Biophysical Journal Volume 95 November 2008 4183–4192

Systematic Multiscale Parameterization of Heterogeneous Elastic Network Models of Proteins

Edward Lyman,* Jim Pfaendtner,*† and Gregory A. Voth*

*Center for Biophysical Modeling and Simulation, University of Utah, Salt Lake City, Utah 84112-0850, and †Computational Science, Department of Chemistry and Applied Biosciences, Swiss Federal Institute of Technology Zürich, Lugano, Switzerland

ABSTRACT We present a method to parameterize heterogeneous elastic network models (heteroENMs) of proteins to reproduce the fluctuations observed in atomistic simulations. Because it is based on atomistic simulation, our method allows the development of elastic coarse-grained models of proteins under different conditions or in different environments. The method is simple and applicable to models at any level of coarse-graining. We validated the method in three systems. First, we computed the persistence length of ADP-bound F-actin, using a heteroENM model. The value of 6.1 $\pm$ 1.6 μm is consistent with the experimentally measured value of 9.0 $\pm$ 0.5 μm. We then compared our method to a uniform elastic network model and a realistic extension algorithm via covariance Hessian (REACH) model of carboxy myoglobin, and found that the heteroENM method more accurately predicted mean-square fluctuations of α-carbon atoms. Finally, we showed that the method captures critical differences in effective harmonic interactions for coarse-grained models of the N-terminal Bin/amphiphysin/Rvs (N-BAR) domain of amphiphysin, by building models of N-BAR both bound to a membrane and free in solution.

INTRODUCTION

In the effort to develop predictive models connecting protein structure with biochemical function, it is often neither possible nor desirable to maintain an atomically detailed description of the studied system. Consequently, a wide spectrum of coarse-grained (CG) approaches have been developed, according to an equally wide spectrum of motivations and principles (1,2). Among CG approaches, elastic network models (ENMs) have enjoyed considerable popularity, because they are fairly simple to implement and often yield considerable insight into the functional motions of proteins (3,4).

Historically, the ENM approach has followed the original implementation of Halioglu et al., which builds a network by connecting with uniform harmonic spring pairs of α-carbon positions closer than a fixed cutoff distance (5). The use of a uniform spring constant across the whole network was based on earlier work regarding the normal mode analysis of proteins (6,7). The uniform spring constant is chosen to reproduce a chosen observable, usually the per-residue root mean-square fluctuations, as determined by a crystallographic or NMR experiment (8). The use of a uniform spring constant and cutoff is motivated by the idea that the dynamic behavior of a folded protein is determined mainly by the averaged, local packing density (5). This approach adopts a ''universal'' point of view, in which the effects of local chemical differences are reflected only through contact topology, in favor of developing a minimalist model with only one fitted parameter. The ENM approach and its extensions were used to gain insights into the functional motion of a wide variety of proteins and macromolecular complexes (9–13). Discussion continues regarding the general applicability of the single-force parameter elastic network approach, with the success of the method apparently depending on the system under study (14,15). Limitations of the approach notwithstanding, ENMs have provided valuable insights into the function of macromolecular complexes such as GroEL (13,16,17) and the ribosome (12,18).

But can the elastic network approach be pushed to a higher degree of predictive accuracy? We present a new method for developing elastic models of macromolecular complexes, based on earlier work (19,20). Our method develops the spring constants of a *heterogeneous* elastic network by fitting to thermal averages (e.g., fluctuations of CG pair distances) computed from atomistic molecular dynamics. In contrast with the standard ENM approach, our model was developed from an atomistic molecular dynamics trajectory. This allowed us to parameterize different elastic models for the same protein in different environments, or under different conditions, in a multiscale fashion. As an example, we demonstrate that our approach picks up differences in the elastic network induced by the binding of an interfacial membrane protein to a bilayer.

Furthermore, because the spring constants are allowed to adopt any value, it is unnecessary to impose any cutoff on the length-scale of CG interactions. Instead, all pairs of CG sites may be allowed to interact. This is especially important when parameterizing models which are CG beyond the α-carbon level, and for which CG sites may differ in the number of residues which they represent. In such situations, effective harmonic interactions may vary considerably in stiffness.

The difference between our approach and the usual ENM approach is both fundamental as well as practical. Funda-

Submitted June 10, 2008, and accepted for publication July 18, 2008.

Address reprint requests to Edward Lyman, Center for Biophysical Modeling and Simulation, University of Utah, 315 S. 1400, E. Rm. 2020, Salt Lake City, UT 84112-0850.

Editor: Angel E. Garcia.

0006-3495/08/11/4183/10 $2.00

doi: 10.1529/biophysj.108.139733

mentally, we propose that the fluctuations observed in explicitly solvated molecular dynamics simulations are more representative of the thermal fluctuations of solvated proteins than are crystallographic B-factors. Practically, we have at our disposal a wider range of observables against which to parameterize our model. This allows a highly heterogeneous network of spring constants to be parameterized without overfitting, trading extra computation for increased accuracy. Indeed, at the C^{α} level of coarse-graining, we find that the spring constants vary over five orders of magnitude.

What is to be gained by moving away from a uniformly stiff elastic model? In a comparison of a homogeneous ENM model with CG normal-mode and quasiharmonic studies of small RNAs, Van Wynsberghe and Cui found that the homogeneous ENM could not capture motional correlations as well as the normal-mode approaches (which are intrinsically heterogeneous) (21). This, they suggested, was due to the lower packing density of nucleic-acid systems, a situation analogous to the aggressively CG models which we develop here. In a later work, the same authors showed that off-diagonal elements of the covariance matrix depend explicitly on ratios of spring constants (22). It is therefore essential for an elastic model which aims to predict not only directions of motions but motional correlations to allow for variable spring constants.

This line of reasoning has some precedence, insofar as Hinsen and Kneller showed that a distance-dependent force constant was sufficient to reproduce the fluctuation spectrum of an atomistic model (23). More recently, other authors considered heterogeneous ENMs of proteins. Based on an atomistic normal-mode analysis of bovine trypsinogen, Ming and Wall (24) observed a bimodal density of vibrational modes. Unable to reproduce such a density of states with a single-paramter ENM, they found that by introducing two distinct spring constants (to model short-tange and long-range interactions), the bimodal density of states was well-reproduced (24). Moritsugu and Smith introduced a method that computes a heterogenous network of spring constants of an α-carbon ENM from the eigenvalues of the covariance matrix of an atomistic simulation (25). This is an attractive approach, because it recognizes the power of using atomistic-simulation data in parameterizing CG models. It relies, however, on the accurate calculation and diagonalization of the covariance matrix, which is complicated by the need to separate effective interactions into local and nonlocal pairs. No such separation is required by this method. Furthermore, anharmonicity observed in the atomistic simulation yields negative spring constants upon inversion of the covariance matrix. This approach instead determines the optimal effective harmonic interactions, given the (atomistically) observed fluctuations. In a similar spirit, Chen et al. parameterized a continuum-level elastic model based on atomistic normal-mode data (26).

We will discuss the application of our heterogeneous ENM to three proteins. In the first application, we built a heterogeneous elastic network model (heteroENM) of an ADP-bound actin filament. A previous CG model, parameterized from atomistic-simulation data, was able to predict accurately the persistence length of the actin filament (19). In this earlier study, intermonomer interactions were modeled with two-body harmonic terms, whereas intramonomer interactions included harmonic three-body terms and an anharmonic four-body term. Here, we demonstrate the accurate calculation of the persistence length, using a model which contains only pairwise harmonic terms. The only input involves the atomistic-simulation data, after which the model development is fully automatic, i.e., no cutoff or spring constants are adjusted after the fact.

In the second calculation, we develop an α-carbon-based heteroENM of myoglobin, to compare the heteroENM method to a single-parameter ENM and to the realistic extension algorthim via covariance Hessian (REACH) method (25). In this case, we imposed a cutoff of 15 Å on pairwise interactions, to make a fair comparison with previous work. At the cost of increased computation, we found improved correlation between the heteroENM and the molecular dynamics (MD) fluctuations, compared to the other methods.

In the third calculation, we built a CG elastic model of the N-terminal Bin/amphiphysin/Rvs (N-BAR) domain of amphiphysin (27). The N-BAR is known to play a role in membrane-remodeling processes, particularly the tubulating and budding of vesicles (28,29). As a step in the development of a comprehensive model of the membrane-remodeling process, we built a heteroENM of the N-BAR domain. By building two different heteroENMs (one for an N-BAR bound to a membrane, and one unbound), we gain insight into how the slow-timescale, long-wavelength dynamics of the N-BAR are modified by its interaction with the membrane. This is a strength of our approach, which follows directly from the ''bottom-up'' philosophy of parameterization based on atomistic simulation. This advantage could readily be extended to other scenarios, allowing, for example, observation of subtle changes in flexibility upon binding of a ligand.

METHODS

Coarse-grained model development

Before parameterizing the interactions of our heteroENM, we must decide on a mapping of atomistic into CG sites. In the first example, the four domains of each actin monomer are CG to their centers of mass. In the second example, we reduce each residue of myoglobin to an interaction site at the α-carbon. To develop the 20-site model of N-BAR, we used the method developed by Zhang et al., in which CG sites are chosen such that the essential low-frequency dynamics observed in the atomistic simulation are captured (30). Here, we work with models with resolutions ranging from 5–140 residues per CG site. The atomistic trajectory is then projected onto a CG trajectory. The CG sites of the heteroENM are then located at their average positions, as computed from the CG trajectory.

The harmonic interactions between CG sites are determined by a procedure similar to that used by Chu and Voth (19), with some important differences. For the sake of completeness, we describe the algorithm step-by-step. The target data are the mean-square distance fluctuations ($\overline{\Delta x^2}$) between all

pairs of CG sites, as observed in the CG trajectory. That is, for each pair of CG sites, we compute

$$\overline{\Delta x_{ij,MD}^2} = \overline{(x_{ij} - \overline{x_{ij}})^2}, \quad (1.1)$$

where x_{ij} is the distance between a pair of CG sites labeled by i and j in a single frame of the CG trajectory, the subscript *MD* indicates that this quantity is computed from the atomistic molecular-dynamics trajectory, and the overbar denotes an average over sampled configurations. We then determine the spring constants of the elastic network by the following iterative algorithm:

1. Begin with all pairs of interacting sites connected by a uniform spring constant, k_{ij}.
2. Relax the positions of the CG sites by a steepest-descent search, and then compute the normal modes of the model. Scale the amplitudes according to equipartition of energy to reflect the temperature of the atomistic calculation.
3. For each pair of sites ij, compute $\Delta x_{ij,NMA}^2$ by projecting the (harmonic) fluctuations computed from the normal-mode analysis onto the intersite distance vectors.
4. Update all spring constants simultaneously according to

$$\frac{1}{4k_{ij}^{n+1}} = \frac{1}{4k_{ij}^{n}} - \alpha(\Delta x_{ij,NMA}^2 - \Delta x_{ij,MD}^2), \quad (1.2)$$

   where n labels the iteration.
5. Repeat steps 2 to 4 until the difference between the normal-mode fluctuations and the target (MD) fluctuations is below a predefined threshold.

The search algorithm defined in Eq. 1.2 seeks the set of spring constants that minimize the difference between the MD and heteroENM pair distance fluctuations. It uses information about the deviation between the MD and heteroENM pair fluctuations to guide the next iteration, stiffening springs that are too soft, and vice versa. The parameter α controls the size of the steps in the spring constants by scaling the difference of the fluctuations. In all cases we have studied, the algorithm is robust to the initial values of the spring constants: for uniform initial values ranging from 1–100 kcal/molÅ^{-2}, and for random initial spring constants in the same range, the iterative procedure converges to the same final spring constants, to within 0.1%. Also, because the system is already at the bottom of the native basin on the potential energy surface, the structure distorts very little as a result of the minimization in step ii. The root mean-square (RMS) difference before and after minimization of the myoglobin model was 0.1 Å, and 0.6 Å for the N-BAR model. The minimization is necessary only to ensure a meaningful normal-mode analysis.

We found that iterating in $1/4k$ instead of k greatly improves the convergence of densely connected networks. This is because the fluctuation of a particular bond depends most directly on its own spring constant, and for a harmonic oscillator $\Delta x^2 \approx 1/4k$. This is in contrast to previous work (19), where the iterative procedure was carried out in k. We found that, for densely connected networks, iterating in k results in very erratic convergence. This is not necessarily at odds with the results reported previously, because this seems to be a problem only for highly heterogeneous, densely connected networks.

Step iii requires the calculation of anisotropic fluctuations, and is therefore an application of the anisotropic network model (31), albeit with heterogeneous spring constants. In this work, we used CHARMM version c32b2 to perform the normal-mode analysis and compute the intersite distance fluctuations. The amplitude of the (mass-weighted) fluctuations were determined by equipartition of energy, with the energy scale set to k_BT, and T set to the temperature of the atomistic molecular-dynamics simulation.

We do not need to impose any cutoff on the CG network interactions beforehand. Instead, the iterative procedure ''decides'' which interactions are unimportant, by driving them to zero. This is quite different from the approach used by Bahar and Jernigan to parameterize α-carbon-based models, in which the optimal cutoff was determined by considering a large set of nonhomologous proteins in the Protein Data Bank (PDB) (8). The approach of Bahar and Jernigan (8) is sensible, because it makes the results of α-carbon-based ENM calculations considerably more universal. Here, however, we are interested not in a universal model which does reasonably well for a wide range of proteins modeled at the α-carbon level. Rather, we want a specific model that quantitatively captures the fluctuations of a specific, solvated protein, and that may be much more aggressively CG. However, in one case studied below, we followed the standard methodology and applied a cutoff of 15 Å, for the sake of making a fair comparison to two other α-carbon-based ENM models.

Atomistic-simulation details

The myoglobin atomistic simulation data and per-residue mean-square fluctuations computed with the REACH method were provided by Moritsugu and Smith (25,32). Details of their method and their atomistic simulation were described previously (25,32).

Atomistic MD simulations of the actin filament and the N-BAR domain of amphiphysin were performed using the NAMD software package (33). Initial structures for the F-actin system were based on monomeric actin in the ADP state (PDB entry 1J6Z). Using the filament model of Holmes et al. (34), an actin filament of 13 subunits was generated in the same manner previously reported by Chu and Voth (20). Longitudinal periodicity was assumed for all atomistic F-actin simulations. The CHARM22 force field (35) was used in conjunction with the particle mesh Ewald algorithm (36) for long-range electrostatic interactions. After an initial 20-ps heating period, the system was preequilibrated in the constant particle number, volume, and temperature ensemble for 50 ps by velocity rescaling. Simulations were then performed in the constant particle number, pressure, and temperature ensemble (310 K and 1 atm), controlling pressure with the Langevin piston Nose-Hoover method, as implemented in NAMD (37,38). After equilibration, a single production trajectory of 40 ns was generated. The normal-mode analysis of the CG elastic network was performed with CHARMM version c32b2 (39), and the iterative fluctuation matching procedure was implemented in a PERL script. A trajectory of the CG elastic network was generated with CHARMM version c32b2 by integrating the Langevin equation of motion, with the temperature of the bath set at 310 K.

The details of membrane-bound N-BAR simulations were reported previously (40). Equilibration of unbound N-BAR simulations was performed using the same protocol as for the bound N-BAR simulation, after which 12-ns production trajectories were generated, again with the same protocol as for the bound N-BAR simulations. The last 10 ns of the production runs were used as the target data for fluctuation matching. The iterative fluctuation matching calculation was performed with the PERL script used for the F-actin calculation.

RESULTS

Calculation of ADP bound F-actin persistence length

Filaments of actin constitute the cytoskeleton of eukaryotic cells (41). The mechanical properties of actin filaments (F-actin), and therefore those of the cell, are modulated by the hydrolysis of bound ATP: one molecule of ATP may be bound per actin monomer. Upon hydrolysis of ATP to ADP, the actin filament softens, as measured by a decrease in persistence length by about a factor of ~50% (42). Previously, Chu and Voth showed that a CG model of F-actin, parameterized by fluctuation matching, can predict the persistence length quite well (19). Here we demonstrate that a heteroENM model of F-actin (parameterized with atomistic data on nm

length-scales) can also predict the persistence length, a μm length-scale property.

The actin monomer was CG into four sites, using the standard definitions of the four domains that make up the monomer (43). A filament of 13 ADP-bound monomers was simulated in atomistic detail with explicit solvent, and projected onto the CG model (a total of 52 sites) to generate the target data for building the heteroENM (details of atomistic simulation are found in Methods). The parameterization resulted in harmonic bonds between all pairs of 52 sites, ranging in stiffness from 10^{-3} to 10 kcal mol^{-1}Å^{-2}. We emphasize that the parameterization of the heteroENM is automated, requiring no after-the-fact tuning of model parameters. Also, this CG model is quite different from the one developed by Chu and Voth (19). Within a single monomer, their model included bond stretching, angle bending, and torsion terms, whereas the model presented here contains only bond-stretching terms. Furthermore, the harmonic bonds between monomers are determined automatically by our procedure, whereas the previous model required selection of a convergent set of intermonomer bonds to reproduce well the atomistic data.

A trajectory of the 52 CG-site heteroENM was then generated by integrating the Langevin equation; simulation details are provided in Methods. The harmonic bonds in our F-actin model are considerably softer than those found in an atomistic model, which allows the use of a much longer time-step than is typically used in atomistic simulation. Indeed, the softness of the springs almost necessitates the use of a long time-step, because to sample thoroughly the fluctuations of the filament, each CG bond must oscillate many times. Attempting to sample these oscillations with femtosecond time-steps and correspondingly tiny displacements results in very inefficient sampling.

The persistence length measures the length-scale of correlations in the direction of a filament, and was measured experimentally for F-actin by video microscopy (42). As described previously (20), we can also compute this quantity from simulation, allowing a stringent test of our CG model. The persistence length of the heteroENM model (6.1 $\pm$ 1.6 μm) was found to agree well with the atomistic MD simulation (7.9 $\pm$ 3.7 μm), and reasonably well with the experimentally measured value for ADP-bound F-actin (9.0 $\pm$ 0.5 μm) (42). However, several points deserve emphasis. First, it is reasonable that the simulations predict a filament that is softer than observed in experiments. Experimentally, it is not clear to what degree the measured F-actin is completely in the ADP-bound form, and some degree of ATP-binding will tend to stiffen the filament (44). It is remarkable that, despite the limited length-scale (35 nm) of the simulation, we are able to predict a μm length-scale property to within 30% of the experimental value. Second, no periodic boundaries were used in the CG simulation: it was of a ''stand-alone'' filament 35 nm in length. This was not the case for the calculation of the persistence length from the atomistic simulation, where periodic boundaries had to be enforced.

HeteroENM model of myoglobin

Myoglobin was the first globular protein structure to be solved at atomic resolution (45). Small and structurally quite stable, it has been the subject of a great deal of study ever since, both experimental and computational. Several α-helices enclose a heme group, which reversibly binds a single O_2, CO, or NO. Following Moritsugu and Smith (25,32), we used myoglobin to validate our novel ENM methodology, to compare the strengths of each method in a well-understood context.

Fig. 1 depicts an α-carbon-level heteroENM model of myoglobin, which clearly traces the α-helices that form the tertiary structure. The CG sites of ENM models are placed at the average positions of the α-carbons, as computed from the MD trajectory. Elastic bonds connect all pairs of α-carbons that are separated, on average, by $<$15 Å over the course of the atomistic MD simulation. Initially, a uniform spring constant of 10 kcal mol^{-1}Å^{-2} is assigned to all interactions. After 2000 iterations of the heteroENM algorithm (Eq. 1.2), the relative root mean-square fluctuations of all bonds computed from the heteroENM model converge to within 1.0% of the values observed in the MD trajectory, after which the model is a highly heterogeneous network of elastic springs. The stiffness of each spring is color-coded in Fig. 1 according to a logarithmic scale. The final values range from 0.01 to 100 kcal mol^{-1}Å^{-2}. The stiffest springs clearly link neighboring α-carbons along the backbone, with longer-range interactions modeled by softer springs. This representation makes plain the regions of dense, soft tertiary contacts, like those in Fig. 1 (*left*). These softer interactions might allow a ''clamshell''-like opening of the molecule, perhaps allowing ligands access to the heme group. In fact, Elber and Gibson described evidence of a pathway for CO in the vicinity of the N-terminus (46) (Fig. 1, *red*).

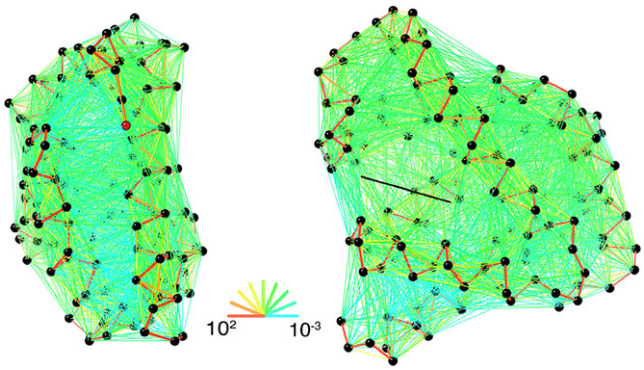

FIGURE 1 Two views of an α-carbon heteroENM model of myoglobin. Pairs of C^{α} atoms separated on average by $<$15 Å are connected by a harmonic spring, indicated by colored lines. Stiffness of the spring is represented by its color on a logarithmic scale, from ~100 kcal mol^{-1}Å^{-2} (*red*) to 0.01 kcal mol^{-1}Å^{-2} (*blue*). The N-terminus is indicated by a red ball (*at left*), and the plane of the heme is indicated by a black line (*right*). The image was rendered with Kinemage, next generation (54).

A different look at the pattern of spring constants is provided in Fig. 2, which demonstrates a rich structure in the scatter of spring constants plotted as a function of distance. The most striking feature is the gap between the very stiff nearest-neighbor C^{α} interactions and the softer, longer-range interactions. Also remarkable is that within the group of longer-range interactions, clear clustering is observed even among C^{α} sites separated by five covalent bonds. This finding echoes those of Moritsugu and Smith (25,32), who found it necessary to fit separately the harmonic interactions corresponding to four distinct groups, distinguished by the number of covalent bonds between sites. In this method, no such assumption is needed a priori. Indeed, though the spring constants tend to cluster, the clustering is far from perfect, and outlier spring constants are found for every group. This finding also echoes those of Ming and Wall, who used a stiffer spring to connect neighboring α-carbons to reproduce a bimodal density of states (24).

Myoglobin: comparison to REACH and single-parameter ENM

Here, we demonstrate that heteroENM does a significantly better job of reproducing per-residue mean-square fluctuations than either the standard, single-parameter ENM or the REACH method, albeit at greater computational expense. To make a fair comparison with the single-parameter ENM and REACH, we coarse-grained myoglobin at the C^{α} level, and enforced a cutoff of 15 Å for harmonic interactions. Regarding the single-parameter ENM, we departed in one significant respect from the standard parameterization protocol. To make a fair comparison with the heteroENM methodology, which is parameterized against atomistic MD data, we parameterized the single-parameter ENM against the same MD data. However, in contrast with heteroENM, which is parameterized against site-site fluctuations, the single-parameter ENM spring constant was fitted to best reproduce the pattern of residue fluctuations (B-factors) by optimizing the correlation coefficient between the B-factors computed from the MD simulation and those computed by the single-parameter ENM. To be fair, it must be noted that heteroENM requires ~2000 iterations to converge, though this is a very reasonable computational load, requiring only a few hours of single-processor CPU time.

The per-residue mean-square fluctuations are plotted in Fig. 3 for the three ENM models, as well as for the MD simulation used to parameterize both models. It appears in Fig. 3 that heteroENM follows more closely the pattern of fluctuations. This observation is borne out by computing the linear correlation between MD fluctuations and those computed by each ENM model. The correlation is strongest between the heteroENM and the MD with a correlation coefficient of 0.93. The correlation of the REACH model and the single-parameter ENM model are considerably weaker, at 0.71 and 0.63, respectively. But we can go further and ask, ''How well are the amplitudes of the residue fluctuations predicted in each case?'' This information is contained in the slope of the linear fit: a slope of 1 indicates that the amplitudes are reproduced perfectly. The slope of the regression is 0.44 for heteroENM, in contrast with a slope of 0.28 for the REACH model and 0.10 for the single-parameter model. Apparently, constraining the ENM to a homogeneous spring constant has the effect of reducing the amplitude of the residue fluctuations, compared with heteroENM.

To assess the predictive power of our method, we parameterized two additional heteroENM models: the first based only on the initial 2.5 ns of the atomistic MD trajectory, and the second based only on the initial 5 ns of the atomistic MD trajectory. We then compared the per-residue mean-square

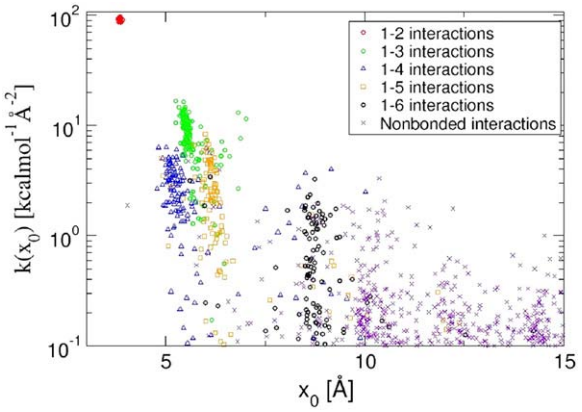

FIGURE 2 Spring constant as a function of equilibrium length x_0 for the myoglobin heteroENM model in Fig. 1. Points are colored to indicate whether they are between neighboring α-carbons (1–2 interactions), next-neighbor α-carbons (1–3 interactions), etc. Interactions cluster into distinct groups, even as far out as 1–6 related α-carbons.

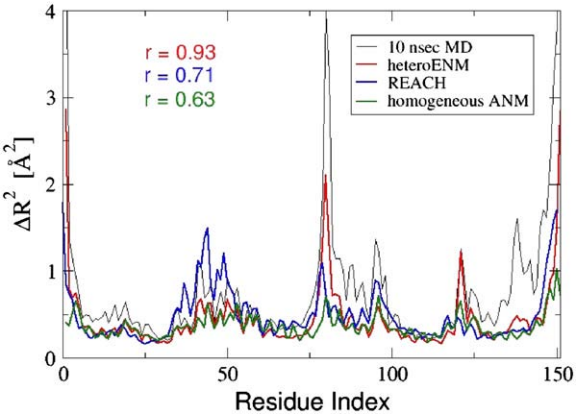

FIGURE 3 Per-residue mean-square fluctuations computed from atomistic MD simulation, heteroENM, REACH, and single-parameter ENM. Neither the heteroENM nor REACH models were parameterized to reproduce these data. Agreement between the MD simulation and ENM models, as measured by linear correlation coefficient r, is best for the heteroENM model.

fluctuations computed from these two heteroENM models with those of the full 10-ns MD trajectory. The correlation coefficient between the MD and the 2.5-ns heteroENM fluctuations was 0.93, and between the MD and 5-ns heteroENM, it was 0.91. The quantitative agreement between the heteroENM and MD fluctuations was maintained, even when the heteroENM was parameterized using MD data of a considerably shorter timescale. Interestingly, some of the spring constants differed between the three heteroENM models (2.5, 5, and 10 ns of MD data used) by as much as 30%. This suggests that high precision in the spring constants is not essential. Instead, the topology of the network plus spring constants, which are approximately correct, seems sufficient. However, even with the differences between the three heteroENMs, the networks are still highly heterogeneous, which seems essential for reaping the benefits of the new methodology. The improved agreement is encouraging, insofar as the purpose of this method is to exchange computational effort for accuracy. Furthermore, the per-residue fluctuations are predicted by the heteroENM method, whereas both the REACH method (through the diagonal elements of the covariance matrix) and the standard ENM approach are parameterized using these data.

Parameterization of a heteroENM N-BAR domain model

The banana-shaped N-BAR domain is a homodimer thought to be involved in the remodeling of membranes during various cellular processes, such as exocytosis and the invagination of vesicles (27,29). There is considerable debate over whether the N-BAR domain actively bends membranes during these processes, or simply senses curved membranes and localizes other proteins to these areas (47,48). Induction of curvature by N-BAR domains was observed in vitro, and on the length-scale of single N-BAR domains by atomically detailed simulations (40). The mechanism by which many N-BAR domains collectively induce curvature is also an unresolved subject, with some authors arguing for a dominant role of the N-terminal amphipathic helices (49), and others advocating a ''scaffold'' model, in which the curved shape of the N-BAR domain plays a central role (50,51). In either case, the collective phenomenon is beyond the scope of atomistic simulation. A field-theoretic model can account for both tubulation and vesiculation of membranes (52) at the level of vesicles, but a truly multiscale description is needed, in which models at different length-scales and timescales are rigorously connected. As a step in this effort, we developed an elastic model of a single N-BAR domain.

We developed two different CG models of the N-terminal N-BAR domain of *Drosophilia* amphiphysin, based on different atomistic-simulation conditions. In one case, N-BAR is ''free'' in aqueous solution, and in the other case, it is bound to a lipid bilayer. We expected that the elastic properties and therefore fluctuations of the N-BAR domain would be modified by its interaction with the bilayer, and were interested to see whether the heteroENM approach could distinguish between these two cases, by building in different effective interactions. More generally, it is of interest to investigate the role that the membrane plays in modulating the behavior of membrane-associated proteins.

In developing our model, we considered only the residues numbered 26–245 by Peter et al. (27). This excludes the N-terminal helix from each monomer, which is thought to undergo a disorder → helix transition upon binding, when the amphipathic helix inserts into the bilayer. We do not expect an elastic network approach, as based on a fixed contact topology, to treat correctly such local folding events. Although it is known that the N-terminal helix plays an important role in binding the N-BAR domain to the bilayer, here we focus our analysis on the elastic properties of the crescent-shaped N-BAR domain scaffold.

For each case (bound and unbound), a 20-site CG model (Fig. 4) was constructed by the essential-dynamics coarse-graining (EDCG) method of Zhang et al. (30). Their method defines ''dynamic domains'' based on the principal components of atomistic simulation, so that CG sites reflect regions of the protein that fluctuate in concert. Because this approach uses only correlations at the pairwise level, it is well-suited for defining CG sites that will interact via harmonic pairwise

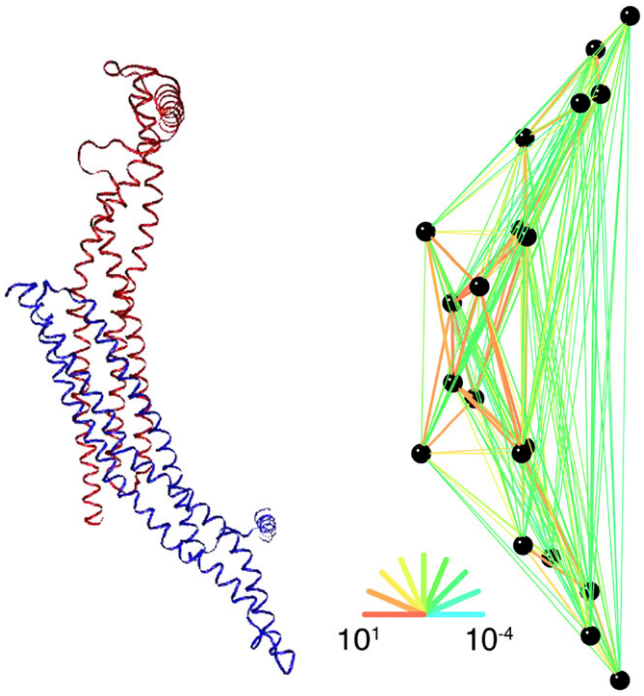

FIGURE 4 Side view of bound N-BAR domain model, comparing backbone level structure (*left*) with heteroENM model (*right*). The CG-site definitions are symmetric across the homodimer. Residues corresponding to CG sites are 26–31, 32–55, 56–103, 104–132, 133–150, 151–163, 164–173, 174–192, 193–218, and 219–244. Ribbon diagram was rendered using visual molecular dynamics (55), and network diagram was rendered using Kinemage, next generation (54).

springs. Our CG model of the N-BAR domain is based on the first 20 principal components that captured ~90% of the fluctuation in atomistic simulations (30). We based our definition of CG sites on the principal components of a simulation of N-BAR free in solution. We then parameterized the harmonic interactions for two different sets of atomistic-simulation data, i.e., the same N-BAR simulation, and also a simulation of an N-BAR bound to a membrane. By using the same model for each case (membrane-bound and unbound), we could compare the harmonic interactions and investigate the effect that the membrane has on the CG protein model.

After ~20,000 iterations, the heteroENM model in Fig. 4 was judged to be converged, as assessed by observing the behavior of the per-bond deviation as a function of iteration:

$$\overline{\delta(\sigma_x)} \equiv \frac{1}{N_{\text{bonds}}} \sum_{i=1}^{N_{\text{bonds}}} |\sigma_{i,\text{NMA}}(x_i) - \sigma_{i,\text{MD}}(x_i)|. \qquad (1.3)$$

After 20,000 iterations, $\overline{\delta(\sigma_x)}$ had decreased from an initial value of ~1 Å/bond to 0.08 Å/bond. The calculation was extended for another 30,000 iterations, over which time $\overline{\delta(\sigma_x)}$ only decreased by another 0.01 Å/bond. In the future, we will improve on the current method by implementing a more sophisticated search algorithm, which might both improve the accuracy of the model and reduce the computational expense. The current algorithm modifies one spring constant at a time, although in reality, the fluctuation of a particular bond depends on the whole network. Nonetheless, it is encouraging that such a simple protocol finds a set of spring constants that reproduce the target data with such accuracy.

The effective harmonic interactions differ significantly between the membrane-bound and unbound cases, as shown in Fig. 5. First, there is an overall stiffening of the N-BAR, upon binding to the membrane, of ~0.25 kcal mol^{-1}Å^{-2}. This effect is reproducible, insofar as it was also observed in independent CG models of the bound and unbound N-BAR, developed from independent atomistic simulations. Because the spring constants are determined directly by fluctuations observed in atomistic simulations, this stiffening suggests a subtle interplay between the dynamics of the protein and those of the membrane. Apart from the overall ''background'' stiffness imparted by the membrane, certain specific interactions are significantly stiffened upon binding, whereas others are actually softened, as shown in Fig. 5. These very specific changes are essential for building a quantitative elastic model of the N-BAR domain in these two different environments, and are not likely to be known in advance. Arkhipov et al. presented a multiscale description of tubulation by N-BAR domains that captures the qualitative features of tubulation (53). Our eventual goal is to develop a multiscale description of membrane-remodeling by N-BAR domains that is carefully parameterized to capture the sorts of differences described here, and that will contribute to understanding the remodeling process quantitatively.

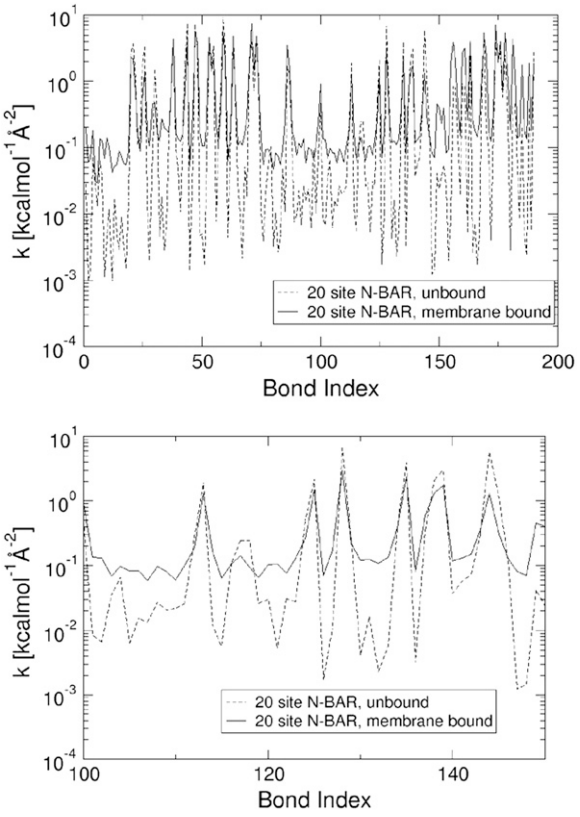

FIGURE 5 (*Top*) Comparison of spring constants determined by a heteroENM calculation between membrane-bound and unbound N-BAR domains. The horizontal axis labels the CG harmonic bonds; the vertical axis is logarithmic to emphasize the difference between the two cases. (*Bottom*) Close-up of *A* for bonds numbered 100–150. To test whether observed differences were significant, error bars were calculated by performing the heteroENM calculation separately for five adjacent 2-ns blocks of atomistic simulations, and computing the standard deviation of resultant spring constants. Error bars are typically on the order of 1–5%. Many of the spring constants differ by far more than the range of their error bars. Moreover, error bars address the convergence of the heteroENM model by demonstrating that *k* values computed from 2-ns blocks agree with those computed from the full 10-ns trajectory.

We computed error bars for the spring constants in Fig. 5 by parameterizing five different heteroENM models based on five independent blocks MD data of 2 ns each. We then compared the averages and standard deviations of the spring constants with the values obtained for a heteroENM based on a single, independent 10-ns MD simulation. The error bars are typically no more than 5% of the value of the spring constant, and agree within the errors with the longer timescale model. The extent to which the spring constants can be converged is the extent to which the pair fluctuations may be converged, and is likely system-dependent. Clearly, the method as implemented here is limited to systems with a well-defined average structure, but is also quite effective for such systems. This observation led us to pursue a harmonic approach to developing a multiscale description of remodeling by N-BAR domains.

DISCUSSION

We present a new multiscale method for parameterizing elastic models with heterogeneous networks of springs, based on atomistic-simulation data. Importantly, this method allows us to parameterize elastic models for the same protein in different environments. This we demonstrated for the N-terminal N-BAR domain of *Drosophilia* amphiphysin, by building models of the N-BAR domain membrane-bound and free in aqueous solvent. The heteroENM method revealed subtle differences in the effective harmonic interactions between the two cases, of the sort that are important for capturing motional correlations. In future work, we will base a much longer length-scale and timescale model of hundreds of oligomerized N-BAR domains on these elastic models, and we expect that capturing such differences will be crucial to the quantitative success of the model. Furthermore, the model may be constructed at whatever resolution is required. The parameterization ensures that the resulting model will capture the fluctuations observed in atomistic simulations. Importantly, no cutoff on the effective interaction range is enforced, so that the method may uncover effective interactions of quite long range, which is especially important when building models of an aggressively CG nature.

The heteroENM method is based on the assumption that the fluctuations observed in solvated molecular-dynamics simulations are more representative of thermal motions of solvated proteins than are crystallographic B-factors. This is a departure from the standard method used to parameterize homogeneous elastic networks, which chooses a uniform spring constant for all harmonic interactions to best-fit experimentally measured, per-residue mean-square fluctuations. Ours seems a reasonable approach, given that state-of-the-art, explicitly solvated simulations of proteins are routinely relied upon to draw inferences about the functional dynamics of proteins. We demonstrated that heteroENM can more accurately reproduce the fluctuations of solvated myoglobin than either the REACH method (25) or the more standard ENM approach (31).

However, the heteroENM method requires considerably more computation than the other two approaches discussed here. Aside from the atomistic MD simulation, each iteration requires a normal-mode calculation. For the case of myoglobin (151 α-carbon sites), $\sim$2000 iterations were required, and in the case of the 20-site N-BAR domain model, $\sim$20,000 iterations were required. One might object that such effort is not justified to model accurately, for example, an overall stiffening of 0.25 kcal mol^{-1}Å^{-2} of the N-BAR domain upon binding of the membrane. However, the method is intended for accurate parameterization of pieces that will be used to model processes on much longer length-scales and timescales. In the case of the N-BAR domain, an error of 0.25 kcal mol^{-1}Å^{-2} per dimer (in a model that may ultimately contain hundreds or thousands of dimers) will result in a complex that is far too soft.

We also demonstrated that heteroENM has predictive validity beyond the regime in which the atomistic simulation was performed. This was shown for the F-actin filament, where a model, parameterized as based on a nanometer-scale atomistic simulation, predicted with remarkable accuracy the experimental persistence length of $\sim$8.5 μm. This was achieved using a 35-nm CG filament without periodic boundary conditions. Periodic boundaries were essential in accurately computing the persistence length from the atomistic simulation. This calculation serves as a validation of this method, which differs from previous work (19) in that 1), only pairwise harmonic interactions are included, and 2), all interactions are allowed, resulting in a ''hands-off'' model parameterization.

Apart from providing a simple way to determine effective harmonic interactions for CG elastic models, the results of a heteroENM calculation function as an analysis tool. Models of the N-BAR domain bound and unbound provided insights into the effects of the membrane on the protein's structural properties. In other cases, a similar approach could yield insights into functional changes brought about upon ligand or substrate binding, revealing unexpected changes in elasticity far from the active site. Automated ''elasticity screening'' of this kind might reveal sites for mutagenesis studies or novel drug targets. Because this idea relies on changes in the stiffness of spring constants, it is only possible within a heteroENM-type approach.

Finally, the method is straightforward, simple, and readily applicable to the development of quasiharmonic CG models of any size or resolution, provided the system lends itself to such a description. The resulting models are easily incorporated into standard molecular-dynamics software, which ought to facilitate the development of multiscale models of long length and timescale processes incorporating CG elastic models. We intend to pursue this approach in future work.

E.L. benefited from numerous discussions with members of Ivet Bahar's group, past and present, especially Chakra Chennubhotla and Ivet Bahar. Kei Moritsugu and Jeremy Smith graciously provided myoglobin simulation data for comparison with REACH, and Richard Swenson provided atomistic N-BAR data.

Funding was provided by the National Science Foundation through a Collaborative Research in Chemistry grant (CHE-0628257) and by the National Institutes of Health (R01 GM063796).

REFERENCES

1. Tozzini, V. 2005. Coarse-grained models for proteins. *Curr. Opin. Struct. Biol.* 15:144–150.
2. Ayton, G. S., W. G. Noid, and G. A. Voth. 2007. Multiscale modeling of biomolecular systems in serial and in parallel. *Curr. Opin. Struct. Biol.* 17:192–198.
3. Ma, J. 2005. Usefulness and limitations of normal mode analysis in modeling dynamics of biomolecular complexes. *Structure*. 13:373–380.
4. Rader, A. J., C. Chennubhotla, L.-W. Yang, and I. Bahar. 2006. The gaussian network model: theory and applications. *In* Normal Mode Analysis. Theory and Applications to Biological and Chemical

Systems. Q. Cui, and I. Bahar, editors. New York, Taylor and Francis Group. 41–64.

5. Halioglu, T., I. Bahar, and B. Erman. 1997. Gaussian dynamics of folded proteins. *Phys. Rev. Lett.* 79:3090–3094.
6. ben-Avraham, D. 1993. Vibrational normal mode spectrum of globular proteins. Phys. Rev. [B.] 47:14559–14560.
7. Tirion, M. M. 1996. Large amplitude elastic motions of proteins from single parameter, atomic analysis. *Phys. Rev. Lett.* 77:1905–1908.
8. Bahar, I., and R. L. Jernigan. 1997. Inter-residue potentials in globular proteins and the dominance of highly specific hydrophilic interactions at close separation. *J. Mol. Biol.* 266:195–214.
9. Tobi, D., and I. Bahar. 2005. Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state. *Proc. Natl. Acad. Sci. USA.* 102:18908–18913.
10. Bahar, I., A. R. Atilgan, M. C. Demirel, and B. Erman. 1998. Vibrational dynamics of folded proteins: significance of slow and fast motions in relation to function and stability. *Phys. Rev. Lett.* 80: 2733–2736.
11. Yang, L.-W., X. Liu, C. J. Jursa, M. Holliman, A. J. Rader, H. A. Karimi, and I. Bahar. 2005. iGNM: a database of protein functional motions based on the gaussian network model. *Bioinformatics.* 21: 2978–2987.
12. Wang, Y., A. J. Rader, I. Bahar, and R. L. Jernigan. 2004. Global ribosome motions revealed with elastic network model. *J. Struct. Biol.* 147:302–314.
13. Maragakis, P., and M. Karplus. 2005. *J. Mol. Biol.* 352:807–822.
14. Petrone, P., and V. S. Pande. 2006. Can conformational change be described by only a few normal modes? *Biophys. J.* 90:1583–1593.
15. Yang, L., G. Song, and R. L. Jernigan. 2007. How well can we understand large-scale protein motions using normal modes of elastic network models? *Biophys. J.* 93:920–929.
16. Keskin, O., I. Bahar, D. Flatow, D. G. Covell, and R. L. Jernigan. 2002. Molecular mechanisms of chaperonin GroEL-GroES function. *Biochemistry.* 41:491–501.
17. Chennubhotla, C., Z. Yang, and I. Bahar. 2008. Coupling between global dynamics and signal transduction pathways: a mechanism of allostery for chaperonin GroEL. *Mol. Biosyst.* 4:787–792.
18. Tama, F., M. Valle, J. Frank, and C. L. Brooks III. 2003. Dynamic reorganization of the functionally active ribosome explored by normal mode analysis and cryo-EM microscopy. *Proc. Natl. Acad. Sci. USA.* 100:9319–9323.
19. Chu, J.-W., and G. A. Voth. 2006. Coarse-grained model of the actin filament derived from atomistic-scale simulations. *Biophys. J.* 90:1572–1582.
20. Chu, J.-W., and G. A. Voth. 2005. Allostery of actin filaments: molecular dynamics simulations and coarse-grained analysis. *Proc. Natl. Acad. Sci. USA.* 102:13111–13116.
21. Van Wynsberghe, A. W., and Q. Cui. 2005. Comparison of mode analyses at different resolutions applied to nucleic acid systems. *Biophys. J.* 89:2939–2949.
22. Van Wynsberghe, A. W., and Q. Cui. 2006. Interpreting correlated motions using normal mode analysis. *Structure.* 14:1647–1653.
23. Hinsen, K., and G. R. Kneller. 1999. A simplified force field for describing vibrational protein dynamics over the whole frequency range. *J. Chem. Phys.* 111:10766–10769.
24. Ming, D., and M. E. Wall. 2005. Allostery in a coarse-grained model of protein dynamics. *Phys. Rev. Lett.* 95:198103.
25. Moritsugu, K., and J. C. Smith. 2007. Coarse-grained biomolecular simulation with REACH: realistic extension algorithm via covariance Hessian. *Biophys. J.* 93:3460–3469.
26. Chen, X., Q. Cui, Y. Tang, J. Yoo, and A. Yethiraj. 2008. Gating mechanisms of mechanosensitive channels of large conductance, I: a continuum mechanics-based hierarchical framework. *Biophys. J.* 95: 563–580.
27. Peter, B. J., H. M. Kent, I. G. Mills, Y. Vallis, P. J. G. Butler, P. R. Evans, and H. T. McMahon. 2004. BAR domains as sensors of membrane curvature: The amphiphysin BAR structure. *Science.* 303: 495–499.
28. Zimmerberg, J., and S. McLaughlin. 2004. Membrane curvature: how BAR domains bend bilayers. *Curr. Biol.* 14:R250–R252.
29. Dawson, J. C., J. A. Legg, and L. A. Machesky. 2006. Bar domain proteins: a role in tubulation, cission, and actin assembly in clathrin-mediated endocytosis. *Trends Cell Biol.* 16:493–498.
30. Zhang, Z., L. Lu, W. G. Noid, V. Krishna, J. Pfaendtner, and G. A. Voth. 2008. A systematic methodology for defining coarse-grained sites in large biomolecules. *Biophys. J.* In press.
31. Atilgan, A. R., S. R. Durell, R. L. Jernigan, M. C. Demirel, O. Keskin, and I. Bahar. 2001. Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys. J.* 80:505–515.
32. Moritsugu, K., and J. C. Smith. 2005. Langevin model of the temperature and hydration dependence of protein vibrational normal modes. *J. Phys. Chem. B.* 109:12182–12194.
33. Phillips, J. C., R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, and K. Schulten. 2005. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* 26:1781–1802.
34. Holmes, K. C., D. Popp, W. Gebhard, and W. Kabsch. 1990. Atomic model of the actin filament. *Nature.* 347:44–49.
35. MacKerell, A. D., Jr., D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher III, B. Roux, M. Schklenkrich, J. C. Smith, R. Stote, J. Straub, M. Wantanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus. 1998. All-atom empirical potential for molecular dynamics studies of proteins. *J. Phys. Chem. B.* 102:3586–3616.
36. Allen, M. P., and D. J. Tildesley. 1987. Computer Simulation of Liquids. Oxford University Press, New York.
37. Martyna, G. J., D. J. Tobias, and M. L. Klein. 1994. Constant pressure molecular dynamics algorithms. *J. Chem. Phys.* 101:4177–4189.
38. Feller, S. E., Y. Zhang, R. W. Pastor, and B. R. Brooks. 1995. Constant pressure molecular dynamics simulation: the Langevin piston method. *J. Chem. Phys.* 103:4613–4621.
39. Brooks, B. R., R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. 1983. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* 4:187–217.
40. Blood, P. D., and G. A. Voth. 2006. Direct observation of Bin/amphiphysin/Rvs (BAR) domain-induced membrane curvature by means of molecular dynamics simulations. *Proc. Natl. Acad. Sci. USA.* 103:15068–15072.
41. Sheterline, P., J. Clayton, and J. C. Sparrow. 1998. Actin. Oxford University Press, New York.
42. Isambert, H., P. Venier, A. C. Maggs, A. Fattoum, R. Kassab, D. Pantaloni, and M.-F. Carlier. 1995. Flexibility of actin filaments derived from thermal fluctuations. *J. Biol. Chem.* 270:11437–11444.
43. Kabsch, W., H. G. Mannherz, D. Suck, E. F. Pai, and K. C. Holmes. 1990. Atomic structure of the actin:DNase I complex. *Nature.* 347: 37–43.
44. Vavylonis, D., Q. Yang, and B. O'Shaughnessy. 2005. Actin polymerization kinetics, cap structure, and fluctuations. *Proc. Natl. Acad. Sci. USA.* 102:8543–8548.
45. Kendrew, J. C., G. Bodo, H. M. Dintzis, R. G. Parrish, H. Wyckoff, and D. C. Phillips. 1958. A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature.* 181:662–666.
46. Elber, R., and Q. H. Gibson. 2008. Toward quantitative simulations of carbon monoxide escape pathways in myoglobin. *J. Phys. Chem. B.* 112:6147–6154.
47. McMahon, H. T., and J. L. Gallop. 2005. Membrane curvature and mechanisms of dynamic cell membrane remodelling. *Nature.* 438:590–596.

48. Blood, P. D., R. D. Swenson, and G. A. Voth. 2008. Factors influencing local membrane curvature induction by N-BAR domains as revealed by molecular dynamics simulations. *Biophys. J.* 95:1866–1876.

49. Farsad, K., and P. D. Camilli. 2003. Mechanisms of membrane deformation. *Curr. Opin. Cell Biol.* 15:372–381.

50. Zimmerberg, J., and M. M. Kozlov. 2006. How proteins produce cellular membrane curvature. *Nat. Rev. Mol. Cell Biol.* 7:9–19.

51. Fernandes, F., L. M. S. Loura, F. J. Chichón, J. L. Carrascosa, A. Federov, and M. Prieto. 2008. Role of helix 0 of the N-BAR domain in membrane curvature generation. *Biophys. J.* 94:3065–3073.

52. Ayton, G. S., P. D. Blood, and G. A. Voth. 2007. Membrane remodeling from N-BAR domain interactions: insights from multi-scale simulation. *Biophys. J.* 92:3595–3602.

53. Arkhipov, A., Y. Yin, and K. Schulten. 2008. Four-scale description of membrane sculpting by BAR domains. *Biophys. J.* 95:2806–2821.

54. Davis, I. W., W. B. I. Arendall, D. C. Richardson, and J. S. Richardson. 2006. The backrub motion: how a protein backbone shrugs when a sidechain dances. *Structure*. 14:265–274.

55. Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD—visual molecular dynamics. *J. Mol. Graph.* 14:33–38.