable—on the order of −300 kcal/mol, that is, much larger in magnitude than the experimentally determined activation energy associated with early stages of collagenolysis (~26–49 kcal/mol for type-I collagen in solution[23] and 101 kcal/mol for the fibrillar state[42]). The conclusion being that the conformational changes themselves are energetically favorable and consequently, the helicase activity of MMPs can occur without additional energy input. However, it is important to note that the HADDOCK energy function was designed to be a "scoring function" and it has only been shown to differentiate between well-docked protein–protein complexes and complexes that are poorly docked.[41] Hence, there are little, if any, data to suggest that values arising from the HADDOCK scoring function have any relationship to measured free energies, including the activation free energy. This is especially poignant because the formalism employed in HADDOCK ignores the conformational entropy of the system. Moreover, the relevance of these energetic data to the activation energy is premised on the assumption that the chosen set of intermediate structures actually corresponds to a physically meaningful pathway. As the intermediate structures were chosen by visual inspection, as opposed to rigorous transition path sampling on a more realistic potential energy surface, it is an open question whether the chosen intermediates are physically relevant. Hence, while the authors' analysis is interesting, it is not clear that their proposed mechanism is physically viable.

One of the strongest arguments in favor of the helicase argument is that a catalytically inactive full-length form of MMP-1 (E200A) can enable MMP mutants that only contain the catalytic domain to cleave collagen molecules in solution.[9] If full-length MMP can act on collagen to dislodge the cleavage site, then it is not surprising that catalytically inactive full-length enzyme can still act as a helicase, thereby leading to exposure of the scissile bond for other enzymes to cleave the molecule. However, a kinetic analysis of these results demonstrates that collagen degradation in the presence of catalytically inactive full-length enzyme and the catalytic domain alone is equally well explained by a mechanism where full-length enzyme binds and stabilizes preexisting partially unfolded states.[20,24] After binding, the catalytic domain is free to cleave a relatively exposed cleavage site.

Our studies focused on the interaction between type-I collagen and MMP-8. While MMP-8, like most other interstitial collagenases, contains a catalytic domain and a hemopexin-like domain, which is required for efficient collagenolysis, it has slightly different cleavage rates and substrate specificity than other MMPs such as MMP-1.[24] Therefore, while our results may have implications for collagenolysis of fibrillar collagen in general, we cannot exclude the possibility that our proposed mechanism is not applicable to other MMPs.

Lastly, we reiterate that several theories have been proposed to explain the mechanism of collagenolysis. A major difference between the prevailing theories is the beginning state of substrate, collagen. One theory proposes that collagen exists

as a well-folded structure and that MMPs unwind this triple-helical conformation. In the other, the substrate adopts partially unfolded states that can be recognized by the enzyme. While there has been considerable debate regarding which theory is correct, our data suggest that the different theories are, in principle, not as contradictory as originally thought. Thermal fluctuations at the cleavage site in the absence of the enzyme are different than thermal fluctuations when the enzyme is near the cleavage site, and it is the combination of both types of motion that lead to successful collagenolysis.

The question of whether energy is required for collagenolysis is based on the assumption that partially unfolded states of collagen do not occur in the fibrillar state or in solution. Indeed, published estimates of the activation barrier for collagenolysis are computed with this assumption.[20,23,42] However, this large energy barrier does not exist if the region near the cleavage site can sample partially unfolded states. Our major point is that thermal fluctuations serve as the driving force for a degradative process that requires both an unfolded cleavage site and the presence of the enzyme.

Methods

Initial model construction

We used C^{α} coordinates from a X-ray fiber diffraction study of native rat tail collagen type-I (PDB 3HQV)[31,33] as the starting point for constructing our model. The initial model was constructed using the SABBAC tool based on these C^{α} coordinates.[34] The inter-chain hydrogen bonds were constrained, and the structure was minimized for 100 steps of steepest descent to relieve bad contacts using CHARMM[43] version 36b2 with the CHARMM19 extended-atom force field. The EEF1 implicit solvent model was employed for these minimizations.[44] We then took this model of triple-helical collagen and constructed a microfibril subunit made up of five triple-helical molecules, using the spatial information from Orgel *et al.*'s fiber diffraction study.[31,32] To simulate a section of the fibril surface, we arranged three microfibrils in the hexagonal packing pattern seen in collagen fibrils.[30] The constructed fibril surface was then minimized with another 100 steps of steepest descent in EEF1, to relieve inter-microfibrillar bad contacts. This is the initial structure that was used for all molecular dynamics simulations.

Molecular dynamics simulations and umbrella sampling

For the "T_E" and "I" sites, we took cross-sections of the initial fibril surface model for a stochastic boundary approach.[45] For each cross-section, 2937 residues comprising the fibrillar surface were included in the simulation, but only a subset of residues underwent full molecular dynamics (i.e., the reaction region). Regions of the system that lay outside the reaction region were constrained to their initial coordinates using average *B*-factors, as per a stochastic boundary approach.[45] For the "T_E" site, we used a cross-section of the entire fibrillar surface model that surrounded the scissile bond for a triple-helical collagen close to solvent. This cross-section model included 12 triple-helical collagen molecules closest to the central triple-helical molecule of interest. For the triple helix that includes the scissile bond, the cross-section is composed of residues 751–832 on the first α1 chain, residues 745–825 on the α2 chain, and residues 753–834 on the second α1 chain. For the other 12 triple helices, we included the same number of residues, in the same local region as the scissile bond triple helix. For the "I" site, we used the same procedure as for the "T_E" site, with a central triple helix that is composed of residues 690–769 on the first α1 chain, residues 683–763 on the α2 chain, and residues 691–771 on the second α1 chain. The same number of triple-helical collagen molecules was used for both "T_E" and "I" sites. These two regions were solvated with 35 Å×40 Å×28 Å water boxes of TIP3P molecules that include 4552 and 4058 water molecules for sites "T_E" and "I", respectively. The region of the simulation that was free to move, or the reaction region, was defined as solvent-accessible residues inside the water box (Fig. S1). Water molecules were kept within the reaction region using the Miscellaneous Mean Field Potential CHARMM command with six planar restraints for each face of a rectangular box. The solvated system was heated to 310 K and equilibrated for 2 ns at the same temperature prior to any production dynamics.

The pmf was calculated for unfolding the α2 chain of both "I" and "T_E" sites using umbrella sampling. The umbrella sampling method we used was similar to that of our previous work.[10] The reaction coordinate we used was the average N–O distance between backbone atoms of the α2 chain and the two α1 chains. Simulation windows were centered on 2.64–6.5 Å in 0.14 Å intervals, using a harmonic biasing potential with a force constant of 196 kcal/mol/$Å^2$ to bias the simulation to the center of each window. Each window was simulated for 2 ns, with the first 1.4 ns of each window as equilibration. The pmf for each site was calculated using the weighted histogram analysis method,[46] and the resulting pmf was smoothed using a moving average window of five adjacent points.

Docking simulations

We attempted to dock 3000 structures from the lowest-energy window of site "T_E" into a structure of MMP-8 (PDB 2OY4; resolution, 1.7 Å).[5,21,33] We used the six hydrogen bonds between the cleaved hexapeptide seen in the active site and the backbone atoms of MMP-8 as a guide for docking. Each of the 3000 structures was docked into the active site of MMP-8 to try and recreate the hydrogen-bonding pattern seen in the crystal structure. Rigid-body docking of the fibrillar structures and MMP-8 was initiated by constraining the arrangement of atoms within each molecule using the CONS HARM BESTFIT CHARMM command, with a force constant of 1000 kcal/mol/$Å^2$. This ensures that the fibrillar and MMP-8 structures are constrained to their initial conformations while being able to translate and rotate in space. We then started the docking simulation by putting the fibrillar and MMP-8 structures at a distance of 30 Å from each other and used a harmonic biasing potential with a force constant of

36 kcal/mol/Å^2 to recreate the crystal structure's six hydrogen bonds between the backbone atoms of the residues surrounding the scissile bond and MMP-8. The simulation was performed for 40 ps at 310 K. The simulation time was chosen because, for most of the 3000 structures, the distance between the fibril and MMP-8 was seen to stabilize at around 10 ps.

For the subsequent pmf calculations (with the catalytic domain present), we picked one of the 3000 structures that had both a low Zn^{2+} to scissile bond distance and a low deviation from the hydrogen-bonding pattern seen in PDB 2OY4. We then constructed a pmf of the complex, using a reaction coordinate of the hydrogen bonds between the backbone atoms of MMP-8 and the atoms on the fibril that correspond to that of the cleaved hexapeptide. An additional constraint between Leu785 and the hydrophobic site S1′ was added to the reaction coordinate to ensure that Leu785 samples the S1′ site. Umbrella sampling was performed with the same protocol for the previous section, with 2 ns simulation windows centered on 2.57–8.83 Å in 0.14 Å intervals, and the first 1.4 ns of each window used as equilibration. The pmf for the docked complex was again calculated using weighted histogram analysis method and smoothed with a moving average window consisting of five adjacent points.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jmb.2013.02.002

Received 17 November 2012;
Received in revised form 1 February 2013;
Accepted 4 February 2013
Available online 8 February 2013

Keywords:
collagen;
extracellular matrix;
collagen fibrils;
molecular simulations

Abbreviation used:
PDB; Protein Data Bank.

References

1. Adiguzel, E., Ahmad, P. J., Franco, C. & Bendeck, M. P. (2009). Collagens in the progression and complications of atherosclerosis. *Vasc. Med.* **14**, 73–89.
2. Cretu, A. & Brooks, P. C. (2007). Impact of the non-cellular tumor microenvironment on metastasis: potential therapeutic and imaging opportunities. *J. Cell. Physiol.* **213**, 391–402.
3. Cremer, M. A., Rosloniec, E. F. & Kang, A. H. (1998). The cartilage collagens: a review of their structure, organization, and role in the pathogenesis of experimental arthritis in animals and in human rheumatic disease. *J. Mol. Med. (Berlin, Ger.)*, **76**, 275–288.
4. Gelse, K., Poschl, E. & Aigner, T. (2003). Collagens—structure, function, and biosynthesis. *Adv. Drug Delivery Rev.* **55**, 1531–1546.
5. Shoulders, M. D. & Raines, R. T. (2009). Collagen structure and stability. *Annu. Rev. Biochem.* **78**, 929–958.
6. Soderhall, C., Marenholz, I., Kerscher, T., Ruschendorf, F., Esparza-Gordillo, J., Worm, M. *et al.* (2007). Variants in a novel epidermal collagen gene (COL29A1) are associated with atopic dermatitis. *PLoS Biol.* **5**, e242.
7. Evanson, J. M., Jeffrey, J. J. & Krane, S. M. (1967). Human collagenase: identification and characterization of an enzyme from rheumatoid synovium in culture. *Science*, **158**, 499–502.
8. Rosenblum, G., Van den Steen, P. E., Cohen, S. R., Bitler, A., Brand, D. D., Opdenakker, G. & Sagi, I. (2010). Direct visualization of protease action on collagen triple helical structure. *PLoS One*, **5**, e11043.
9. 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.
10. Nerenberg, P. S. & Stultz, C. M. (2008). Differential unfolding of α1 and α2 chains in type I collagen and collagenolysis. *J. Mol. Biol.* **382**, 246–256.
11. Sodek, J. & Overall, C. M. (1992). Matrix metalloproteinases in periodontal tissue remodelling. *Matrix, Suppl.* **1**, 352–362.
12. 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.
13. Kramer, R. Z., Bella, J., Mayville, P., Brodsky, B. & Berman, H. M. (1999). Sequence dependent conformational variations of collagen triple-helical structure. *Nat. Struct. Biol.* **6**, 454–457.
14. Fields, G. B. (1991). A model for interstitial collagen catabolism by mammalian collagenases. *J. Theor. Biol.* **153**, 585–602.
15. 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.
16. Stultz, C. M. (2002). Localized unfolding of collagen explains collagenase cleavage near imino-poor sites. *J. Mol. Biol.* **319**, 997–1003.
17. Overall, C. M. (2002). Molecular determinants of metalloproteinase substrate specificity: matrix metalloproteinase substrate binding domains, modules, and exosites. *Mol. Biotechnol.* **22**, 51–86.
18. Tam, E. M., Wu, Y. I., Butler, G. S., Stack, M. S. & Overall, C. M. (2002). Collagen binding properties of the membrane type-1 matrix metalloproteinase (MT1-MMP) hemopexin C domain. The ectodomain of the 44-kDa autocatalytic product of MT1-MMP inhibits cell invasion by disrupting native type I collagen cleavage. *J. Biol. Chem.* **277**, 39005–39014.
19. Pelman, G. R., Morrison, C. J. & Overall, C. M. (2005). Pivotal molecular determinants of peptidic and collagen triple helicase activities reside in the S_3′ subsite of matrix metalloproteinase 8 (MMP-8). The role of hydrogen bonding potential of Asn^{188} and Tyr^{189} and the connecting *cis* bond. *J. Biol. Chem.* **280**, 2370–2377.
20. 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*, **70**, 1154–1161.

21. Bertini, I., Calderone, V., Fragai, M., Luchinat, C., Maletta, M. & Yeo, K. J. (2006). Snapshots of the reaction mechanism of matrix metalloproteinases. *Angew. Chem., Int. Ed. Engl.* **45**, 7952–7955.
22. Bertini, I., Fragai, M., Luchinat, C., Melikian, M., Toccafondi, M., Lauer, J. L. & Fields, G. B. (2012). Structural basis for matrix metalloproteinase 1-catalyzed collagenolysis. *J. Am. Chem. Soc.* **134**, 2100–2110.
23. Han, S., Makareeva, E., Kuznetsova, N. V., DeRidder, A. M., Sutter, M. B., Losert, W. *et al.* (2010). Molecular mechanism of type I collagen homotrimer resistance to mammalian collagenases. *J. Biol. Chem.* **285**, 22276–22281.
24. Salsas-Escat, R., Nerenberg, P. S. & Stultz, C. M. (2010). Cleavage site specificity and conformational selection in type I collagen degradation. *Biochemistry*, **49**, 4147–4158.
25. Salsas-Escat, R. & Stultz, C. M. (2010). Conformational selection and collagenolysis in type III collagen. *Proteins*, **78**, 325–335.
26. Manka, S. W., Carafoli, F., Visse, R., Bihan, D., Raynal, N., Farndale, R. W. *et al.* (2012). Structural insights into triple-helical collagen cleavage by matrix metalloproteinase 1. *Proc. Natl Acad. Sci. USA*, **109**, 12461–12466.
27. Veis, A. & Yuan, L. (1975). Structure of the collagen microfibril. A four-strand overlap model. *Biopolymers*, **14**, 895–900.
28. Holmes, D. F. & Kadler, K. E. (2006). The 10+4 microfibril structure of thin cartilage fibrils. *Proc. Natl Acad. Sci. USA*, **103**, 17249–17254.
29. Wess, T. J., Hammersley, A., Wess, L. & Miller, A. (1995). Type I collagen packing, conformation of the triclinic unit cell. *J. Mol. Biol.* **248**, 487–493.
30. Hulmes, D. J., Wess, T. J., Prockop, D. J. & Fratzl, P. (1995). Radial packing, order, and disorder in collagen fibrils. *Biophys. J.* **68**, 1661–1670.
31. Orgel, J. P., Irving, T. C., Miller, A. & Wess, T. J. (2006). Microfibrillar structure of type I collagen in situ. *Proc. Natl Acad. Sci. USA*, **103**, 9001–9005.
32. 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.
33. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H. *et al.* (2000). The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242.
34. Maupetit, J., Gautier, R. & Tuffery, P. (2006). SABBAC: online Structural Alphabet-based protein BackBone reconstruction from Alpha-Carbon trace. *Nucleic Acids Res.* **34**, W147–W151.
35. Hulmes, D. J. (2002). Building collagen molecules, fibrils, and suprafibrillar structures. *J. Struct. Biol.* **137**, 2–10.
36. Brooks, C. L., Karplus, M. & Pettitt, B. M. (1988). Advances in Chemical Physics: Proteins: A Theoretical Perspective of Dynamics, Structure, and Thermodynamics J. Wiley, New York, NY.
37. Murphy, G. & Knauper, V. (1997). Relating matrix metalloproteinase structure to function: why the "hemopexin" domain? *Matrix Biol.* **15**, 511–518.
38. Lovejoy, B., Cleasby, A., Hassell, A. M., Longley, K., Luther, M. A., Weigl, D. *et al.* (1994). Structure of the catalytic domain of fibroblast collagenase complexed with an inhibitor. *Science*, **263**, 375–377.
39. Sarkar, S. K., Marmer, B., Goldberg, G. & Neuman, K. C. (2012). Single-molecule tracking of collagenase on native type I collagen fibrils reveals degradation mechanism. *Curr. Biol.* **22**, 1047–1056.
40. Lee, G. & Ha, T. (2012). Single-molecule imaging: a collagenase pauses before embarking on a killing spree. *Curr. Biol.* **22**, R499–R501.
41. Dominguez, C., Boelens, R. & Bonvin, A. M. (2003). HADDOCK: a protein–protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* **125**, 1731–1737.
42. Welgus, H. G., Jeffrey, J. J. & Eisen, A. Z. (1981). Human skin fibroblast collagenase. Assessment of activation energy and deuterium isotope effect with collagenous substrates.. *J. Biol. Chem.* **256**, 9516–9521.
43. Brooks, B. R., Brooks, C. L., 3rd, Mackerell, A. D., Jr, Nilsson, L., Petrella, R. J., Roux, B. *et al.* (2009). CHARMM: the biomolecular simulation program. *J. Comput. Chem.* **30**, 1545–1614.
44. Lazaridis, T. & Karplus, M. (1999). Effective energy function for proteins in solution. *Proteins: Struct., Funct., Genet.* **35**, 133–152.
45. Brooks, C. L., 3rd, Brunger, A. & Karplus, M. (1985). Active site dynamics in protein molecules: a stochastic boundary molecular dynamics approach. *Biopolymers*, **24**, 843–865.
46. 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.