Bend-twist-stretch model for coarse elastic network simulation of biomolecular motion

Joseph N. Stember and Willy Wriggers[a)]
Department of Physiology and Biophysics and Institute for Computational Biomedicine, Weill Medical College of Cornell University, 1300 York Ave., New York, New York 10065, USA

(Received 24 April 2009; accepted 30 May 2009; published online 20 August 2009)

The empirical harmonic potential function of elastic network models (ENMs) is augmented by three- and four-body interactions as well as by a parameter-free connection rule. In the new bend-twist-stretch (BTS) model the complexity of the parametrization is shifted from the spatial level of detail to the potential function, enabling an arbitrary coarse graining of the network. Compared to distance cutoff-based Hookean springs, the approach yields a more stable parametrization of coarse-grained ENMs for biomolecular dynamics. Traditional ENMs give rise to unbounded zero-frequency vibrations when (pseudo)atoms are connected to fewer than three neighbors. A large cutoff is therefore chosen in an ENM (about twice the average nearest-neighbor distance), resulting in many false-positive connections that reduce the spatial detail that can be resolved. More importantly, the required three-neighbor connectedness also limits the coarse graining, i.e., the network must be dense, even in the case of low-resolution structures that exhibit few spatial features. The new BTS model achieves such coarse graining by extending the ENM potential to include three-and four-atom interactions (bending and twisting, respectively) in addition to the traditional two-atom stretching. Thus, the BTS model enables reliable modeling of any three-dimensional graph irrespective of the atom connectedness. The additional potential terms were parametrized using continuum elastic theory of elastic rods, and the distance cutoff was replaced by a competitive Hebb connection rule, setting all free parameters in the model. We validate the approach on a carbon-alpha representation of adenylate kinase and illustrate its use with electron microscopy maps of *E. coli* RNA polymerase, *E. coli* ribosome, and eukaryotic chaperonin containing T-complex polypeptide 1, which were difficult to model with traditional ENMs. For adenylate kinase, we find excellent reproduction (>90% overlap) of the ENM modes and *B* factors when BTS is applied to the carbon-alpha representation as well as to coarser descriptions. For the volumetric maps, coarse BTS yields similar motions (70%–90% overlap) to those obtained from significantly denser representations with ENM. Our Python-based algorithms of ENM and BTS implementations are freely available. © *2009 American Institute of Physics*.
[DOI: 10.1063/1.3167410]

I. INTRODUCTION

In recent years elastic network models (ENMs) have been widely needed for mechanistic predictions of large-scale biomolecular motions.[1] Rooted in the observation that many biomolecules behave, more than expected, as if the energy surface were parabolic, the basic assumption (and limitation) of the vibrational analysis afforded by ENMs is that the potential energy of the system varies quadratically about a single minimum energy conformation. Although first principles of physics and chemistry dictate that such energy landscapes contain many local minima,[2] the harmonic approximation of biomolecular motion seems to be in excellent agreement with many observations of structural polymorphism, where domains "move in relation to one another with only small expenditures of energy."[3]

The representation of this type of motion by vibrational analysis from diagonalization of a Hessian matrix can be traced back more than a hundred years to the harmonic analysis popular in Europe of the 19th and 20th centuries.[4] Vibrational modes of chemical molecules became a focus in the 1950s with the advent of spectroscopy[5] and have been established in computational biology in the early 1980s (Refs. 6–8) under the term "normal mode analysis" (NMA). This methodology of NMA has been discussed in many excellent textbooks and reviews[9,10] and has spurred on the more recent efforts to extend the method to large biomolecular structures by coarse graining. A first step in the simplification of NMA is the replacement of the chemical details of the physics-based atomic force field by a simplified Hookean spring potential between neighboring atoms. This approach, pioneered by Tirion,[11] showed that low-frequency modes depend more on the global character of the deformations than on the precise form of the potential. This realization opened the door to entirely empirical models that reduce the amount of spatial detail by using the simplified harmonic interaction force field developed by Tirion. Excellent agreement with experimental temperature factors obtained with a harmonic

[a)]Permanent address: D. E. Shaw Research, 120 W. 45th St., New York, NY 10036, USA. Electronic mail: willy.wriggers@deshawresearch.com.

model including only the C_α positions[12,13] or with even sparser models[14,15] suggests that it is the global shape and not the resolution or local detail of the force field that governs the low-frequency modes observed with NMA.[16,17]

As illustrated by the accompanying ENM software, basic solvers for coarse models at C_α level of detail or below can be implemented using the Tirion potential with only about a hundred lines of code. The simple implementation lends itself well to integration within a web interface and several popular servers are available.[18,19] It is therefore not surprising that the accessibility of the approach fueled a significant rise in applications in recent years.[1] Known domain movements of large biomolecular machines such as the ribosome, chaperonin CCT [chaperonin containing T-complex polypeptide 1 (TCP1)], and RNA polymerase have been found to be well reproduced in the lowest-frequency normal modes.[20]

Although ENMs have been very successful, the quality of their modes and the predictive power are sometimes limited. Some general limitations of harmonic analysis such as arbitrary amplitudes of the predicted directional motions or the uncertain significance of the mode order have been discussed elsewhere.[20] In the following we address two particular limitations of the functional form of the potential that limit the practical capacity of ENMs for very coarse models of large biomolecular assemblies: (1) the dense oversampling required to capture modes reliably and (2) the distance-dependent step function for assigning the connectedness of "pseudoatoms." Our model enables the simulation of "molecular stick figures" and in fact of any connected three-dimensional (3D) graph by shifting complexity from the spatial resolution of the coarse graining to that of the potential function. We also eliminate the arbitrary distance cutoff of the Tirion potential by automatically assigning the connectedness of nearest neighbor pseudoatoms using computational geometry considerations. The coarse graining of ENMs was introduced before,[14,15,20,21] but to our knowledge this paper is the first that modifies the Tirion functional form based on concepts from continuum elastic theory and neurocomputing.

Our work is motivated by the fact that many systems of biological interest are not amenable to crystallization and are too large for solution state nuclear magnetic resonance analysis. This leaves a vast array of biological systems that escape any kind of mechanistic elucidation. Lower resolution structures, however, are often obtainable for such systems with a variety of biophysical techniques. Cryogenic electron microscopy (cryo-EM) and small angle scattering are two prominent low resolution methods.[22,23] They are complementary in that the former is appropriate for large ($>1/3$ mega-Dalton) biomolecular systems while the latter is limited to smaller systems that can be suspended in solution without aggregation. Coarse-grained NMA can partially bridge the gap from missing experimental structures to functional interpretation in that it provides a lowest level approximation for biomolecular motion. However, ENMs are unforgiving in the presence of false negatives in the connectivity, giving rise to unbounded zero-frequency vibrations when (pseudo)atoms are connected to fewer than three neighbors. In practical applications a long cutoff distance and a dense sampling must be chosen, leading to models comprised of hundreds or thousands of atoms and whose spatial density is one to two orders of magnitude above the nominal resolution of the data (see, for example, Fig. 1a in Ref. 20). To ameliorate such limitations to the desired coarse graining and to eliminate the three-neighbor connectedness requirement, we propose below the novel bend-twist-stretch (BTS) model. After introducing the functional form of the BTS model, we describe its parametrization as well as its validation runs on C_α and lower level of detail representations of adenylate kinase (ADK). We then test and optimize our continuum mechanics model empirically and demonstrate the method's descriptive aptitude by applying it to three cryo-EM structures that had previously been studied with the classic ENM approach. We conclude with a discussion of usability and limitations.

II. METHODS

A. Functional form of the BTS potential

We consider an undirected graph of N connected nodes. Each node corresponds to a (pseudo)atom in the coarse representation and is connected to at least one other node. We do not impose any other restrictions on the connectedness pattern. The graph yields N_r bonds, N_θ angles, and N_ϕ dihedrals, which are extracted from all connected pairs, triplets, and quadruplets of nodes as follows. The lists of unique bonds, angles, and dihedrals are straightforward to generate for a particular pattern, except care must be taken to avoid mirror duplications and cyclic subgraphs. We also count dihedrals only once per central atom pair. Triplets such as ABC and CBA or quadruplets such as ACDE, BCDE, EDCB, and EDCA are thus counted only as one angle or one dihedral, respectively. Also, cyclic patterns such as ABA and ABCA are excluded. This definition is motivated by the mechanistic role of each unique two-, three-, and four-atom interactions in the following Hookean parameterization.

Our proposed BTS potential takes the form

$$V = \sum_{i=1}^{N_r} \frac{1}{2} k_s (r_i - r_{i,\mathrm{eq}})^2 + \sum_{j=1}^{N_\theta} \frac{1}{2} k_b (\theta_j - \theta_{j,\mathrm{eq}})^2 + \sum_{k=1}^{N_\phi} \frac{1}{2} k_t (\phi_k - \phi_{k,\mathrm{eq}})^2, \quad (1)$$

where r_i is the length of bond i, θ_j is the magnitude of angle j, ϕ_k is the magnitude of dihedral k, and $r_{i,\mathrm{eq}}$, $\theta_{j,\mathrm{eq}}$, and $\phi_{k,\mathrm{eq}}$ are the corresponding equilibrium values. k_s, k_b, and k_t are the force constants for bond stretching, angle bending, and dihedral twisting, respectively. These force constants will be defined in Sec. II B.

Having a potential function in hand, the next step is calculate the Hessian matrix, whose ijth element is given by

$$H_{ij} = \frac{\partial^2 V(\boldsymbol{q}_{\mathrm{eq}})}{\partial q_i \, \partial q_j} = \frac{\partial^2 V(\boldsymbol{q}_{\mathrm{eq}})}{\partial q_j \, \partial q_i}, \quad (2)$$

where $\boldsymbol{q}_{\mathrm{eq}} = \{q_{1,\mathrm{eq}}, q_{2,\mathrm{eq}}, \ldots, q_{3N,\mathrm{eq}}\}$ is the equilibrium configuration. In our current implementation of the algorithm we compute second derivatives numerically according to the central difference approximation

$$\frac{\partial^2 V(\boldsymbol{q}_{\mathrm{eq}})}{\partial q_i\,\partial q_j} \approx \frac{1}{(2\Delta q^2)}\{V(\boldsymbol{q}_{\mathrm{eq}}+\boldsymbol{e}_i\Delta q+\boldsymbol{e}_j\Delta q) - V(\boldsymbol{q}_{\mathrm{eq}}+\boldsymbol{e}_i\Delta q) - V(\boldsymbol{q}_{\mathrm{eq}}+\boldsymbol{e}_j\Delta q) + V(\boldsymbol{q}_{\mathrm{eq}}-\boldsymbol{e}_i\Delta q-\boldsymbol{e}_j\Delta q) - V(\boldsymbol{q}_{\mathrm{eq}}-\boldsymbol{e}_i\Delta q) - V(\boldsymbol{q}_{\mathrm{eq}}-\boldsymbol{e}_j\Delta q)\}. \tag{3}$$

Here $\Delta q \ll 1$ Å is a small displacement and $\boldsymbol{e}_k$ is the kth direction unit vector. We note that the usual $2V(\boldsymbol{q}_{\mathrm{eq}})$ term, being identically zero, has vanished.

Solving for the eigensystem of H, the eigenvectors represent normal mode directions, and the $3N$ eigenvalues are the normal mode frequencies squared.[24] When ordered by ascending frequency/eigenvalues, the first six modes are the zero frequency center of mass translations and rotations. Of the $3N-6$ nontrivial modes, the initial lowest-frequency modes represent the dominant motion of functional significance.

All calculations are implemented in PYTHON 2.5 with NUMPY version 1.0.4. Due to our finite-difference approach the vast majority of BTS computing time is devoted to calculating Hessian elements. To improve speed, we employ a hash table that assigns to each degree of freedom i only those two-, three-, and four-body interactions that are affected by i. We denote the sets of indices of the bonds, angles, and dihedrals that include i by hr_i, $h\theta_i$, and $h\phi_i$, respectively. One may then compute the terms in Eq. (1) as sums over only the interactions that are altered when degrees of freedom i, j, or both shift in value, since all other contributions to the potential are zero. For instance, the term $V(\boldsymbol{q}_{\mathrm{eq}}+\boldsymbol{e}_i\Delta q)$ is computed as

$$V(\boldsymbol{q}_{\mathrm{eq}}+\boldsymbol{e}_i\Delta q) = \sum_{l\in hr_i}\frac{1}{2}k_s(r_l-r_{l,\mathrm{eq}})^2 + \sum_{m\in h\theta_i}\frac{1}{2}k_b(\theta_m-\theta_{m,\mathrm{eq}})^2 + \sum_{n\in h\phi_i}\frac{1}{2}k_t(\phi_n-\phi_{n,\mathrm{eq}})^2. \tag{4}$$

B. Determination of force constant ratios

To determine suitable ratios of the force constants k_s, k_b, and k_t, we turn to the theory of continuum mechanics.[25,26] An important consideration is the choice of mechanistic model for a graph. Since we wish to keep our method generally applicable to any arbitrary graph we cannot make assumptions about the underlying biological structure. Instead we choose to model each of the graph edges as an elastic rod. Elastic rods provide straightforward physical interpretations for stretching and torsional twist, and angle changes can be modeled efficiently in the form of rod bending. In the following we assume an elastic circular rod (cylinder) consisting of a homogeneous isotropic material.

For a radius r and length L, the energy of stretching the rod along its length[26] is given by

$$E_s = \left(\frac{Y\pi r^2}{2}\right)\left(\frac{1}{L}\right)(\Delta L)^2, \tag{5}$$

where Y is the Young modulus of the material. Equation (5) implies that we may extract the effective stretching force constant as $k_s = Y\pi r^2/L$.

The energy required to bend the rod to a radius of curvature R (Ref. 25) is given by

$$E_b = \left(\frac{B}{2}\right)L\left(\frac{1}{R^2}\right), \tag{6}$$

where $B=Y(\pi r^4/4)$ is the bending stiffness.[25] We note that the angle subtended by the bent rod is related to the radius of curvature by $\theta = S/R$, where S is the arclength of the portion of a circle of radius R that corresponds to the bent rod. We can see that $S=L$, so that

$$E_b = \left(\frac{B}{2}\right)\left(\frac{1}{L}\right)\theta^2. \tag{7}$$

Hence, we can extract the bending force constant as $k_b = B/L$.

Finally we turn to the torsional motion. Assuming that the volume stays constant, the torsional energy is given by

$$E_t = \left(\frac{C}{2}\right)\left(\frac{1}{L}\right)\phi^2, \tag{8}$$

where $C=Y(\pi r^4/6)$ is the torsional stiffness.[25,27] We extract $k_t = C/L$.

We take the twist to bend ratio to cancel the undetermined modulus Y and the rod radius r,

$$\frac{k_t}{k_b} = \frac{C}{B} = \frac{\frac{1}{6}Y\pi r^4}{\frac{1}{4}Y\pi r^4} = \frac{2}{3}. \tag{9}$$

We note that for realistic biological "materials" this estimate of the twist to bend ratio is only approximate to the order of magnitude. For example, some experimental measurements of actin filament rigidity yield $B>C$, while others have given $B<C$ with all the measured values of B and C on the same order of magnitude.[27] We thus use $k_t/k_b=1$ for our initial calculations and validate/optimize this ratio further by comparisons with ENM results.

Taking the bend to stretch ratio we have

$$\frac{k_b}{k_s} = \frac{\frac{B}{L}}{\frac{Y\pi r^2}{L}} = \frac{B}{Y\pi r^2} = \frac{Y\frac{\pi r^4}{4}}{Y\pi r^2} = \frac{r^2}{4}. \tag{10}$$

Again, the undetermined modulus Y cancels as before, but the ratio k_b/k_s still depends on the radius of the elastic rod. As above we consider this estimate to be approximate to the order of magnitude, so for our initial calculations we make the *ad hoc* assertion that rods do not overlap, i.e., $r=\frac{1}{2}\langle d\rangle$, where $\langle d\rangle$ is the average bond distance in the graph. With this assumption we are able to tailor k_b/k_s to a particular graph through

$$\frac{k_b}{k_s} = \frac{1}{16}\langle d\rangle^2. \tag{11}$$

As before, we will validate/optimize this initial estimate further by comparisons with ENM results. One benefit of the refinement against ENM will be that we can determine the effective radius of our rod model. For the model to make physical sense the spacing $\langle d\rangle$ of the graph should form an upper bound for the rod diameter.

Finally, we arbitrarily set k_s equal to one, hence, determining k_b and k_t from the ratios. As in ENMs, we are not concerned about the overall global normalization of the spring constants since it does not affect the eigenvectors. A scaling of the solutions can be performed (after the analysis) by fitting of the crystallographic temperature factors (see below).

C. Coarse graining and connectedness

The coarse graining of biomolecular structures can be accomplished with vector quantization implemented in the program package SITUS.[28,29] Among other utilities, SITUS provides tools for generating coarse representations given an input structure, whether volumetric or atomic. The vector quantization approach[30,31] uses a so-called "neural gas" network to represent a structure by a set of N pseudoatoms. Recently, this coarse-graining approach was also implemented in the "cgtools" plugin of the popular VMD program.[32] The SITUS implementation currently differs from the VMD approach by providing an additional clustering of multiple statistically independent neural gas runs, yielding a statistically reproducible assignment of pseudoatoms. More importantly, the neural gas approach in SITUS is also able to establish the connectedness of adjacent pseudoatoms by employing the so-called "competitive Hebb" rule.[33] This idea derived from computational geometry is closely related to the well known Delaunay tessellation.[34] If necessary, a density mask is first created from an atomic structure using the SITUS "pdb2vol" utility. Then the density mask (or alternatively, volumetric data from low-resolution structures) is processed with the SITUS "qvol" tool to generate the coarse model and the connectedness in the form of PDB (Protein Data Bank) and PSF (Protein Structure File, a chemical file format containing bond information) files, respectively. The competitive Hebb rule yields a Delaunay tessellation that is masked by the input density[33] with the number N of pseudoatoms as sole input parameter. We describe in Sec. III how this number was determined for each test system.

Although the tessellation typically produces a connectedness with three neighbors for each node, the number of neighbors is not guaranteed due to the masking. Especially for very sparse models the representation becomes skeleton-like and resembles a "molecular ball-and-stick figure."[30] This (desired) strong coarse graining rules out ENM for solving the normal modes, but the above BTS approach is perfectly capable of handling any type of graph. Compared to ENMs, much sparser representations, both in terms of connectedness and in terms of atom density, can be solved by BTS.

III. RESULTS

To validate BTS we have determined the mode overlap with ENM for several test systems. We chose ADK at various levels of course graining as well as cryo-EM reconstructions of RNA polymerase, ribosome, and chaperonin CCT because these systems had been studied well in earlier ENM analysis work[15,20] and thus provide a good standard for comparison.

The validation required in some cases that a coarse BTS model be brought to the level of detail of the ENM. If necessary, the BTS infinitesimal displacements were extended to the ENM node positions by interpolation with the 3D kernel of the thin-plate spline method.[35] If we represent the ith ENM normal mode by a $3N$-vector $\Psi^{(i)}_{\mathrm{ENM}}$ and the jth (possibly interpolated) BTS mode by the $3N$-vector $\Psi^{(j)}_{\mathrm{BTS}}$, then the squared inner product

$$\mathcal{O}_{ij} = \left(\frac{\Psi^{(i)}_{\mathrm{ENM}} \cdot \Psi^{(j)}_{\mathrm{BTS}}}{\|\Psi^{(i)}_{\mathrm{ENM}}\| \, \|\Psi^{(j)}_{\mathrm{BTS}}\|} \right)^2, \tag{12}$$

$0 \leq \mathcal{O}_{ij} \leq 1$ provides a measure of mode similarity. Typically, overlap values above 0.5 are considered good and above 0.7 are considered excellent.[15]

An alternative method for comparison is the crystallographic temperature or B-factor. The B-factor B_i of a specific atom i is a measure of conformational variability. Assuming that this variability originates from the internal motions, B_i can be shown to be proportional to the root mean square displacement of atom i over the modes.[36,37] Consistent with the first 14 dominant modes in our overlap matrices we set

$$B_i = C \sum_{n=7}^{20} \frac{\Psi^2_{in}}{\sigma_n}, \tag{13}$$

where Ψ_{in} is the 3D displacement for atom i and mode n, σ_n is the eigenvalue of mode n, and C is a scaling factor. All B-factor profiles in the following are normalized to have equivalent areas with an (arbitrary) maximum value of 100.

A. Adenylate kinase at variable level of detail

ADK is a phosphotransferase that helps to maintain cellular energy homeostasis. ADK from PDB entry 4AKE (Ref. 38) consists of 214 amino acids and has been shown via ENM and crystallographic studies to exhibit a cleft-closure motion.[15] This open-close motion is also prominently featured at the C_α level in BTS mode 7 depicted in Fig. 1(a). To perform a more systematic test of our BTS approach, we computed the overlap matrix between ENM and BTS [Eq. (12)] at comparable level of detail using the C_α atoms (ENM distance cutoff: 12 Å). For the BTS parametrization we used the initial, continuum mechanics derived approximations of the force constants as described in Sec. II. The diagonal structure of the overlap in Fig. 1(b) demonstrates that the first six nontrivial modes are identical between BTS and ENM, and the higher modes are generally overlapping only within a very small subspace spanned by neighboring modes. Also, the B-factors shown in Fig. 1(c) are very similar (Pear-

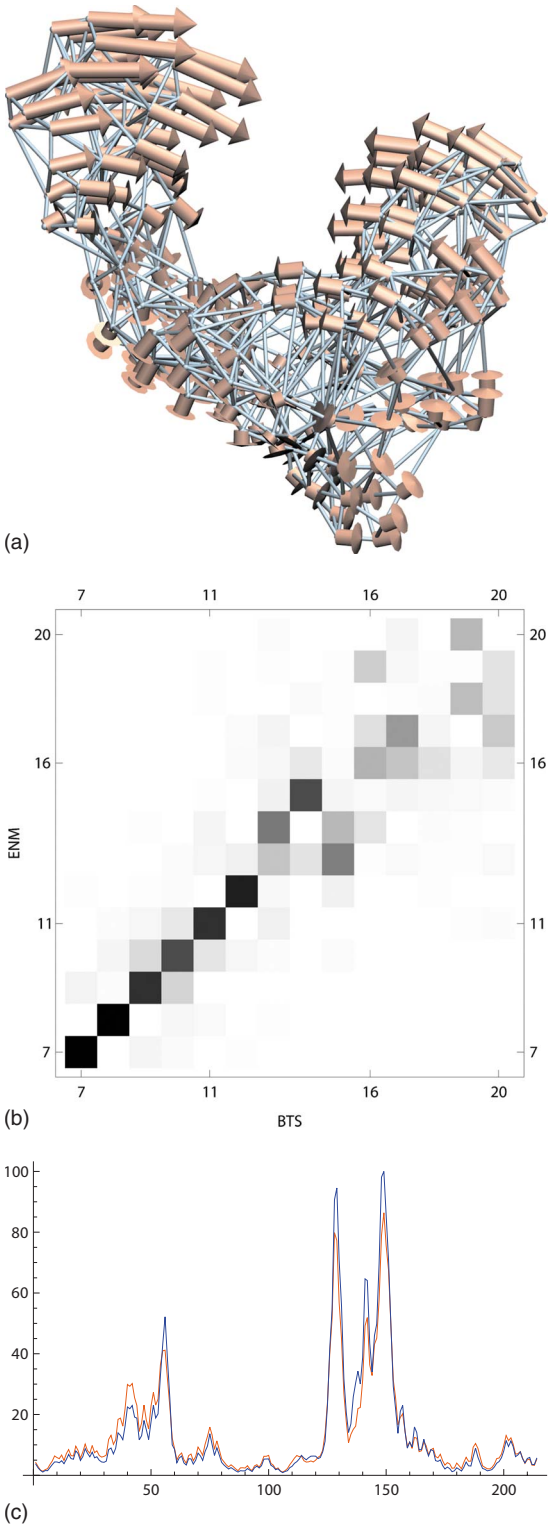

FIG. 1. ADK, $N=214$ (C_α model). (a) BTS mode 7 (golden arrows) and corresponding graph edges (blue rods). All molecular graphic figures in this paper were created with the visualization program VMD (Ref. 50). (b) Overlap matrix of the first 14 nontrivial normal modes for ENM and BTS. (c) *B*-factor as a function of C_α index (amino acid sequence) for ENM (blue) and BTS (red).

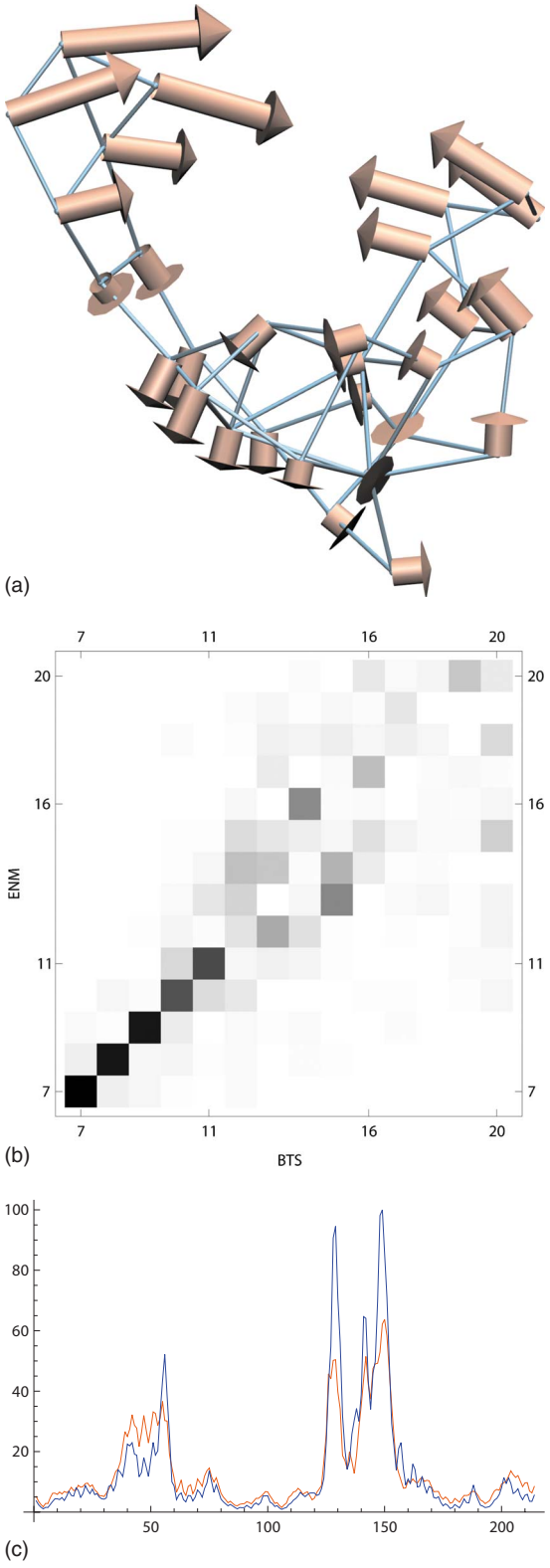

FIG. 2. ADK, $N=30$. (a) BTS mode 7 (golden arrows) and corresponding graph edges (blue rods). (b) Overlap matrix of the first 14 nontrivial normal modes for ENM and BTS. (c) *B*-factor as a function of C_α index (amino acid sequence) for ENM (blue) and BTS (red).

son correlation coefficient: 0.98). Hence, for practical applications at the C_α level, our results suggest that BTS and ENM can be used interchangeably.

To test the robustness of BTS at a low level of detail we coarse grained the ADK system, reducing the number of pseudoatoms roughly by an order of magnitude. When compressing a structure drastically, we found a certain variability in the agreement in overlap and *B*-factor upon small changes in the number *N*. Figures 2 and 3 show two examples: $N=30$ which gives an excellent agreement of modes and

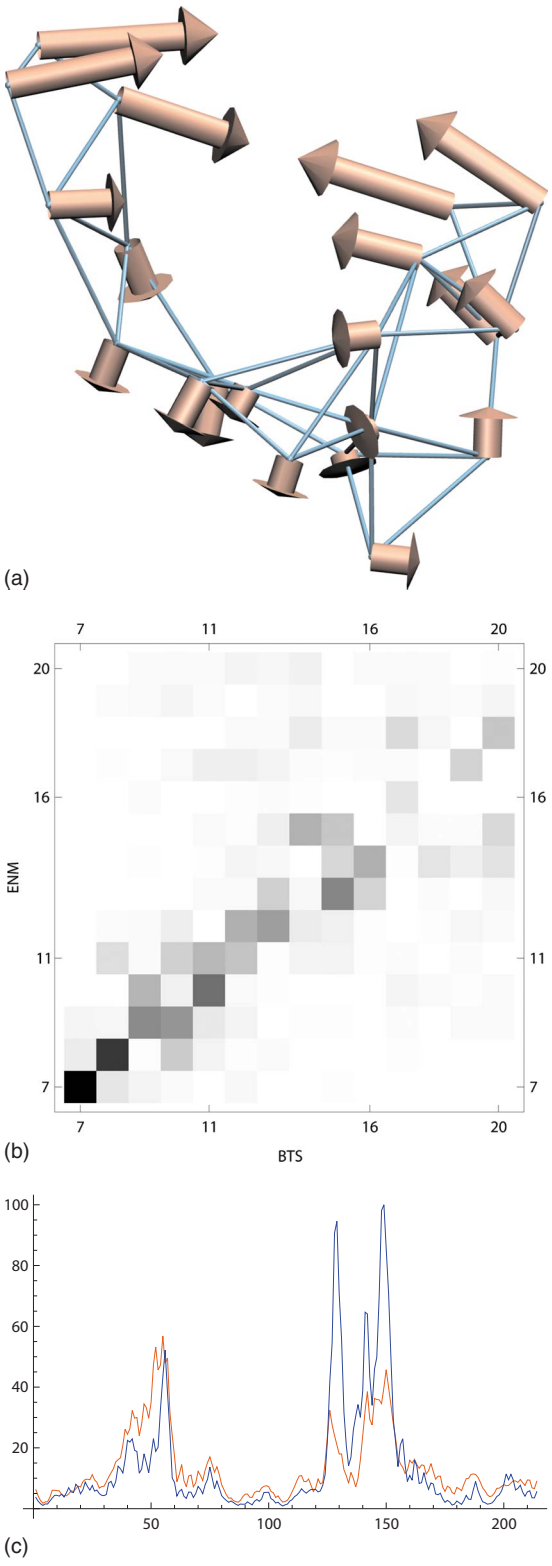

FIG. 3. ADK, $N=20$. (a) BTS mode 7 (golden arrows) and corresponding graph edges (blue rods). (b) Overlap matrix of the first 14 nontrivial normal modes for ENM and BTS. (c) B-factor as a function of C_α index (amino acid sequence) for ENM (blue) and BTS (red).

$N=20$ which provides a worst-case scenario within the range $20 \leq N \leq 30$. As can be seen in panels (a) the reduced structures still provide an excellent reproduction of mode 7, even in the worst case $N=20$. Panels (b) show that the case $N=30$ faithfully reproduces the first five nontrivial modes from

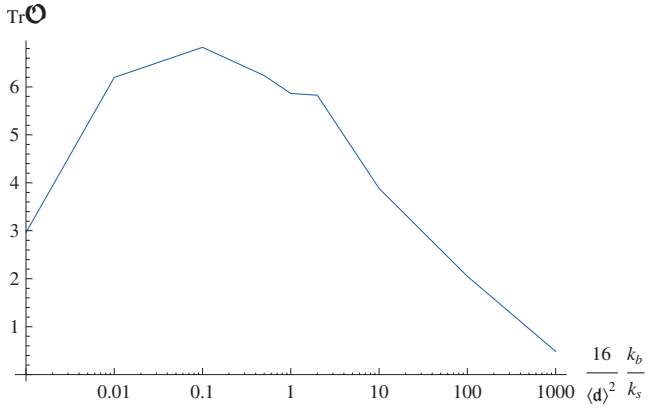

FIG. 4. ADK, trace of the overlap matrix from the first 14 nontrivial modes as a function of $\frac{16}{\langle d\rangle^2}\left(\frac{k_b}{k_s}\right)$ for $\frac{k_t}{k_b}=1$.

ENM, whereas $N=20$ only reproduces the first two. Nevertheless, the structure of the overlap matrix is near diagonal even in the worst-case scenario, demonstrating that modes tend to mix only with their immediate neighbors. The B-factors in panels (c) show that the mobile regions are correctly identified, but the patterns exhibit a degradation consistent with the results of the overlap analysis with Pearson correlation coefficients relative to ENM of 0.92 and 0.71 for $N=30$ and $N=20$, respectively.

B. Empirical validation of force constants

As we have seen, the overlap matrices in panels (b) of Figs. 1–3 provide a way to estimate the quality of a particular BTS model. In the present work we consider the well established C_α ENM as a gold standard for comparison. We can test and optimize our BTS parametrization by maximizing the overlap between both models. To explore the robustness of the BTS parametrization, we varied the force constant ratios derived in Sec. II by several orders of magnitude and measured the effect of the parameter change on the trace of the overlap matrix with the C_α ENM (distance cutoff: 12 Å).

Figure 4 shows the trace as a function of the normalized [Eq. (11)] bend to stretch ratio $\frac{16}{\langle d\rangle^2}\left(\frac{k_b}{k_s}\right)$ (a value of one corresponds to the initial approximation from continuum mechanics). Our initial estimate of the ratio gives near-maximal overlaps. For small, submaximal ratios in Fig. 4 the angle term in our potential essentially vanishes, reducing the similarity between BTS and ENM. The maximum at $\frac{16}{\langle d\rangle^2}\left(\frac{k_b}{k_s}\right) \approx 0.1$ allows one to compute the effective diameter $2r$ of the optimal elastic rod model using Eq. (10): $2r \approx \sqrt{0.1}\langle d\rangle \approx 0.32\langle d\rangle$. This result makes perfect physical sense: the virtual rods in our optimal model exhibit a thickness of order $\langle d\rangle$ but they are thin enough to avoid occupying the same space. For larger ratios in Fig. 4 the similarity between BTS and ENM is decaying once the virtual rod thickness grows beyond the available space.

Figure 5 shows the trace of the overlap as a function of the twist to bend ratio k_t/k_b (a value of one again corresponds to the initial approximation from continuum mechanics), for the optimal bend to stretch ratio $\frac{16}{\langle d\rangle^2}\left(\frac{k_b}{k_s}\right)=0.1$, and for

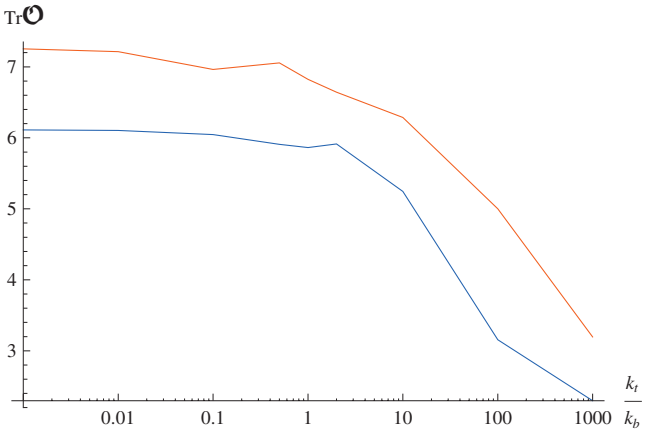

FIG. 5. ADK, trace of the overlap matrix from the first 14 nontrivial modes as a function of $\frac{k_t}{k_b}$ for $\frac{16}{\langle d\rangle^2}\left(\frac{k_b}{k_s}\right)=1$ (blue) and $\frac{16}{\langle d\rangle^2}\left(\frac{k_b}{k_s}\right)=0.1$ (red).

the earlier $\frac{16}{\langle d\rangle^2}\left(\frac{k_b}{k_s}\right)=1$. The figure shows that the initial estimate $k_t/k_b=1$ forms an upper bound of possible values, whereas stronger torsion contributions $k_t/k_b>1$ reduce the agreement with ENM. On the other hand, smaller ratios $k_t/k_b<1$ are not affecting the high similarity between ENM and BTS. It is justified, therefore, to keep the maximum viable $k_t/k_b=1$ to facilitate the calculation of modes for skeletonlike graphs where the torsional term is important to avoid nontrivial zero frequency modes.

In the following (and in our disseminated program) we use the empirically validated ratios $\frac{16}{\langle d\rangle^2}\left(\frac{k_b}{k_s}\right)=0.1$ and $k_t/k_b=1$. We also repeated the calculations of the modes in Figs. 1–3 with the new $\frac{16}{\langle d\rangle^2}\left(\frac{k_b}{k_s}\right)=0.1$ ratio, but the observed changes were insignificant (data not shown).

C. Application to three low-resolution cryo-EM structures

One of the benefits and intended uses of BTS is its ability to characterize low-resolution biophysical data. The intrinsic flexibility of three quintessential macromolecular assemblies was investigated in Ref. 20 with dense (high N) ENMs based on relatively low-resolution (15–27 Å) cryo-EM resolution data. Although much higher resolution cryo-EM maps of these and similar systems are available,[39] it is still typical for most cryo-EM studies to bootstrap the reconstruction and analysis from such low-resolution models as more images are collected. Also, it makes sense for this proof of concept paper to focus on systems that were already well characterized by ENM. In addition to the ENM study,[20] the opening and closing motion of the *E. coli* RNA polymerase (Fig. 6) and the ratcheting motion of the ribosome (Fig. 7) were observed experimentally at atomic and intermediate resolution,[40–43] and the functional mechanisms of these molecular machines continue to be of significant interest. Likewise, the interaction of the eukaryotic chaperonin CCT (Fig. 8) with partially folded substrates is believed to involve significant flexibility.[44]

In Ref. 20 the significant conformational change corresponded to a single, low-frequency mode of its open form,[45] namely, mode 7. Therefore, we focus in the following on the lowest-frequency mode 7 only. The ENM study[20] also found a significant mismatch of resolution between the cryo-EM data and that of the dense representation required to capture the functional motion by mode 7. One can count the maximum number of independent pieces of information in a

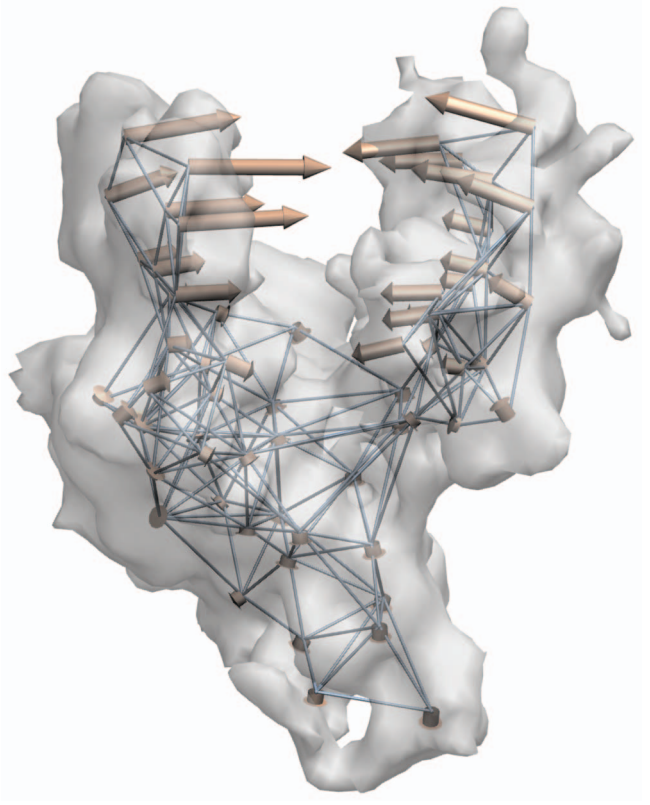

FIG. 6. *E. coli* RNA polymerase, electron microscopy map (transparent isocontour surface), represented by $N=61$ pseudoatoms connected by bonds (blue rods). BTS mode 7 is shown by golden arrows.

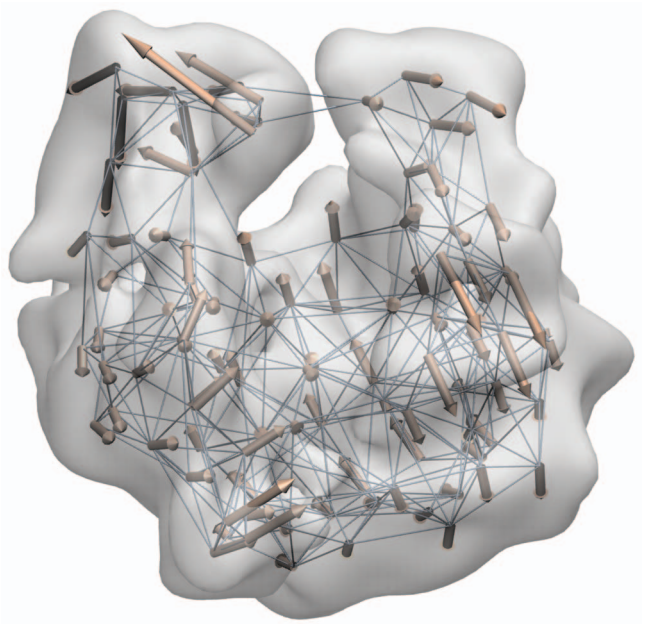

FIG. 7. *E. coli* ribosome, electron microscopy map (transparent isocontour surface), represented by $N=98$ pseudoatoms connected by bonds (blue rods). BTS mode 7 is shown by golden arrows.

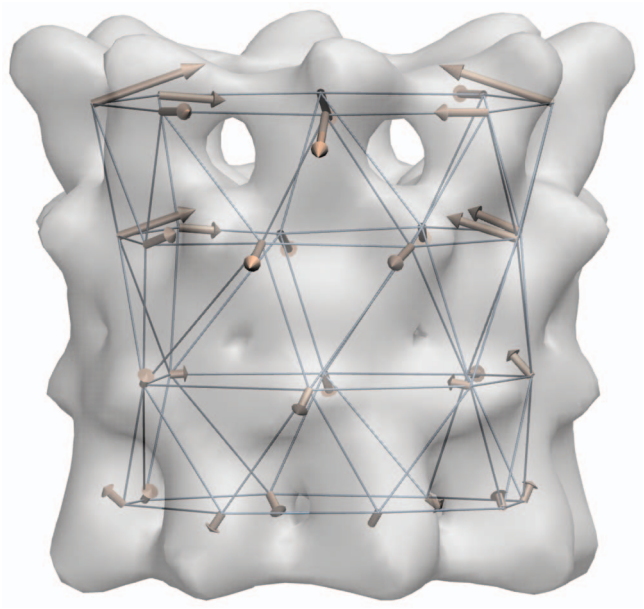

FIG. 8. Eukaryotic chaperonin CCT, electron microscopy map (transparent isocontour surface), represented by $N=32$ pseudoatoms connected by bonds (blue rods). BTS mode 7 is shown by golden arrows.

cryo-EM map by dividing the volume of the structure by the volume of a resolution element, i.e., a cube whose length corresponds to the spatial resolution. For low-resolution (~10–30 Å) cryo-EM maps this maximum "number of resolved features" is surprisingly small, on the order of a few dozen.[46] In contrast, the number N of pseudoatoms in the ENM is on the order of 1000 (Ref. 20) due to the required oversampling of density and connectedness. Here, we chose for the BTS a much smaller number N, namely, half of the maximum number of resolved features in the cryo-EM map, to ensure that we are not oversampling the data.

For RNA polymerase (15 Å resolution) the ENM (distance cutoff: 14 Å) required originally $N=993$ pseudoatoms[20] to describe the open-close motion of the structure (Fig. 6). By applying BTS with half-maximum number of features $N=61$ we obtain an overlap of 0.92 of mode 7 with the dense ENM. This high overlap remains robust for a range of numbers of pseudoatoms, only for very small numbers $N<15$ does the overlap fall below a value of 0.5.

For the ribosome (25 Å resolution) the ENM (distance cutoff: 18 Å) required originally $N=1481$ pseudoatoms[20] to describe the ratcheting motion of the two ribosomal subunits about their mutually adjoining stalk (Fig. 7). Applying BTS with a half-max feature number of $N=98$ we obtain an overlap of 0.82 for ENM mode 7. Overlap values remain above 0.5 down to $N=20$.

Chaperonin CCT (27 Å resolution) forms a hollow cylinder with eightfold symmetry for which the ENM (distance cutoff: 14 Å) originally required $N=1992$ pseudoatoms.[20] The motion of mode 7 consists of alternating elliptical deformations of the apical domains in the opposing cylinder walls (Fig. 8). Applying BTS with a half-max feature number of $N=32$ we obtain an overlap of 0.75 for the symmetry-related mode 7 in the ENM. Overlap values remain above 0.5 down to $N=9$.

IV. DISCUSSION AND CONCLUSIONS

We introduced a novel BTS functional form for coarse elastic network simulation and demonstrated its application to atomic and volumetric data at various level of detail. The sole input parameters are the graph nodes (pseudoatoms) and their connectedness which can be defined by the user or computed with existing tools. We optimized two parameters in our model based on continuum elastics and empirical validations such that they can remain hidden in future practical applications. All steps in Sec. II are automated including the calculations of the list of bonds, angles, dihedrals, numerical calculation and diagonalization of the Hessian matrix, interpolation, and data file output. The tools described in this article will be documented and freely distributed as part of the Python-based "MODEHUNTER" package at URL http://modehunter.biomachina.org.

BTS affords a compression in the spatial level of detail by one to two orders of magnitude compared to ENM with little loss in accuracy. With regard to low-resolution structures, we are able to recreate biomolecular motion with coarser delineations whose level of detail reflects the experimental resolution. As a beneficial side effect, the coarse graining reduces dramatically the computational complexity for solving the eigensystem of the Hessian. The remaining bottleneck in our scheme is due to the numerical computation of the Hessian [cf. Eq. (4)]. Although we have not explored this further in the present work, we expect that the construction of the Hessian could be sped up by using analytic expressions of the second derivatives.

Our empirical validations showed that the additional angle term of the BTS potential compensates for the loss of long-range contacts in our network. In ENMs the cutoff is typically twice the nearest-neighbor distance,[47] resulting in a large number of redundant connections (another pitfall of long-range ENM connections is that these may traverse regions of negligible molecular density so as to constrain certain molecular motions artificially). When we replace the ENM connections with the sparser and shorter connections afforded by the competitive Hebb rule (a masked version of Delaunay tessellation, see Sec. II), the angle term maintains the similarity of the resulting BTS modes with those expected from an ENM (Fig. 4). Although the additional torsion term of the BTS potential was not needed for the well-connected C_α model we investigated (Fig. 1), our tests at the C_α level show that the torsion term is safe to use up to the continuum mechanics limit (Fig. 5).

Recently, two related modifications of the Tirion potential have been proposed: Lu *et al.*[48] performed NMA in an internal coordinate framework to alleviate an observed ENM problem termed the "tip effect." This effect, sometimes observed at protruding loops, is due to unrestricted motions caused by poor connectivity of the ENM. Their solution is closely related to the angle term used in our parametrization, but it required a chain connectivity and is thus limited to proteins at the C_α level. Separately, Jeong *et al.*[47] proposed an interesting connection rule to eliminate the ENM distance cutoff. Their rule used chemical bonding information (both covalent and weak interactions) and was also limited to

atomic detail protein structures. This stands in contrast to the competitive Hebb connectivity rule in BTS that relies on computational geometry alone. Our BTS model builds upon the benefits of both improvements,[47,48] but it is generally applicable to any graph or network whose edges can be physically modeled as elastic rods.

The novel torsional term in BTS is critical for BTS to model a graph irrespective of its degree distribution. The degree of a graph node is the number of edges connected to that node. In a coarse biomolecular model, for example, the degree would be the number of bonds projecting from each pseudoatom. This number is at least one, and indeed many nodes in realistic networks have a low degree, i.e., just one or two connections.[49] In contrast to the BTS model, ENMs require a degree of at least three to avoid nontrivial zero-frequency modes. In other words, the ENM must be a supergraph of a so-called three-nearest-neighbor graph. This entails both a high density of linking partners and requires the mentioned long-range cutoff that in turn leads to many redundant connections. BTS does not suffer from the three-degree constraint and can be made arbitrarily sparse. At a high spatial compression,[30] BTS thus provides the biomolecular analog to "stick figure" animation.

ACKNOWLEDGMENTS

We thank Harel Weinstein for mentorship and support and Rhys Adams and Sebastian Stolzenberg for discussions. This work was supported in part by NIH (Grant Nos. R01GM62968, DA012923 and DA012408).

[1] Q. Cui and I. Bahar, *Normal Mode Analysis* (Chapman and Hall/CRC, London/Boca Raton, 2006).
[2] T. Horiuchi and N. Go, Proteins: Struct., Funct., Genet. **10**, 106 (1991).
[3] M. Gerstein, A. M. Lesk, and C. Chothia, Biochemistry **33**, 6739 (1994).
[4] J. S. Byrnes, *Twentieth Century Harmonic Analysis - A Celebration* (Kluwer, Dordrecht, 2001).
[5] E. B. Wilson, J. C. Decius, and P. C. Cross, *Molecular Vibrations* (McGraw-Hill, New York, 1955).
[6] M. Tasumi, H. Takeuchi, S. Ataka, A. M. Dwivedi, and S. Krimm, Biopolymers **21**, 711 (1982).
[7] T. Noguti and N. Go, Nature (London) **296**, 776 (1982).
[8] M. Levitt, C. Sander, and P. S. Stern, J. Mol. Biol. **181**, 423 (1985).
[9] D. A. Case, in *Computer Simulation of Biomolecular Systems*, edited by W. F. van Gunsteren, P. K. Weiner, and A. J. Wilkinson (Kluwer, Dordrecht, 1997), Vol. 3, pp. 284–301.
[10] D. A. McQuarrie, *Statistical Mechanics* (Harper, New York, 1976).
[11] M. M. Tirion, Phys. Rev. Lett. **77**, 1905 (1996).
[12] I. Bahar, A. R. Atilgan, and B. Erman, Folding Des. **2**, 173 (1997).
[13] K. Hinsen, Proteins: Struct., Funct., Bioinf. **33**, 417 (1998).
[14] P. Doruker, R. L. Jernigan, and I. Bahar, J. Comput. Chem. **23**, 119 (2002).
[15] F. Tama, W. Wriggers, and C. L. Brooks, J. Mol. Biol. **321**, 297 (2002).
[16] F. Tama and C. L. Brooks, Annu. Rev. Biophys. Biomol. Struct. **35**, 115 (2006).
[17] M. Lu and J. Ma, Biophys. J. **89**, 2395 (2005).
[18] http://igs-server.cnrs-mrs.fr/elnemo. A server for creating and solving elastic network models.
[19] http://enm.lobos.nih.gov. A server for creating and solving elastic network models
[20] P. Chacón, F. Tama, and W. Wriggers, J. Mol. Biol. **326**, 485 (2003).
[21] D. Ming, Y. Kong, M. A. Lambert, Z. Huang, and J. Ma, Proc. Natl. Acad. Sci. U.S.A. **99**, 8620 (2002).
[22] J. Frank, Annu. Rev. Biophys. Biomol. Struct. **31**, 303 (2002).
[23] H. H. Niemann, M. V. Petoukhov, M. Härtlein, M. Moulin, E. Gherardi, P. Timmins, D. W. Heinz, and D. I. Svergun, J. Mol. Biol. **377**, 489 (2008).
[24] W. Wriggers, Z. Zhang, M. Shah, and D. C. Sorensen, Mol. Simul. **32**, 803 (2006).
[25] L. D. Landau and E. M. Lifshitz, *Mechanics - Course of Theoretical Physics*, 3rd ed. (Butterworth, Washington, DC, 1982), Vol. 1.
[26] B. Lautrup, *Physics of Continuous Matter* (Institute of Physics, University of Reading, Berkshire, 2005).
[27] P. A. Janmey, J. X. Tang, and C. F. Schmidt, Biophysics Textbook Online (BTOL), 1999, http://www.biophysics.org/education/topics.htm.
[28] W. Wriggers, R. A. Milligan, and J. A. McCammon, J. Struct. Biol. **125**, 185 (1999).
[29] http://situs.biomachina.org. Situs: A package for the modeling of atomic resolution structures into low-resolution density maps.
[30] W. Wriggers, R. A. Milligan, K. Schulten, and J. A. McCammon, J. Mol. Biol. **284**, 1247 (1998).
[31] W. Wriggers, P. Chacón, J. Kovacs, F. Tama, and S. Birmanns, Neurocomputing **56**, 365 (2004).
[32] http://www.ks.uiuc.edu/Research/vmd/plugins/cgtools. Coarse grained modeling plugin for the molecular graphics program VMD.
[33] T. M. Martinetz, S. G. Berkovich, and K. Schulten, IEEE Trans. Neural Netw. **4**, 558 (1993).
[34] M. de Berg, M. van Kreveld, M. Overmars, and O. Schwarzkopf, *Computational Geometry: Algorithms and Applications*, 2nd ed. (Springer-Verlag, Berlin, 2000).
[35] F. L. Bookstein, *Morphometric Tools for Landmark Data* (Cambridge University Press, Cambridge, 1991).
[36] W. C. Lu, C. Z. Wang, E. W. Yu, and K. M. Ho, Proteins: Struct., Funct., Bioinf. **62**, 152 (2006).
[37] S. E. Dobbins, V. I. Lesk, and M. J. E. Sternberg, Proc. Natl. Acad. Sci. U.S.A. **105**, 10390 (2008).
[38] C. W. Müller, G. J. Schlauderer, J. Reinstein, and G. E. Schulz, Structure (London) **4**, 147 (1996).
[39] X. Yu, L. Jin, and Z. H. Zhou, Nature (London) **453**, 415 (2008).
[40] S. A. Darst, N. Opalka, P. Chacón, A. Polyakov, C. Richter, G. Zhang, and W. Wriggers, Proc. Natl. Acad. Sci. U.S.A. **99**, 4296 (2002).
[41] G. Zhang, E. A. Campbell, L. Minakhin, C. Richter, K. Severinov, and S. A. Darst, Cell **98**, 811 (1999).
[42] J. Frank and R. K. Agrawal, Nature (London) **406**, 318 (2000).
[43] F. Tama, M. Valle, J. Frank, and C. L. Brooks, Proc. Natl. Acad. Sci. U.S.A. **100**, 9319 (2003).
[44] X. Zhang, F. Beuron, and P. S. Freemont, Curr. Opin. Struct. Biol. **12**, 231 (2002).
[45] F. Tama and Y.-H. Sanejouand, Protein Eng. **14**, 1 (2001).
[46] W. Wriggers and P. Chacón, Structure (London) **9**, 779 (2001).
[47] J. I. Jeong, Y. Jang, and M. K. Kim, J. Mol. Graphics Modell. **24**, 296 (2006).
[48] M. Lu, B. Poon, and J. Ma, J. Chem. Theory Comput. **2**, 464 (2006).
[49] M. Newman, Phys. Today **61**, 33 (2008).
[50] W. F. Humphrey, A. Dalke, and K. Schulten, J. Mol. Graphics **14**, 33 (1996).