Nonuniform elastic properties of macromolecules and effect of prestrain on their continuum nature

Ankush Aggarwal, Eric R. May, Charles L. Brooks, III, and William S. Klug

Phys. Rev. E **93**, 012417 — Published 28 January 2016

DOI: 10.1103/PhysRevE.93.012417

On Non-Uniform Elastic Properties of Macromolecules and Effect of Pre-Strain on their Continuum Nature

Ankush Aggarwal,[1,*] Eric R. May,[2] Charles L. Brooks III,[3] and William S. Klug[4]

[1] *Zienkiewicz Centre for Computational Engineering, Swansea University, Swansea, UK SA1 8EN*
[2] *Department of Molecular & Cell Biology, University of Connecticut*
[3] *Department of Chemistry & Department of Biophysics, University of Michigan*
[4] *Department of Mechanical & Aerospace Engineering, University of California, Los Angeles, CA 90095*

Many experimental and theoretical methods have been developed to calculate the coarse-grained continuum elastic properties of macromolecules. However, all of those methods assume uniform elastic properties. Following the continuum mechanics framework, we present a systematic way of calculating the non-uniform effective elastic properties from atomic thermal fluctuations obtained from molecular dynamics simulation at any coarse-grained scale using a potential of mean force approach. We present the results for a mutant of sesbania mosaic virus capsid, where we calculate the elastic moduli at different scales and observe an *apparent* problem with the chosen reference configuration in some cases. We present a possible explanation using an elastic network model, where inducing random pre-strain results in a similar behavior. This phenomenon provides a novel insight into the continuum nature of macromolecules and defines the limits on details that the elasticity theory can capture. Further investigation into pre-strains could elucidate important aspects of conformational dynamics of macromolecules.

PACS numbers: 87.10.Pq, 87.14.E-, 87.15.A-, 87.15.La
Keywords: Macromolecules, continuum properties, potential of mean force, reference configuration, virus capsid

I. INTRODUCTION

Biological macromolecules can be highly dynamic entities, capable of undergoing self-assembly, disassembly and configurational changes, and binding to specific sites, either spontaneously under specific conditions or driven by certain agents. These configurational changes and binding interactions also affect their mechanical behavior, which, in many cases, are closely related to their function [1, 2]. Various experimental and theoretical/numerical methods have been used to assess the mechanical properties of macromolecules [3, 4]. In principle, atomic potentials can be used to simulate the dynamics of proteins under any mechanical loading condition. However, the time step in such calculations is usually limited to femtoseconds and, as the system size grows, it becomes prohibitively expensive to calculate molecular dynamics trajectories on timescales beyond microseconds. Moreover, to keep the computation time within reasonable range, in many situations, the simulated loading rates have to be increased to a non-physical level. This is especially true in the case of quasi-static problems, which have an infinitesimal loading rate.

To solve these problems, a number of coarse-graining methods have been developed, e.g., elastic network [5], Gaussian network [6], structure-based (Gō-like) models [7–9], and continuum mechanical models [10–13]. In addition to being computationally faster, these techniques also provide simplified and generalized organizational principles. For example, the elasticity models have provided insight into the shape [14] and buckling behavior [13] of icosahedral viral capsids, which hold true irrespective of the molecular details of the constitutive proteins. These methods all make use of some simplified, approximate description of both the atomic structure and the molecular interactions, and have some free fitting parameters – e.g., spring constants in elastic network models, and shear and bulk moduli in continuum mechanics. These parameters, although having some dependence on the molecular details, are either scaled to the available experiments or obtained by comparing with molecular dynamics (MD) trajectory [4, 15].

Despite the fact that interactions within macromolecular structures are highly heterogeneous (e.g., including covalent and non-covalent types), the common approximation of uniform elastic properties employed by these methods has been surprisingly successful in predicting and recapitulating their large-scale, overall mechanical behavior [10, 11, 16]. Such results are suggestive that the *structural* details of macromolecules are more important than the nature of the atomic interactions in determining their mechanical properties [16]. Others have reported significant change in the stiffness of viral capsid from a single point mutation [3], which, in terms of continuum elasticity, is best interpreted as a change in the constitutive relationship with the structure unaltered. These experiments provide evidence that local constitutive details, i.e., changes in energetics, are important and therefore should be captured in a coarse-grained model. The existing methods for computing continuum-scale elastic properties of macromolecules assume uniform properties

* a.aggarwal@swansea.ac.uk

[4, 15]. Hence, there is a need for a systematic framework that can be used to calculate *non-uniform* elastic properties and, thus, capture the effect of local energetic variations. Using such a framework, one would then be able to quantitatively address the related overarching question: how non-uniform are the effective elastic properties of macromolecules and their assemblies? Or put differently, what are the length scales that are characteristic of heterogeneity in elastic properties? How do they compare to the characteristic structural length scales – primary through quaternary?

Moreover, the validity of continuum approximation to inherently discrete structures, like macromolecules, is arguable. Continuum mechanics assumes that all points in the domain are filled with material where the lowest energy reference configuration is well-defined and known a priori. As we push the limit of continuum mechanics to nanoscale structures [17], it remains to be determined how valid these assumptions are, how many details can be captured with this model, and what do we exactly mean by limit of continuum? One key feature to observe is the heterogeneity of the system. That is, if we need a different stress-strain relation at every point of the continuum, then it could be argued that a continuum description is not suitable. However, there could be more fundamental deviations from the continuum behavior which we seek to explore in this paper.

In particular, one fundamental assumption of continuum mechanics which poses problems at near molecular-scales is that of a reference configuration. The strain and strain energy of a system upon deformation are calculated with respect to the reference state, which is also considered stress-free. Thus, if there is a change in the reference configuration, it intrinsically modifies the governing equations of the continuum model. In macromolecules, because of configurational changes, it becomes a challenge to identify an appropriate reference configuration. In addition, the assumption of a stress-free reference configuration requires additional considerations because in many cases the minimum energy state may not be stress-free: Pre-stress has been quantified in folded proteins and expected to play a role in their unfolding dynamics [18]. In the case of icosahedral viruses, the 5-fold sites have been shown to act as stress concentration points – even for the minimum energy configuration [19]. The reference configuration in that case is a flat sheet with edges cut at the five-fold sites and similar theory has been proposed for other platonic solids [20]. Previously, we have presented an elasticity theory for conformational changes in macromolecules which takes into account the change in reference configuration due to active nature of biomolecules [12]. In those situations, the reference configuration was not the *intact* minimum energy configuration; instead it consisted of incompatible or broken parts where the components had to be glued together to restore compatibility – either at the 5-fold sites or along the hexamer edges. Any such incompatabilities in the constituents of a structure lead to strains and stresses in the minimum energy configuration of the system – we call these pre-strain and pre-stress respectively. In those situations, the knowledge of a "true" reference configurations made it possible to analyze the system using continuum mechanics principles. Could there be situations where such a reference configuration does not exist at all, defying the applicability of continuum mechanics altogether?

In the present work, we seek answers to these questions about non-uniformity of elastic properties and effect of reference configuration on continuum nature in macromolecules. First, we present a method to systematically calculate the non-uniform elastic properties of macromolecules at any selected coarsening scale. We use a spherical virus capsid as an example because of the hierarchical organization and symmetry in the structure. Capsids possess the primary–secondary–tertiary hierarchy of proteins followed by scales of multiple proteins and, then, that of a spherical shell structure. Also, they have the symmetry of an icosahedron which can be exploited in our analysis to improve the results. We study a $T = 1$ mutant of wild type Sesbania Mosaic Virus (SeMV) (Fig. 1), where we apply our framework at various coarsening levels to determine the scale of inhomogeneity. Then, we demonstrate the role of reference configuration in our analysis using an elastic network model, which presents an explanation of the results we observe from MD. Finally, we discuss the importance of the results presented here, and consider generalizations to other biomolecular structures and the future directions in this area.

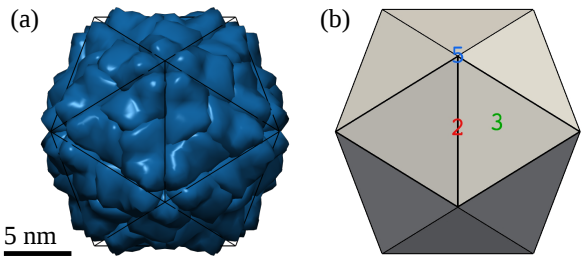

FIG. 1. (Color online) (a) SeMV coordinates obtained from viperDB [21] and visualized in Chimera [22], and (b) an icosahedron shape showing 2–, 3–, and 5–fold symmetry axes.

II. METHOD FOR CALCULATING NON-UNIFORM ELASTIC PROPERTIES

In principle, the continuum-scale elastic properties can be derived directly from atomic potentials. However, direct theoretical calculations starting from atomic potentials miss the entropic component of free energy and the effect of solvent on atomic interactions. In this section, we present a systematic way to extract the spatial distribution of non-uniform elastic properties from a molecular dynamics (MD) trajectory and, thus, take into

account the entropic component and solvent effects. In addition, our approach allows a flexibility of choosing different coarsening scales and combining different regions of macromolecule based upon a priori homogeneity assumptions or symmetry considerations.

The prerequisite in this framework is an MD trajectory of equilibrium fluctuations at constant temperature. The equilibrium fluctuations at finite temperature include the entropic part of the free energy, which can not be obtained directly from the potential energy of the system. Also the effect of solvent is incorporated explicitly in MD simulations. It can be argued that if an MD simulation is required to extract the mechanical properties, there is no gain in computational efficiency. However, here we need only equilibrium fluctuations which mitigates the problem of unrealistically large loading rates required for performing deformation inducing trajectories (e.g. simulated AFM nanoindentation [23]).

A. Continuum Model

We model the coarse-grained mechanics by continuum elasticity, defining point-wise a strain energy density function dependent only on the local strain. A material point of the system in its reference configuration is denoted by its position coordinates $\boldsymbol{X}$, which upon deformation moves to $\boldsymbol{x}(\boldsymbol{X},t)$ at time t. The Green-Lagrange strain tensor $\mathbf{C}=\mathbf{F}^T\cdot\mathbf{F}$ gives a proper measure of the local deformation, where $\mathbf{F}(\boldsymbol{X},t)=\partial\boldsymbol{x}(\boldsymbol{X},t)/\boldsymbol{X}$ is the deformation gradient.

We explore two continuum models: three-dimensional (3D) bulk elasticity and two-dimensional (2D) shell elasticity. For both of the models we assume isotropic elasticity, so the strain energy is invariant under rigid-body rotations of the reference state. This implies that the strain energy depends only on the isotropic invariants of the strain tensor $\mathbf{C}$.

For 3D bulk elasticity, we choose the three isotropic strain invariants:

$$J=\sqrt{\det\left(\mathbf{C}\right)},\quad \bar{I}_1=\mathrm{tr}\left(\mathbf{C}\right)J^{-2/3},\ \text{and}$$
$$\bar{I}_2=\frac{1}{2}\left[\mathrm{tr}^2\left(\mathbf{C}\right)-\mathrm{tr}\left(\mathbf{C}^2\right)\right]J^{-4/3}. \tag{1}$$

Here J represents the deformed volume ratio, and $\bar{I}_1$ and $\bar{I}_2$ are insensitive to volume change and, therefore, measure shearing of the material. By construction, $\bar{I}_1\geq 3$ and $\bar{I}_2\geq 3$ for all deformation mappings, and for the reference configuration the strain tensor is an identity $\mathbf{C}=\mathbf{F}=\mathbf{I}$. Thus, the reference configuration with zero strain corresponds to $\bar{I}_1=\bar{I}_2=3$ and $J=1$. Total elastic strain energy of the body is written as:

$$\mathcal{U}\left(\mathbf{C}\right)=\overline{\mathcal{U}}_0+\int_V W(\bar{I}_1,\bar{I}_2,J)dV. \tag{2}$$

Here, $\overline{\mathcal{U}}_0$ is arbitrary reference energy of the system and does not affect its elastic properties. The exact form of the elastic strain energy density function W defines the constitutive relation between stresses and strains, and will be chosen based upon the results from the MD simulation. If the reference configuration is stress-free, the strain energy density function W should have a minimum at the reference configuration, that is, at $\bar{I}_1=\bar{I}_2=3$ and $J=1$.

In 2D shell elasticity theory, the deformation is defined in terms of a mid-surface. Therefore, the deformation gradient $\mathbf{F}$ and the strain $\mathbf{C}$ are 2-rank tensors, which define only the in-plane deformation. The two isotropic strain invariants in this case are:

$$J^{\mathrm{2D}}=\sqrt{\frac{1}{2}\left[\mathrm{tr}^2\left(\mathbf{C}\right)-\mathrm{tr}\left(\mathbf{C}^2\right)\right]}\ \text{and}\ \bar{I}_1^{\mathrm{2D}}=\frac{\mathrm{tr}\left(\mathbf{C}\right)}{J^{\mathrm{2D}}}. \tag{3}$$

Here, J^{2D} is the deformed surface area ratio, and $\bar{I}_1^{\mathrm{2D}}$ measures shearing of the surface and is insensitive to area change. Analogous to the 3D bulk elasticity, $\bar{I}_1^{\mathrm{2D}}\geq 2$ for all in-plane deformations. In addition to the in-plane deformation, the out-of-plane or bending deformation is defined by curvature tensor $\mathbf{B}$, which has two isotropic invariants – the mean curvature $H=1/2\,\mathrm{tr}(\mathbf{B})$ and Gaussian curvature $K=\det(\mathbf{B})$ (for details on definition of the curvature tensor and related shell kinematics, see, e.g., [24]). However, since the dependence of bending energy on Gaussian curvature K is not clear for non-planar reference surfaces, we simplify it by assuming that bending energy depends only on the mean curvature H. Following the assumption that stretching and bending energies are independent, the total elastic energy of shell is:

$$\mathcal{U}\left(\mathbf{C},\mathbf{B}\right)=\overline{\mathcal{U}}_0+\int_{\mathcal{S}}\Big[W^{\mathrm{b}}(H)+W^{\mathrm{s}}(\bar{I}_1^{\mathrm{2D}},J^{\mathrm{2D}})\Big]d\mathcal{S}. \tag{4}$$

Similar to the 3D bulk elasticity model, the exact forms of the elastic strain energy density functions W^{b} and W^{s} will be chosen based upon the MD results. However, based upon the assumption of reference configuration being stress-free, it is a required that the stretching energy density function W^{s} is minimum at $\bar{I}_1^{\mathrm{2D}}=2$ and $J^{\mathrm{2D}}=1$.

B. Discretized Continuum Model

Before connecting the atomic model to the continuum one, we discretize the continuum model using displacement at finite number of points termed as nodes. The position coordinates of these nodes are denoted by $\boldsymbol{x}_I$, for $I=1,\ldots,N_{\mathrm{nodes}}$ – the total number of nodes in the system. Calculation of the strain field from discrete nodal positions requires an interpolation function. For the 3D bulk elasticity model, we use meshfree approximation [25] which we have previously used to create a framework for connecting continuum and atomic models [26, 27]. In the case of 2D shell elasticity model, we use different interpolations for the two parts of elastic energy. For calculating in-plane strain tensor $\mathbf{C}$ we use C^0-continuous Lagrange

linear polynomial interpolation, while for calculating the curvature tensor $\mathbf{B}$ we use C^1-continuous Loop's subdivision approximation [28]. We have previously proposed an elasticity theory of macromolecules based on this mixed formulation [12].

Using the discretized degrees of freedom, the integrals in strain energy Eqs. (2) and (4) can be converted into summations. There is a subtle difference in the numerical implementation of discretized version of two continuum models, because of which the summations are carried out over different entities in two cases. For 3D bulk elasticity model, it is carried out over nodes:

$$\mathcal{U}(\mathbf{C}) = \overline{\mathcal{U}}_0 + \sum_{I=1}^{N_{\text{nodes}}} W_I(\bar{I}_1, \bar{I}_2, J) V_I, \tag{5}$$

while for 2D shell elasticity model, it is carried out over triangular faces (elements) of the surface mesh:

$$\mathcal{U}(\mathbf{C}, \mathbf{B}) = \overline{\mathcal{U}}_0 + \sum_M W_M \mathcal{S}_M = \overline{\mathcal{U}}_0 + \sum_{M=1}^{N_{\text{triangles}}} \Big[W_M^{\text{b}}(H) + W_M^{\text{s}}(\bar{I}_1^{\text{2D}}, J^{\text{2D}}) \Big] \mathcal{S}_M. \tag{6}$$

Here V_I is the volume corresponding to node I and $\mathcal{S}_M$ is the surface area of triangle M. V_I is taken as the volume of the Voronoi cell corresponding to the node, which in turn is calculated from its dual graph – the Delaunay tessellation [27][29]. For 2D case, $\mathcal{S}_M$ is simply the area of triangle M. All the evaluations in the above summations are done at a single point – either the node or the centroid of the triangle. The reason behind the difference in integration schemes is purely numerical, and interested readers are referred to related literature [30, 31]. The main approximation here is that, in the above equations, the strain energy density functions (W_I, W_M^{b} and W_M^{s}) are constant over each node/triangle, but are allowed to vary from one node/triangle to the other. Thus, the number of nodes/triangles sets the degree of heterogeneity allowed in our models.

C. Connecting Atomic and Discretized Continuum Models

Consider an MD trajectory over a set of time steps $t_0, t_1, \ldots, t_n = t_{n-1} + \Delta t$. Let the positions of the atoms at time t_i be denoted by $\boldsymbol{r}_a(t_i)$, $a = 1, \ldots, N_{\text{atoms}}$. For each frame of the MD trajectory, we define a set of representative nodes as the degrees of freedom of our continuum elasticity model by coarse-grained mapping from the atomic positions. Specifically, we consider linear combinations

$$\boldsymbol{x}_I(t_i) = \sum_{a=1}^{N_{\text{atoms}}} A_{Ia} \boldsymbol{r}_a(t_i), \; I = 1, \ldots, N_{\text{nodes}}, \tag{7}$$

where A is a connectivity matrix of dimension $N_{\text{nodes}} \times N_{\text{atoms}}$ and $\sum_a A_{Ia} = 1 \;\; \forall I$. Furthermore, exactly one element of each column can be non-zero to make sure that the mapping is single-valued. Or in other words, each atom should contribute to define the position of one and only one node.

In this work, we define the mapping coefficients A_{Ia} in two ways: a) by placing a node at the center-of-mass of amino acids (also called residues) [32] bringing down the number of total degrees of freedom by a factor of $N_{\text{atoms}}/N_{\text{nodes}} \sim \mathcal{O}(10)$ and giving us a three dimensional bulk elasticity model; and b) by averaging through the thickness of the virus capsid, thus giving us a two dimensional shell elasticity model and a reduction in the degrees of freedom of the order of $N_{\text{atoms}}/N_{\text{nodes}} \sim \mathcal{O}(100)$.

While calculating the center of masses, each amino acid can be assigned a single node or multiple amino acids can be grouped together to create one node. When combining multiple amino acids, those next to each other on the protein sequence are considered neighbors and thus merged together. Here, we consider cases of two, three, and six amino acids grouped together, in addition to one node per amino acid. To create 2D shell model by averaging through thickness, the averaging zones (of conical shape) are obtained from subdivision of an icosahedral net. The number of subdivisions can be set as desired. Thus, this step of defining coarse-grained nodes provides us with the flexibility of choosing a desired level of coarse-graining, which we will exploit later. The essential idea in this step is to extract continuum-scale deformation metrics at each time frame of the MD trajectory that can be used in the elastic energy expressions (5,6). From the trajectory of deformation metrics, free energy of the system is derived using a potential of mean force, which is described next.

D. Potential of Mean Force

Originally proposed by Kirkwood to evaluate chemical potentials of fluid mixtures [33], potential of mean force is based upon Boltzmann's law. In thermodynamic equilibrium, the system fluctuates around its reference configuration and explores all of the available microstates. These microstates of the system obey Boltzmann's law, which gives the local probability distribution of states, defined in terms of deformation measure $\mathcal{D}$, at point $\boldsymbol{X}$ as

$$p(\mathcal{D}; \boldsymbol{X}) = \frac{1}{\mathcal{Z}} \exp\left(-\frac{\mathcal{U}(\mathcal{D}; \boldsymbol{X})}{k_{\text{B}} T} \right), \tag{8}$$

where $\mathcal{U}(\mathcal{D}; \boldsymbol{X})$ is internal energy, $\mathcal{Z}(\boldsymbol{X}) = \int d\mathcal{D} \exp\left[-\mathcal{U}(\mathcal{D}; \boldsymbol{X}) / k_{\text{B}} T \right]$ is the partition function, and k_{B} and T are Boltzmann's constant and temperature respectively. The deformation measure $\mathcal{D}$ represents the set of deformation metrics chosen to represent the system. Specifically, it represents the strain tensor $\mathcal{D} = \{\mathbf{C}\}$ in the case of 3D bulk elasticity, and combination of strain and curvature tensors $\mathcal{D} = \{\mathbf{C}, \mathbf{B}\}$ in the case of 2D shell elasticity. Or, if isotropic assumption is

made beforehand, $\mathcal{D} = \left\{J, \bar{I}_1, \bar{I}_2\right\}$ in the case of 3D bulk elasticity, and $\mathcal{D} = \left\{J^{\mathrm{2D}}, \bar{I}_1^{\mathrm{2D}}, H\right\}$ in the case of 2D shell elasticity. Therefore, we introduce common symbols for the set of deformation measures $\mathcal{D} = \{\mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3\}$.

Inverting Eq. (8), we obtain potential of mean force representing the effective internal energy or, in the present context, the elastic strain energy function

$$\mathcal{U}\left(\mathcal{D}; \boldsymbol{X}\right) = \mathcal{U}_0\left(\boldsymbol{X}\right) - k_{\mathrm{B}}T \log p\left(\mathcal{D}; \boldsymbol{X}\right). \tag{9}$$

Here $\mathcal{U}_0\left(\boldsymbol{X}\right) = -k_{\mathrm{B}}T \log \mathcal{Z}\left(\boldsymbol{X}\right)$ is the reference energy. Since the reference energy value does not affect the elastic moduli of the system, we do not need to calculate the partition function in this framework. Analogously, one does not need to normalize the probability distribution and can directly work with the number of frames in MD. Assuming that the deformation measures are statistically independent at each point $\boldsymbol{X}$, we obtain $p\left(\mathcal{D}\right) = p\left(\mathcal{D}_1\right) p\left(\mathcal{D}_2\right) p\left(\mathcal{D}_3\right)$ which decouples the energy in Eq. (9). Hence, in discrete form, the total energy obtained from potential of mean force is simply a sum over all nodes/triangles

$$\mathcal{U}\left(\mathcal{D}\right) = \sum_I \left[\mathcal{U}_0\left(\boldsymbol{X}_I\right) - \sum_{i=1}^{3} k_{\mathrm{B}}T \log p\left(\mathcal{D}_i(\boldsymbol{X}_I)\right)\right]. \tag{10}$$

Comparing with continuum models Eqs. (5, 6), we observe that the continuum strain energy densities W for our two models can be decoupled in strain invariants and written as sum of three parts:

$$\text{3D: } \mathcal{U}\left(\mathbf{C}\right) = \overline{\mathcal{U}}_0 + \sum_I \left[W_I^1(\bar{I}_1) + W_I^2(\bar{I}_2) + W_I^3(J)\right] V_I$$

$$\text{2D: } \mathcal{U}\left(\mathbf{C}, \mathbf{B}\right) = \overline{\mathcal{U}}_0 + \sum_M \Big[W_M^{\mathrm{b}}(H) + W_M^{\mathrm{s},1}(\bar{I}_1^{\mathrm{2D}}) + W_M^{\mathrm{s},2}(J^{\mathrm{2D}})\Big] \mathcal{S}_M.$$

In common notation, we write the continuum energy

$$\mathcal{U}\left(\mathcal{D}\right) = \overline{\mathcal{U}}_0 + \sum_I \sum_{i=1}^{3} W_I^i(\mathcal{D}_i) V_I. \tag{11}$$

Comparing the potential of mean force energy Eq. (10) and continuum energy Eq. (11), we observe that the strain energy density function $W_I^i(\mathcal{D}_i)V_I \simeq -k_{\mathrm{B}}T \log p\left(\mathcal{D}_i(\boldsymbol{X}_I)\right)$. Furthermore, given that we are calculating fluctuations around the equilibrium, we consider only the lowest order functions for the strain energies. Therefore, there is one elastic modulus $\mathcal{M}_i$ associated with each of the deformation measure $\mathcal{D}_i$, and collectively denoted as $\mathcal{M} = \{\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3\}$. These elastic constants are derived from the probability distribution of the corresponding strain invariant by fitting the lowest order polynomial. That is, $-\log p\left(\mathcal{D}_i\right) \approx m_i \left(\mathcal{D}_i - \mathcal{D}_i^0\right)^{\alpha} + b_i$, where $\mathcal{D}_i^0$ is the *expected* minimum value of $\mathcal{D}_i$, α is the lowest polynomial order, and m_i and b_i are the fitting parameters. With these notations, the elastic constants are $\mathcal{M}_i^I = -\frac{k_{\mathrm{B}}T m_i}{V_I}$ for 3D and $\mathcal{M}_i^I = -\frac{k_{\mathrm{B}}T m_i}{\mathcal{S}_I}$ for 2D.

E. Combining Equivalent Nodes

Under homogeneous assumption, one might consider a set of nodes $\mathcal{I} = \left\{I^1, I^2, \ldots, I^n\right\}$ being equivalent. In that case, we combine the probabilities of those nodes and calculate the potential of mean force. This improves the statistics as well as ensures that the resulting moduli of equivalent nodes come out to be identical. Therefore, the relation for elastic moduli becomes $\mathcal{M}_i^{\mathcal{I}} = -\frac{k_{\mathrm{B}}T m_i}{\sum_{\mathcal{I}} V_I}$.

F. Exploiting Icosahedral Symmetry

If a macromolecule is symmetric, its trajectory should also follow the same symmetry as the observation time increases. Hence, the icosahedral symmetry of the spherical virus capsids can be exploited in various ways. One way is to simulate the full capsid using MD and then generate icosahedral rotations of the resulting trajectory, $\boldsymbol{x}_I^{\beta} = R^{\beta} \cdot \boldsymbol{x}_I$ for $\beta \in [1, 60]$, to obtain 60 times larger statistics dataset. A different way could be to apply icosahedral symmetry boundary conditions on a single asymmetric unit for the MD run, thus reducing the computational cost of MD. However, it is not entirely obvious that the latter technique would provide the same results as MD simulation of full capsid. Therefore, we employ the first technique for mutant SeMV capsid and leave the latter for future work. Details of the SeMV MD simulation can be found in [4]; Total MD simulation length was 28 nanoseconds with 1 femtosecond of time step and a sampling frequency of 1 picosecond. We assumed that the last 2 nanoseconds of the trajectory were equilibrium fluctuations and hence used that for the results presented. Therefore, a total of 2000 microstates were sampled, which using the symmetry of capsid provided us with an effective 120 ns long trajectory sampled with 120000 microstates.

Lastly, we need reference configuration $\boldsymbol{X}_I$ for calculating the strains in continuum models. To that end, the rigid body motion – both translational and rotational – are removed from the atomic trajectory $\boldsymbol{r}_a^{\beta}(t_i) = R^{\beta} \cdot \boldsymbol{r}_a(t_i)$ and then time averaged to obtain the average position $\left\langle \boldsymbol{r}_a^{\beta}(t_i) \right\rangle_{t,\beta}$. These average atomic coordinates are then mapped to the coarse-grained nodes to define the reference configuration $\boldsymbol{X}_I = \sum_a A_{Ia} \left\langle \boldsymbol{r}_a^{\beta}(t_i) \right\rangle_{t,\beta}$.

G. Steps

The overall flow of the framework presented here can be summarized in the following way:

1. Coarse-grained nodes are defined from the atomic coordinates using a map, resulting in either 3D or 2D model.

2. Icosahedral averaging is performed by generating 60 copies of the MD trajectory through icosahedral

rotations.

3. Deformation metrics are calculated for each frame of the MD trajectory with respect to the reference (i.e. average) configuration. The deformation metrics are defined using either 3D or 2D shell kinematics under isotropic assumption.

4. Probability distribution is calculated for the three deformation invariants at each coarse-grained node.

5. Using these probability distributions of each strain invariant and Boltzmann's inversion (9), the potential of mean force is calculated for each node.

6. Lowest order strain energy functions are fit to these potentials giving us an estimate of the elastic constants which vary spatially.

The results using this framework are presented next.

III. RESULTS

A. Assuming Homogeneous Elasticity

As a first step, we assume homogeneous elasticity and, thus, treat all nodes as equivalent. Therefore, we combine statistics for all nodes to a single probability. This is calculated at four different levels of coarsening – one node per residue, one node per two residues, one node per three residues and one node per six residues – all using 3D bulk elasticity model. The resulting energy density functions show a decrease in the slope (for $\bar{I}_1$ and $\bar{I}_2$) and curvature (against J) as we move away from the zero strain (Fig. 2). As the local slope and curvature are related to the instantaneous stiffness, the results show that the stiffness decreases at higher strains amounting to a "softening" behavior.

If our assumption of homogeneity was correct, the results would have been independent of the level of coarsening. However, we see a dramatic variation in the resulting energy density as we coarsen our model (Fig. 2). This indicates that the assumption of homogeneity is not justified (see Appendix A for a simplified demonstration of this result). Henceforth, we can reject the hypothesis of homogeneity and move on to the heterogeneous treatment where all the nodes are treated independently.

Furthermore, since we lumped together stiff and soft nodes, the strain energy response at lower strains is dominated by stiff nodes and that at the higher strains is dominated by the soft nodes. Therefore, the apparent softening response is only an artifact of the spurious homogeneous assumption. Nevertheless, some interesting observations can be made from these results: a) mere thermal fluctuations produce rather large strains at this scale, and b) the minimum energy coincides with the zero strain ($\bar{I}_1 = \bar{I}_2 = 3,\ J = 1$). This suggests that, globally, the choice of reference configuration is justified.

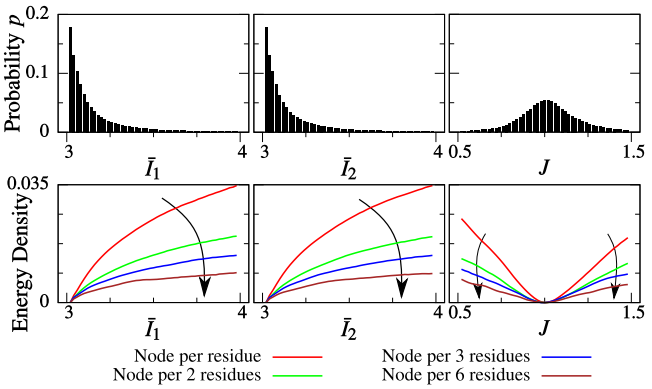

FIG. 2. (Color online) We start by hypothesizing that all the nodes have same elastic properties (i.e. homogeneous). This leads to probability distributions (top) that have a maximum at zero strains. The corresponding energy density plots (bottom) show a softening response which also varies dramatically as the model is coarsened successively.

B. 3D Bulk Elasticity

Next, we assign one node to each amino acid and consider all of the nodes as independent. It is observed that different nodes exhibit widely varying strain probability distributions indicating a highly heterogeneous behavior. As an example, Fig. 3 shows the probability distributions and effective energy terms at one chosen node of a coarse-grained model of SeMV. We observe that the energy matches quite well with the low-order Mooney-Rivlin constitutive law developed for rubber elasticity [34][35], which is of the form

$$\mathcal{U}(\mathbf{C}) = \overline{\mathcal{U}}_0 + \sum_{I=1}^{N_{\text{nodes}}} \Big[\gamma_{10}^I \left(\bar{I}_1 - 3\right) + \gamma_{01}^I \left(\bar{I}_2 - 3\right) + \delta_1^I \left(J-1\right)^2 \Big] V_I. \tag{12}$$

After least-squares fit, the slopes of $-\log p\left(\bar{I}_1\right)$ and $-\log p\left(\bar{I}_2\right)$, and the curvature of $-\log p\left(J\right)$ produce estimates of moduli γ_{10}^I, γ_{01}^I and δ_1^I at node I respectively.

Furthermore, we observe that several nodes have an energy landscape such that the lowest energy is manifested at a deformation $\mathbf{C} \neq \mathbf{I}$. The deviation of minimum energy strain from zero strain is small in most of the cases (Fig. 3), but large in others (Fig. 4). This is in contrast to our observation during homogeneous treatment (i.e., when all the nodes were considered equivalent); in that case, the system had energy minimized at zero strain. This problem of minimum energy at a non-zero strain violates the assumption that a reference configuration is well-defined and known a priori. This issue, although indicative of some problem in the analysis presented, is puzzling and will be discussed in detail in the next section.

Suppressing, for the moment, issues with the reference configuration, we fit the material constants for all nodes

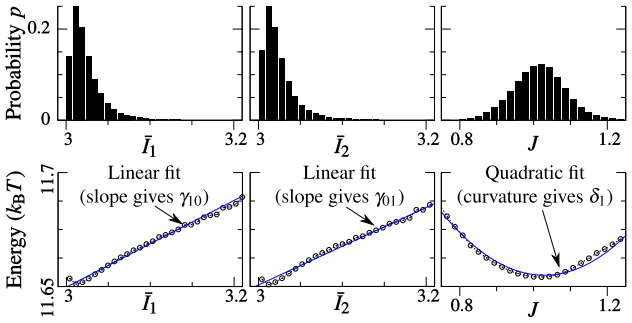

FIG. 3. (Top row) Probability distribution of strain invariants at a representative node. (Bottom row) Energy of that node in the continuum description. Points (circles) are obtained from the probability distributions in top row by (9). Solid (blue) curves are fits to the Mooney-Rivlin constitutive law (12).

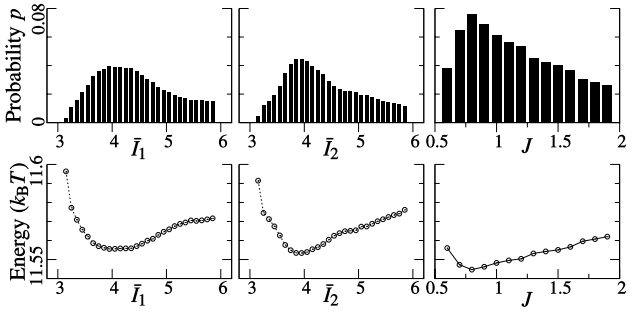

FIG. 4. (Top row) Probability distributions at an example node, where the most probable state is observed at large strains ($\bar{I}_1 \approx 4$, $\bar{I}_2 \approx 4$ and $J \approx 0.8$) severly voilating the assumption of known reference configuration. (Bottom row) Corresponding energy plots of the node.

(amino acids) in SeMV, and the resulting elastic moduli γ_{10}, γ_{01} and δ_1 are shown in Fig. 5. The range of elastic moduli are in agreement with previous experimental estimates on similar-sized virus capsids [3, 17, 36]. These results also provide other useful information: the ratio of shear and bulk modulus is approximately 10, while the values of γ_{10} and γ_{01} are comparable. In addition, all the elastic moduli are highly heterogeneous and their distributions have a large tail with a few points showing a stiffness as high as 10 times the mean value (Fig. 6). The underestimation of surface nodal volumes V_I [29] might lead to a slight overestimation of the modulus at the surface nodes. However, if we use the presented meshfree model with the calculated elastic moduli for simulating, for example, indentation of capsid, the exact energy expression is recovered for all the nodes.

Returning to the issue of reference configuration, we hypothesize that the problem is because continuum elasticity is not able to capture the atomic motions at the scale of individual amino acids. Although the reason of such a limit on the applicability of continuum elasticity is not clear yet, we test our hypothesis by further coarsening our continuum model.

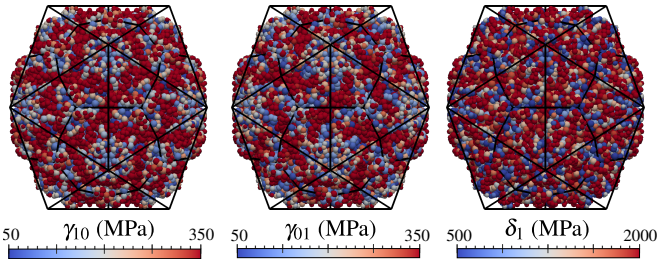

FIG. 5. (Color online) 3D bulk elasticity moduli distribution on SeMV capsid calculated using last 2 ns of full capsid MD trajectory and icosahedral symmetry imposed (with a net overlay showing the symmetry of its constituent proteins).

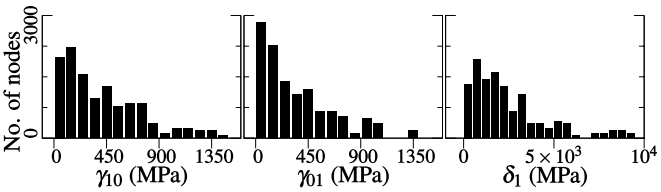

FIG. 6. Histograms of the distribution of all three elastic moduli over SeMV using 3D continuum model and full MD show a large tail. Similar distributions are obtained for all the cases.

C. Effect of Coarsening

We successively coarsen our model by choosing centers of every two, three and six residues as the continuum node. The feature of minimum energies at non-zero strain still remains prevalent. The resulting moduli are within the same range as the previous section, don't provide us with any extra information and, thus, are not shown here for brevity. To provide an intuitive idea about the difference in the analysis as we coarsen our model consider the following. A larger averaging zone leads to higher smoothening of the thermal fluctuations, thus making higher strains less probable. Using Boltzmann's law, this leads to higher energies at large strains and, thus, the energy-strain curves become steeper (Eq. 9). At the same time, larger averaging zones have a higher volume in the strain energy integration term (Eqs. 12 and 13). Hence, the elastic moduli $\mathcal{M}_i^I \propto \frac{m_i}{V_I}$ are expected to remain same upon coarsening – only "homogenized" or averaged over the amino acids that are lumped together to form the nodes. However, in our SeMV example this trend is not observed. That is, the resulting coarsened moduli here are not an average of the moduli at the associated nodes calculated previously.

If we coarsen more than one node per six residues with our 3D bulk elasticity model, the numerical calculations become erroneous. This is for two reasons – 1) the nodes become sparse, especially through the thickness, making the field approximation poor, and 2) the points become

very unevenly spaced since residues on a protein chain are not necessarily spatially close. Therefore, to analyze another level of coarsening, we turn to the 2D shell approximation. The idea is to average through the thickness and get a surface description of the capsid motion, thus eliminating those errors as we coarsen further.

D. Coarsen to 2D Shell Elasticity

As explained above, the relation between coarsening zone and volume used to derive elastic moduli is critical for obtaining consistent results. This should be kept in mind while designing the averaging procedure, because coarsening using an arbitrary smoothing parameter will make the results inconsistent. The averaging procedure used here is shown in Fig. 7: to calculate the connectivity matrix A_{Ia}, center of each amino acid is radially projected onto a subdivision of an icosahedron. The projection points a that lie within the red (light gray) colored polygon around each vertex (black color) are averaged to obtain the position of node I. This gives us a mapping that satisfies the condition that one, and only one, element of each column of the connectivity matrix can be non-zero. The average mean surface or shell representation of SeMV is thus calculated (Fig. 7b), and its limit surface is calculated using recursive Loop's subdivision (Fig. 7c). The limit surface is used to calculate the curvature tensor for shell kinematics [28]. This model provides a coarseing of approximately 18 amino acids per node on an average.

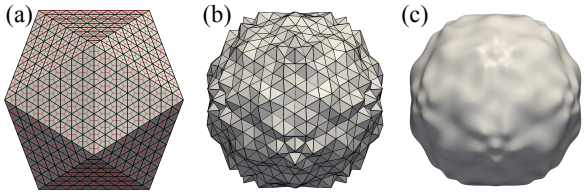

FIG. 7. (Color online) (a) Subdivision of an icosahedron used for calculating the mean surface, (b) the average mean surface and (c) its limit surface used to calculate the curvature tensor.

An example of probabilities of 2D strain invariants are shown in Fig. 8. We note that the strain energy density matches quite well with the form as reported in Evans and Skalak [37]:

$$\mathcal{U}(\mathbf{C},\mathbf{B}) = \overline{\mathcal{U}}_0 + \frac{1}{2}\sum_{M=1}^{N_{\text{triangles}}} \Big[\kappa_C^M \left(H - H_0\right)^2 + \kappa_S^M \left(J^{2\text{D}} - 1\right)^2 + \mu^M \left(\bar{I}_1^{2\text{D}} - 2\right)\Big]\mathcal{S}_M. \quad (13)$$

The mean curvature H_0 is defined as that having the highest probability. Importantly, in this case, the minimum energy corresponds to zero strains and the *apparent* problem of wrong reference configuration disappears at this scale *for all the nodes/elements*. The resulting elastic moduli distributions are shown in Fig. 9. The area modulus κ_S is small compared to shearing modulus μ. Moreover, the ratio of bending to shearing modulus $\mu/\kappa_C \approx 1/3$ nm^{-2}, which is consistent with the previous experimental estimate for spherical capsids [14].

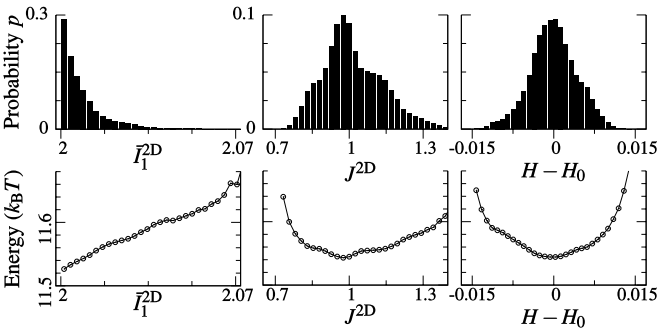

FIG. 8. Averaging through the surface provides a coarsening of ≈18 residues per node, and the problem of wrong reference configuration disappears completely.

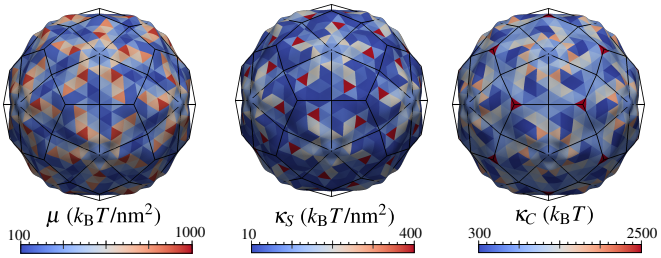

FIG. 9. (Color online) 2D shell elasticity moduli distribution on SeMV capsid calculated using last 2 ns of full capsid MD trajectory and icosahedral symmetry imposed (with a net overlay showing the symmetry of its constituent proteins).

E. Validation Using AFM Indentation

The most commonly used experimental method of determining the elastic properties of virus shells is indentation with an atomic force microscope (AFM). We simulate the AFM experiment using the calculated elastic moduli via both 3D and 2D models. Resulting force-height relationships for indenting along three different symmetry axes, which do not involve *any* scaling parameter, are plotted in Fig. 10. The 2D shell model results are consistently softer compared to the 3D bulk elasticity model. Since there are no experimental results available for SeMV, we compare the effective stiffness with that of Cowpea Chlorotic Mottle Virus (CCMV). The effective stiffness is the force divided by change in height and varies between 300–800 pN/nm for our SeMV calculations. For CCMV, the experimental results reported in [3] show an effective stiffness of 200 pN/nm, so that SeMV capsid is approximately ∼1.5–4 times stiffer. According to thin-shell theory the effective spring constant

of shell $k_{\text{shell}} \propto \frac{h^2}{R}$ if the elastic modulus is kept constant. For SeMV $R \approx 9.1$ nm, while for CCMV $R \approx 14$ nm, and both have similar thicknesses of approximately 4 nm. Thus, thin-shell theory predicts the ratio of effective stiffness of SeMV to that of CCMV to be 1.5. As stiffness of smaller SeMV capsid using 2D model is found to be higher than that of larger CCMV capsid by similar factor, present results are consistent with the experimental observations.

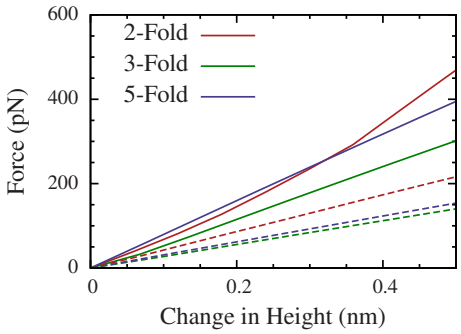

FIG. 10. (Color online) Force response of SeMV capsid to AFM indentation along three different axes. The results are calculated using non-uniform elastic properties without any scaling parameter with both 3D elasticity (solid lines) and 2D shell model (dashed lines).

IV. EFFECT OF REFERENCE CONFIGURATION

Strains and the elastic moduli calculated using the presented framework depend intrinsically on the choice of reference configuration. For the results presented in the last section, an average was performed over the equilibrium part of the trajectory to obtain the most common configuration during the MD run, and this was used as the reference state. As pointed out in the previous section, several points using the 3D model gave us an energy landscape which had a minima at non-zero strains – suggesting a wrong choice for the reference configuration. The deviation of the minima from zero strains was small at some points (Fig. 3) while large at others. One example of the latter is shown in Fig. 4 where the minimum energy is observed at rather high strains ($\bar{I}_1 \approx 4$, $\bar{I}_2 \approx 4$ and $J \approx 0.8$).

This is not only a source of error in the fitting process but a fundamental problem – suggesting that the chosen reference configuration does not depict the lowest energy state – thus violating our assumption of a well-defined reference configuration with minimum energy. If the most probable or average configuration is not the correct reference state, it is not clear what a suitable reference configuration should be or how to calculate it.

As an alternative explanation, one could speculate that the duration of MD run was not long enough to equilibrate and/or the sampling time was insufficient to obtain proper averaging of the lowest frequency modes of motion. If equilibrium is not attained, Boltzmann's law does not hold. And if the lower frequency modes of protein vibration are missing from the trajectory, then the average of the trajectory yields a biased configuration resulting in incorrect reference configuration.

On the other hand, we noted that this problem did not arise in the case when all the nodes were considered equivalent (Fig. 2). In that case, the reference configuration did appear to be a global energy minimum. Additionally, this problem did not show up in the analysis when the motions were averaged enough, at a scale of $\approx$18 amino acids or $\mathcal{O}(100)$ atoms per node, resulting in a 2D shell elasticity model (Fig. 8). In that case, the energy landscapes for all the nodes had minima at zero strains (within numerical accuracy), and, thus, the reference configuration seemed to be properly chosen. Furthermore, this same MD trajectory was used by May and Brooks [4], where they successfully calculated the homogeneous properties by comparing equilibrium fluctuations to the global spherical harmonic modes – thus indicating that the equilibrium was, indeed, achieved. All of these observations indicate that the reason of minimum energy at non-zero strains is something other than insufficient duration of MD or erroneous reference configuration. We propose an explanation for these observations using the idea of pre-strain as described next.

A. 1D Springs Example

As the most simplified case, consider two springs in parallel (Fig. 11a) with equilibrium lengths l_1 and l_2 and stiffnesses $k_1 = k_2 = k$. Under a force f the springs are stretched to a length of x, so that the elastic energy of system is $E = 1/2k(x - l_1)^2 + 1/2k(x - l_2)^2$ and the governing equation is $2kx - k(l_1 + l_2) = f$. Therefore, in the absence of external force, the equilibrium length of two springs combined together is $l_{12} = \left(\frac{l_1+l_2}{2}\right)$. When l_{12} is used as the reference length to define strain, the energy-strain curve of the two springs combined is a quadratic around zero strain (Fig. 11b). Furthermore, if we can only observe the combined system, and not isolated springs, we also choose the reference length of a single spring from the global equilibrium (i.e. l_{12}). In that case, the energy-strain curve of each spring is a quadratic around a non-zero strain $\pm\epsilon_0$ (Fig. 11b), as long as $l_1 \neq l_2$.

Extending this example, now consider a one-dimensional network of several springs arranged into n columns and m rows (Fig. 12). Springs in each column are constrained to have the same length. Also, each spring is "pre-strained" randomly around a mean length l, i.e., the zero-energy or reference length of isolated spring (i, j) is $l_{ij} = l\,(1 + \text{rand}[-0.2, 0.2])$; here, first index i is column and j is the row. Each spring has the same stiffness $k_{ij} = k$ and the energy is quadratic in

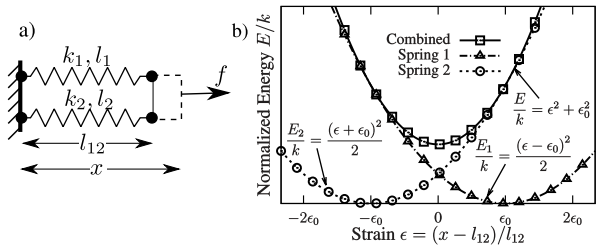

FIG. 11. Two-springs model, $\epsilon_0 = \frac{l_1 - l_2}{l_1 + l_2}$ is the pre-strain

displacement

$$E_{ij} = 1/2k\Delta x_{ij}^2 = 1/2k\left(x_i - x_{i-1} - l_{ij}\right)^2. \qquad (14)$$

Therefore, the governing equations of the system are $\sum_j k_{ij}(x_i - x_{i-1}) - \sum_j k_{ij} l_{ij} = f \; \forall i$. We solve this system of equations at various values of force f, and use the displacements thus obtained to calculate the spring energies (Eq. 14). We define strain based on the spring lengths at global equilibrium to obtain energy-strain relation for each spring. Summing up all the spring energies, we also obtain a "global" energy-strain relation. In the absence of pre-strain, or, in other words, if the zero-energy lengths of all isolated springs are the same ($l_{ij} = l$), the normalized energy-strain relation of all the springs as well as combined global system overlap – it is a quadratic relation around zero strain (Fig. 12). With the introduction of a random pre-strain, minimum of the total energy of the network remains at zero strain, but the minimum energy is non-zero positive. This is caused by the incompatibility among springs of each column due to their different lengths. In order to obtain energy-strain relation of individual spring, we note that, for a single spring, its "apparent" reference length from the global equilibrium is different from its "actual" natural length. Using the length from global equilibrium as the reference, "local" energy landscape of a single spring is shifted such that its energy is minimized at a finite non-zero strain. Furthermore, if we sum together a sufficiently large number of springs, by virtue of the random nature of pre-strain, the energy minimum returns to zero strain (the term "sufficiently large" is used because of the random function making these calculations stochastic). We draw an analogy to the results for SeMV in the last section, where sufficient averaging via the 2D shell model resulted in energy landscapes with energy minima at zero strains.

In this one-dimensional example, the pre-strain results in merely a shift in the energy landscape and the apparent stiffness of springs remain unaffected. Also, this "shift" can be rectified in the following way: once we plot the energy-strain curve of any spring, or collection of springs, we note the strain $\bar{\epsilon}$ at which energy is minimized. If we, then, use the length corresponding to $\bar{\epsilon}$ as the reference and redefine our strain using that, we will obtain strain-energy curves centered around zero strain, thus correcting the shift. However, the effect of such a pre-strain becomes much more complicated in the three-dimensional case which we investigate next.

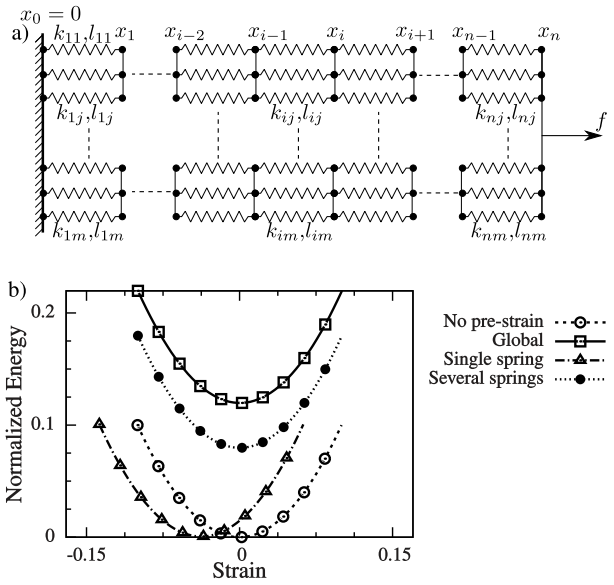

FIG. 12. (a) 1D spring network model for pre-strain and (b) its energy-strain relationship. Combining several springs is equivalent to sufficient coarse-graining that nullifies the effect of random pre-strain.

B. SeMV Spring Network

As a three-dimensional extension of the one-dimensional spring network, an elastic network model is constructed from SeMV atomic coordinates by connecting atoms i and j with a linear spring (i, j). A cut-off distance of 5 Å is used, i.e., $k_{ij} = k$ if $\|\boldsymbol{x}_i - \boldsymbol{x}_j\| < 5$ Å and $= 0$ otherwise. The reference configuration of this system is calculated by equilibrating in the absence of any applied forces or displacements – the true global minimum energy state. Thus, the reference configuration in this model is known a priori without any ambiguity.

In order to obtain strain-energy relationship, a numerical experiment akin to an extension/inflation test is performed. Although, instead of applying forces, deformation modes are excited by applying displacement boundary conditions (BCs) on the atoms on the outer most surface. Spherical harmonics with gradually increasing amplitude are used to define these displacements on the outer surface. The choice of spherical harmonics is not unique and any other boundary conditions could be used as long as they induce global displacements resulting in deformations in all the springs.

The total elastic energy of the spring network is minimized at every deformation step. In order to connect the SeMV spring network model to the 3D bulk elasticity model, for each equilibrated configuration, a mapping from atomic positions to coarse-grained nodes is performed as previously defined (Eq. 7). The nodal positions are used to define deformation metrics with respect to the

well-defined reference configuration. The energies in the continuum model are defined as a sum of spring energies; each spring's energy is divided equally into the connecting atoms, which are then summed according to the connectivity matrix A_{Ia} to obtain energy of the continuum nodes. Thereby, energy for each node as a function of the deformation metric is obtained. Without any pre-strain, the original coordinates of SeMV represent the global reference configuration and all three energy landscapes have minima have at a zero strain (Fig. 13a).

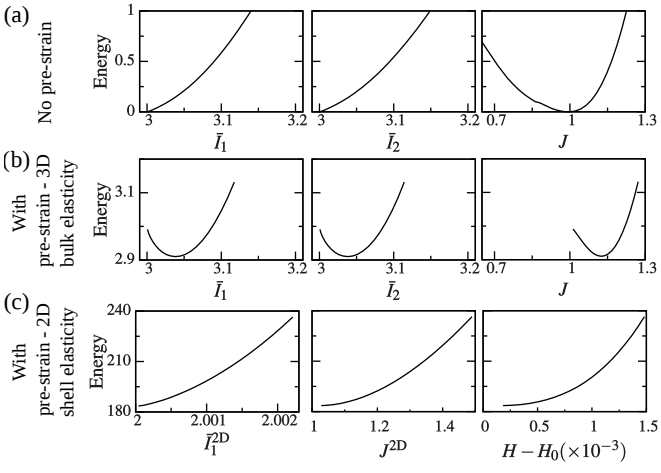

FIG. 13. Energy landscapes of SeMV elastic network (a) without pre-strain, (b) with pre-strain in 3D bulk elasticity model and (c) with pre-strain in 2D shell elasticity model (after averaging through shell thickness).

Next, a random pre-strain is introduced: $l_{ij} = \|\boldsymbol{x}_i - \boldsymbol{x}_j\|(1 + \text{rand}[-0.2, 0.2])$. Then, the above procedure of calculating strain-energy response at each coarse-grained node of 3D bulk elasticity model is repeated. The resulting energy landscapes show similar behavior as those from the SeMV MD trajectory, where many of the nodes have an energy minima at non-zero strain (example in Fig. 13b). Although this result suggests a wrong choice of reference state, we note that an unambiguous reference state (i.e. the state obtained after equilibrating without any boundary displacements or forces, so that the total energy of the system is minimized) was used. Instead, this discrepancy is a result of the incompatibility between springs caused by the pre-strain, so that even in the reference state the springs have internal forces (or stresses in continuum sense). The pre-strain at the atomic level (i.e. linear springs) manifests in the continuum model in a complicated way which can not be fixed by the method proposed in the 1D case. Moreover, a simple split of the total deformation into elastic contribution and pre-strain contribution, as used in [12], would also not solve this problem because the induced pre-strain is not at the continuum scale. The only conceivable way to fix this issue would be to consider all the springs separately and thus not make the continuum assumption in the first place. This is a novel observation which not only demonstrates a fundamental challenge for coarse-graining with a continuum model but also might have implications related to stress state of proteins.

Lastly, the continuum model is coarsened further by averaging through the capsid thickness and creating a 2D shell elasticity model. In this case, the spring energies are summed within each triangular mesh element and we observe that the strain-energy functions are minimized at zero strain again (Fig. 13c). We note an analogy to the 1D spring network case where summing up the energy and displacement of several springs nullified the pre-strain effect. Exactly the same effect was observed in our analysis of SeMV MD trajectory in the last section, where energy minima at zero-strains were observed for all the shell elements. This strongly suggests that the pre-strain model presented here is a possible explanation for our observations during analysis of MD trajectories. The significance of these findings are discussed next.

V. DISCUSSION

In this paper, we presented a general framework for calculating non-uniform elastic properties at the continuum level from the atomic equilibrium dynamics for biomolecules with the flexibility of choosing any desired scale. The prerequisite for this framework is to have an MD trajectory where equilibrium has been attained, and we utilized the potential of mean force technique for deriving strain energy as a function of deformation metrics based on their probability. We applied this framework to the SeMV virus capsid. The mutant of SeMV used here is a small virus which allowed us to run an MD simulation of the complete capsid in explicit solvent. Although the utility of the presented framework is limited by our ability to run MD simulations that reach equilibrium, in the future this method will be extended such that it can be applied to MD trajectories of a single asymmetric unit using symmetric boundary conditions. Another possibility would be using coarse-grained MD, which has much lower computational cost and can simulate larger molecular systems. As long as one can obtain thermal fluctuations of a system, the presented framework would be able to extract non-homogeneous continuum elastic properties.

The framework produced highly non-uniform elastic moduli for SeMV – both in the 3D bulk elasticity model where each residue has a different elastic modulus (Fig. 5), and in the 2D shell elasticity model where the elastic properties vary on a shell surface (Fig. 9). It also allows us to simulate AFM indentation without any scaling parameter (Fig. 10). The range of elastic moduli obtained here are in agreement with previous works on various virus capsids of similar or slightly larger sizes which estimate their Young's modulus in the 100–1000 MPa range [3, 17, 36]. Although the calculations by May and Brooks [4] resulted in a much softer elastic modulus, that discrepancy can be resolved if we consider all

the displacements instead of only radial ones (see Appendix B). The presented framework is an important technique with potential application to various biomolecular systems. Building such continuum models will provide us with insights into the elastic properties of macromolecules at various scales.

While analyzing the results, we observed another important aspect of the biomolecular systems. Continuum models derived from MD demonstrated an interesting behavior where the strain-energy curves obtained from potential of mean force calculations suggested a wrong choice reference configuration in some cases but not in others. In particular, when all the coarse-grained nodes were considered equivalent, the energy was minimum at zero strain (Fig. 2). When the nodes were considered to have independent elastic properties in the 3D model, several nodes had energy minima at a non-zero finite strain (Fig. 4). However, when we further averaged through the capsid thickness to build a 2D shell model, the strain-energy curves were back to normal, i.e. the energy was minimum at zero strain for all the elements (Fig. 8).

We presented a possible explanation of this behavior using spring network model. An elastic network from SeMV atomic coordinates was constructed, and it was deformed using displacement boundary conditions on the outer surface. Strain-energy curves were calculated directly using the spring energies corresponding to each deformed state. This was done with and without a pre-strain induced into springs. In this system, the global minimum-energy reference configuration was known *a priori* and, therefore, carried no ambiguity. When strain-energy curves were calculated in the presence of a randomly induced pre-strain, it produced exactly the same behavior as our analysis of SeMV MD trajectory (Fig. 13).

This resemblance in the behavior supports our argument that a pre-strain is the cause for energy minima at non-zero strains observed in our calculations. A pre-strain-like effect in molecular systems can be expected; each atomic bond has different characteristics, e.g., bond lengths and angles, in the presence of other atomic interactions compared to when it is isolated. Or in other words, in the global equilibrium of biomolecule, each of its atomic bonds is not in a state that would resemble the bond's isolated minimum-energy state and, thus, carry some "internal forces". Similar would be the case for non-bonded interactions. At small enough scales, this may lead to problems in applicability of continuum mechanics to macromolecules as we noticed in the results presented here. This is an important observation which warrants further discussion.

At a coarse scale, continuum mechanics is an important tool for analyzing systems. However, as we push the limit of continuum mechanics to smaller scales, new challenges emerge. One of such challenges is the knowledge about reference configuration. The results presented here indicate that in the presence of an atomic-scale pre-strain it *may* not be possible to identify a reference configuration in the continuum sense. The true reference configuration exists only if we isolate all the atomic bonds or springs, and treat them as discrete – something we see as a breakdown in the continuum nature of macromolecules. May and Brooks [4] previously proposed that macromolecules behave in a way that does not fit the elasticity theory, particularly for spherical harmonic modes (l, m) with $l = 0, 1$ and $l > 6$. $l > 6$ modes are low wavelength and equivalent to applying continuum models at the scale of amino acids, and the difficulty in applying continuum elasticity at that scale is consistent with our observation here. However, the discrepancy of modes $l = 0$ and 1 observed by May and Brooks can be resolved by using all the displacements rather than using only radial ones used in their calculations (see Appendix B). In addition, here, we also propose the idea of atomic-scale pre-strain to describe the reason behind deviation from continuum elasticity.

The results presented here also elucidate the success of isotropic elasticity in describing strongly directional interactions in protein assemblies. Most of the continuum models in literature that successfully elucidated the mechanics of viral capsids were formulated at the scale of surface motions of the shell – which, as we saw, can be correctly captured by continuum mechanics. At that scale, averaging through the thickness smears out all the anisotropic interactions, thus making isotropic treatment valid. However, this also points out the limitations of continuum models when applied to viral capsids, in specific, and macromolecules, in general. It would be unreasonable to expect correct predictions of finer scale motions, e.g., those of amino acids and atoms, using continuum models unless appropriate adjustments are made in the formulation.

It is worth pointing out that the pre-strain in this case varies randomly at the scale of atomic interactions. In the case where pre-strains lead to coordinated large scale motions, e.g., those in the conformational changes, it becomes possible to define a continuum scale pre-strain and formulate an elasticity theory. We have previously presented an elasticity theory for such a case where the pre-strains (or more appropriately termed as conformation strains in that case) were constant for each hexamer of the capsid, and only varied from hexamer to hexamer [12]. However, for the present case, it is not clear how something similar could be achieved. Specifically, there are two problems: 1) the pre-strain values are unknown. Although the strain invariants at minimum energy can be identified, the deformation gradient corresponding to those invariants is not unique. 2) As the pre-strain varies for every atomic interaction, it is not clear how to derive continuum scale pre-strain from them.

This calls for further analysis which is beyond the scope of the present manuscript. In future, we will analyze the effect of pre-strain in greater details and compare it to the MD trajectory in a statistical sense. Following the approach of [18], we will also calculate the atomic scale pre-strains from MD simulations. That will provide us

with more information for formulating a theory of heterogeneous elasticity at the scale of amino acids as well as elucidating the relation of pre-strain with conformational changes in macromolecules. In the meanwhile, in the light of current results, the applicability of continuum theory to macromolecules is limited to elucidating their coarse-scale motions and deformations.

ACKNOWLEDGMENTS

This work was partially supported by National Science Foundation (DMR-1309423 to W.S.K., MCB-1121575 to C.L.B., and DBI-0905773 to E.R.M.), American Heart Association Southwest (14POST18720037 to A.A.), and Sêr Cymru National Research Network (F28 to A.A.).

Appendix A: Artificial softening response due to spurious homogeneous assumption

In order to demonstrate the effect of the assumption of homogeneity while calculating the potential of mean force for a system which, in actuality, is heterogeneous, we start by considering a system with n linear springs of stiffnesses $k_i, i = 1, \ldots, n$. The values of these k_i are unknown, while one can observe the strain energy states of all the springs. We assume that all energies and spring constants have been normalized with respect to the thermal energy $k_{\mathrm{B}}T$, thus, simplifying the governing equations. Following the Boltzmann's law, under thermal energy, each spring experiences strain ϵ with a probability

$$p(\epsilon) \propto e^{-k_i \epsilon^2/2}. \tag{A1}$$

Therefore, if we observe a finite (but large) number of states, the number of states with strain ϵ in spring i is

$$\phi(\epsilon) = \Phi e^{-k_i \epsilon^2/2}, \tag{A2}$$

where Φ is the proportionality constant. If we a priori assume that springs have the same stiffness (which may or may not be correct), we sum the number of states of strain ϵ from all of the springs

$$\phi(\epsilon) = \Phi \sum_{i=1}^{n} e^{-k_i \epsilon^2/2} \tag{A3}$$

and equate it to the "effective" energy under the homogeneous treatment (which should be the same for each spring)

$$\phi(\epsilon) = \Phi \sum_{i=1}^{n} e^{-k_i \epsilon^2/2} = \Phi \sum_{i=1}^{n} e^{-E_{\mathrm{eff}}} = \Phi n e^{-E_{\mathrm{eff}}}. \tag{A4}$$

Therefore, the effective energy in this case is given by

$$E_{\mathrm{eff}}(\epsilon) = -\log \left(\frac{\sum_{i=1}^{n} \exp\left(-k_i \epsilon^2/2\right)}{n} \right). \tag{A5}$$

Next, we add the strain energies of every two consecutive springs before applying potential of mean force (which is similar to coarsening the continuum model) while keeping the homogeneous assumption. For two springs with stiffness k_1 and k_2 in series, their energy is

$$E_{12} = \frac{k_1 k_2}{k_1 + k_2} (2\epsilon)^2/2 = 2\frac{k_1 k_2}{k_1 + k_2} \epsilon^2. \tag{A6}$$

Therefore, in this case the effective energy of *a single spring* under homogeneous assumption is given by (assuming n is even):

$$\overline{E}_{\mathrm{eff}}(\epsilon) = -\frac{1}{2} \log \left(\frac{\sum_{j=1}^{n/2} \exp\left(-2\frac{k_{2j-1}k_{2j}}{k_{2j-1}+k_{2j}} \epsilon^2\right)}{n/2} \right). \tag{A7}$$

If the homogeneous assumption is valid (i.e. $k_i = k\ \forall i$), we obtain $E_{\mathrm{eff}} = \overline{E}_{\mathrm{eff}} = k\epsilon^2/2$. However, the behavior is more complicated when the a priori homogeneity assumption is incorrect. We demonstrate this using an example of ten springs ($n = 10$) with stiffnesses $k_i = i^2$ (Fig. 14). It is clear that the wrong homogeneous assumption leads to a softening behavior and change of effective energy when we combine springs (i.e. coarsen).

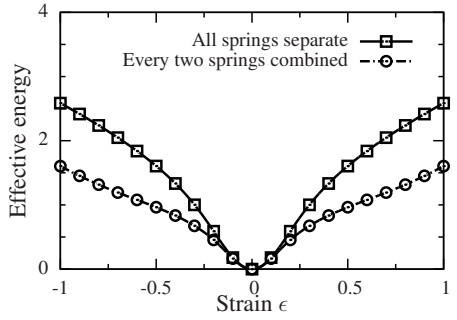

FIG. 14. The effective energy plot of a series of springs under spurious homogeneous assumption at two levels: when all springs are considered separately and when every two consecutive springs are combined. We observe the same softening behavior as we saw in the homogenization of SeMV.

Appendix B: Spherical Harmonic Decomposition of Elastic Shell

May and Brooks [4] decomposed the fluctuations of SeMV virus capsids obtained from molecular dynamics simulation into spherical harmonics. When compared to the analytical solution for a spherical shell with radial-only displacements, it was found that the first two modes $l = 0, 1$ cannot be matched. Here we show that this can be resolved by allowing all displacements for the spherical shell. Since in the case where non-radial displacements are allowed, analytical solution for spherical harmonics

is not possible, we turn to numerical solution using finite element method.

Energy of a thin elastic shell can be written as the sum of bending and stretching energies

$$\mathcal{U} = \frac{1}{2}\int_{\mathcal{S}} d\mathcal{S}\Big[\kappa_C(H-H_0)^2+\kappa_S(J^{2\mathrm{D}}-1)^2+\mu\Big(\frac{\mathrm{tr}\,\mathbf{C}}{J^{2\mathrm{D}}}-2\Big)\Big]. \tag{B1}$$

Here the symbols have the same meaning as in the main text. Using finite element approximation, the energy form can be discretized and linearized to the quadratic form $\mathcal{U} = \frac{1}{2}\boldsymbol{u}^T\cdot\mathbf{K}\cdot\boldsymbol{u}$, where $\mathbf{K}$ is the Hessian of the energy functional $\mathcal{U}$, termed as the stiffness matrix, and $\boldsymbol{u} = \boldsymbol{x} - \boldsymbol{X}$ is the displacement vector. First n eigenvectors $\boldsymbol{V}_i$ and eigenvalues λ_i of the stiffness matrix $\mathbf{K}$ are calculated by solving the eigenvalue problem $\mathbf{K}\boldsymbol{V}_i = \lambda_i\boldsymbol{V}_i$. The eigenvectors satisfy the relation $\boldsymbol{V}_i\cdot\boldsymbol{V}_j = 0 \;\; \forall i \neq j$, and thus form an orthonormal set. Therefore, the average displacement can be written in terms of the eigenvectors $\langle\boldsymbol{u}\rangle = \sum_i \eta_i\boldsymbol{V}_i$. Thus, the ensemble average of the energy expression can be written as

$$\langle\mathcal{U}\rangle = \sum_i \frac{1}{2}\left(\eta_i\boldsymbol{V}_i\right)^T\cdot\mathbf{K}\cdot\left(\eta_i\boldsymbol{V}_i\right) = \sum_i \frac{1}{2}\eta_i^2\lambda_i\left(\boldsymbol{V}_i\cdot\boldsymbol{V}_i\right). \tag{B2}$$

Using the theorem of equipartition, i.e. each mode carries an energy of $k_{\mathrm{B}}T/2$, we obtain the expression for mode amplitude

$$\eta_i = \sqrt{\frac{k_{\mathrm{B}}T}{\lambda_i\left(\boldsymbol{V}_i\cdot\boldsymbol{V}_i\right)}}. \tag{B3}$$

These average displacements are then projected onto the spherical harmonic basis:

$$\langle\boldsymbol{u}\rangle = \sum_i \eta_i\boldsymbol{V}_i = \sum_{l,m} a_{lm}\boldsymbol{Y}_{lm}, \tag{B4}$$

where $\boldsymbol{Y}_{lm}$ are also orthonormal set. Therefore, the amplitudes of spherical harmonics are obtained

$$a_{lm} = \sum_{i=1}^{n}\sqrt{\frac{k_{\mathrm{B}}T}{\lambda_i\left(\boldsymbol{V}_i\cdot\boldsymbol{V}_i\right)}}\frac{|\boldsymbol{V}_i\cdot\boldsymbol{Y}_{lm}|}{\boldsymbol{Y}_{lm}\cdot\boldsymbol{Y}_{lm}}. \tag{B5}$$

For a perfect sphere of radius $R = 9$ nm, Poisson's ratio $\nu = 1/3$, shear modulus $\mu = 330\ k_{\mathrm{B}}T/\mathrm{nm}^2$, bending modulus $\kappa_C = 990\ k_{\mathrm{B}}T$, discretized with 4131 nodes and 8258 triangular finite elements (C^0 Lagrange interpolation for the stretching part and C^1 Loop's subdivision for the bending part), the amplitude of spherical harmonics $|a_l|^2 = \sum_m a_{lm}^2$ are plotted in Fig. 15.

For the case, where we consider only radial displacements, an analytical expression for the amplitude of the spherical harmonics was derived by May and Brooks [4]

$$\left\langle|a_l|^2\right\rangle = \frac{k_{\mathrm{B}}T}{8\mu\frac{1+\nu}{1-\nu} + \kappa_C\frac{l(l-1)(l+1)(l+2)}{R^2}}. \tag{B6}$$

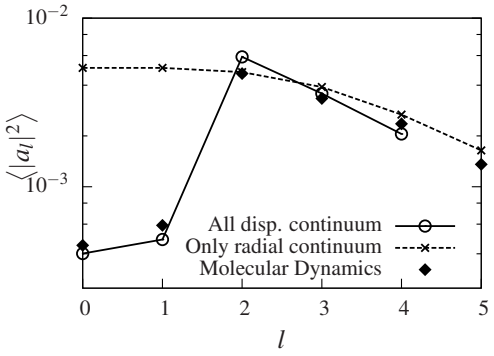

FIG. 15. Average amplitude of spherical harmonic decomposition for a spherical shell with all the displacements allowed (solid line – eq. B5) and with only radial displacements allowed (dashed line – eq. B6).

This case is plotted with the values obtained by May and Brooks when they fitted the modes $l = 2, \ldots, 6$ to the MD trajectory ($\mu = 13.275\ k_{\mathrm{B}}T/\mathrm{nm}^2$, $\nu = 1/3$ and $\kappa_C = 39.7\ k_{\mathrm{B}}T$). From these results it is clear that when we include the non-radial displacements, the spherical harmonics for $l = 0, 1$ can also be matched to the MD trajectory. It should be noted that the elastic parameters for all displacements case were chosen roughly and were not fit systematically to the MD data. Thus, the material parameters here do not represent exact fitted values. The results here are presented only to demonstrate that it is possible to capture the large scale spherical harmonic modes using continuum elasticity. Furthermore, even though the exact values of elastic constants κ_C and μ obtained using the two approaches are different, their ratio is conserved whether only radial or all displacements are considered. May and Brooks have used the calculated ratio of these elastic properties based upon radial displacements to predict the shapes of virus capsids and these results would not be effected [38].

[1] A. K. Dunker, I. Silman, V. N. Uversky, and J. L. Sussman, Current opinion in structural biology **18**, 756 (2008).

[2] A. J. Roberts, T. Kon, P. J. Knight, K. Sutoh, and S. A. Burgess, Nature Reviews Molecular Cell Biology **14**, 713 (2013).

[3] J. P. Michel, I. L. Ivanovska, M. M. Gibbons, W. S. Klug, C. M. Knobler, G. J. L. Wuite, and C. F. Schmidt, PNAS **103**, 6184 (2006).

[4] E. R. May and C. L. Brooks III, Physical Review Letters **106**, 188101 (2011).

[5] F. Tama and C. L. Brooks III, Journal of molecular biology **345**, 299 (2005).

[6] A. R. Atilgan, S. R. Durell, R. L. Jernigan, M. C. Demirel, O. Keskin, and I. Bahar, Biophysical Journal **80**, 505 (2001).

[7] E. R. May, J. Feng, and C. L. Brooks, Biophysical journal **102**, 606 (2012).

[8] O. Kononova, J. Snijder, M. Brasch, J. Cornelissen, R. I. Dima, K. A. Marx, G. J. Wuite, W. H. Roos, and V. Barsegov, Biophysical journal **105**, 1893 (2013).

[9] K. J. Boyd, P. Bansal, J. Feng, and E. R. May, Frontiers in bioengineering and biotechnology **3** (2015).

[10] M. Bathe, Proteins: Structure, Function, and Bioinformatics **70**, 1595 (2008).

[11] M. M. Gibbons and W. S. Klug, Biophysical journal **95**, 3640 (2008).

[12] A. Aggarwal, J. Rudnick, R. F. Bruinsma, and W. S. Klug, Phys. Rev. Lett. **109**, 148102 (2012).

[13] W. S. Klug, R. F. Bruinsma, J.-P. Michel, C. M. Knobler, I. L. Ivanovska, C. F. Schmidt, and G. J. L. Wuite, Phys. Rev. Lett. **97**, 228101 (2006).

[14] J. Lidmar, L. Mirny, and D. R. Nelson, Physical Review E **68**, 051910 (2003).

[15] E. R. May, A. Aggarwal, W. S. Klug, and C. L. Brooks III, Biophysical Journal **100**, L59 (2011).

[16] M. M. Gibbons and W. S. Klug, Physical Review E **75**, 031901 (2007).

[17] W. H. Roos, M. M. Gibbons, A. Arkhipov, C. Uetrecht, N. R. Watts, P. T. Wingfield, A. C. Steven, A. Heck, K. Schulten, W. S. Klug, *et al.*, Biophysical journal **99**, 1175 (2010).

[18] S. A. Edwards, J. Wagner, and F. Gräter, PLoS Comput Biol **8**, e1002509 (2012).

[19] M. Widom, J. Lidmar, and D. R. Nelson, Physical Review E **76**, 031911 (2007).

[20] E. H. Yong, D. R. Nelson, and L. Mahadevan, Physical review letters **111**, 177801 (2013).

[21] C. M. Shepherd, I. A. Borelli, G. Lander, P. Natarajan, V. Siddavanahalli, C. Bajaj, J. E. Johnson, C. L. Brooks, and V. S. Reddy, Nucleic acids research **34**, D386 (2006).

[22] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin, Journal of computational chemistry **25**, 1605 (2004).

[23] M. Zink and H. Grubmüller, Biophysical journal **96**, 1350 (2009).

[24] F. Feng and W. S. Klug, Journal of Computational Physics **220**, 394 (2006).

[25] J. S. Chen, W. Han, Y. You, and X. Meng, International Journal for Numerical Methods in Engineering **56**, 935 (2003).

[26] A. Aggarwal, J.-S. Chen, and W. S. Klug, Proceedings of EMI Conference (2011).

[27] A. Aggarwal, J.-S. Chen, and W. S. Klug, Computer Modeling in Engineering & Sciences **98**, 69 (2014).

[28] F. Cirak and M. Ortiz, International Journal for Numerical Methods in Engineering **51**, 813 (2001).

[29] After generating Delaunay tessellation of node cloud, the tetrahedral elements outside the solvent excluded surface (SES) are removed. Then, the volume of each tetrahedron is assigned equally to each of its four vertices. Summing over all the elements we obtain the nodal volume V_I. Since we do not account for the volume between the nodes on the surface and its SES, V_I of surface nodes is slightly underestimated.

[30] J.-S. Chen, C.-T. Wu, S. Yoon, and Y. You, International journal for numerical methods in engineering **50**, 435 (2001).

[31] J. Dolbow and T. Belytschko, Computational mechanics **23**, 219 (1999).

[32] In the results presented here, we did not use mass weighting. However, we excluded the hydrogen atoms while calculating the coarse-grained node positions. In practice, we did not observe any significant differences in the node positions and the final results when we used mass weighted mapping. This can be expected as majority of the atoms (carbon, oxygen and nitrogen) are closely similar in their weights, and weight of hydrogen is negligible in comparison.

[33] J. G. Kirkwood, The Journal of Chemical Physics **3**, 300 (1935).

[34] L. R. G. Treloar, H. G. Hopkins, R. S. Rivlin, and J. M. Ball, Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences **351**, 301 (1976).

[35] Conventionally, C_{10}, C_{01}, and D_1 symbols are used for elastic constants in the Mooney-Rivlin law. However to avoid confusion with deformation tensor $\mathbf{C}$ and deformation measure set $\mathcal{D}_1$, we have used Greek equivalents – $C_{10} \rightarrow \gamma_{10}$, $C_{01} \rightarrow \gamma_{01}$ and $D_1 \rightarrow \delta_1$.

[36] N. Kol, M. Gladnikoff, D. Barlam, R. Z. Shneck, A. Rein, and I. Rousso, Biophysical Journal **91**, 767 (2006).

[37] E. A. Evans and R. Skalak, CRC critical reviews in bioengineering **3**, 181 (1979).

[38] E. R. May and C. L. Brooks III, The journal of physical chemistry B **116**, 8604 (2012).