

# A Structural Model that Explains the Effects of Hyperglycemia on Collagenolysis

Collin M. Stultz and Elazer R. Edelman

Harvard-MIT Division of Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139; and Department of Medicine, Division of Cardiovascular Medicine, Brigham and Women's Hospital, and Harvard Medical School, Boston, Massachusetts 02115

**ABSTRACT** Prior investigations into the effects hyperglycemia on collagen degradation have yielded conflicting results. We present a new formalism for understanding the biochemistry of collagenolysis and the effects of hyperglycemia on collagen degradation. The analysis is based on an understanding of environments that affect the conformational stability of collagen. We suggest that collagen can exist in two distinct conformational states—a native state and a vulnerable state. Vulnerable collagen corresponds to a non-native conformation where partially unfolded regions near collagenase cleavage sites enable collagenases to efficiently degrade collagen. Theoretical calculations on collagen-like model peptides suggest that relatively short periods of hyperglycemia can alter the equilibrium distribution of states to favor vulnerable states of collagen. These data provide new insights into the mechanism of collagenolysis and resolve apparently discrepant experimental data on the effects of hyperglycemia on collagen degradation.

## INTRODUCTION

Collagenolysis plays an important role in a number of diseases such as tumor metastasis, arthritis, and atherosclerotic heart disease (Celentano et al., 1997; McDonnell et al., 1999). In particular, rupture of mature atherosclerotic plaques, a process that depends on degradation of the protective collagen layer surrounding the plaque, is the major cause of catastrophic cardiovascular events in the developed world (Gaziano, 2001). Therefore, methods that lead to a deeper understanding of collagenolysis may result in effective therapies for unstable arterial syndromes such as myocardial infarctions.

X-ray crystallographic structures of collagenlike triple-helical peptides suggest that the scissile bond that forms the collagenase cleavage site is concealed from solvent (Kramer et al., 1999; Stultz, 2002). It has been suggested that regions near collagenase cleavage sites must become exposed prior to collagenolysis for collagenases to gain access to their cleavage sites (Fields, 1991). Circular dichroism studies and hydrogen exchange experiments suggest that regions in the vicinity of collagenase cleavage sites are relatively unstable and can adopt alternate conformations in solution (Fan et al., 1993; Long et al., 1993). These data are consistent with the notion that protein flexibility plays a significant role in the process of collagen degradation.

All collagen sequences contain a significant proportion of the imino acids proline and hydroxyproline (Stryer, 1988). However, an analysis of collagenase cleavage sites from a variety of different interstitial collagens reveals that the collagenase scissile bond is located immediately upstream

from a region that is essentially devoid of any imino acids (Fields, 1991). We previously computed the free energy profile for the unfolding of a collagen-like triple-helical peptide, T3-785, that contains a representative imino-poor segment from type 3 collagen (Stultz, 2002) (Fig. 1 *a*). These data suggest that imino-poor segments from collagen can exist in two well-defined states with similar energies—a state **N** that corresponds to the crystallographically observed conformation and a state **V** that corresponds to an alternate structure. In state **V** the imino-poor region that is immediately downstream from the collagenase cleavage site adopts a metastable partially unfolded state that is solvent exposed (Fig. 1 *b*). These data support a model of collagenase cleavage where local unfolding of collagen begins at these imino-poor regions and continues until the collagenase cleavage site is exposed to collagenases (Fields, 1991). We refer to collagen conformations with partially unfolded imino-poor regions as “vulnerable,” since these conformations enable collagenases to gain access to their cleavage sites (Fig. 2).

A number of factors will likely affect the equilibrium distribution of native and vulnerable states. Since vulnerable collagen is, in principle, degraded more efficiently than native collagen, factors that lead to the stabilization of vulnerable conformations are expected to promote collagen degradation. In particular, diabetes, which is characterized by hyperglycemia, is a major risk factor for acute cardiovascular events. Therefore, the effect of hyperglycemia on the distribution of native and vulnerable states is of particular interest.

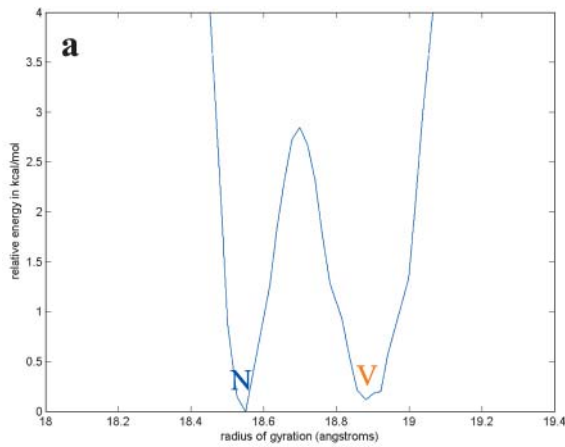
Previous studies on the effects of hyperglycemia on collagen degradation have yielded inconsistent results. Prolonged incubation of collagen with high glucose levels results in collagen fibers that are resistant to cleavage (Paul and Bailey, 1996). By contrast, collagen isolated from the Achilles tendons of rats exposed to a hyperglycemic state for

Submitted May 1, 2003, and accepted for publication June 27, 2003.

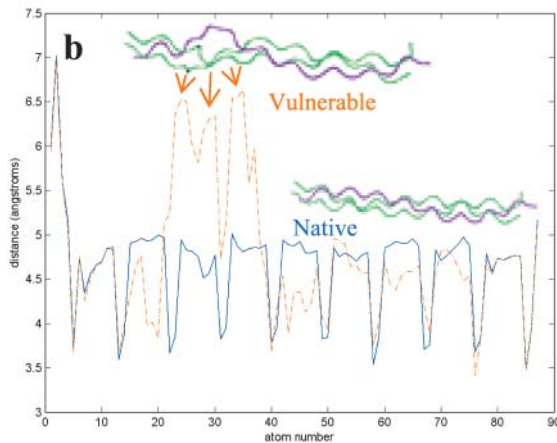
Address reprint requests to Collin Stultz, Bldg. 16-343, MIT, 77 Massachusetts Ave., Cambridge, MA 02139. Tel.: 617-694-1692; E-mail: cmstultz@partners.org.

© 2003 by the Biophysical Society

0006-3495/03/10/2198/07 \$2.00

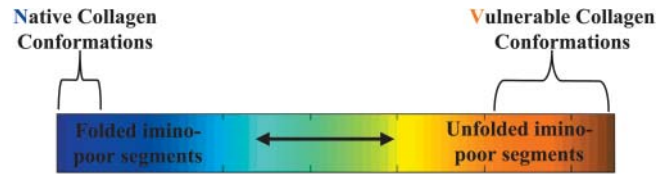


Backbone Separation as a Function of Atom Number



**FIGURE 1** (a) Free energy profile for the unfolding of a single chain in the collagenlike peptide T3-785 (a triple-helical peptide composed of three identical peptide chains labeled A, B, and C). The  $x$  axis denotes the radius of gyration and the free energy, in kcal/mol, is located on the  $y$  axis. The radius of gyration serves as the reaction coordinate for chain unfolding. State **N** denotes the native state and state **V** denotes an alternate state where the proximal portion of the imino-poor region of collagen is solvent exposed. (b) Chain separation as a function of backbone atom number. For each residue in chain A, the distance between each backbone atom (N,  $C\alpha$ , C) and the nearest backbone atom in the other two strands was computed and plotted on the  $y$  axis. Chain A has 29 residues (the first residue was not found in the electron density), therefore the number of backbone atoms go from 1 to 87. All plots made with MATLAB (Mathworks). Representative structures from states **N** and **V** are shown. Molecular figures made with QUANTA (Molecular Simulations).

a relatively brief period (28 days) is more susceptible to cleavage *in vitro* relative to collagen isolated from rats with normal blood glucose levels (Leung et al., 1986). In addition, collagen isolated from chick calvarias that have been exposed to high concentrations of extracellular glucose for 24 h is more readily degraded relative to collagen isolated from media free of glucose (Lien et al., 1992). More recently, it has been suggested that hyperglycemia is an independent predictor of mortality in patients who are hospitalized with



**FIGURE 2** Definition of native and vulnerable collagen conformations. The color scale represents the spectrum of possible collagen conformations. Red represents conformations where imino-poor segments, adjacent to collagenase cleavage sites, are unfolded, and blue denotes native collagen conformations where imino-poor segments are in their folded conformations. Vulnerable collagen corresponds to conformations where the imino-poor sites are partially unfolded. Since the number of unfolded states is in general larger than the number of folded conformations, the number of vulnerable conformations is significantly larger than the number of native conformations.

a cardiovascular event (Wahab et al., 2002), even in patients without a known diagnosis of diabetes. This is consistent with the notion that the protective collagen layer surrounding the atherosclerotic plaque is more readily degraded in the presence of high glucose levels. These data suggest that perhaps paradoxically, unlike prolonged periods of glucose exposure, relatively brief periods of hyperglycemia may make collagen more susceptible to cleavage.

Hyperglycemia affects collagen primarily through non-enzymatic glycosylation (glycation) (Paul and Bailey, 1996). There are two types of advanced glycation end products (AGEs)—those that form intermolecular cross-links and those that remain non-cross-linked. The effects of cross-linking AGEs on collagen have been well described (Paul and Bailey, 1996). Cross-links that form between collagen molecules hinder collagen degradation primarily by decreasing their elasticity and conformational flexibility. Typically, cross-linking AGEs are formed on the timescale of weeks to months, therefore only molecules that are exposed to high glucose levels for an extended period of time are affected by this mechanism of glycation (Aronson and Rayfield, 2002). By contrast, non-cross-linking AGEs form within days (Aronson and Rayfield, 2002). Therefore we hypothesized that the discrepancy between the reported effects of hyperglycemia on collagen can be explained by the fact that non-cross-linking AGEs alter the equilibrium distribution of states to make collagen more vulnerable to collagenolysis.

In this work, we investigate the effects of non-cross-linking AGEs that form in the vicinity of collagenase cleavage sites with the aid of detailed molecular dynamics simulations. Our results support a model of collagenolysis where acute exposure to elevated environmental glucose alters the equilibrium distribution of conformational states to favor vulnerable conformations. In addition, these data demonstrate how the presence of native and vulnerable collagen conformations provide additional insights into the mechanism of collagenolysis.

## METHODS

### Construction of the initial model used in the molecular dynamics simulations

Our calculations utilized the collagen-like peptide T3-785, a triple helical peptide composed of three identical peptide chains labeled A, B, and C (Kramer et al., 1999). Coordinates for the T3-785 peptide, were taken from PDB file 1bvk (Berman et al., 2000). A polar hydrogen model of the peptide was constructed with the CHARMM program (Brooks et al., 1983). All water molecules were deleted from the PDB coordinate file and a polar hydrogen model was constructed using CHARMM. The resulting model was energy minimized for 100 steps of steepest descent minimization with a distance dependent dielectric to relieve any poor contacts within the molecule.

### Determining the preferred conformation of glycosylated collagen

Glycation preferentially occurs at sites containing a terminal amino group; e.g., lysine and arginine residues (Paul and Bailey, 1996). Interestingly, all known collagenase cleavage sites are located upstream from an imino-poor region that always contains an arginine residue (Fields, 1991). Therefore we focused on the effects of glycation on the central arginine located in one of the T3-785 peptide chains.

Arginine can be glycosylated at two distinct sites (Fig. 3 a). However, since one site is relatively hidden from solvent in the T3-785 structure, we focused on glycation at the other site, nitrogen N<sub>1</sub> (Fig. 3 b). The resulting advanced glycation end product, carboxymethyl arginine (CMA), can exist in two distinct conformations; a *cis* and a *trans* structure (Fig. 3 c). To determine the preferred conformation of CMA, we calculated the free energy difference between the *cis* and *trans* conformations (*path 3*, Fig. 3 c) using thermodynamic integration with a linear alchemical path. These simulations employed a dual topology setup. Models for CMA<sup>cis</sup> and CMA<sup>trans</sup> were constructed in CHARMM; i.e., CMA<sup>cis</sup> or CMA<sup>trans</sup> was constructed by replacing the HH11 or HH12 hydrogen atom on NH1 with a carboxymethyl group, respectively. Parameters for CMA were derived from the CHARMM19 parameters for arginine and the methylcarboxyl moiety from a glutamate residue. The replica facility of CHARMM was used to construct a protein structure file where the arginine residue is replaced with two copies of CMA; i.e., one copy corresponds to CMA<sup>cis</sup> and the other to CMA<sup>trans</sup>. The resulting T3-785 structure with replicated CMA residues was overlaid with a set of 1000 equilibrated TIP3P water molecules (Jorgenson et al., 1983). Water molecules that overlapped with the protein structure were removed. The remaining water molecules were equilibrated in the field of the fixed protein structure with 100 steps of steepest descent minimization followed by 5 ps of standard molecular dynamics at 300 K. Molecular dynamics simulations employed a nonbonded cutoff of 13 Å and the van der Waals energy was switched to 0 between 10 Å and 12 Å and the electrostatic energy was shifted to 0 at a distance of 12 Å. A total of three solvent overlays were performed and a total of 1044 water molecules were added to the structure. These were the only simulations that employed a fixed protein structure. All simulations employed a stochastic boundary setup as previously described (Stultz, 2002).

The potential energy used for the alchemical transformation is given by  $V_{\lambda} = \lambda V_{\text{trans}} + (1 - \lambda) V_{\text{cis}}$  where  $V_{\text{trans}}$  or  $V_{\text{cis}}$  denote the potential energy of T3-785 with arginine 14 replaced with CMA<sup>trans</sup> or CMA<sup>cis</sup>, respectively. The free energy difference between the *cis* and *trans* states is given by  $\Delta G_{\text{cis} \rightarrow \text{trans}} = \int_0^1 \langle \Delta V \rangle_{\lambda} d\lambda$  where  $\Delta V = V_{\text{trans}} - V_{\text{cis}}$ . Simulations were performed at  $\lambda = 0.02, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$ , and 0.98. At each  $\lambda$  value, 20 ps of equilibration followed by 20 ps of production dynamics were performed. Simulations were not performed at the end points,  $\lambda = 0$  and  $\lambda = 1$  to avoid the “van der Waals endpoint problem” (Simonson, 1993). The integral of  $\langle \Delta V \rangle_{\lambda}$  was obtained using the calculated values at each  $\lambda$  and the trapezoidal rule. The electrostatic contributions and

van der Waals contributions were calculated separately. The value of the integral for the van der Waals component in the interval  $\lambda = 0 \rightarrow 0.1$  was obtained by fitting the values  $\langle \Delta V_{\text{vdw}} \rangle_{\lambda}$  at  $\lambda = 0.02$  and  $\lambda = 0.1$  to a function of the form (Simonson, 1993)  $A\lambda^{-1/4} + B$ . Similarly, the van der Waals contribution in the interval  $\lambda = 0.9 \rightarrow 1.0$  was obtained by fitting values of  $\langle \Delta V_{\text{vdw}} \rangle_{\lambda}$  at  $\lambda = 0.98$  and  $\lambda = 0.9$  to a function of the form  $A(1 - \lambda)^{-1/4} + B$ . Each function was then integrated to yield the contributions from 0 to 0.1 and 0.9 to 1.0. The electrostatic component in the regions  $\lambda = 0 \rightarrow 0.1$  and  $\lambda = 0.9 \rightarrow 1.0$  were obtained by fitting the values of  $\langle \Delta V_{\text{coul}} \rangle_{\lambda}$  at  $\lambda = 0.02 \rightarrow 0.1$  and  $\lambda = 0.9 \rightarrow 0.98$  to straight lines, respectively. The values at  $\lambda = 0$  and  $\lambda = 1$  were set equal to the corresponding y-intercepts and the value of the integral from  $\lambda = 0$  to 1 was obtained using the trapezoidal rule as mentioned above.

Initial simulations revealed that scaling the internal bond, bond angles and dihedral angles of the CMA replicas resulted in large structural distortions and rotations about the N-CZ bond (Fig. 3 c). This occurs primarily near the endpoints when a given replica is assigned a low  $\lambda$  value. Therefore, the internal bond, bond angle, and dihedral angles of the CMA replicas (i.e., CMA<sup>cis</sup> and CMA<sup>trans</sup>) were not scaled during these simulations. Only the interactions between the two CMA replicas and the surrounding water molecules and the protein were scaled. This corresponds to a free energy simulation where the endpoints are gas-phase molecules. To obtain the correct free energy difference, additional simulations were performed to compute the free energy needed to transform each CMA replica from a gas-phase molecule to a collection of gas-phase atoms. The final free energy reported in this work includes this additional correction factor. All bond, bond angle, electrostatic, and van der Waals components were obtained by numerical integration as previously described (Archontis et al., 1998).

### Calculating the potential of mean force for glycosylated collagen

The free energy simulations suggest that CMA<sup>trans</sup> is the dominant form of glycosylated arginine in T3-785. Therefore, for these calculations we focused on the structure of T3-785 with the central arginine replaced with CMA<sup>trans</sup>. We refer to the glycosylated form of collagen as T3-785\*. The reaction region consisted of a sphere of radius 30 Å centered at the center of mass and included all residues between the 5th and the 28th residues in each chain. As in our previous work, the free energy of unfolding a single chain was computed; i.e., simulations with all three chains unfolding simultaneously led to high barriers suggesting that this was an unlikely mechanism for collagen unfolding (Stultz, 2002). The reaction region for the unfolding reaction was the radius of gyration of the C $\alpha$  atoms in chain A; i.e., the chain containing CMA<sup>trans</sup>. Windows from 17.7 Å to 20.0 Å in increments of 0.1 Å were used; i.e., a total of 24 windows. Simulations for the T3-785\* peptide began with the 18.5 Å window (the value corresponding to the x-ray crystal structure) and proceeded in both the forward and reverse directions. The starting structure for each window was the final structure from the prior window. Each simulation consisted of 20 ps of equilibration followed by 20 ps of production dynamics. All atoms within the reaction region, including all three peptide chains of T3-785\*, underwent full molecular dynamics without restraints and all residues outside the reaction region underwent molecular dynamics while being harmonically constrained to their crystallographic coordinates. Harmonic constraints were derived from average B-factors as previously described (Brooks and Karplus, 1989). At each window the reaction coordinate was restrained to the specified center point using a harmonic biasing potential with a force constant of 500 kcal/mol/Å<sup>2</sup>. The value of the reaction coordinate during the simulations was saved every 0.01 ps; this yielded 2000 data points per window. The potential of mean force at window  $i$ ,  $W_i(\xi)$ , was calculated from the resulting frequency distribution,  $p_i$ , and the biasing potential,  $V_i(\xi)$ , using the equation  $W_i(\xi) = -kT \ln p_i(\xi) - V_i(\xi) + C_i$  where  $C_i$  is a constant that is a function of the temperature,  $T$ , and the biasing potential,  $V_i(\xi)$ . The results of the simulations yielded 24 separate potentials of mean force; i.e., one for

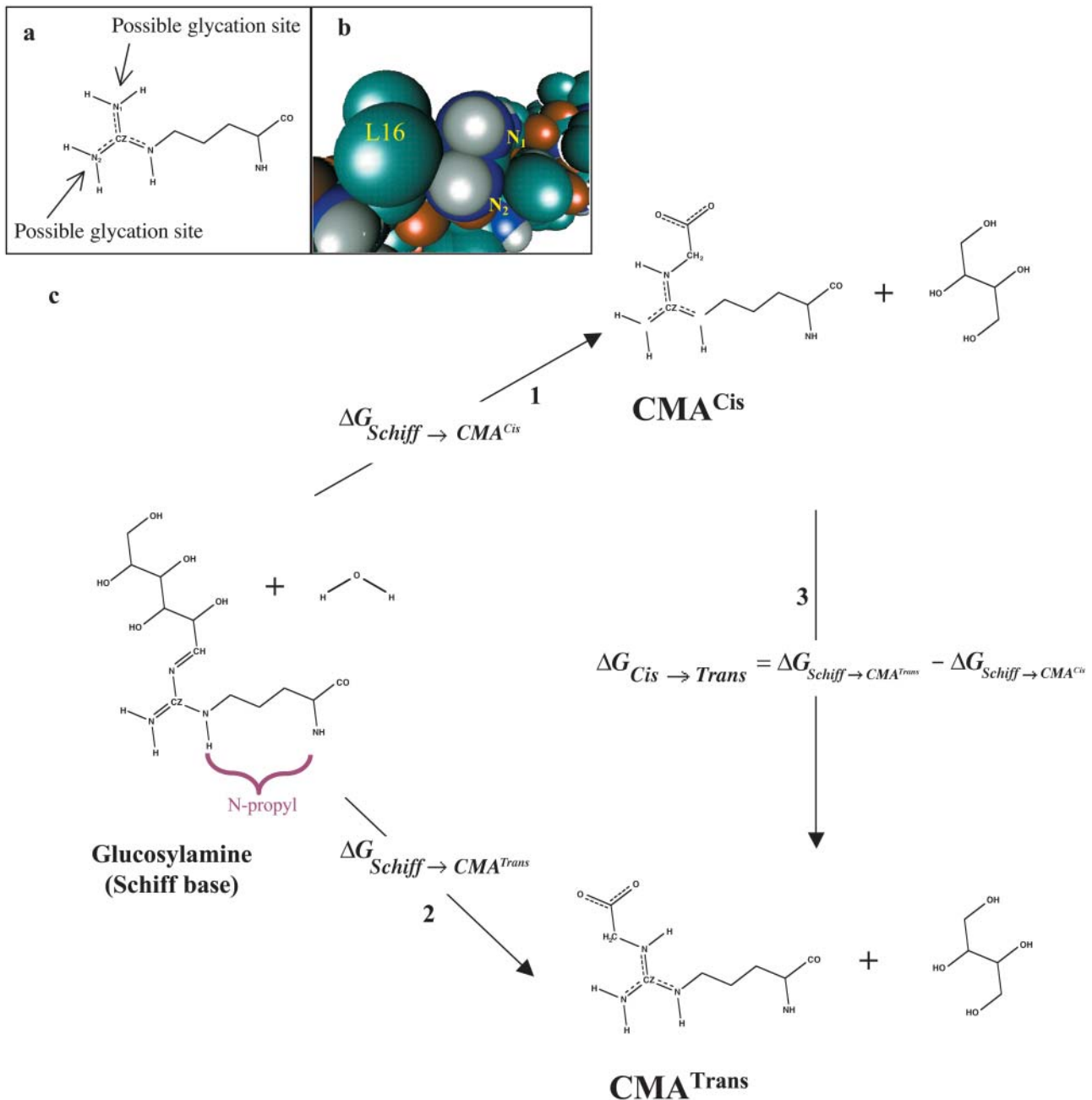


FIGURE 3 (a) Schematic of an arginine residue showing the two possible glycation sites. (b) Van der Waals surface in the vicinity of arginine 14 in chain A of T3-785. The two guanidino nitrogens are denoted by  $N_1$  and  $N_2$ . (c) Thermodynamic cycle used to calculate the preferred conformation of the carboxymethyl moiety in CMA. (path 1 or 2). The carboxymethyl group can exist in a position that is either *cis* or *trans* to the *N*-propyl group of the arginine side chain. The *N*-propyl moiety is labeled with a purple brace. The free energy difference along path 3 is equal to the difference between the free energies along paths 1 and 2.

each window. To create one continuous potential of mean force, the individual potentials need to be linked together. The potential of mean forces from two sequential windows were linked by first comparing the overlapping regions to find a common point where the potentials had similar slopes. The second potential was then adjusted to agree with the first at this common point. The procedure for linking potentials together to form one continuous potential of mean force has been automated in the program SPLICE (Stultz, 2002). The final potential of mean force was smoothed using the moving average window method with a window span of 3.

## RESULTS

### The structure of glycated collagen

In the current formalism, vulnerable collagen is defined as conformations that contain partially unfolded imino-poor regions that are adjacent to collagenase cleavage sites (Fig. 2). Diabetes is associated with accelerated atherosclerosis, and diabetic patients have a high incidence of cardiovascular

events that is not entirely explained by any one specific mechanism. Therefore, the effect of hyperglycemia on the conformational free energy landscape of collagen is of particular interest and may well clarify the propensity of atherosclerotic lesions in diabetics for rupture.

Recently, a new non-cross-linking AGE, carboxymethyl-arginine, has been detected in large quantities within serum proteins of diabetic patients (Iijima et al., 2000; Odani et al., 2001). Interestingly, imino-poor regions of collagen that are adjacent to collagenase cleavage sites all contain a conserved arginine residue (Fields, 1991) and previous simulations on the accessible conformational states of a collagenlike model peptide suggest that this arginine residue plays a major role in stabilizing vulnerable conformations of collagen (Stultz, 2002). Therefore, the effect of arginine glycation on the conformational free energy landscape of collagen is of particular interest. To investigate the effect of CMA on the structure of collagen, we computed the free energy profile for the unfolding of a collagenlike triple-helical peptide, T3-785, that has this central arginine residue replaced with CMA (Kramer et al., 1999, 2001).

We note that CMA is formed when one of the terminal guanidino nitrogens in the side chain of arginine is glycated (Fig. 3 *a*). Ordinarily, in solution both guanidino nitrogens are chemically indistinguishable and both can be glucosylated. However, in the structure of T3-785, one guanidino nitrogen interacts with atoms within the other peptide chains. In particular, guanidino nitrogen N<sub>2</sub>, shown in Fig. 3 *b*, is relatively shielded from solvent by the side chain of leucine 16 in chain B; i.e., the solvent accessible surface of atom N<sub>2</sub> is 30% less than that of nitrogen N<sub>1</sub>. Since glycation is a nonspecific, nonenzymatic process, it preferentially occurs at those positions that are the most solvent exposed because these sites make the most contact with glucose molecules in the surrounding solvent. Therefore our calculations focused on arginine glycation at nitrogen N<sub>1</sub> (Fig. 3 *a*).

Glycation of arginine yields a glucosylamine (a Schiff base), which then undergoes an Amadori rearrangement to form an intermediate that is then subjected to nucleophilic attack from a water molecule to form CMA (Aronson and Rayfield, 2002) (Fig. 3 *c*). In principle, this glycation reaction can yield two conformationally distinct species. The initial glucosylamine can either form a glycation end product where the carboxymethyl moiety is *cis* (*path 1*, Fig. 3 *c*) or *trans* (*path 2*, Fig. 3 *c*) to the *N*-propyl group. The two conformations, CMA<sup>cis</sup> and CMA<sup>trans</sup>, are not interchangeable because the N-CZ bond in the guanidino moiety has partial double character and, therefore, rotations about this bond are energetically unfavorable (Fig. 3 *c*). To determine the preferred conformation of the carboxymethyl moiety, we used the thermodynamic cycle shown in Fig. 3 *c*. The free energy difference between the *cis* and *trans* conformations (*path 3*,  $\Delta G_{\text{cis} \rightarrow \text{trans}}$ ) is equivalent to the difference between the free energy for the *cis* reaction (*path 1*, Fig. 3) and the *trans* reaction (*path 2*, Fig. 3 *c*). Therefore, the sign of

$\Delta G_{\text{cis} \rightarrow \text{trans}}$  determines the preferred conformation of CMA. The free energy difference along *path 3* was computed using thermodynamic integration as described in the methods. The overall free energy change was calculated to be approximately  $-39$  kcal/mol suggesting that the glucosylamine to CMA<sup>trans</sup> reaction (*path 2*, Fig. 3 *c*) is overwhelmingly favored and hence CMA<sup>trans</sup> is the preferred conformation.

A free energy component analysis suggests that, for the given simulation pathway, a significant portion of this free energy difference arises from poor steric contacts between the carboxymethyl moiety and the *N*-propyl group in the CMA<sup>cis</sup> structure. The van der Waals contribution to the overall free energy change favors the CMA<sup>trans</sup> structure by approximately  $-11$  kcal/mol. Virtually all of this contribution arises from interactions between the carboxymethyl moiety and the *N*-propyl moiety (Fig. 3 *c*); i.e., in the CMA<sup>cis</sup> structure the *N*-propyl group makes unfavorable van der Waals interactions with the *N*-propyl moiety relative to the CMA<sup>trans</sup> structure. In representative structures from each simulation endpoint, ( $\lambda = 0$ ,  $\lambda = 1$ ) the carboxymethyl group undergoes a rotation away from the *N*-propyl group to avoid these unfavorable van der Waals interactions in the CMA<sup>cis</sup> structure. This results in the carboxymethyl group reorienting itself in a position that is close to the backbone of the peptide. The result is that the carboxymethyl moiety forms fewer hydrogen bonds with solvent relative to structures obtained from simulations with CMA<sup>trans</sup>. The relatively sparse interactions between CMA<sup>cis</sup> and the surrounding solvent also result in a large electrostatic contribution of approximately  $-22$  kcal/mol that favors the CMA<sup>trans</sup> structure. These contributions explain why CMA<sup>trans</sup> is the overwhelmingly favored conformation.

### The free energy landscape of glycated collagen favors vulnerable conformations

To determine how glycation affects the accessible conformational states of collagen we computed the free energy for the unfolding of the T3-785 peptide with the central arginine replaced with the dominant conformation of CMA, CMA<sup>trans</sup>. We refer to this mutated peptide as T3-785\*. Unlike the free energy profile of the original peptide, T3-785 (Fig. 1 *a*), the lowest energy state of T3-785\*, **V**, has a larger radius of gyration relative to the native state, **N** (Fig. 4 *a*); i.e., the lowest energy state is partially unfolded. In addition, the energy of state **V** is almost 1.3 kcal/mol lower than that of state **N**, a greater than 2 kT difference at room temperature. This suggests that the majority of T3-785\* molecules exist in a vulnerable state. There are three regions where chain A separates from the other two chains in the peptide and the areas of greatest “unfolding” lie near the two termini of the molecule (Fig. 4 *b*).

An analysis of representative structures from states **N** and **V** reveals the physical basis behind the different stabilities of the two conformations (Fig. 4 *c*). In state **N**, the side chain of

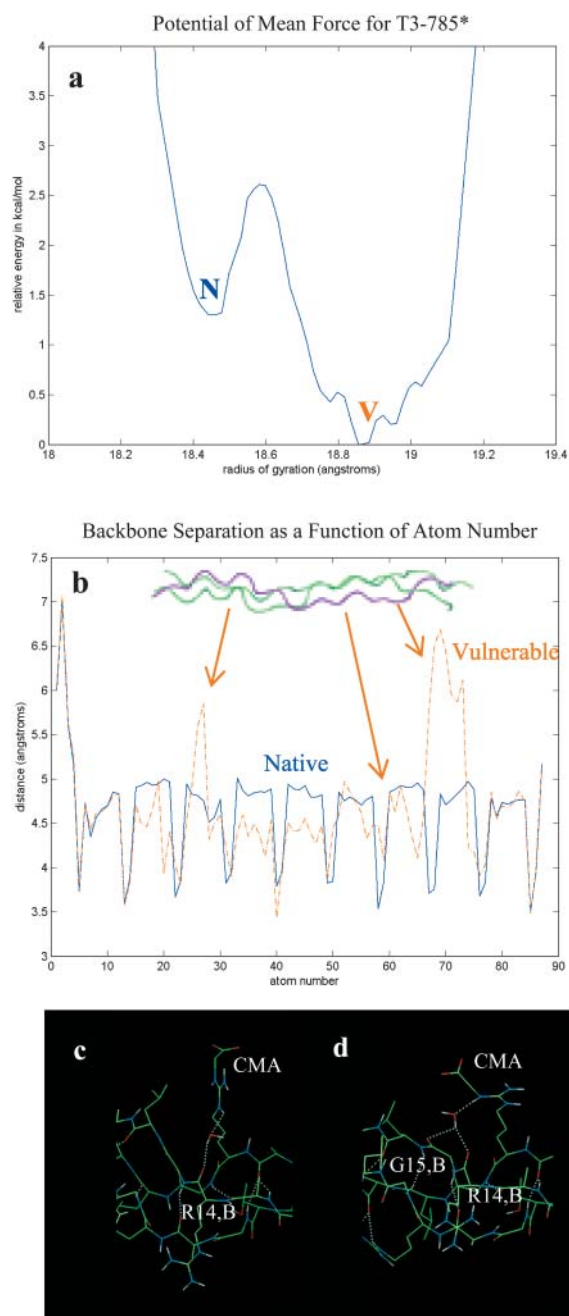


FIGURE 4 (a) Free energy profile for the unfolding of T3-785\*. (b) Chain separation as a function of distance. (c) Representative structure from state N in the vicinity of CMA. (d) Representative structure from state V in the vicinity of CMA.

CMA<sup>trans</sup> hydrogen bonds to a water molecule that then hydrogen bonds to the backbone carbonyl of arginine 14 from chain B. This bridging water molecule is present in 96% of the structures that comprise the trajectory corresponding to state N. By contrast, in state V, the same bridging water molecule is present, but it now makes two hydrogen bonds to the backbone carbonyl groups of the arginine 14 and glycine 15 in the adjacent chain (Fig. 4 d).

Moreover, this water molecule is present in 80% of the structures comprising the production dynamics run of an the trajectory corresponding to state V. That the energy of state V is  $\sim 1.3$  kcal/mol lower than state N suggests that this additional hydrogen bond contributes to the stability of state V. This is consistent with prior data that an uncharged hydrogen bond donor-acceptor pair can contribute  $\sim 1$  kcal/mol to the stability of a protein complex (Fersht et al., 1985).

## DISCUSSION

Collagen, the most abundant protein in mammals, plays an integral role in a number of disease processes (Stryer, 1988). Therefore significant effort has been directed at understanding the fundamental interactions involved in collagen degradation (Celentano and Frishman, 1997; McDonnell et al., 1999). Prior studies suggest that regions near collagenase cleavage sites are conformationally labile and that this lability is important for collagenolysis (Fields, 1991; Fan et al., 1993; Stultz, 2002). In this study, we investigated the effect of hyperglycemia on the conformational free energy landscape of collagen to understand mechanisms whereby hyperglycemia, and hence diabetes, may affect collagenolysis. The calculated free energy profiles suggest that glycation of the conserved arginine residue within imino-poor regions near collagenase cleavage sites, results in a shift in the equilibrium distribution of states to more vulnerable conformations. Our analysis resolves apparently discrepant data regarding the effects of glucose on collagenolysis. Long-term exposure to glucose leads to the formation of cross-linking AGEs that have well-established effects on collagen metabolism (Aronson and Rayfield, 2002). By contrast, relatively short term exposure to glucose results in the formation of non-cross-linking AGEs (Aronson and Rayfield, 2002), and the effects of such molecules on collagen degradation have not been well studied. Our data suggests that the effects of relatively short-term glucose exposure can be explained by conformational changes in regions near collagenase cleavage sites. These data are consistent with prior experimental studies that collagen exposed to high concentrations of glucose for relatively short periods of time is more susceptible to collagenolysis (Leung et al., 1986; Lien et al., 1992).

These observations have implications for the pathogenesis of atherosclerosis and atherosclerotic plaque rupture. Collagen is responsible for maintaining the structural integrity of the atherosclerotic plaque, hence excessive collagenolysis often leads to plaque rupture and acute myocardial infarctions (Celentano and Frishman, 1997). Diabetes, which is characterized in part by hyperglycemia, is a known risk factor for atherosclerotic heart disease (Nesto and Libby, 2001). Our data suggests that diabetes not only promotes the progression of atherosclerosis, but that hyperglycemia can also make preexistent atherosclerotic plaques prone to rupture by making the collagen within the

plaque more vulnerable to cleavage by interstitial collagenases. These data are consistent with recent observations that hyperglycemia, even in the absence of overt diabetes, is associated with a worse prognosis in patients who present with an acute cardiovascular event (Wahab et al., 2002).

Until recently, the effects of post-translational modifications on the conformational free energy landscape of collagen have been under appreciated. However, glycation, in addition to other modifications, alters the charge distribution and solvent accessible surface of proteins. Therefore, it is not surprising that such modifications can significantly alter the equilibrium distribution of conformational states in flexible regions. Since conformational changes are essential to many enzymatic processes, including enzyme-substrate recognition, the equilibrium distribution of states will affect the ability of enzymes to interact with and metabolize their substrates. As such, we expect that our formalism will be a useful means to evaluate the effects of post-translational modifications on collagen as well as on a number of different proteins.

This work was supported by grants GM049039 and HL67246 from the National Institutes of Health.

## REFERENCES

- Archontis, G., T. Simonson, D. Moras, and M. Karplus. 1998. Specific amino acid recognition by aspartyl-tRNA synthetase studied by free energy simulations. *J. Mol. Biol.* 275:823–846.
- Aronson, D., and E. Rayfield. 2002. How hyperglycemia promotes atherosclerosis: molecular mechanisms. *Cardiovasc. Diabetol.* 1:1–10.
- Berman, H. M., J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne. 2000. The Protein Data Bank. *Nucleic Acids Res.* 28:235–242.
- Brooks, B. R., R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. 1983. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* 4:187–217.
- Brooks, C., and M. Karplus. 1989. Solvent effects on protein motion and protein effects on solvent motion. Dynamics of the active site region of lysozyme. *J. Mol. Biol.* 208:159–181.
- Celentano, D. C., and W. H. Frishman. 1997. Matrix metalloproteinases and coronary artery disease: a novel therapeutic target. *J. Clin. Pharmacol.* 37:991–1000.
- Fan, P., M. Li, B. Brodsky, and J. Baum. 1993. Backbone dynamics of (Pro-Hyp-Gly)<sub>10</sub> and a designed collagen-like triple helical peptide by <sup>15</sup>N NMR relaxation and hydrogen-exchange measurements. *Biochemistry.* 32:13299–13309.
- Fersht, A. R., J. Shi, J. Knill-Jones, D. M. Lowe, A. J. Wilkinson, D. M. Blow, P. Brick, P. Carter, M. M. Y. Waye, and G. Winter. 1985. Hydrogen bonding and biological specificity analysed by protein engineering. *Nature.* 314:235–239.
- Fields, G. F. 1991. A model for interstitial collagen catabolism by mammalian collagenases. *J. Theor. Biol.* 153:585–602.
- Gaziano, J. M. 2001. Global burden of cardiovascular disease. In *Heart Disease*. E. Braunwald, D. P. Zipes, P. Libby, editors. Saunders, Philadelphia. 1–18.
- Iijima, K., M. Murata, H. Takahara, S. Irie, and D. Fujimoto. 2000. Identification of N<sup>ω</sup>-carboxymethylarginine as a novel acid-labile advanced glycation end product in collagen. *Biochem. J.* 347:23–27.
- Jorgenson, W. L., J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926–935.
- Kramer, R. Z., J. Bella, P. Mayville, B. Brodsky, and H. M. Berman. 1999. Sequence dependent conformational variations of collagen triple-helical structure. *Nat. Struct. Biol.* 6:454–457.
- Kramer, R. Z., J. Bella, B. Brodsky, and H. M. Berman. 2001. The crystal and molecular structure of a collagen-like peptide with a biological relevant sequence. *J. Mol. Biol.* 311:131–147.
- Leung, M. K., G. A. Folkes, N. S. Ramamurthy, M. Schneir, and L. M. Golub. 1986. Diabetes stimulates procollagen degradation in rat tendon in vitro. *Biochim. Biophys. Acta.* 880:147–152.
- Lien, Y., M. M. Tseng, and R. Stern. 1992. Glucose and glucose analogs modulate collagen metabolism. *Exp. Mol. Pathol.* 57:215–221.
- Long, C. G., E. Braswell, D. Zhu, J. Apigo, J. Baum, and B. Brodsky. 1993. Characterization of collagen-like peptides containing interruptions in the repeating Gly-X-Y sequence. *Biochemistry.* 32:11688–11695.
- McDonnell, S., M. Morgan, and C. Lynch. 1999. Role of matrix metalloproteinases in normal and disease processes. *Biochem. Soc. Trans.* 27:734–740.
- Nesto, R. W., and P. Libby. 2001. Diabetes mellitus and the cardiovascular system. In *Heart Disease Braunwald, E., Zipes, D. P., Libby, P., editors*. Saunders, Philadelphia. 2133–2150.
- Odani, H., K. Iijima, M. Nakata, S. Miyata, H. Kusunoki, Y. Yasuda, Y. Hiki, S. Irie, K. Maeda, and D. Fujimoto. 2001. Identification of N<sup>ω</sup>-carboxymethylarginine, a new advanced glycation endproduct in serum proteins of diabetic patients: possibility of a new marker of aging and diabetes. *Biochem. Biophys. Res. Commun.* 285:1232–1236.
- Paul, R. G., and A. J. Bailey. 1996. Glycation of collagen: the basis of its central role in the late complications of ageing and diabetes. *Int. J. Biochem. Cell Biol.* 28:1297–1310.
- Simonson, T. 1993. Free energy of particle insertion: an exact analysis of the origin singularity in simple liquids. *Mol. Phys.* 80:441–447.
- Stryer, L. 1988. *Biochemistry*. W. H. Freeman and Company, New York. 1089
- Stultz, C. M. 2002. Localized unfolding of collagen explains collagenase cleavage near imino-poor sites. *J. Mol. Biol.* 319:997–1003.
- Wahab, N. N., E. A. Cowden, N. J. Pearce, N. J. Gardner, H. Merry, and J. L. Cox. 2002. Is blood glucose an independent predictor of mortality in acute myocardial infarction in the thrombolytic era? *J. Am. Coll. Cardiol.* 40:1748–1754.