Molecular modeling of mechanical stresses on proteins in glassy matrices: Formalism

Harold W. Hatch and Pablo G. Debenedetti[a)]
Department of Chemical and Biological Engineering, Princeton University, Princeton, New Jersey 08544, USA

(Received 10 January 2012; accepted 22 June 2012; published online 18 July 2012)

We present an expression for the calculation of microscopic stresses in molecular simulation, which is compatible with the use of electrostatic lattice sums such as the Ewald sum, with the presence of many-body interactions, and which allows local stresses to be calculated on surfaces of arbitrarily complex shape. The ultimate goal of this work is to investigate microscopic stresses on proteins in glassy matrices, which are used in the pharmaceutical industry for the long-term storage and stabilization of labile biomolecules. We demonstrate the formalism's usefulness through selected results on ubiquitin and an α-keratin fragment, in liquid and glassy states. We find that atomic-level normal stresses on hydrophilic side-chains exhibit a similar fingerprint in both proteins, and protein-level normal stresses increase upon vitrification. Both proteins experience compressive stresses of the order of 10^2 bar in the glassy state. © *2012 American Institute of Physics*. [http://dx.doi.org/10.1063/1.4734007]

I. INTRODUCTION

One of the most common techniques for preserving the stability of medicinal biochemicals such as proteins, peptides, and small molecules during long-term storage is freeze-drying (lyophilization).[1] Lyophilization involves the freezing of ice from an initially dilute carbohydrate solution, followed by vitrification and further lowering of the water content by sublimation.[2,3] The preserved substance is stabilized because the glassy state greatly reduces molecular mobility. Consequently, the rates of potentially damaging processes, such as aggregation and reaction, are dramatically slowed down.[4,5] Freeze-drying has become the industry standard method to preserve labile biomolecules due to its high production yield, low product variability, and approval by regulatory agencies.[6]

Although lyophilization is the preferred process for preserving labile biochemicals, including therapeutic proteins, many such molecules do not survive the freezing and drying stresses experienced during the process.[7] These include low temperature stresses, ice formation, large ionic concentration and *p*H changes, phase separation, and water removal from the protein hydration shell.[1] The use of carbohydrates mitigates these stresses,[8,9] but there are significant gaps in basic understanding of the molecular mechanisms of stabilization. As a result, the engineering of lyophilization processes remains largely empirical. It is therefore important to acquire microscopic, fundamental understanding of stabilization mechanisms of labile biochemicals in glassy matrices.

Numerous experimental and computational investigations relevant to liquid- and solid-phase protein stabilization have been reported, including experiments on proteins in amorphous solids by differential scanning calorimetry and Fourier transform infrared spectroscopy (see, e.g., Refs. 10–13), computer simulations of the structure and dynamics of carbohydrate-water solutions,[14–20] and simulations of the effect of bioprotectants on proteins.[21–24] An experimental study[25] of annealing following lyophilization[26–28] suggests that residual mechanical stresses may play an important role in causing the unfolding of some proteins in the glassy state, and their subsequent aggregation upon reconstitution. The addition of an annealing step resulted in less cracking of dried protein cakes, more native protein secondary structure and less aggregation upon reconstitution of the lyophilized cake.[25] Mechanical stresses that cause cracking may act at the microscopic level, and these stresses are difficult to probe experimentally. In this work we present a statistical mechanical and computational formalism for investigating mechanical stresses on proteins at the microscopic level, and we demonstrate the usefulness of the approach through selected initial results. Although our specific interest is on the mechanical response of proteins to vitrification, the approach is general, and should be broadly applicable to the molecular-level investigation of stresses in biomolecules in solution.

The microscopic stress tensor used in molecular simulation has been derived by the following methods: combining hydrodynamic and statistical mechanical concepts,[29–32] Fourier analysis of the aforementioned hydrodynamic equations,[33–35] spatial averaging,[36,37] volume perturbation,[38] and invariance properties of the partition function.[39–42] The resulting expressions for the local stress are invalid for a protein model that includes long-range electrostatics,[43,44] and cannot be used to compute stresses on arbitrary surfaces, where planar expressions[45,46] do not apply. Furthermore, it is an approximation to decompose the virial stress into contributions from local atoms,[47,48] as described in Sec. II A, and such an approximation has been shown to lead to spurious results.[45] Numerous studies investigate stresses with long range electrostatics (e.g., Refs. 49 and 50). Important progress toward addressing problems arising from the presence of long-range electrostatic interactions in stress tensor

a)Electronic mail: pdebene@princeton.edu.

calculations was made in a study of lipid bilayers,[51] but consistency with the Irving-Kirkwood approach[29] was not attained for such long-range forces.[51] In another pioneering work, non-planar expressions for the pressure field were derived in the study of lipid membranes and membrane-protein complexes,[52,53] but long-range electrostatics were not considered. Although significant progress has been made in this body of literature, new computer simulation methods must be developed to investigate microscopic mechanical stresses on proteins.

Here, we present an expression for microscopic stresses, applicable to protein simulations including electrostatic interactions, and apply it to length scales varying from atoms to individual residues, and to whole proteins. To our knowledge, this is the first computer simulation of microscopic mechanical stresses on a protein that is compatible with the use of electrostatic lattice methods such as the Ewald sum.[43,44] The derivation of local stress is a generalization of the work of Morante, Rossi, and Testa[41,42] and is presented in Sec. II. Computer simulation methods are outlined in Sec. III. We then demonstrate the usefulness of the approach by applying it to two model proteins: ubiquitin,[54,55] a well-characterized globular protein, and a coiled-coil fragment from the intermediate filament of α-keratin.[56,57] In particular, in Sec. IV we present initial results on the effects of vitrification and of the presence of the carbohydrates trehalose and maltose on the local stresses on these proteins. These results are meant to illustrate the method's suitability; they do not represent a comprehensive computational investigation of mechanical stresses on lyophilized proteins. This will be the subject of future work. A summary of conclusions from this work and a discussion of future research directions are included in Sec. V.

II. LOCAL STRESS TENSOR

A. Local stress tensor derivation

In this section, the statistical mechanical work of Morante, Rossi, and Testa[41,42] is generalized in order to derive a microscopic expression for the local stress tensor, which is valid for electrostatic potentials with explicit volume dependence, as described in Sec. II B and Appendix A. The local stress tensor, $\tau(\vec{r})$ at a point, $\vec{r}$ is given by[41,42,58]

$$\tau^{ab}(\vec{r}) = \left.\frac{\delta A}{\delta \eta^{ab}(\vec{r})}\right|_T, \tag{1}$$

where A is the Helmholtz free energy, T is the temperature, $\frac{\delta}{\delta \eta^{ab}(\vec{r})}|_T$ is the functional derivative[59,60] at constant temperature with respect to the local deformation tensor, η^{ab} and indices a, b represent coordinate directions. For an infinitesimal local deformation, $\vec{\epsilon}(\vec{r})$,

$$\vec{r} \to \vec{r}' = \vec{r} + \vec{\epsilon}(\vec{r}), \tag{2}$$

the deformation tensor is given by

$$\eta^{ab}(\vec{r}) = \frac{1}{2}\left[\nabla_r^a \epsilon^b(\vec{r}) + \nabla_r^b \epsilon^a(\vec{r})\right], \tag{3}$$

where $\vec{\epsilon}(\vec{r})$ will be given the additional meaning of a canonical point transformation and $\nabla_r^a = \frac{\partial}{\partial r^a}$.

It is difficult to evaluate $\tau(\vec{r})$ with Eq. (1) because of the constraint of constant temperature. The important insight of Morante *et al.*[41] was to exploit the invariance of the augmented canonical partition function, Z_c^{aug} with respect to a canonical point transformation, $\vec{\epsilon}(\vec{r})$, which leads to a local conservation equation for the stress tensor. For a given deformation caused by an external force field (force per unit volume), $\vec{f}_{ext}(\vec{r})$, internal forces characterized by the local stress tensor arise, which at equilibrium, satisfy the condition[58]

$$\nabla_r^b \tau^{ab}(\vec{r}) + f_{ext}^a(\vec{r}) = 0. \tag{4}$$

Equation (4) is obtained by explicitly demonstrating the invariance of the partition function via functional derivative with respect to the infinitesimal canonical transformation,

$$-\frac{1}{\beta}\left.\frac{\delta \log Z_c^{aug}}{\delta \vec{\epsilon}(\vec{r})}\right|_{\vec{\epsilon}(\vec{r})=\vec{0}} = \left\langle \left.\frac{\delta H_{ext}}{\delta \vec{\epsilon}(\vec{r})}\right|_{\vec{\epsilon}(\vec{r})=\vec{0}} \right\rangle_c = 0, \tag{5}$$

where $\beta = \frac{1}{k_B T}$, k_B is Boltzmann's constant, H_{ext} is the "external" Hamiltonian defined below, $|_{\vec{\epsilon}(\vec{r})=\vec{0}}$ denotes that the expression is evaluated at $\vec{\epsilon}(\vec{r}) = \vec{0}$, and $\langle \ldots \rangle_c$ is a canonical ensemble average. In summary, the approach to deriving the local stress tensor is to evaluate the functional derivative of the augmented Hamiltonian in Eq. (5), and to compare with Eq. (4) in order to read off $\tau(\vec{r})$ in terms of molecular variables. The local stress tensor, $\tau(\vec{r})$ may also be given the elegant interpretation as the set of Lagrangian multipliers[42] that enforce the relationship between infinitesimal displacements, $\vec{\epsilon}(\vec{r})$ and the deformation tensor, $\eta(\vec{r})$ given by Eq. (3). Appendix B demonstrates how the same equation for the stress tensor is obtained by this alternative derivation.

The Hamiltonian of the perturbed system, H_{ext} with respect to the transformed coordinates, $\{\vec{q}'\}$ and momenta, $\{\vec{p}'\}$, is given by

$$H_{ext} = \sum_i^N \left(\frac{(\vec{p}'_i)^2}{2m_i}\right) + U(\{\vec{q}'_{ij}\}, V') + U_{ext}(\vec{q}'_i), \tag{6}$$

where N is the number of particles in a simulation cell subject to cubic periodic boundary conditions, m_i is the particle mass and U_{ext} is some external potential. The following derivation requires that the potential energy, U be expressed as a function of the set of pairwise separation vectors, $\vec{q}_{ij}$ and the volume of the unit cell, V. The electrostatic potential is specifically discussed for the Ewald sum in Sec. II B and the Coulomb sum in Appendix A.

The canonical transformation is given by

$$[\{\vec{q}\}, \{\vec{p}\}] \to [\{\vec{q}'\}, \{\vec{p}'\}], \tag{7}$$

$$q_i'^a = q_i^a + \epsilon^a(\vec{q}_i), \tag{8}$$

$$p_i^a = \left[\delta^{ab} + \nabla_q^b \epsilon^a(\vec{q})|_{\vec{q}=\vec{q}_i}\right] p_i'^b, \tag{9}$$

where δ is the Kronecker delta ($\delta^{ab} = 1$ if $a = b$ and 0 otherwise). The generalization presented in this work is for the case of an electrostatic potential function that depends explicitly on the volume. Consider a local volume, Ω_i, about

atom i, such that the total volume, $V = \sum_i^N \Omega_i$. The relative change in the local volume upon an infinitesimal canonical point transformation is given by the sum of the principal components of the deformation tensor[58]

$$\Omega_i' = [1 + \delta^{ab}\eta^{ab}(\vec{q})|_{\vec{q}=\vec{q}_i}]\Omega_i. \tag{10}$$

The functional derivatives of the transformed variables are now obtained in order to evaluate Eq. (5) with the chain rule. The functional derivative of Eq. (8) is given by

$$\frac{\delta q_i'^b}{\delta\epsilon^a(\vec{r})} = \delta^{ab}\delta(\vec{r} - \vec{q}_i), \tag{11}$$

where $\delta(\vec{r} - \vec{q}_i)$ is the d-dimensional Dirac delta distribution. The functional derivative of the squared momenta is obtained from Eq. (9),

$$\frac{\delta(p_i')^2}{\delta\epsilon^a(\vec{r})} = 2p_i^a p_i^b \nabla_r^b \delta(\vec{r} - \vec{q}_i), \tag{12}$$

where $\nabla_q^a \delta(\vec{r} - \vec{q}) = -\nabla_r^a \delta(\vec{r} - \vec{q})$. Similarly, the functional derivative of Eq. (10) is

$$\frac{\delta\Omega_i'}{\delta\epsilon^a(\vec{r})} = -\delta^{ab}\Omega_i \nabla_r^b \delta(\vec{r} - \vec{q}_i). \tag{13}$$

The functional derivative of the transformed pairwise separation vector leads to a difference of Dirac delta distributions, which may be Taylor-expanded, leading to the Irving-Kirkwood formula,[29] or cast into the closed-form expression,

$$\frac{\delta q_{ij}'^c}{\delta\epsilon^a(\vec{r})} = -\delta^{ac} q_{ij}^b \nabla_r^b \int_0^1 dt\, \delta[\vec{r} - \vec{q}_i + t\vec{q}_{ij}], \tag{14}$$

by a similar procedure as Lemma 1 of Noll.[30] The non-uniqueness of the stress tensor due to the choice of contour of integration is the subject of much debate.[31,61,62] Previous studies tend to recommend the Irving-Kirkwood contour, given by the straight line path between pairs of particles, over other alternatives. In a derivation employing the invariance properties of the partition function, Mistura argues that the stress tensor is uniquely defined by the Irving-Kirkwood contour by definition of the distance between particle pairs.[40] Rossi and Testa claim that the stress tensor may also be defined uniquely as the set of Lagrange multipliers which enforce the relation between the displacement vector and deformation tensor for short-range potentials (see Appendix B).[42] In the special case of a planar interface, observables such as the surface tension are shown to be invariant to the choice of contour.[63,64] For spherical interfaces, however, the Irving-Kirkwood contour agrees with microscopic sum rules for pressure difference, surface tension and Tolman length, while an alternative contour does not.[65] While the Irving-Kirkwood contour gives the same local pressure of a hard-sphere fluid in Cartesian and spherical coordinates, the Harasima contour[32] leads to unphysical results in spherical coordinates.[66] Finally, there is no reasonable definition of the Harasima contour for arbitrary shapes, such as in this work, because, for example, the Harasima contour is defined with respect to slab geometry (see Figure 1 of Sonne *et al.*[51]) or spherical geometry (see Figure 1 of Thompson *et al.*[67]).

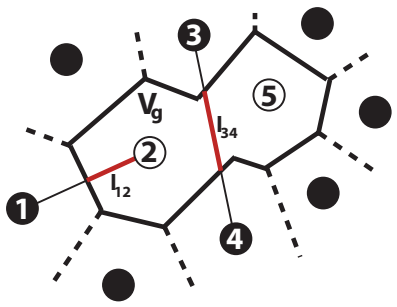

FIG. 1. Illustration of the local stress tensor calculation, Eq. (19), for a group containing particles 2 and 5. The volume of the group, V_g, is the sum of the Voronoi volumes of particles 2 and 5 shown by thick black lines. By definition of the group, $\Lambda_2 = \Lambda_5 = 1$ while $\Lambda_1 = \Lambda_3 = \Lambda_4 = 0$. The dimensionless quantity, $l_{12} = 0.5$, is the fraction of the line segment connecting particles 1 and 2 that is within V_g, shown by the thick red line. Note that the double sum of Eq. (19) involves all pairs of particles, which may lead to contributions from particles not within the group (e.g., $l_{34} \simeq 0.4$).

Finally, $\tau(\vec{r})$ is identified by comparison of Eq. (4) with the evaluation of Eq. (5), applying the chain rule to Eq. (6), and invoking Eqs. (11)–(14). The resulting, new expression is

$$\tau^{ab}(\vec{r}) = -\left\langle \sum_i^N \left[\left(\frac{p_i^a p_i^b}{m_i} - \delta^{ab}\Omega_i \frac{\partial U}{\partial \Omega_i} \right) \delta(\vec{r} - \vec{q}_i) - \frac{1}{2} \sum_j{}' \frac{\partial U}{\partial q_{ij}^a} q_{ij}^b \int_0^1 dt\, \delta[\vec{r} - \vec{q}_i + t\vec{q}_{ij}] \right] \right\rangle_c, \tag{15}$$

where $\sum_j'$ is a sum over all particles that interact with i including periodic images and many-body terms. The partial derivative $\frac{\partial}{\partial\Omega_i}$ is evaluated at constant particle pair separation due to the chain rule, and only contributes for potentials with explicit volume dependence, such as electrostatic potentials. In addition, $\frac{\partial}{\partial\Omega_i}$ may be replaced with $\frac{\partial}{\partial V}$ by the chain rule.

The contribution of many-body potentials to the local stress tensor is included in Eq. (15) where the pairwise force for many-body potentials is calculated via the equation $\vec{F}_{ij} = -\frac{\partial U}{\partial \vec{q}_{ij}}$, in agreement with previous studies.[68–71] Note, however, that in the present work, following the statistical mechanical approach of Morante *et al.*,[41] we do not invoke an *a priori* decomposition of many-body interactions into pairwise forces. Instead, the pairwise force, $-\frac{\partial U}{\partial \vec{q}_{ij}}$, naturally arises as a result of the functional differentiation of Eq. (6). For the special case of angle potentials, a strict numerical test of the many-body interaction calculation was performed, based on the fact that the contribution of angle-dependent potentials to the local stress tensor is traceless. Intuitively, this is expected because angle potentials are invariant to isotropic volume scaling, which is the thermodynamic route to calculating the pressure.[72] This result was observed numerically for configurations described in Sec. III A, where each configuration contained tens of thousands of angle potential terms with traceless stress tensor contributions. Note that normal stresses (the focus of results presented in Sec. IV) are given by the trace of the stress tensor, and thus angle potentials do not contribute.

The contribution of holonomic constraints (e.g., rigidity) to the local stress tensor may be included in Eq. (15) by

utilizing the pairwise constraint forces and momenta obtained from typical constraint algorithms (e.g., SHAKE (Ref. 73)). Insertion of the constrained Hamiltonian[74,75] into Eq. (5) leads to an additional term in Eq. (15) for the pairwise constraint forces, $\vec{F}_{ij} = -\sum_{k=1}^{M} \lambda_k \frac{\partial \sigma_k}{\partial \vec{q}_{ij}}$, where λ_k are a set of M Lagrangian multipliers, and the holonomic constraints, $\sigma_k(\{\vec{q}_{ij}\})$, depend on pairwise separation vectors, $\vec{q}_{ij}$.

To obtain the local stress tensor, τ_g^{ab} on an atom or an arbitrary group of atoms, one averages over the volume of this group, $V_g = \sum_i^N \Omega_i \Lambda_i$ with the "volume average" formula,[35,46]

$$\tau_g^{ab} = \frac{1}{V_g} \int_{V_g} d^3 r \tau^{ab}(\vec{r}), \tag{16}$$

where Λ_i is unity if particle i is contained in the group and zero otherwise. When Eq. (16) is applied to Eq. (15), the following two integrals appear

$$\Lambda_i = \int_{V_g} d^3 r \delta(\vec{r} - \vec{q}_i), \tag{17}$$

$$l_{ij} = \int_{V_g} d^3 r \int_0^1 dt \delta[\vec{r} - \vec{q}_i + t\vec{q}_{ij}], \tag{18}$$

where the geometric interpretation of the dimensionless quantity l_{ij}, Eq. (18), is the fraction of the line segment connecting particles i and j that is within V_g ($0 \leq l_{ij} \leq 1$), in agreement with previous studies.[35,36] As illustrated in Figure 1, the l_{ij} term may lead to stress contributions from atoms not within the group.

The local stress tensor, τ_g^{ab} on an atom or an arbitrary group of atoms is obtained from Eqs. (15)–(18):

$$\tau_g^{ab} = -\left\langle \frac{1}{V_g} \sum_i^N \left(\frac{p_i^a p_i^b}{m_i} \Lambda_i + \frac{1}{2} {\sum_j}' F_{ij}^a q_{ij}^b l_{ij} \right) - \delta^{ab} \frac{\partial U}{\partial V} \right\rangle_c. \tag{19}$$

Throughout this work, we use the terminology "approximate local stress" to denote Eq. (19) in the limit $l_{ij} \to \Lambda_i$. In the last term of Eq. (19), $\frac{\partial}{\partial \Omega_i}$ is replaced with $\frac{\partial}{\partial V}$ by the chain rule. For molecular simulations, Ω_i is the Voronoi volume[76,77] of particle i, and $\vec{p}_i$ is the momentum of particle i relative to the center of mass velocity of the group. Although the derivation was performed in the canonical ensemble, Eq. (19) is also valid for the microcanonical ensemble[41] and an ensemble of mechanically stable configurations obtained by energy minimization, commonly referred to as inherent structures.[78] The formula for inherent structures may be obtained by employing the partition function for inherent structures[79] in Eq. (5). For inherent structures, the kinetic term of Eq. (19) is zero.

It is useful to consider the limits in which Eq. (19) is identical to the pressure and the "approximate local stress" ($l_{ij} \to \Lambda_i$). The first limit is the case where the volume average is computed over the entire system, $V_g \to V$, such that the line segment connecting particles i and j is always contained within the system volume. In this case, the pressure, $P = -trace(\tau_g^{ab})/3$. For large V_g, the "local stress approximation" ($l_{ij} \to \Lambda_i$) approaches Eq. (19) when the contribution of l_{ij} terms internal to V_g is larger than the contribution of l_{ij} terms which span the boundary of V_g.[48] The second limit is the case where a potential is sufficiently short-ranged such that each particle only interacts with its nearest-neighbors. In this case, each interacting particle pair shares a Voronoi half-plane, which is equivalent to $l_{ij} \to \Lambda_i$ and consistent with the limit $\vec{q}_{ij} \to \vec{0}$ in Eq. (14).

Zhou[80] has questioned the inclusion of the kinetic term involving momenta, $\vec{p}_i$, in Eq. (19). This matter has become the subject of controversy.[48,81–87] In the present work, the kinetic contribution to the stress, being a direct consequence of the canonical transformation, Eqs. (7), (9) and (12), is included. We note, furthermore, that inclusion of the kinetic term has been found to be essential in numerous studies. For example, Admal and Tadmor[48] showed that Zhou's conclusions result from not taking into account the difference between absolute and relative velocities. Hoover *et. al.* showed excellent agreement between atomistic and continuum perspectives provided that both the kinetic and interaction contributions to the stress tensor are included.[83] Subramaniyan and Sun used molecular dynamics simulations to show that only the full tensor (kinetic and potential terms) is consistent with macroscopic thermodynamics, and that very significant errors can result from neglecting the kinetic energy term.[82] Finally, we note that the kinetic term is required for thermodynamic consistency (e.g., $P = -trace(\tau_g)/3$ when $V_g \to V$).

Two coordinate-independent quantities can be extracted from the local stress tensor; the local normal stress, p, and the von Mises shear stress, κ. These invariants are given by[88]

$$p = -trace(\tau_g)/3 = -(\tau_1 + \tau_2 + \tau_3)/3, \tag{20}$$

$$\kappa = \sqrt{\frac{1}{6}[(\tau_1 - \tau_2)^2 + (\tau_1 - \tau_3)^2 + (\tau_2 - \tau_3)^2]}, \tag{21}$$

where τ_1, τ_2, τ_3 are the principal components of the stress tensor, and p is defined such that compressive normal stresses are positive. In what follows, we focus on normal stresses.

B. Electrostatic potentials: The Ewald sum case

In this work, charge-charge interactions were treated with the Ewald sum,[44,89] a common method for efficiently calculating the long-range electrostatic potential. In addition, the Ewald sum was checked by comparison with the less efficient Coulomb sum in Appendix A. In order for the Ewald sum to be compatible with the local stress tensor, Eq. (19), the Ewald sum potential energy, $U^e = U^e(\epsilon_s = \infty)$, is expressed in the explicit pairwise form for conducting boundary conditions of dielectric constant, $\epsilon_s = \infty$,

$$U^e = \frac{1}{2} \sum_i^N \sum_j^N U_{ij}^k + \frac{1}{2} \sum_i^N {\sum_j}' U_{ij}^r + \frac{1}{2} \sum_{(i,j)\epsilon M} U_{ij}^s,$$

$$U_{ij}^k = \frac{z_i z_j}{V} \sum_{k \neq 0} A_k e^{i\vec{k} \cdot \vec{q}_{ij}},$$

$$A_k = \frac{4\pi}{k^2} e^{-\frac{k^2}{4\alpha}},$$

$$\vec{k} = 2\pi(n_x/L_x, n_y/L_y, n_z/L_z),$$
$$U^r_{ij} = z_i z_j \frac{erfc(\sqrt{\alpha}|q_{ij}|)}{|q_{ij}|},$$
$$U^s_{ij} = -z_i z_j \left(\frac{erf(\sqrt{\alpha}|q_{ij}|)}{|q_{ij}|} + 2\delta_{ij}\sqrt{\frac{\alpha}{\pi}}\right), \quad (22)$$

where z_i is the charge on particle i with Gaussian units that normalize the charge by a factor of $1/\sqrt{4\pi\epsilon_0}$, ϵ_0 is the permittivity of free space, the orthorhombic cell volume, $V = L_x L_y L_z$, α is a parameter associated with the Gaussian width of screening charge, n_x, n_y, $n_z = 0, \pm1, \pm2, \ldots \pm\infty$, $n_x^2 + n_y^2 + n_z^2 \leq n_{max}^2$, and *erf*() and *erfc*() are the error function and the complementary error function, respectively. The sum over particle pairs (i, j) in Eq. (22) within the set, M, of excluded intramolecular interactions and self-interactions is the real-space correction for the spurious inclusion of these interactions in Fourier space.

The necessary derivatives for the local stress tensor, Eq. (19), may be obtained from Eq. (22),

$$\vec{F}^k_{ij} = \frac{z_i z_j}{V}\sum_{k\neq0} A_k \vec{k}\sin(\vec{k}\cdot\vec{q}_{ij}),$$
$$\vec{F}^r_{ij} = z_i z_j \left(\frac{erfc(\sqrt{\alpha}|q_{ij}|)}{|q_{ij}|^3} + \frac{2\sqrt{\alpha}}{|q_{ij}|^2\sqrt{\pi}}e^{-\alpha|q_{ij}|^2}\right)\vec{q}_{ij},$$
$$\vec{F}^s_{ij} = -z_i z_j \left(\frac{erf(\sqrt{\alpha}|q_{ij}|)}{|q_{ij}|^3} - \frac{2\sqrt{\alpha}}{|q_{ij}|^2\sqrt{\pi}}e^{-\alpha|q_{ij}|^2}\right)\vec{q}_{ij},$$
$$\frac{\partial U^e}{\partial V} = \frac{z_i z_j}{6V^2}\sum_i^N\sum_j^N\sum_{k\neq0} A_k[(k^2/2\alpha - 1)\cos(\vec{k}\cdot\vec{q}_{ij}) + \vec{k}\cdot\vec{q}_{ij}\sin(\vec{k}\cdot\vec{q}_{ij})], \quad (23)$$

where $\vec{F}_{ij} = -\frac{\partial U_{ij}}{\partial \vec{q}_{ij}}$. It is important to consider the constraints of $\frac{\partial U^e}{\partial V}$ for Eq. (23). First, particle pair separation vectors, $\vec{q}_{ij}$, are held constant due to the chain rule applied to obtain Eq. (15). This leads to a $\sin(\vec{k}\cdot\vec{q}_{ij})$ term not present when scaled coordinates are held constant when deriving the pressure[72,90,91] (note, however, that this sine term in $\frac{\partial U^e}{\partial V}$, last line of Eq. (23), cancels the identical contribution from $\vec{F}^k_{ij}$, first line of Eq. (23), in the limit $V_g \to V$, and hence the Ewald pressure obtained in the usual way, by calculating $\frac{\partial U^e}{\partial V}$ at constant scaled coordinates, is identical to the pressure obtained via Eq. (19) in the limit $V_g \to V$, with $\frac{\partial U^e}{\partial V}$ and $\vec{F}^k_{ij}$ contributions as per Eq. (23)). Second, the ratios L_x/L_y, L_x/L_z are fixed in $\frac{\partial U^e}{\partial V}$ because an arbitrary deformation of an orthorhombic cell, $V = L_x L_y L_z$, can always be written as a combination of isotropic (L_x/L_y, L_x/L_z fixed) and anisotropic (V fixed) deformations, with the latter not contributing to $\frac{\partial U^e}{\partial V}$.

III. SIMULATION METHODS

A. Molecular dynamics

Molecular dynamics was used to simulate a protein in water and carbohydrate-water solutions. Ubiquitin, a 76-residue globular protein (see Fig. 2), was chosen as one of two model proteins. Ubiquitin is found in all eukaryotic cells and is highly conserved,[54] perhaps due to its central role in protein degradation.[55] The CHARMM force field[92] was used to model ubiquitin, and the initial structure was taken from the protein data bank (PDB ID code 1UBQ (Ref. 93)). The VMD program[94] determined the coordinates of the missing hydrogen atoms in the PDB, and the protein was solvated with SPC/E water molecules[95] in cubic periodic boundary conditions. For simulations of ubiquitin in 20 wt. % carbohydrate-water solutions, the CHARMM force field[96] modeled 57 $\alpha-\alpha$-trehalose molecules with initial structures from the trehalose dihydrate crystal structure,[97] or 57 maltose molecules.[98] Carbohydrate molecules were placed in a grid with random orientations, and the carbohydrate molecules that overlapped with the protein were removed. An equilibrated water configuration was then inserted while removing water molecules that overlapped with the protein and carbohydrate. The final number of water molecules, N_w, reported in Table I, was constrained by the weight fraction of carbohydrate (20%) and setting the target total equilibrated volume equal to the equilibrated volume of the protein in water.

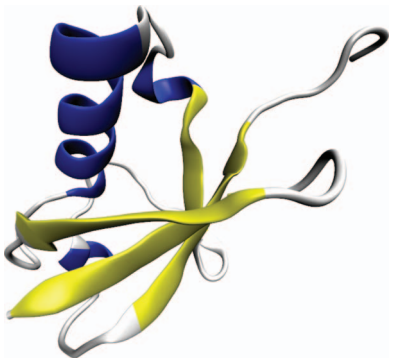

FIG. 2. The STRIDE algorithm[120] was used to make secondary structure assignments of ubiquitin, and a typical liquid configuration in water at ambient conditions is shown. The α-helix is the largest helical region shown in blue. The other two blue regions are 3_{10} helices. β-strands are colored yellow. The C-terminus is the coil at the top right.

The second model protein used in this work was an α-keratin fragment. Due to the difficulty of simulating an entire intermediate filament of α-keratin at atomic resolution,[99] only a fragment composed of the coiled-coil 1a region was simulated,[57] shown in Figure 3. The coiled-coil is composed of two α-helical monomers taken from both human keratin genes k5 and k14.[56,100] The coiled-coil fragment was built using the backbone coordinates of leucine zipper protein, an ideal coiled-coil (PDB ID 2ZTA (Ref. 101)), and mutating the side chains to residues 171-201 of gene k5 and residues 118-148 of gene k14 with the Swiss PDB viewer.[102] The

TABLE I. Summary of simulations performed in this work.

System	N_w[a]	V(nm^3)	t_{eq}(ns)	RMSD(Å)
Ubiquitin(ubq)	5000	161.4	15	0.8(0.1)
ubq + 20 wt. % trehalose	4326	162.3	20	0.8(0.1)
ubq + 20 wt. % maltose	4324	162.2	20	0.8(0.1)
Keratin	10310	319.8	15	1.4(0.3)

[a]Number of water molecules.

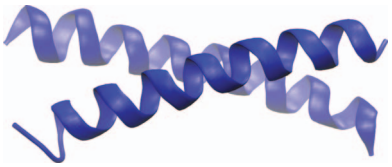

FIG. 3. The 1a coiled-coil region of α-keratin was simulated as a fragment solvated in water. The coiled-coil structure is stabilized by a stripe of hydrophobic side chains which run along the length of both α-helical monomers.

protein was solvated in a cubic periodic box with SPC/E water molecules[95] and 2 chloride ions to maintain charge neutrality.

All simulations were performed with the LAMMPS (Ref. 103) program, and the ch2lmp tool converted the PDB and protein structure files from VMD into LAMMPS data format. A 1-femtosecond integration time step was used, and the rigidity of water molecules was enforced with the SHAKE algorithm.[73] Non-bonded interactions were smoothly turned off from 9 Å to 10 Å with the CHARMM switching function. Coulombic interactions were calculated with the particle-particle particle-mesh solver[104] with the optimal number of wave vectors to obtain a fractional energy tolerance of 10^{-4}.

The initial configuration was equilibrated to ambient conditions of atmospheric pressure and a temperature of 300 K through a series of steps designed to minimize disturbance of the original protein structure. First, the solution was relaxed for a very brief time of 0.25 ps with a constant, rescaled temperature of 300 K, keeping the protein rigid. Energy minimization was then used for 5 steps in order to remove overlaps between protein atoms. Next, most of the free volume created during the simulation setup was removed by ramping down the barostat[105] target pressure from 50 kbars to 1 bar over 30 ps, at 300 K while holding the protein rigid. Finally, the entire system was equilibrated at atmospheric pressure and a temperature of 300 K (Ref. 105) for t_{eq} nanoseconds (see Table I). Pressure was then passively maintained at atmospheric pressure without a barostat by setting the volume of the system to the average equilibrated volume. Production simulations were performed for 50 ns in the canonical ensemble with Nosé-Hoover integration[106] and a 2.5 ps coupling time, with configurations stored every picosecond for analysis. For every 100 ps, the liquid configuration was enthalpy-minimized and the stress tensor was calculated for both liquid (MD; molecular dynamics) and glass (IS; inherent structures) configurations.

During production simulation, the structural stability of the proteins was quantified by the root-mean-squared deviation (RMSD) of backbone α-carbons, reported in Table I. The RMSD was calculated with VMD (Ref. 94) by alignment of an ensemble of protein configurations with the first configuration to correct for displacement and rotation of the molecule. Less structured protein regions were not considered in the RMSD calculation. Residues 73-76 of the C-terminus of ubiquitin, and residues within 2 amino acids of the ends of the keratin monomer fragments, did not contribute to the RMSD in Table I.

B. Inherent structures

In this work, the glassy state is represented by inherent structures[78] obtained by enthalpy minimization of an ensemble of configurations from the liquid state. Enthalpy minimization[107] is similar to energy minimization, but the density is allowed to relax to achieve a target pressure, P^{target}. Physically, enthalpy minimization may be thought of as an infinitely fast quench under isobaric conditions, and is more appropriate than energy minimization when comparing with experiments. The enthalpy, H, may be expressed as a function of its independent variables with the rigid body framework,

$$H = U + P^{target}V = H[\{\vec{r}/L\}, \{\vec{r}_{COM}/L\}, \{\vec{\theta}\}, V], \tag{24}$$

where $\vec{r}$ is the position of atoms belonging to flexible (non-rigid-body) molecules (e.g., protein atoms), $V = L^3$ is the cubic simulation volume, $\vec{r}_{COM}$ is the center-of-mass position of the rigid-body molecules (e.g., water) and $\vec{\theta}$ are the Euler angles[108] associated with the orientation of the rigid bodies. During minimization with the conjugate-gradient Polak-Ribeire algorithm,[109,110] the independent degrees of freedom of the enthalpy, Eq. (24), are successively line minimized along directions determined by the gradient, until the local minimum is reached. The gradient of Eq. (24) is given by the forces of non-rigid-body atoms, center of mass forces and torques of rigid-body molecules, and $\frac{\partial H}{\partial V} = \frac{\partial U}{\partial V} + P^{target}$, calculated by finite difference[107] of the change in energy due to a small, 1 Å^3, isotropic volume contraction. Minimization was deemed complete when the fractional enthalpy difference between successive line minimization steps was 10^{-15} with a Brent's method[111] parabolic interpolation tolerance of 10^{-8} which utilizes force units of kcal/mole/Å.

C. Optimization of local stress algorithm

The most computationally demanding part of the local stress tensor calculation is the l_{ij} term of Eq. (19) in the inner loop of the pairwise Ewald sum, Eq. (23), where one ubiquitin configuration takes about 30 min on one core of a 3.0 GHz processor using the LAMMPS program with pairwise Ewald sum and Voronoi intersection algorithm added in-house. In the pairwise Ewald form, Eq. (22), it is optimal to reduce the number of wave vectors, $\vec{k}$, by employing a real-space cutoff of nearly half the box length. The optimal number of wave vectors to obtain a fractional energy tolerance of 10^{-4} was determined by LAMMPS. The number of trigonometric operations was reduced by precomputation of sine and cosine of $(\vec{k} \cdot \vec{q}_i)$ and employing sum-difference identities.

The algorithm to calculate l_{ij} proceeds as follows. First, generate a list of coordinates for the Voronoi diagram by wrapping the coordinates in the periodic box, and then separately include periodic images which are within some threshold of the center of the box for each dimension (0.7L was sufficient for this work). Using an efficient calculation of the regular Voronoi polyhedra,[112,113] store, for each particle, the Voronoi half-planes and the associated nearest neighbors of each half-plane. For a given pair of particles labeled as i and j, parameterize the line segment with a variable which is zero at

the position of particle i and unity at j. Starting with the cell of particle i, calculate the intersection of the line segment with each half-plane using parametric equations which avoid the square root operation. After the intersection with the nearest half-plane is found, store the stress contribution of the current Voronoi cell, move to the adjacent Voronoi cell which shares that half-plane, and repeat the search for the nearest half-plane intersection until a Voronoi cell of a periodic image or the Voronoi cell of particle j is reached. If a Voronoi cell of a periodic image is reached, move to the original cell of that periodic image and parameterize the line segment with a variable that is now zero at the position of a periodic image of particle i. If the Voronoi cell of particle j is reached, consider the next pair of particles until all particle pairs are exhausted.

IV. RESULTS AND DISCUSSION

In this section we demonstrate the application of our formalism for the calculation of molecular-level stresses to the particular case of proteins in glassy matrices. The results are simply meant to illustrate the technique's capabilities, rather than to present a comprehensive molecular-level picture of freeze-drying. Such studies will be the subject of future publications. In what follows, local normal stress calculations, using Eqs. (19) and (20), will be presented across length scales ranging from individual protein atoms to residues, and to the entire protein. A hierarchy of local normal stress, according to the chemical nature of protein atoms, is shown in Figure 4 for ubiquitin, and in Figure 5 for the keratin fragment. Atoms with the largest normal stresses belong to the side chains of the hydrophilic residues lysine, aspartic acid, and glutamic acid. The backbone carbonyl carbons experience large compressive normal stresses, and α-carbons experience large tensile normal stresses. A similar hierarchy may be seen in the liquid where normal stresses on protein atoms tend to be less compressive than in the glass; the same trend occurs in keratin (not shown). In spite of the structural differences between the two proteins, there is a remarkable similarity between their atomic-level hierarchies of local normal stresses. We plan to explore the generality of this observation in future studies. We note that, when the "local stress approximation" ($l_{ij} \to \Lambda_i$) is applied to Eq. (19), stresses are substantially different (absolute deviations up to 463 kbars).

The normal stresses on the length scale of entire residues, shown in Figure 6 for ubiquitin and the keratin fragment, have a smaller range of magnitudes than do the normal stresses on individual protein atoms. Normal stresses are expected to be reduced for larger groups of atoms, and ultimately converge to ambient pressure for the entire system. The hydrophilic glutamic and aspartic acid residues experience the largest, tensile normal stresses, and are highlighted in color in Figure 6. The "local stress approximation" ($l_{ij} \to \Lambda_i$) assigns large compressive normal stresses on aspartic and glutamic acid residues (not shown), as opposed to the large tensile normal stresses observed in Figure 6.

For the normal stresses on the entire ubiquitin protein, the largest length scale considered in this work, the ensemble averaged stress is on the order of magnitude of hundreds of bar (see Figure 7 and Table II). For all simulations performed, the protein experiences compressive normal stress in the glass, and tensile normal stress in the liquid (see Table II). The "local stress approximation" ($l_{ij} \to \Lambda_i$ in Eq. (19)), however, indicates compressive normal stresses for the protein in both the glass and liquid (see Table II). The fluctuation in the

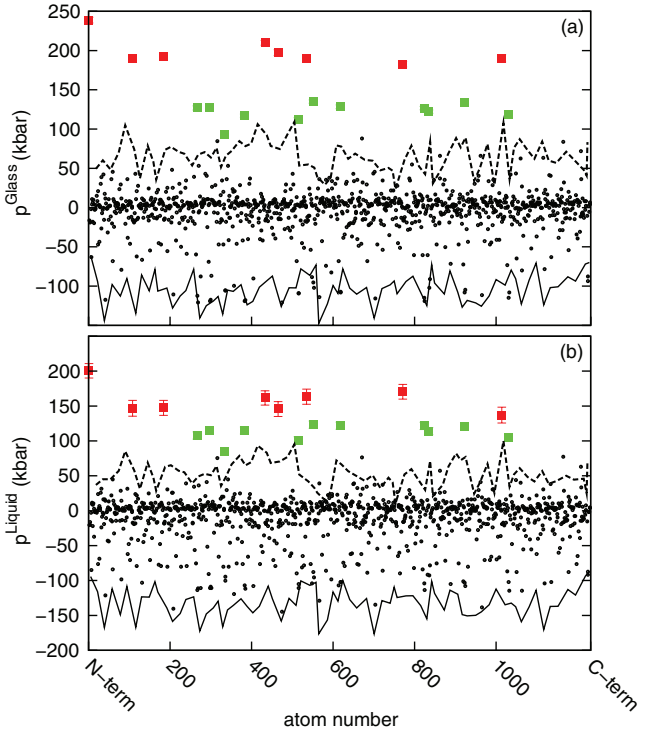

FIG. 4. (a) Atomic-level normal stresses on ubiquitin in glassy water at 1 bar. Colored squares represent the following hydrophilic side chain atoms: (red) nitrogen on ammonium ($-N^+H_3$) groups in lysine and the N-terminus, and (green) carbon on carboxylate ($-COO^-$) groups in aspartic and glutamic acid. Backbone carbonyl carbons are represented by the dashed line, α-carbons by the solid line, and all other atoms appear as dots. Atoms are listed from N-terminus to C-terminus. (b) Atomic-level normal stresses on ubiquitin in liquid water at 1 bar and 300 K. Error bars are the standard deviation of the mean from 500 samples in both liquid and glass.

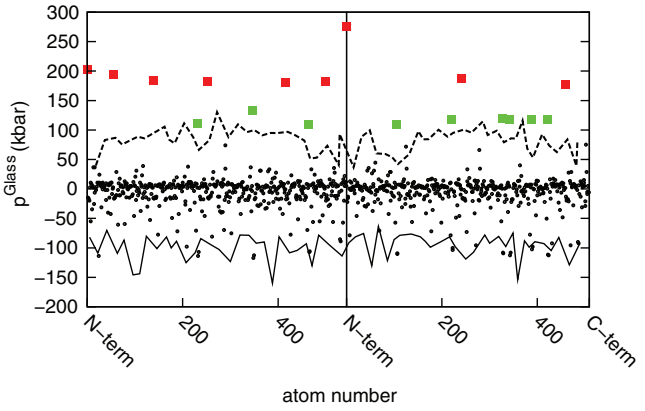

FIG. 5. Atomic-level normal stresses for keratin fragment in glassy water at 1 bar. Symbols are as in Figure 4. Atoms are listed from the N-terminus to the C-terminus for both coils in succession. The first coil ends and the second coil begins in the middle of the figure, denoted by the vertical line. Error bars are the standard deviation of the mean from 500 samples.

TABLE II. Summary of the normal stress on whole proteins in the glass and liquid.

System	p^{Glass}(kbar)	p^{Liquid}(kbar)	$p^{Virial,Glass}$(kbar)[a]	$p^{Virial,Liquid}$(kbar)
Ubiquitin(ubq)	0.43(0.03)	−0.7(0.1)	1.71(0.02)	0.4 (0.1)
ubq + 20 wt. % trehalose	0.42(0.03)	−1(0.1)	1.68(0.02)	0 (0.1)
ubq + 20 wt. % maltose	0.5(0.03)	−0.7(0.1)	1.72(0.03)	0.4(0.1)
Keratin	0.15(0.04)	−0.84(0.07)	1.38(0.05)	0.26(0.08)

[a]Virial superscript indicates the "local stress approximation" ($l_{ij} \rightarrow \Lambda_i$) in Eq. (19).

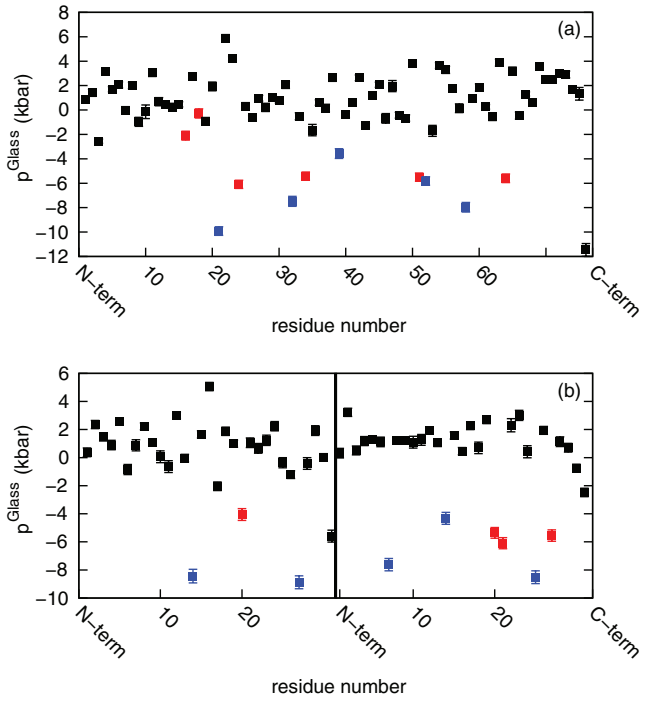

FIG. 6. (a) Residue-level normal stresses for ubiquitin in glassy water at 1 bar. Colored squares represent (blue) aspartic acid and (red) glutamic acid. Residues are listed from N-terminus to C-terminus. Note that the C-terminus residues also possess large, tensile normal stresses, perhaps due to the carboxylate group which is also present in aspartic and glutamic acid. (b) Residue-level normal stresses for keratin fragment in glassy water at 1 bar. The first coil ends and the second coil begins in the middle of the figure, denoted by the vertical line. Error bars are the standard deviation of the mean from 500 samples.

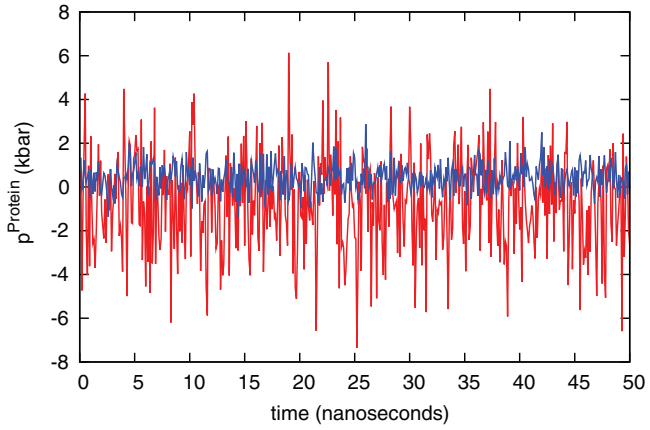

FIG. 7. The instantaneous normal stress on ubiquitin at 1 bar in the glass (blue) and liquid at 300 K (red). On average, ubiquitin is under compression in the glass and under tension in the liquid (see Table II for ensemble averages).

local normal stress in the liquid is greater than in the glass, as expected due to thermal motion in the liquid.

Table II includes illustrative results of normal stress calculations in the presence of carbohydrates. Stresses on ubiquitin are not significantly altered by addition of 20 wt. % trehalose or maltose. Studies across a range of sugar weight fractions, currently in progress, are needed in order to understand the effect of co-solutes on the mechanical response of proteins to dehydration and vitrification.

V. CONCLUSIONS

In this work, we have developed a statistical mechanical formalism that enables the calculation of microscopic stresses on proteins, with an eye toward a comprehensive investigation of freeze-drying at the molecular level. The stress tensor equation presented in this paper is, to our knowledge, the first such expression that is valid for molecular simulation of biological molecules that can accommodate electrostatic lattice sums, many-body interactions, and complex (non-planar) shapes. Selected results have been presented that demonstrate the method's ability to compute the quantities of interest for a molecular-level understanding of the effects of freeze-drying on protein conformational stability. In general, the magnitude of the local normal stress decreases with increasing length scale (atom, residue, whole protein), increases upon entering the glassy state (the protein experiences increased compressive stress upon vitrification), and is qualitatively different from the "approximate local stress" ($l_{ij} \rightarrow \Lambda_i$ in Eq. (19)). The present approach can also accommodate the presence of complex co-solutes, such as carbohydrates, which are commonly used in the solid-state stabilization of proteins.

While the initial results presented here demonstrate the use of the microscopic stress expression on two model proteins, future investigations will be performed to explore systematically the behavior of proteins in glassy matrices by expanding the range of proteins, sugar concentration, temperature, and pressure to be investigated. Furthermore, another freeze-drying process parameter will be explored by performing simulations at varying quench rates and comparing the results with those obtained via enthalpy minimization. Finally, local stresses will be decomposed so as to isolate the contributions arising from the solvent, from co-solutes (e.g., carbohydrates), and from the protein itself (intramolecular contributions). These investigations, as well as corresponding studies of shear stresses (the results presented in this paper having been focused on normal stresses), will be the subject of future publications.

ACKNOWLEDGMENTS

The financial support of Unilever U.K. Central Resources and the National Science Foundation Graduate Fellowship to H.W.H. is gratefully acknowledged. Computations were performed at the Terascale Infrastructure for Groundbreaking Research in Engineering and Science (TIGRESS) facility at Princeton University.

APPENDIX A: ELECTROSTATIC POTENTIALS: THE COULOMB SUM CASE

Although the results presented in this paper use the Ewald sum to calculate Coulombic contributions to the stress tensor (see Sec. II B), the equivalence of the Ewald sum, and the conditionally convergent Coulomb sum is demonstrated for two reasons. First, the Coulomb sum serves as a check for the Ewald sum.[114] Second, the Coulomb sum serves as a check for the explicit volume dependent term $\frac{\partial U}{\partial V}$ in Eq. (19) for the local stress tensor. The volume dependent term in question is further justified if identical results may be obtained from the Coulomb sum (no volume dependence) and Ewald sum (volume dependent).

The Coulomb sum potential energy, $U^c = U^c(\epsilon_s = 1)$ for vacuum boundary conditions, $\epsilon_S = 1$, is given by[43]

$$U^c = \frac{1}{2}\sum_{\vec{n}}{}' \left(\sum_{i=1}^{N}\sum_{j=1}^{N} z_i z_j |\vec{q}_{ij} + \vec{n}|^{-1} \right), \tag{A1}$$

where U^c is the Coulomb energy, z_i is the charge of particle i, $\vec{n} = (n_x L, n_y L, n_z L)$, and L is the length of the cubic simulation cell. The sum over $\vec{n}$ runs over all simple cubic lattice points of $|n_x, n_y, n_z| \leq n_c$, where the prime indicates omission of $i = j$ for $|\vec{n}| = 0$. The roughly spherical build-up of simulation cell images about the central cell facilitates convergence of the Coulomb sum, and n_c is chosen such that the energy and pressure computed via the Coulomb sum are equivalent to the Ewald sum, where $n_c = 2$ was found to be sufficient for our system. The Coulomb sum implicitly assumes that the simulation is surrounded by vacuum, or dielectric constant, $\epsilon_s = 1$. The Ewald sum, however, assumes conducting boundary condition, or dielectric constant $\epsilon_s = \infty$. In order to demonstrate equivalence of the Ewald and Coulomb sums, a polarization correction term to the Coulomb sum, $U^p = U^c(\epsilon_s = \infty) - U^c(\epsilon_s = 1)$ is given by[115,116]

$$\begin{aligned} U^p &= -\frac{2\pi}{3L^3}\left|\sum_i^N z_i \vec{q}_i\right|^2 \\ &= \frac{1}{2}\sum_i^N\sum_j^N \frac{2\pi}{3V} z_i z_j q_{ij}^2 \\ &= \frac{1}{2}\sum_i^N\sum_j^N U_{ij}, \end{aligned} \tag{A2}$$

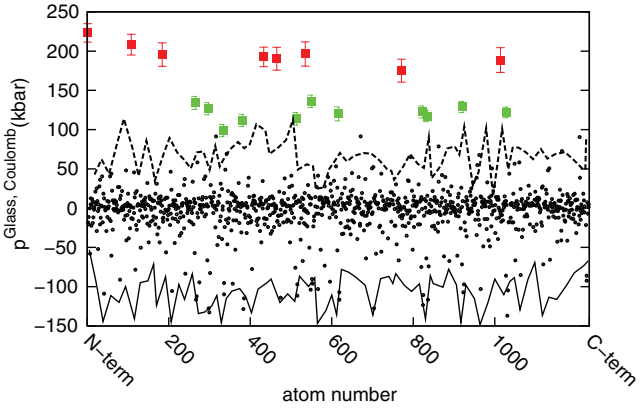

FIG. 8. Atomic-level normal stresses on ubiquitin in glassy water at 1 bar calculated with the Coulomb sum, $U^c(\epsilon_s = \infty) = U^c(\epsilon_s = 1) + U^p$. Symbols are as in Figure 4. Error bars are the standard deviation of the mean from 100 samples.

where the necessary derivatives for Eq. (19) are

$$\begin{aligned} \frac{\partial U_{ij}^p}{\partial \vec{q}_{ij}} &= \frac{4\pi}{3V} z_i z_j \vec{q}_{ij}, \\ \frac{\partial U^p}{\partial V} &= -\frac{U^p}{V}. \end{aligned} \tag{A3}$$

The atomic-level normal stresses on ubiquitin atoms in glassy water at 1 bar calculated with the Coulomb sum $U^c(\epsilon_s = \infty) = U^c(\epsilon_s = 1) + U^p$ are shown in Figure 8. The Coulomb sum gives the same results as the Ewald sum for atomic-level normal stresses, Figure 4. A quantitative comparison of the statistical differences between the Coulomb and Ewald sum using the Z-test found all but 14 of 1231 protein atoms to lie within two standard deviations of the null result (no statistical differences with 95% confidence).[117] The differences in the Coulomb and Ewald sum for the other 14 atoms lie within three standard deviations of the null result. Having shown that the Coulomb and Ewald summations for $\epsilon_s = \infty$ conducting boundary condition are equivalent, it is trivial to see that the two summations are also equivalent for $\epsilon_s = 1$.

APPENDIX B: ALTERNATIVE LOCAL STRESS TENSOR DERIVATION

The local stress tensor given by Eq. (15) may be obtained by an alternative procedure[42] which interprets $\tau(\vec{r})$ as the Lagrange multipliers that enforce the relation between displacements and the local deformation tensor given by Eq. (3). Employing the principle of virtual work for an infinitesimal local transformation,[118] and assuming that the transformation is reversible,[119] the variation of free energy, dA is the negative of the reversible work, δW_{rev}

$$dA = \int_V d^3r[\tau^{ab}(\vec{r})\eta^{ab}(\vec{r}) - F^a_{ext}(\vec{r})\epsilon^a(\vec{r})]. \tag{B1}$$

Taylor-expanding the transformed Hamiltonian about the untransformed variables, one obtains with the help of Eqs. (6),

(8) and (10),

$$\begin{aligned} H_{ext} &= H^0_{ext}[\{\vec{q}\},\{\vec{p}\}], \\ &+ \sum_i^N \eta^{ab}(\vec{q}_i)\left(-\frac{p_i^a p_i^b}{m_i} + \delta^{ab}\Omega_i \frac{\partial U}{\partial \Omega_i}\right) \\ &+ \frac{1}{2}\sum_i^N \sum_j{}' [\epsilon^a(\vec{q}_i) - \epsilon^a(\vec{q}_j)]\frac{\partial U_{ij}}{\partial q_{ij}^a} \\ &+ \sum_i^N \epsilon^a(\vec{q}_i)\frac{\partial U_{ext}}{\partial q_i^a}. \end{aligned} \tag{B2}$$

Inserting the Taylor expansion of $\epsilon(\vec{q}_j)$ about $\vec{q}_i$ in the third term of Eq. (B2),

$$\epsilon^a(\vec{q}_i) - \epsilon^a(\vec{q}_j) = q_{ij}^b \sum_{n=1}^{\infty} \frac{1}{n!}(-\vec{q}_{ij}\cdot\vec{\nabla}_r)^{n-1}\eta^{ab}(\vec{q}_i), \tag{B3}$$

and making use of the identity $\int f(x)(\nabla_x)^n \delta(x-a)dx = (-\nabla_x)^n f|_{x=a}$, the variation of the free energy is

$$\begin{aligned} dA &= \int_V d^3 r \Bigg[\eta^{ab}(\vec{r}) \\ &\times \Bigg\langle -\sum_i^N \delta(\vec{r}-\vec{q}_i)\left(\frac{p_i^a p_i^b}{m_i} - \delta^{ab}\Omega_i\frac{\partial U}{\partial \Omega_i}\right) \\ &+ \frac{1}{2}\sum_i^N \sum_j{}' \frac{\partial U}{\partial q_{ij}^a} q_{ij}^b \sum_{n=1}^{\infty}\frac{1}{n!}(\vec{q}_{ij}\cdot\vec{\nabla}_r)^{n-1}\delta(\vec{r}-\vec{q}_i)\Bigg\rangle_c \\ &- \epsilon(\vec{r})\left\langle \sum_i^N F_i^{a,ext}\delta(\vec{r}-\vec{q}_i)\right\rangle_c \Bigg], \end{aligned} \tag{B4}$$

which is compared with Eq. (B1) to obtain $\tau^{ab}(\vec{r})$, noting that the summation over n in the third term of Eq. (B4) is related to the Taylor series expansion of the difference of Dirac delta distributions and may be cast in the closed-form integral expression appearing in Eq. (15).

[1] W. Wang, Int. J. Pharm. **203**, 1 (2000).
[2] C. A. Oksanen and G. Zografi, Pharm. Res. **7**, 654 (1990).
[3] J. A. Searles, J. F. Carpenter, and T. W. Randolph, J. Pharm. Sci. **90**, 860 (2001).
[4] S. Yoshioka and Y. Aso, J. Pharm. Sci. **96**, 960 (2007).
[5] V. Molinero and W. A. Goddard, Phys. Rev. Lett. **95**, 045701 (2005).
[6] C. J. Roberts and P. G. Debenedetti, AIChE J. **48**, 1140 (2002).
[7] L. Yu, Adv. Drug Delivery Rev. **48**, 27 (2001).
[8] S. D. Allison, M. C. Manning, T. W. Randolph, K. Middleton, A. Davis, and J. F. Carpenter, J. Pharm. Sci. **89**, 199 (2000).
[9] W. Garzon-Rodriguez, R. L. Koval, S. Chongprasert, S. Krishnan, T. W. Randolph, N. W. Warne, and J. F. Carpenter, J. Pharm. Sci. **93**, 684 (2004).
[10] M. J. Pikal, D. R. Rigsbee, and M. L. Roy, J. Pharm. Sci. **96**, 2765 (2007).
[11] M. J. Pikal, D. R. Rigsbee, and M. L. Roy, J. Pharm. Sci. **97**, 5122 (2008).
[12] S. Giuffrida, G. Cottone, E. Vitrano, and L. Cordone, J. Non-Cryst. Solids **357**, 677 (2011).
[13] G. Bellavia, S. Giuffrida, G. Cottone, A. Cupane, and L. Cordone, J. Phys. Chem. B **115**, 6340 (2011).
[14] C. J. Roberts and P. G. Debenedetti, J. Phys. Chem. B **103**, 7308 (1999).
[15] P. B. Conrad and J. J. de Pablo, J. Phys. Chem. A **103**, 4049 (1999).
[16] N. C. Ekdawi-Sever, P. B. Conrad, and J. J. de Pablo, J. Phys. Chem. A **105**, 734 (2001).
[17] S. L. Lee, P. G. Debenedetti, and J. R. Errington, J. Chem. Phys. **122**, 204511 (2005).
[18] A. Lerbret, P. Bordat, F. Affouard, M. Descamps, and F. Migliardo, J. Phys. Chem. B **109**, 11046 (2005).
[19] R. Politi, L. Sapir, and D. Harries, J. Phys. Chem. A **113**, 7548 (2009).
[20] A. Magno and P. Gallo, J. Phys. Chem. Lett. **2**, 977 (2011).
[21] G. Cottone, L. Cordone, and G. Ciccotti, Biophys. J. **80**, 931 (2001).
[22] G. Cottone, S. Giuffrida, G. Ciccotti, and L. Cordone, Proteins: Struct., Funct., Bioinf. **59**, 291 (2005).
[23] A. Lerbret, P. Bordat, F. Affouard, A. Hedoux, Y. Guinet, and M. Descamps, J. Phys. Chem. B **111**, 9410 (2007).
[24] F. Liu, L. Ji, L. Zhang, X. Dong, and Y. Sun, J. Chem. Phys. **132**, 225103 (2010).
[25] S. D. Webb, J. L. Cleland, J. F. Carpenter, and T. W. Randolph, J. Pharm. Sci. **92**, 715 (2003).
[26] J. A. Searles, J. F. Carpenter, and T. W. Randolph, J. Pharm. Sci. **90**, 872 (2001).
[27] B. Wang and M. J. Pikal, J. Pharm. Sci. **99**, 663 (2010).
[28] B. Wang, M. T. Cicerone, Y. Aso, and M. J. Pikal, J. Pharm. Sci. **99**, 683 (2010).
[29] J. H. Irving and J. G. Kirkwood, J. Chem. Phys. **18**, 817 (1950).
[30] R. B. Lehoucq and A. Lilienfeld-Toal, J. Elast. **100**, 5 (2010).
[31] P. Schofield and J. R. Henderson, Proc. R. Soc. London, Ser. A **379**, 231 (1982).
[32] A. Harasima, Adv. Chem. Phys. **1**, 203 (1958).
[33] J. F. Lutsko, J. Appl. Phys. **64**, 1152 (1988).
[34] J. F. Lutsko, J. Appl. Phys. **65**, 2991 (1989).
[35] J. Cormier, J. M. Rickman, and T. J. Delph, J. Appl. Phys. **89**, 99 (2001).
[36] R. J. Hardy, J. Chem. Phys. **76**, 622 (1982).
[37] A. I. Murdoch and D. Bedeaux, Proc. R. Soc. London, Ser. A **445**, 157 (1994).
[38] A. Ghoufi and P. Malfreyt, J. Chem. Phys. **135**, 104105 (2011).
[39] L. Mistura, J. Chem. Phys. **83**, 3633 (1985).
[40] L. Mistura, Int. J. Thermophys. **8**, 397 (1987).
[41] S. Morante, G. C. Rossi, and M. Testa, J. Chem. Phys. **125**, 034101 (2006).
[42] G. C. Rossi and M. Testa, J. Chem. Phys. **132**, 074902 (2010).
[43] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Clarendon, 1989), pp. 156–157.
[44] D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications* (Academic, 2002).
[45] B. D. Todd, D. J. Evans, and P. J. Daivis, Phys. Rev. E **52**, 1627 (1995).
[46] D. M. Heyes, E. R. Smith, D. Dini, and T. A. Zaki, J. Chem. Phys. **135**, 024512 (2011).
[47] A. I. Murdoch, J. Elast. **88**, 113 (2007).
[48] N. C. Admal and E. B. Tadmor, J. Elast. **100**, 63 (2010).
[49] K. Van Workum, K. Yoshimoto, J. J. de Pablo, and J. F. Douglas, Phys. Rev. E **71**, 061102 (2005).
[50] S. G. Moore and D. R. Wheeler, J. Chem. Phys. **136**, 164503 (2012).
[51] J. Sonne, F. Y. Hansen, and G. H. Peters, J. Chem. Phys. **122**, 124903 (2005).
[52] O. H. S. Ollila, H. J. Risselada, M. Louhivuori, E. Lindahl, I. Vattulainen, and S. J. Marrink, Phys. Rev. Lett. **102**, 078101 (2009).
[53] O. H. S. Ollila, M. Louhivuori, S. Marrink, and I. Vattulainen, Biophys. J. **100**, 1651 (2011).
[54] G. Goldstein, M. Scheid, U. Hammerling, D. H. Schlesinger, H. D. Niall, and E. A. Boyse, Proc. Natl. Acad. Sci. U.S.A. **72**, 11 (1975).
[55] M. Hochstrasser, Nature (London) **458**, 422 (2009).
[56] R. Lersch, V. Stellmach, C. Stocks, G. Giudice, and E. Fuchs, Mol. Cell. Biol. **9**, 3685 (1989).
[57] C. Danciulescu, B. Nick, and F. Wortmann, Biomacromolecules **5**, 2165 (2004).
[58] L. D. Landau and E. M. Lifshitz, *Theory of Elasticity* (Pergamon, 1986), Vol. **7**.
[59] H. T. Davis, *Statistical Mechanics of Phases, Interfaces, and Thin Films* (Wiley-VCH, 1996).
[60] Following Morante *et al.* (Ref. 41), we adopt the functional derivative as a convenient alternative to Landau and Lifshitz's (Ref. 58) per unit volume notation.
[61] J. S. Rowlinson, Pure Appl. Chem. **65**, 873 (1993).
[62] M. Baus and R. Lovett, Phys. Rev. Lett. **67**, 407 (1991).
[63] J. Walton and K. Gubbins, Mol. Phys. **55**, 679 (1985).
[64] C. G. Gray, K. E. Gubbins, and C. G. Joslin, *Theory of Molecular Fluids* (Oxford University Press, 2011), Vol. 2.
[65] E. M. Blokhuis and D. Bedeaux, J. Chem. Phys. **97**, 3576 (1992).
[66] B. Hafskjold and T. Ikeshoji, Phys. Rev. E **66**, 011203 (2002).

[67]S. M. Thompson, K. E. Gubbins, J. P. R. B. Walton, R. A. R. Chantry, and J. S. Rowlinson, J. Chem. Phys. **81**, 530 (1984).
[68]J. Zhang and B. D. Todd, Phys. Rev. E **69**, 031111 (2004).
[69]H. Heinz, W. Paul, and K. Binder, Phys. Rev. E **72**, 066704 (2005).
[70]Y. Chen, J. Chem. Phys. **124**, 054113 (2006).
[71]N. C. Admal and E. B. Tadmor, J. Chem. Phys. **134**, 184106 (2011).
[72]G. Hummer, N. Gronbech-Jensen, and M. Neumann, J. Chem. Phys. **109**, 2791 (1998).
[73]J. Ryckaert, G. Ciccotti, and H. J. Berendsen, J. Comput. Phys. **23**, 327 (1977).
[74]G. Ciccotti and J. Ryckaert, Comput. Phys. Rep. **4**, 345 (1986).
[75]S. W. de Leeuw, J. W. Perram, and H. G. Petersen, J. Stat. Phys. **61**, 1203 (1990).
[76]M. de Berg, M. van Kreveld, M. Overmars, and O. Schwarzkopf, *Computational Geometry: Algorithms and Applications* (Springer, 2000).
[77]M. Gerstein, J. Tsai, and M. Levitt, J. Mol. Biol. **249**, 955 (1995).
[78]F. H. Stillinger, Science **267**, 1935 (1995).
[79]P. G. Debenedetti, F. H. Stillinger, T. M. Truskett, and C. J. Roberts, J. Phys. Chem. B **103**, 7390 (1999).
[80]M. Zhou, Proc. R. Soc. London, Ser. A **459**, 2347 (2003).
[81]J. A. Zimmerman, E. B. Webb III, J. J. Hoyt, R. E. Jones, P. A. Klein, and D. J. Bammann, Modell. Simul. Mater. Sci. Eng. **12**, S319 (2004).
[82]A. K. Subramaniyan and C. Sun, Int. J. Solids Struct. **45**, 4340 (2008).
[83]W. G. Hoover, C. G. Hoover, and J. F. Lutsko, Phys. Rev. E **79**, 036709 (2009).
[84]B. Liu and X. Qiu, J. Comput. Theor. Nanos. **6**, 1081 (2009).
[85]J. Wang, C. Lu, Q. Wang, P. Xiao, F. Ke, Y. Bai, Y. Shen, X. Liao, and H. Gao, Nanotechnology **23**, 025703 (2012).
[86]G. J. Tucker, S. Tiwari, J. A. Zimmerman, and D. L. McDowell, J. Mech. Phys. Solids **60**, 471 (2012).
[87]H. Cheng, C. Yu, and W. Chen, J. Mater. Sci. **47**, 3103 (2012).
[88]D. Srolovitz, K. Maeda, V. Vitek, and T. Egami, Philos. Mag. A **44**, 847 (1981).
[89]U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, J. Chem. Phys. **103**, 8577 (1995).
[90]S. Nose and M. Klein, Mol. Phys. **50**, 1055 (1983).
[91]W. Smith, CCP5 Info. Quart. **26**, 43 (1987).
[92]A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wirkiewicz-Kuczera, D. Yin, and M. Karplus, J. Phys. Chem. B **102**, 3586 (1998).
[93]S. Vijay-Kumar, C. E. Bugg, and W. J. Cook, J. Mol. Biol. **194**, 531 (1987).
[94]W. Humphrey, A. Dalke, and K. Schulten, J. Mol. Graphics **14**, 33 (1996).
[95]H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma, J. Phys. Chem. **91**, 6269 (1987).
[96]O. Guvench, E. Hatcher, R. M. Venable, R. W. Pastor, and A. D. MacKerell, J. Chem. Theory Comput. **5**, 2353 (2009).
[97]T. Taga, M. Senma, and K. Osaki, Acta Cryst. B **28**, 3258 (1972).
[98]S. M. E. Gres and G. A. Jeffrey, Acta Cryst. B **33**, 2490 (1977).
[99]Z. Qin and M. J. Buehler, J. Mol. Model. **17**, 37 (2010).
[100]D. Marchuk, S. McCrohon, and E. Fuchs, Cell **39**, 491 (1984).
[101]E. O'Shea, J. Klemm, P. Kim, and T. Alber, Science **254**, 539 (1991).
[102]N. Guex and M. C. Peitsch, Electrophoresis **18**, 2714 (1997).
[103]S. Plimpton, J. Comput. Phys. **117**, 1 (1995).
[104]R. W. Hockney and J. W. Eastwood, *Computer Simulation Using Particles* (Hilger, 1989).
[105]H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, J. Chem. Phys. **81**, 3684 (1984).
[106]W. G. Hoover, Phys. Rev. A **31**, 1695 (1985).
[107]T. F. Middleton and D. J. Wales, J. Chem. Phys. **118**, 4583 (2003).
[108]H. Goldstein, *Classical Mechanics* (Addison-Wesley, 1980).
[109]W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes: The Art of Scientific Computing* (Cambridge University Press, 2007).
[110]C. Chakravarty, P. G. Debenedetti, and F. H. Stillinger, J. Chem. Phys. **123**, 206101 (2005).
[111]R. P. Brent, *Algorithms for Minimization Without Derivatives* (Prentice-Hall, 1973).
[112]H. Edelsbrunner and N. R. Shah, Algorithmica **15**, 223 (1996).
[113]J. Bernal, http://math.nist.gov/~JBernal/JBernal_Sft.html.
[114]T. M. Nymand and P. Linse, J. Chem. Phys. **112**, 6152 (2000).
[115]S. W. de Leeuw, J. W. Perram, and E. R. Smith, Proc. R. Soc. London, Ser. A **373**, 27 (1980).
[116]E. R. Smith, J. Stat. Phys. **77**, 449 (1994).
[117]J. L. Devore, *Probability and Statistics for Engineering and the Sciences* (Thomson/Brooks/Cole, 2008).
[118]L. D. Landau and E. Lifshitz, *Mechanics* (Pergamon, 1969), Vol. 1.
[119]L. D. Landau and E. Lifshitz, *Statistical Physics* (Pergamon, 1978), Vol. 5.
[120]D. Frishman and P. Argos, Proteins: Struct., Funct., Genet. **23**, 566 (1995).