JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Elastic Network Models Are Robust to Variations in Formalism

Nicholas Leioatts, Tod D. Romo, and Alan Grossfield*

Department of Biochemistry and Biophysics, University of Rochester Medical Center, Rochester, New York 14642, United States

Supporting Information

ABSTRACT: Understanding the functions of biomolecules requires insight not only from structures but from dynamics as well. Often, the most interesting processes occur on time scales too slow for exploration by conventional molecular dynamics (MD) simulations. For this reason, alternative computational methods such as elastic network models (ENMs) have become increasingly popular. These simple, coarse-grained models represent molecules as beads connected by harmonic springs; the system's motions are solved analytically by normal-mode analysis. In the past few years, many different formalisms for performing ENM calculations have emerged, and several have been optimized using all-atom MD simulations. In contrast to other studies, we have compared the various formalisms in a systematic, quantitative way. In this study, we optimize many ENM functional forms using a uniform data set containing only long (>1 μs) all-atom MD simulations. Our results show that all models once optimized produce spring constants for immediate neighboring residues that are orders of magnitude stiffer than more distal contacts. In addition, the statistical significance of ENM performance varied with model resolution. We also show that fitting long trajectories does not improve ENM performance due to a problem inherent in all network models tested: they underestimate the relative importance of the most concerted motions. Finally, we characterize ENMs' resilience by tessellating the parameter space to show that broad ranges of parameters produce similar quality predictions. Taken together, our data reveal that the choice of spring function and parameters are not vital to the performance of a network model and that simple parameters can by derived "by hand" when no data are available for fitting, thus illustrating the robustness of these models.

1. INTRODUCTION

Structural changes in proteins are vital to activation and regulation, and many proteins function via conformational rearrangements that involve displacing large pieces of the molecule. These concerted or global motions often occur on the microsecond to millisecond time scale, making their study by all-atom computational methods intractable to all but a few research groups with highly specialized resources.[1−3] However, insight into protein dynamics is valuable, and even approximate interactions can help guide experiments and tailor hypotheses. For this reason, elastic network models (ENMs) have become very popular.[4] ENMs are a class of simple models for protein motion, where the system is represented as a network of beads interconnected by harmonic springs with lengths matching those found in the protein's native state; usually, the system is coarse-grained by only representing the C_α's of a protein. By applying normal-mode analysis (NMA) to this model, one can extract the collective motions of the system with minimal computational investment. Network models have garnered much attention due to their success in predicting global motions.[5−13] They can also be a practical aid in designing and addressing the feasibility of experiments; for example, these models have proven useful in evaluating the functional motions of proteins[6−9,14−16] as well as refining structural data.[17,19] Notably, because the normal modes of motion are solved analytically, there is no statistical error in their predictions, a problem that generally plagues molecular dynamics.[20]

ENMs' ease of implementation and computational efficiency makes them a suitable starting point for many structural studies; however, their utility is limited in some areas due to their simplicity.[19,21,22] To counter this lack of detail, numerous groups have developed features designed to augment the predictive power of ENMs.[23−34] We explore several such methods of improving ENMs while utilizing long molecular dynamics (MD) trajectories to evaluate each model's efficacy. Specifically, we examine the performance of different spring functional forms and resolutions, systematically analyzing their performance after fitting to simulation. For this study, we will use the anisotropic form of the network model first described by Atilgan et al.[23]

Historically, the performance of ENMs has been validated by comparison to crystallographic B-factors. However, this is a limited comparison since B-factors at best describe the flexibility of different regions of the proteins, usually isotropically. By contrast, the ENM-NMA provides far more information, describing the set of coherent motions available to the system, as well as their relative amplitudes. For this reason, we prefer to compare to molecular dynamics simulations, applying principal component analysis (PCA) to the trajectories to extract analogous information;[35] the eigenvectors have an analogous meaning in PCA and ENM-NMA. The eigenvalues have different meanings—magnitude in PCA and frequency in ENM-NMA—but given the harmonic nature of ENMs, one can simply invert the frequencies obtained from ENM-NMA to give a unitless measure of the relative magnitude of motion. In the current study, we will compare harmonic NMA and quasi-harmonic PCA eigendecompositions[15,36,37] (details of this comparison are outlined in section 2.2).

Recently, there have been many articles aimed at refining the functional form of the spring contacts in network models using

Received: January 17, 2012
Published: June 5, 2012

   dx.doi.org/10.1021/ct3000316 | J. Chem. Theory Comput. 2012, 8, 2424−2434

MD. This is a worthwhile undertaking as all-atom MD is considered the gold-standard among molecular simulation techniques. However, in each case, short simulations were used, leading to concern about the statistical quality of the data[20,35,38−40] as well as systematic errors.[35] For this evaluation, we chose to use a smaller parametrization set of three proteins for which we had long trajectories available (each greater than 1 μs in length). We hypothesized that fitting long trajectories was vital since it is precisely the slow motions that we extract from the NMA of ENMs. In a recent publication, however, Liu et al. suggested that simulations spanning a range of 2 orders of magnitude produce similar dominant motions.[41] This is an interesting result that cannot be overlooked; however, they consider only the most dominant motion, as opposed to comparing the whole fluctuation space, which was done in the current study. We previously showed that fitting to short simulations leads to systematic errors which manifest in the eigenvalue spectra.[35] This result will be further explored in the present manuscript.

We also evaluate the effects of varying model resolution. ENMs were originally developed as a lower-resolution alternative to all-atom NMA.[37,42,43] Each residue is typically represented by one bead placed at the C_α position, but occasionally the system size or level of detail restrains this decision. Several models exist for altering resolution,[19,26,44] but for this study we decided on two simple variations to extract a trend. While performance varied, the new resolutions were capable of predicting motions with a similar fidelity to the standard model.

2. METHODS

Many variants of the basic ENM method have been published, including a variety of functional forms to define the spring constants,[24,25,27,30−33,45] several levels of resolution,[19,26,44] and methods of interfacing with the surrounding environment.[21] There have also been many attempts to refine these models based on crystallographic B-factors[16,28,46] and by comparison to all-atom MD.[30−33,35,36] In the present work, we parametrize spring functions using only long MD simulations.

2.1. Elastic Network Models. ENMs are constructed from a network of beads connected by Hookean springs. These beads are typically located at the C_α's of each residue when studying proteins. In the current study, average structures were calculated from the MD trajectories using an iterative alignment.[47] Formally, this network of interactions is described by the following potential:

$$U_{ij} = k(R_{ij}^\circ)(R_{ij} - R_{ij}^\circ)^2 \tag{1}$$

where R_{ij} is the distance between two particular beads, i and j, R_{ij}° is their equilibrium distance, and $k(R_{ij}^\circ)$ is a variable spring function. The input structure is designated as the relaxed position of all the springs, making it the global minimum by definition. The functional form of $k(R_{ij}^\circ)$, or the spring function, can be defined by one of several formalisms. In addition to the descriptions below, explicit definitions of all spring functions can be found in the Supporting Information.

In the standard ENM, this spring function simply identifies connections. Two beads are designated as connected if they are within a cutoff distance, R_c, and not connected otherwise. Formally, this is defined by the Heaviside function:

$$k(R_{ij}^\circ) = \begin{cases} 1 & : R_{ij}^\circ < R_c \\ 0 & : R_{ij}^\circ \geq R_c \end{cases} \tag{2}$$

This is the simplest spring function, but it has a few significant weaknesses. First, we intuitively expect that the interactions between distant residues should be weaker than interactions between adjacent ones. Although this is captured in an approximate way by the use of a simple cutoff, one could arguably improve the behavior by dispensing with the cutoff and making the spring constant a smoothly varying function of the minimum energy distance, R_{ij}°. Moreover, this resolves a numerical instability found in the pure cutoff approach: if the cutoff is too short, then certain beads may not have enough interactions resulting in underdetermined dynamics. In that case, the resulting eigendecomposition will fail.[23]

One example of a spring that varies smoothly is an exponentially distance-dependent function:

$$k(R_{ij}^\circ) = \exp(-aR_{ij}^\circ) \tag{3}$$

Here, α is an adjustable scaling factor. This function connects all beads in the model while weighting the "stiffness" of the spring by the distance between beads. Several network models use this form (in part), including those introduced by Hinsen,[48] Kovacs et al.,[25] Moritsugu and Smith,[30] Lyman et al.,[31] Orellana et al.,[33] and Yang et al.[32]

A similar, more detailed model was proposed by Hinsen and co-workers.[24] In this formalism, referred to as HCA, beads closer than the distance R_c are connected by a linearly distance-dependent spring, while beads outside the cutoff are scaled by a power decay:

$$k(R_{ij}^\circ) = \begin{cases} a(R_{ij}^\circ) + b & : R_{ij}^\circ < R_c \\ c(R_{ij}^\circ)^{-d} & : R_{ij}^\circ \geq R_c \end{cases} \tag{4}$$

Here the parameters a, b, c, and d in addition to R_c are all adjustable.

Other methods explicitly take the chemical structure of the molecule into account by using different spring functions for virtual "bonded" interactions—those between residues adjacent in sequence—than for "nonbonded" interactions. An example similar to that of Hinsen et al. would be

$$k(R_{ij}^\circ) = \begin{cases} k_1 & : \text{bonded} \\ c(R_{ij}^\circ)^{-d} & : \text{nonbonded} \end{cases} \tag{5}$$

Here, one other small difference was introduced; instead of scaling "bonded" terms by distance, they are all given the same stiffness, k_1. In addition, multiple levels of connectivity can be implemented. For instance, the REACH[30] method uses virtual bonds, angles, and torsions, connecting first, second, and third nearest neighboring C_α's, respectively, followed by a nonbonded term.

Any of the above spring functions can be applied to eq 1, defining the potential energy between two beads. This is done for all pairs of beads in the system, and the resulting Hessian is then computed. The normal modes of this system can be solved by an eigendecomposition; the eigenvectors represent the directions of motion, and the eigenvalues represent their associated frequencies. A more detailed description of this diagonalization has been outlined previously (see for example refs 4, 23, and 49). In this work, all network models were

implemented using the LOOS "ElasticNetworks" package (freely available at http://loos.sourceforge.net).[50,51] LOOS is a lightweight structure analysis library that reads all major MD model and trajectory file formats.

2.1.1. Varying the Resolution of ENMs. While network models are generally abstracted to C_α resolution, we implemented and validated a higher and lower resolution as well. These can be used to fit the scope of a particular project, either looking for more subtle details[19] or predicting the motions of a very large system.[6−8,14]

Our higher resolution model incorporates a second bead per residue positioned at each side chain's center of mass (CoM) as previously employed by Bahar and Jernigan[52] and Zheng.[19] The lower resolution model uses a single bead for every two residues placed at the center of mass of the pair. Throughout this manuscript, we will refer to these as the two-bead per residue model and one-bead per two-residue model, respectively.

We applied each of the spring functions discussed above to all three resolutions. For models that take connectivity into account, we included bonds between neighboring C_α's and from each C_α to its side chain bead.

2.2. Validating Against Molecular Dynamics Simulations. To evaluate the normal modes of the network models, we compared them to microsecond-scale all-atom MD of three G protein-coupled receptors (GPCRs): rhodopsin (RHOD in figure legends), β_2 adrenergic receptor (β_2AR), and cannabinoid receptor 2 (CB2). These simulations were published separately and were also used in our prior ENM studies.[35,53−56] Briefly, the rhodopsin simulation contained 49 1-stearoyl-2-docosahexaenoyl-phosphatidylcholine (SDPC), 50 1-stearoyl-2-docosahexaenoyl-phosphatidylethanolamine (SDPE), and 24 cholesterol molecules and was conducted using the NVE ensemble. The simulation of β_2AR contained 99 1-palmitoyl-2-oleoyl-phosphatidylcholine (POPC) lipids and was conducted using the NVT ensemble. The CB2 simulation contained 123 POPC and 38 sn-2-arachidonoylglycerol molecules and was conducted using the NVT ensemble. All of these simulations were performed using the CHARMM27r force field for the lipids[57,58] and the CHARMM22 force field for the protein.[59]

In addition, simulations of two other proteins were used just for testing our results. Specifically, we analyze a pair of ∼1 μs trajectories of the A2A adenosine receptor (A2A) in a POPC bilayer, run by Lyman and co-workers.[60] For this analysis, we concatenated the two simulations, for a total simulation time of 2.2 μs. Additionally, we considered a 472 ns trajectory of opsin (the apo form of bovine rhodopsin), containing 123 SDPE lipids; further details of the simulation can be found in the Supporting Information.

As in previous work, we performed principal component analysis (PCA) on the trajectories taking advantage of the fact that PCA produces information analogous to that acquired from ENMs. While the PCA of a simulation depends on the covariance matrix and is explicitly not harmonic, its similarity to harmonic analyses like NMA have been extensively studied in the literature.[15,36,37] The eigenvectors have analogous physical meanings in the two formalisms: the direction of motion associated with a particular eigenvalue. The eigenvalues produce similar, although not identical, information. In the case of simulation PCA, they represent the relative amplitude of motion as sampled from a thermalized trajectory, while in network model NMA, they represent the relative frequencies of the motion. Because ENMs are explicitly harmonic in nature, the frequencies can simply be inverted to get effective amplitudes. In light of the similarities between these two analyses, we will use the quasi-harmonic PCA of a simulation to parametrize ENM-NMA results. We compare the results of PCA and ENM-NMA using the same methods used to compare two PCA results. In particular, we used the covariance overlap, because it uses the eigenvalues to account for the relative contribution of each eigenvector:[35,38]

$$\Omega_{A,B} = 1 - \left[\frac{\sum_i^{3N} (\lambda_i^{A} + \lambda_i^{B}) - \sum_i^{3N} \sum_j^{3N} \sqrt{\lambda_i^{A} \lambda_j^{B}} (\vec{\nu}_i^{A} \cdot \vec{\nu}_j^{B})^2}{\sum_i^{3N} (\lambda_i^{A} + \lambda_i^{B})} \right]^{1/2} \quad (6)$$

Here, λ represents the eigenvalue associated with the eigenvector, ν. For the ENM-NMA normal modes, the eigenvalues were inverted to produce quantities proportional to relative amplitude. The superscripts A and B denote the two results being compared while the subscripts i and j indicate this summation is over all $3N - 6$ modes (excluding the three purely translational and three purely rotational modes which have zero frequency). The inverted eigenvalues are normalized across a single eigendecomposition prior to application of eq 6, such that the total amplitude of predicted motion is identical. This metric scales from 0 to 1, where 0 indicates that the eigenvectors are orthogonal and 1 indicates complete overlap of the eigenspace. It is worth noting that this measure reduces to something close to the subspace overlap[61] if one assumes the eigenvalue spectrum contains only ones and zeroes. The parameters obtained from our fitting (using a Nelder-Mead Simplex[62] optimization) are documented (for all resolutions and spring functions) in Supporting Information Table S4.

2.2.1. Estimation of Equivalent MD Trajectory Length. The physical significance of a particular covariance overlap value is difficult to interpret in the absence of context. Therefore, in order to better understand the comparison between ENM-NMA results and PCA, we compare the performance of our ENM-NMA results to the sampling of an MD trajectory. To do this, we combined the covariance overlap with a block averaging approach as in previous work.[35,40] Block averaging allows assessment of the sampling quality of a small contiguous piece of the trajectory relative to the full simulation. Briefly, this entails splitting the trajectory into many contiguous blocks and computing the principal components of each block individually. The average covariance overlap between the principal components of these blocked trajectories and the full simulation PCA describes how well the short simulations reproduce the conformational space explored by the full simulation. As the length of the blocks increases, the sampling improves, resulting in higher covariance overlaps.

This method facilitates the appraisal of a network model's predicted motions relative to increasing block lengths. The block averaged covariance overlap is plotted as a function of block size, and the covariance overlap between the ENM-NMA and our gold standard (the full-length simulation PCA) is then overlaid.[35,40] The intersection of the block averaging curve and the ENM-NMA is the point where they have the same covariance overlap. Following this value to the x axis gives an estimate of the simulation time required before the trajectory would predict the system's motions as well as the ENM.

2.2.2. Comparing Different Resolution ENMs. In calculating covariance overlaps of different resolution models, we noticed that the dimensionality of the system intrinsically altered the covariance overlap; physically, this is related to the fact that in higher dimensional spaces it is harder to find two vectors with large dot products. Thus, we needed to find a means to compare models with varying resolution on an equal footing.

Our first idea was to compute a Z-score, comparing the eigendecomposition from a given ENM to that of a number of randomly generated covariance matrices. However, the random matrices tended to produce very flat eigenvalue spectra, as opposed to the rapidly decaying ones produced by ENM-NMA and simulation PCA. The eigendecompositions from the random matrices gave large covariance overlaps with each other, as they would with any other eigenset (set of eigenpairs), invalidating their use as a control.

To work around the covariance overlap's sensitivity to the shape of the eigenvalue spectrum, we randomly scrambled the eigenvectors such that they were associated with different eigenvalues. We then performed a bootstrap-like analysis, generating 10 000 eigensets in this manner and computing their covariance overlap to the full trajectory. This was used to compute the mean ($\bar{\Omega}$) and standard deviation (σ) of the resulting distribution. Finally, we calculated the Z-score for the ENM's eigenset using

$$Z = \frac{\Omega - \bar{\Omega}}{\sigma} \tag{7}$$

3. RESULTS

3.1. Any Spring Function can Perform Well. We quantified the performance of several different spring functions and resolutions of network models by fitting to MD simulations. In order to determine the robustness of the parametrization, each spring function was fit to the entire data set (all three proteins) and to each trajectory individually. The results in Figure 1 and Supporting Information Tables S1−S3 show that all spring functions considered yielded models which predict motions with similar fidelity. Figure 1A shows the results for spring functions parametrized using the β_2AR data. The covariance overlap between an ENM using the indicated spring function and the PCA of the MD simulation was plotted. Panels B and C show equivalent data for rhodopsin and CB2, respectively. In each case, the data are clustered by model resolution.

We found that after fitting, all spring functions for a particular protein and resolution produced similar overlaps. This result was quite surprising considering the scope of the literature based on defining and improving spring functions. Comparison of the various potentials revealed the basis of this similar performance; as shown in Figure 2, all of the models produce very strong interactions between each bead and the first surrounding shell of beads, while more distant interactions are roughly an order of magnitude weaker. The standard anisotropic network model, with spring contacts represented by the Heaviside step function, also follows this pattern: the fitting produces a very short cutoff such that only immediately neighboring beads are connected. A close inspection of the literature reveals this same trend of stiffening close contacts in the initial development of these methods.[24,37,30,31,45] In fact, the original Rouse model only connected beads neighboring in sequence.[63] Specifically addressing ENMs, Hinsen et al. found the transition to be 4.0 Å,[24] and Moritsugu and Smith gave

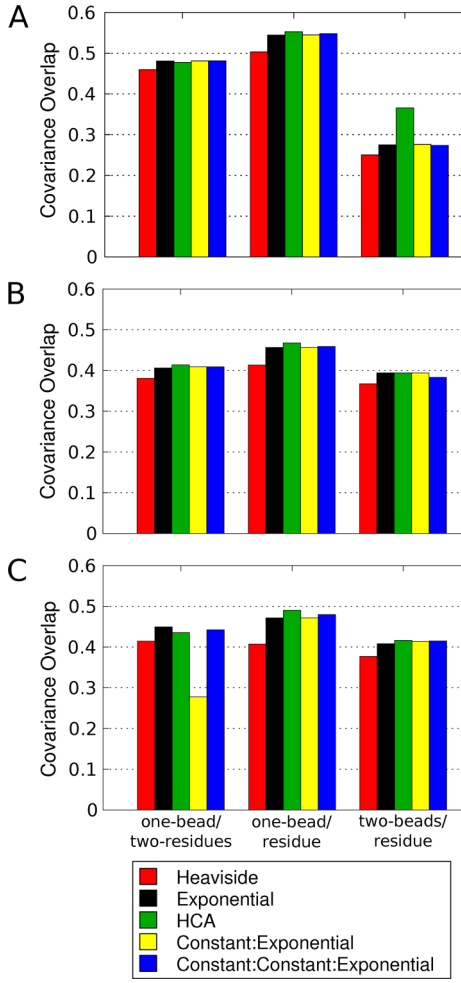

Figure 1. Covariance overlap, eq 6, between an ENM result and the full length simulation PCA for representative spring functions clustered by model resolution. "Heaviside" refers to the standard distance cutoff method described by eq 2. "Exponential" form is defined by eq 3. "HCA" is defined by eq 4, and "constant" implies a scalable constant value, for instance k_1 in eq 5. Spring names separated by colons (:) represent functions where explicit bonding was used. Here, C_α's neighboring in sequence are bonded (n to $n + 1$). When the two-bead per residue resolution is used, each Cα bead is bonded to its own C_{CoM} bead as well. The hierarchy of explicit levels of bonding takes the form (n to $n + 1$:n to $n + 2$:...:n to nonbonded). All values reported were fit to the full simulation using a Nelder-Mead Simplex[62] optimization to maximize the covariance overlap. Data are shown for (A) β_2AR, (B) rhodopsin, and (C) CB2.

virtual bonds a stiffness 2 orders of magnitude higher than virtual angles.[30] The vertical dotted line drawn at 3.8 Å in Figure 2 represents the mean distance of the first C_α shell.

3.2. Performance Is Affected by Model Resolution. In examining Figure 1, it is immediately clear that ENM performance is affected by model resolution. For example, the two-bead per residue model yielded a lower covariance overlap in all cases. The overlaps and spring parameters used are completely documented in Supporting Information Tables S1−S4. The lower overlap appears to be due to three phenomena. First, a higher dimensional system attempts to predict the dynamics with higher detail, in effect giving the ENMs a tougher problem to solve. Second, side chains often have more than one rotameric minimum and thus do not obey the underlying harmonic assumption characteristic of network

dx.doi.org/10.1021/ct3000316 | J. Chem. Theory Comput. 2012, 8, 2424−2434

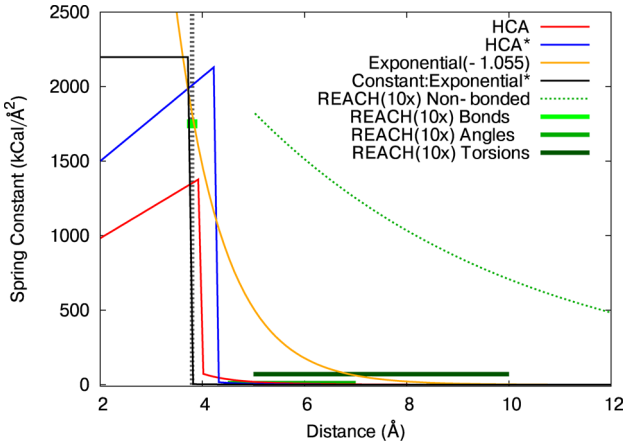

Figure 2. Spring stiffness as a function of distance. Several formalisms are shown. "HCA" refers to the function developed by Hinsen et al.[24] and defined here by eq 4. "HCA*" is the same function but uses values obtained from our parametrization.[35] "Exponential" is defined by eq 3, using a = 1.055, which gave the highest covariance overlap in our simultaneous fit of all three trajectories (see Supporting Information Table S4) when using the standard one-bead per residue resolution. "Constant:Exponential" refers to eq 5, after fitting (parameters in Supporting Information Table S4). Lastly, the parameters from the realistic extension algorithm via the covariance Hessian (REACH)[30] method are shown. Here, springs are split into four hierarchical levels; virtual bonds, angles, and torsions and a nonbonded term. The parameters reported in that study were multiplied by a factor of 10 for easier visualization. The vertical dotted line drawn at x = 3.8 Å represents mean C_α distance.

model normal modes. Finally, working in a higher dimensional space intrinsically reduces the dot products between vectors, since the entropic penalty for making two vectors parallel increases dramatically.

This shows up in the PCA results from the MD as well; Figure 3 shows that the block averaged covariance overlap is lower than in the one-bead per residue resolution PCA. However, the drop in ENM overlap when changing resolution outpaced the new block averaging result. This is illustrated in Figure 4, which shows how the various ENMs perform for rhodopsin. In panel A, the similarity to MD of the one-bead per

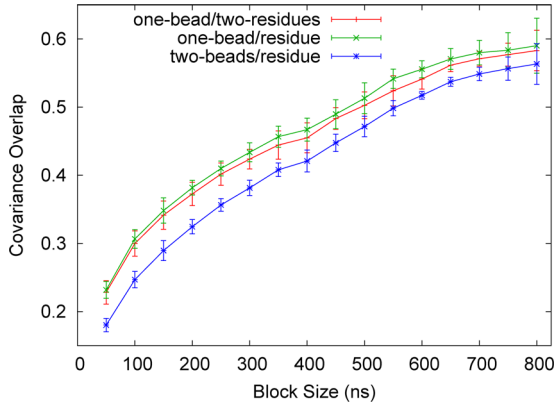

Figure 3. Block covariance overlap of rhodopsin for all three resolutions. The overlap between shorter contiguous simulation blocks are plotted against the full length trajectory with block length increasing along the x axis. Both the higher and lower resolution models (blue and red curves, respectively) converge more slowly than the one-bead per residue resolution (green curve).

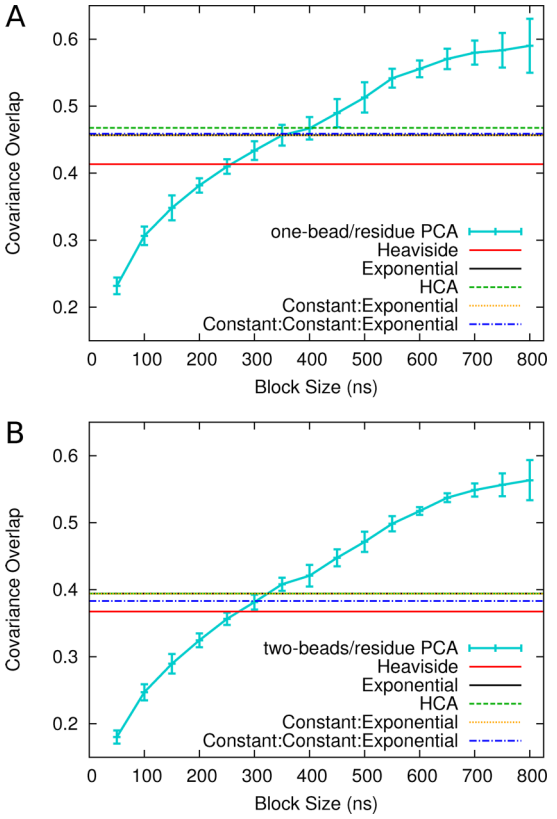

Figure 4. Rhodopsin simulation block covariance overlap plotted with ENM covariance overlaps. As in Figure 3, overlap is computed against the full trajectory. Here, ENM covariance overlaps are overlaid using the same coloring and naming convention as Figure 1. (A) Results of the standard one-bead per residue resolution, as plotted in Figure 1B, middle column. After fitting, all spring functions besides the simple distance cutoff achieved a similar overlap (0.47−0.49) which is equivalent to the block averaging result for a 375−450 ns simulation. (B) Two-bead per residue resolution (described in section 2.1.1), as plotted in Figure 1B, right column. Here, the blocked covariance overlap converges more slowly, but the ENM overlap is also lower than the standard resolution model. The covariance overlaps are again very similar between ENMs, scoring about 0.41, which was comparable to the average overlap of 350 ns trajectory blocks. Again, the distance cutoff method performed slightly worse than the more sophisticated spring functions.

residue models is plotted over the block averaging result. With standard resolution, the ENM normal modes sample approximately as well as a 400−450 ns simulation. This contrasts with panel B, which shows the two-bead per residue result where the ENMs reproduce the dominant fluctuations with the accuracy of about 350 ns of simulation. For all spring functions tested, the higher-resolution model produced modes less similar to the MD.

3.2.1. Resolution Impacts Significance of Covariance Overlap. To assess the statistical significance of the lower covariance overlaps, we used a standard measure, the Z-score. A bootstrapping-like procedure was used to generate a large number of fake data sets with the appropriate dimensionality and eigenvalue spectrum (see section 2.2.2 for details). The Z-score analysis showed that the significance of a particular covariance overlap increased with model resolution. The aggregate Z-scores are plotted against their respective covariance overlaps in Figure 5. Here, each datum is a particular spring formalism. The significance, which is plotted

dx.doi.org/10.1021/ct3000316 | J. Chem. Theory Comput. 2012, 8, 2424−2434

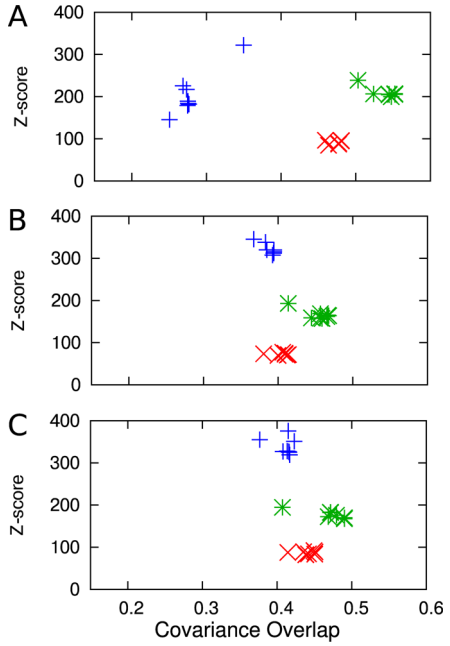

Figure 5. Significances of covariance overlaps are dependent on model resolution. Z-scores are plotted by their covariance overlap with the full simulation PCA. We first computed the covariance overlap between ENM-NMA and the simulation PCA. After this, the eigenpairs were scrambled (as described in section 2.2.2). The new eigenset was then compared to the simulation PCA. This was done for 10 000 iterations, and the mean and standard deviation of the resulting covariance overlaps was used to calculate a Z-score for the original result. Data points in red used the one-bead per two-residues resolution. Green points represent one-bead per residue resolution, and blue points represent two-beads per residue resolution. Data are shown for all three proteins studied: (A) β_2AR, (B) rhodopsin, and (C) CB2.

as a function of covariance overlap, shows that the same overlap is more difficult to achieve in the higher resolution models.

3.3. Fitting to a Single Protein is Sufficient. Our next step was to examine ENM dependence on the data set used for parametrization. To do so, we performed a cross-validation. Specifically, we fit ENMs to each MD trajectory individually and used the resulting spring parameters to compute the motions for the other proteins. For completeness, we also tested the performance of ENMs fit to the three trajectories in aggregate. The results, shown in Figure 6, indicate very little difference in performance regardless of which data set was used for fitting.

The lack of improvement upon individualized fitting stems from homogeneity in the spring parameters. As mentioned in section 3.1, all functional forms tend to favor stiffening the first shell interactions much more than distal contacts. In fitting eq 2, this was realized by a shortened cutoff distance, which can be seen in Table 1. In all cases, using the new parameters increased the covariance overlaps to their respective simulation PCA results.

Although this result may be due in part to the similarity of the three proteins used for fitting (all class A GPCRs), there is some level of generality in fitting network model parameters—at least among the proteins tested. In order to further validate our parameters, we compared them to simulations of two additional GPCRs not used for fitting: A2A (2.2 μs in aggregate) and opsin (472 ns). The opsin case is particularly challenging, since the structure more closely resembles an active GPCR structure, while the others clearly model the inactive forms of the protein. Table 1 lists their covariance overlaps using the typical 15 Å cutoff (before fitting) and the cutoffs we obtained via parametrization. We found that the covariance overlaps for both proteins behave much as the ones used in the fitting: the fit parameters perform significantly better than the standard ones in both the one- and two-beads per residue models, with covariance overlaps comparable to those for the other proteins. The covariance overlaps of the one-bead per two-residues model were similar before and after fitting. This is unsurprising, because fitting only shortened the cutoff by ~1 Å.

3.4. Does the Quality of the Fitting Results Depend on the Trajectory Length? Our previous work suggested that fitting to short MD trajectories was a flawed strategy because, in addition to larger statistical errors, short trajectories also differ systematically from longer ones. Specifically, the eigenvalues computed from PCA on short trajectories decay more slowly. Physically, this means that short trajectories systematically underestimate the importance of the largest motions in the system. In that manuscript (ref 35), we argued that fitting against short trajectories would recapitulate these problems and thus should be avoided.

In the present study, we have shown that the overall quality of the results, as measured by covariance overlap, does not seem to depend sensitively on the data used in the fitting. To clarify the apparent contradiction, we explicitly tested ENM performance while varying the length of the trajectory used for fitting.

To do this, the first microsecond of each long trajectory was split into contiguous pieces, yielding sets of 25, 50, 100, 250, and 500 ns trajectories. These shorter trajectories were then used to parametrize the various ENM formalisms in the same manner described in section 2.2. The resulting parameters were then used to perform ENM calculations that were compared to the entire trajectory: 1.02 μs for β_2AR, 1.6 μs for rhodopsin, and 1.9 μs for CB2. Surprisingly, ENMs parametrized against short trajectories perform nearly as well as those fit to long trajectories, as shown in Figure 7.

Following the results of parametrizing with short trajectories, we wanted to further explore why ENM-NMA performance was not affected by the trajectory length used for fitting. We previously noted that the power spectra from PCA applied to trajectories varied systematically with the length of the trajectory, as seen in Figure 8.[35] Panel A shows that as the simulation time increases, the fractional contribution made by the lowest mode increases. Note that this is not due to larger motions by the protein itself—the RMSD to the starting structure plateaus within the first 100 ns—but rather due to the dynamics, as longer time-scale motions are gradually accessed by the trajectory.

From our previous data, we predicted an improvement of ENM power spectra after fitting. However, in the present work, we found that the difference between ENM power spectra after fitting a short or a long trajectory is negligible. An example is shown in panel B of Figure 8, using the HCA formalism. Here, the power spectra averaged over the entire ensemble were plotted for each of the trajectory lengths tested. This includes ENMs fit to each of the shorter trajectories along with the those fit to the full trajectory. The error bars represent one standard deviation from the ensemble average of shorter trajectories. The fractional contribution of a particular mode to the overall

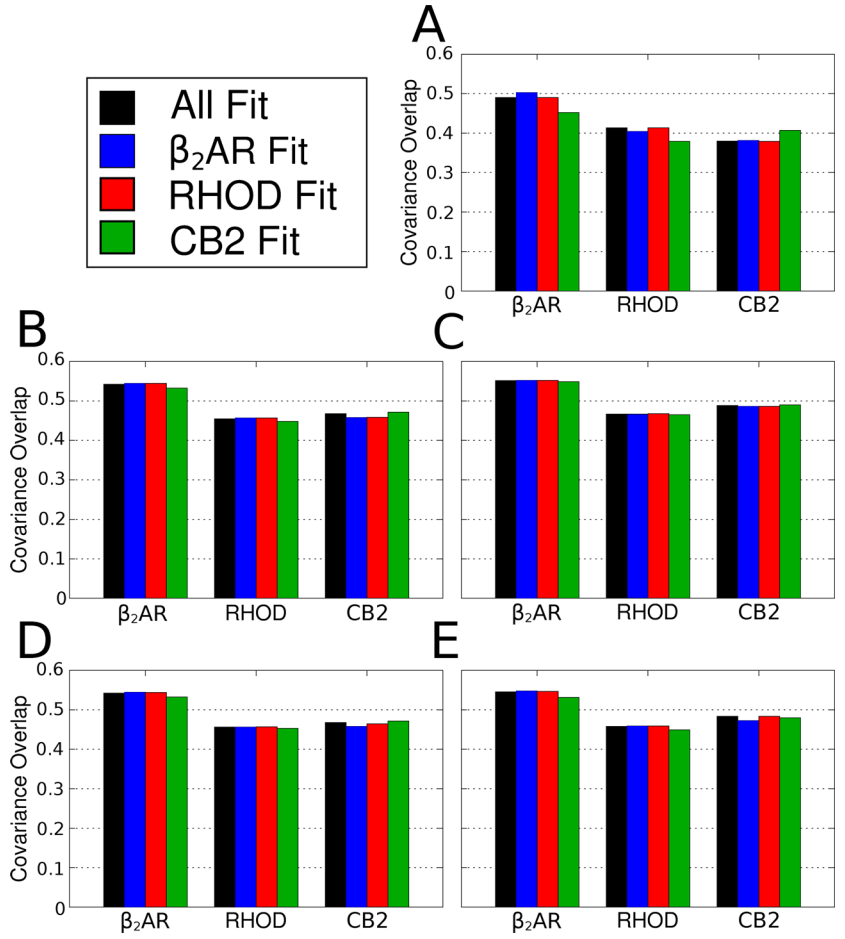

Figure 6. Simultaneous fitting yields similar accuracy as individual fitting. Here, data are clustered by the protein tested. The four bars in each cluster represent the different proteins which are used in parametrization. First, all three proteins are fit simultaneously (black); then individual fits to β_2AR (blue), rhodopsin (red), and CB2 (green) are conducted. In general, testing with the protein used for parametrization did not achieve a significantly higher covariance overlap. Simultaneous fitting also produced ENMs that performed almost as well as individualized fits. The naming convention from Figure 1 is repeated here. Results of (A) the standard distance cutoff ENM defined by the Heaviside function, eq 2, (B) the exponential function described by eq 3, (C) the HCA function defined in eq 4, (D) the spring function explicitly distinguishing "bonded" terms, eq 5, and (E) the spring function distinguishing both "bonded" and "angle" terms from "non-bonded" terms, eq S8.

motion is plotted for the first 10 eigenvalues in the ENM-NMA, with modes increasing along the horizontal axis.

Interestingly, the ENM-NMA power spectrum, shown in panel B, is actually slightly "softer" (the lowest modes contribute less to the overall motion, and the contribution decays more slowly) after fitting to longer trajectories. This trend was seen in all distance-dependent spring functions. However, the change in fractional contribution of the most dominant mode is much smaller than the change seen in the PCA of increasingly longer simulations. By contrast, the power spectrum of a simulation's principal components (Figure 8A) changes drastically with increasing time. Fitting improved the ENM-NMA power spectrum, but not enough to keep up with the changes in fractional contribution of the simulation's principal components. Thus, it appears that this flaw is fundamental to the ENM formalism. One possible explanation is that the lowest frequency modes in the MD trajectories are transitions between similar substates, as opposed to fluctuations about a single minimum. These motions are intrinsically absent in an ENM-NMA formalism, and this absence could account for the lack of improvement in their performance.

3.4.1. Low Frequency PCA Modes Show Transitions between Substates. In order to assess whether transitions appear in the simulation, we histogrammed the projection of our rhodopsin trajectory onto the lowest principal components (also called the right singular vectors, RSV).[35,64] The peaks in such a histogram indicate the highly populated conformational substates along a given mode. Multiple well-distinguished peaks suggest that there is more than one distinct state along that vector, whereas unimodal histograms indicate fluctuations about a single state. The latter case is well-modeled by a harmonic approximation, while the former is not. As can be seen in Figure 9, time series for the slowest modes of a one-bead per residue PCA show evidence of structural transitions—the motions along these modes have multiple peaks. These low-mode transitions indicate the sampling of separate substates within the most concerted motions in the trajectory. These are precisely the kinds of anharmonic motions that ENMs are unable to predict; even if they accurately detect the direction of motion, they will not be "aware" of the second energy well and will thus systematically underestimate the amplitude of the motion along the mode.

3.5. ENMs Are Robust to Changes in Parameters. In order to systematically explore the dependence of ENM performance on the chosen parameters, we explicitly tesselated the parameter space for one model (HCA) using the one-bead

Table 1. Covariance Overlaps Using the Standard Heaviside Function (eq 2) after Parametrization against β_2AR, Rhodopsin, and CB2 in Aggregate[a]

	covariance overlap		
protein	1/2	1/1	2/1
β_2AR	0.442	0.490	0.251
rhodopsin	0.380	0.413	0.358
CB2	0.413	0.380	0.374
before fit		R_c = 15 Å	
A2A	0.361	0.305	0.305
opsin	0.330	0.273	0.222
after fit	R_c = 13.93 Å	R_c = 8.636 Å	R_c = 8.628 Å
A2A	0.362	0.410	0.410
opsin	0.327	0.363	0.296

[a]The data are shown for all three resolutions tested, one-bead/two-residues (1/2), one-bead/residue (1/1), and two-beads/residue (2/1). A2A and opsin were not parametrized. Covariance overlaps between these ENM-NMAs and simulation PCA are reported first using a standard 15 Å cutoff distance and then after using the parameters obtained by fitting the other three proteins (cutoff distances indicated).

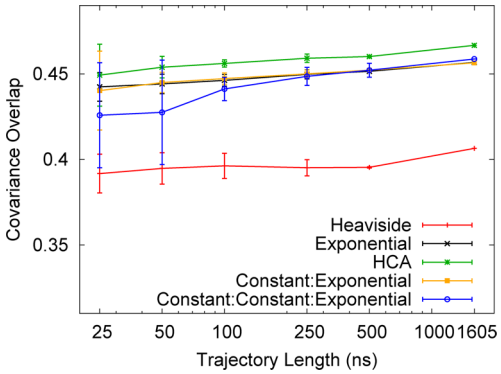

Figure 7. ENM performance does not depend on the length of trajectory used for parametrization. The first microsecond of each MD trajectory was broken into short contiguous pieces (of 25, 50, 100, 250, and 500 ns) and then used to fit the various ENM spring formalisms. The average covariance overlap across an ensemble ± one standard deviation is plotted as a function of the trajectory length used to parametrize the model (note the log-scaled x axis). Here, data are shown for rhodopsin using the typical one-bead per residue resolution. Curves show different spring functions, labeled as described in Figure 1. ENMs fit to longer trajectories do not have significantly higher covariance overlaps.

per residue resolution. We found the two most important parameters were the cutoff distance, R_c, and the exponent weighting, d. Figure 10 demonstrates that the covariance overlap is quite robust to systematic variation of these parameters. In Figure 10, the value of the exponential ("d" in eq 4) is plotted on the vertical axis and cutoff distance ("R_c" in eq 4) on the horizontal axis, with the resulting covariance overlap indicated by the heat map's intensity. This illustrates that a broad range of parameters (d = 7−14 and R_c = 1−5) yielded similar overlap with the simulation's principal components. This degeneracy in parameters partially explains ENM's robustness to the variations we tested, revealing how many spring functions and even shorter trajectory fits arrive at such similar results. Interestingly, at R_c < 3.8 Å, nearly all

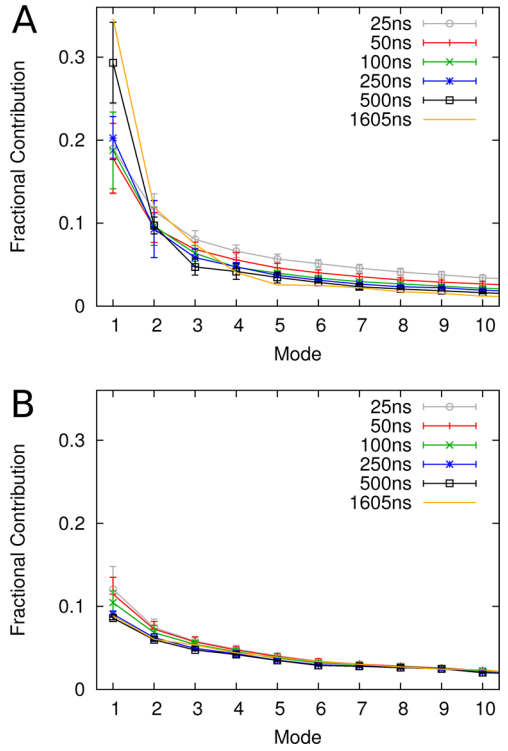

Figure 8. ENMs fail to accurately reproduce the significance of the lowest modes of motion. The fractional contribution of the 10 most significant modes (normalized contribution of the 10 lowest frequency eigenvalues) are plotted for (A) the PCA of the simulation and (B) the HCA formalism using rhodopsin. Data are compiled by averaging the eigenvalues across each ensemble of short simulations, for a given trajectory length. Error bars indicate one standard deviation in our ensemble of shorter trajectories. (A) The fractional contribution of the lowest mode increases with longer simulations. The spectrum from PCA of the full simulation is shown in orange. The first two modes contribute more to the variance as simulation time increases. (B) Network models fail to reproduce the significance of the lowest, most concerted motions. The key denotes the MD trajectory length used to parametrize each ENM. Data from the HCA formalism[24] eq 4 are shown. Similar results were obtained with other functional forms. In all trajectory lengths used for fitting, the ENM under-represents the importance of the lowest modes. In contrast to the molecular dynamics PCA, fitting longer trajectories appears to decrease the significance of these lowest modes.

connections are represented by the exponentially decaying spring. Thus, even changing the spring function does not significantly impact the overlap.

4. DISCUSSION

Through the use of long trajectories and an eigenvalue-sensitive metric, we were able to quantitatively show and clarify the underlying generality of network model parameters. We found that this robustness makes network models relatively easy to parametrize and that the choice of spring function makes little difference on performance. Interestingly, we also found essentially no dependence on the length of trajectory used for parametrization. However, this was largely because the network models were not able to accurately reproduce the eigenvalue spectrum of a simulation. Some other variation in the formalism that is able to better capture this power spectrum might well benefit from being fit only to long trajectories.

dx.doi.org/10.1021/ct3000316 | J. Chem. Theory Comput. 2012, 8, 2424−2434

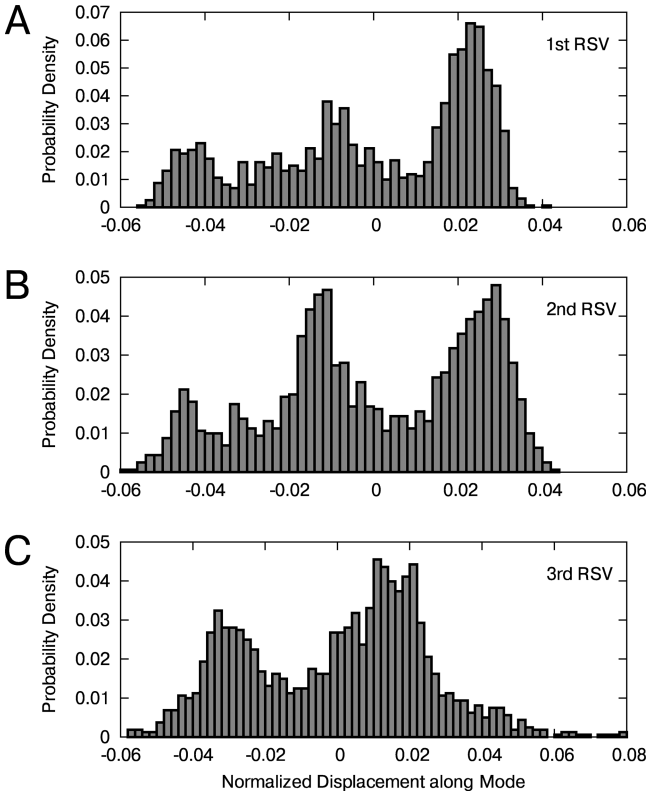

Figure 9. Histograms of right singular vectors (projection of motions onto the eigenvectors) show multiple states in the simulation's principal components. These histograms show the normalized displacement along a given principal component and were constructed using the PCA of C_α's from our rhodopsin simulation. Motions along the (A) first, (B) second, and (C) third components are shown. In all three plots, multiple peaks are present.

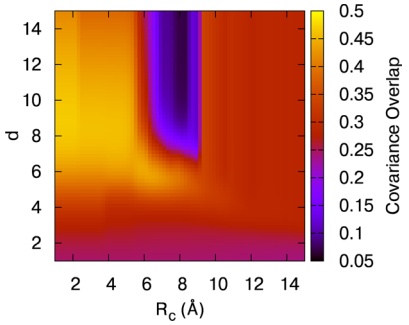

Figure 10. A broad range of parameters are capable of yielding high overlap with simulation. The parameters d and R_c of eq 4 were systematically varied and compared to the PCA of the full length simulation of rhodopsin, using the covariance overlap metric, eq 6. Data shown used the standard one-bead per residue resolution. This revealed a large area of high overlap (from roughly d = 6−14 and R_c < 5 Å).

4.1. ENMs Rely on Short-Range Contacts. With the numerous formalisms in print, we set out to measure the accuracy of each ENM with our set of long simulations. This data set has the advantage of reproducing fluctuations on a much longer time scale than was used for earlier ENM parametrizations, so there is a higher probability of capturing the slow motions predicted by a network model.

As shown in Figures 1 and 4, the choice of spring function has little bearing on the performance of a network model. Figure 2 brings the reason to light: all spring functions tend to favor much stiffer springs in the first C_α shell. Interestingly, Tozzini reported this same trend in a recent review of coarse-grained models for protein motion.[65] This evidence argues against any single functional form and elucidates their most striking commonality, reliance on the short-range contacts.

4.2. Higher Resolution Models Show Lower Covariance Overlap. Upon fitting higher resolution models, the covariance overlap decreased regardless of spring function. This decrease was more rapid than that found in block-averaged PCA results. In order to characterize this dependency on resolution, we implemented a Z-score test (using decoupled eigenpairs to calculate an average overlap as described in section 2.2.2). Although the absolute covariance overlap dropped, the statistical significance generally improved. Moreover, the simulation time required to produce equivalent covariance overlaps is longer for the higher resolution models (see for example Figure 3), suggesting that the higher resolution models do perform well.

4.3. Parameters Are Independent from Protein Parametrized. Although our parametrization set consisted of only three simulations, each of a class A GPCR, the result of cross-validating the parametrization of each protein is important. The high overlap when cross-validating ENM normal modes with various proteins reveals two key details. The first is that there is some level of generality within the proteins we tested, and the results of our parametrization can be used in the future to study GPCRs. To this effect, we tested our shorter Heaviside cutoffs on simulations of two other GPCRs, opsin and A2A. The results in Table 1 show that the fit parameters perform better than the defaults for both proteins, although the change is insignificant for the one-bead per two-residues resolution, and that the performance for these proteins is comparable to that for the ones used in the fitting. Thus, we conclude that ENMs are robust to variations in spring parameters, as illustrated by Figures 6 and 10. That said, we believe it is important in the future to extend this test set to other protein architectures, particularly globular proteins.

Our results explain the ease of finding high overlaps and, considered together with the rigidity of first shell contacts, show that any ENM can be quickly parametrized "by hand" by stiffening those contacts. Another method would be to discretely represent the first shell separately from all other contacts, as in eq 5. This is in fact used as part of the REACH[30] and edENM[33] formalisms, but the separation of more distal contacts presented in these methods was not found to be critical to ENM performance.

Taken together, these two pieces of evidence demonstrate the robustness of network models. In light of these results, we plotted the parameter space of the HCA formalism. Figure 10 illustrates the broad range of ENM parameters that yield high covariance overlaps. This demonstrates ENMs' insensitivity to the choice of parameters and functional form.

4.4. ENMs Are Insensitive to Trajectory Length Used for Parametrization. The results shown in Figure 7 were surprising. Our previous study showed that the fractional contribution of motions in MD simulations had a strong time dependence and that network models better reproduced this curve after fitting.[35] While it is clear from Figure 8A that the power spectrum of a simulation PCA result changes shape with trajectory length, closer inspection of panel B shows that the power spectrum of an ENM-NMA result still underestimates the contribution of the first few modes. Also, Figure 8 showed

that the power spectra of ENM-NMAs fit to different length trajectories varied little. Therefore, it appears that the parametrization improves overlap in a manner independent from the actual trajectory fit. This revealed a flaw inherent in all network models we parametrized: they were unable to correctly predict the relative importance of the lowest-frequency modes of long simulations. One possible explanation is shown in Figure 9. The first three principal components of the simulations contain transitions (data shown for rhodopsin). ENM-NMA, which uses a harmonic representation, is unable to predict such motions. In the future, an ENM that could accurately describe the power spectrum of a long simulation would likely benefit from fitting to only long trajectories. In this respect, the "double-native" Gō model proposed by Zuckerman is quite intriguing.[66]

In addition, incorporating more simulations of diverse molecules into this study without relaxing our restriction of trajectory length may improve our interpretation. GPCR transmembrane regions are quite rigid, so including other proteins would further elucidate the behavior of eigenvalue contributions. Including different classes of proteins would also be significant, as our results showed a very general trend in fitting—namely, stiffening the first shell contacts.

5. CONCLUSIONS

We evaluated many formalisms present in the elastic network model literature using a set of long (each > 1 μs) molecular dynamics simulations. The principal components of these simulations were the basis for comparison, evaluation, and fitting of network models. Fitting was preformed using an eigenvalue-sensitive metric, the covariance overlap, in order to capture the relative amplitude of motions present in both systems. Somewhat surprisingly, all of the formalisms showed roughly equivalent performance regardless of the data used for parametrization. This result could be viewed as disappointing—sophisticated efforts to improve the models are largely unsuccessful—or as informative due to the similarities in parameters between functional forms. Our results showed that fitting caused all formalisms to assign strong interactions between first shell contacts and much weaker interactions for longer range contacts.

This led us to tessellate the parameter space in several models. We found that the spring parameters were fairly robust with broad ranges yielding high covariance overlaps. Therefore, we conclude that while significant improvement can come from fitting the spring parameters, the formalism complexity has little bearing on the actual accuracy of the model. In addition, a high performance parametrization can likely be obtained easily by assigning stiff connections between C_α's neighboring in sequence and much weaker connections to more distal contacts.

■ ASSOCIATED CONTENT

Supporting Information

Definitions of all spring functions used in this study and details of the opsin simulation. In addition, the results of our standard score and covariance overlap analysis are available in Tables S1–S3, and the parameters obtained from fitting all three proteins in aggregate are reported in Table S4. This information is available free of charge via the Internet at http://pubs.acs.org/.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: alan_grossfield@urmc.rochester.edu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors would like to thank Joshua Horn, Dejun Lin, and Jessi Leioatts for their critical review and suggestions to the manuscript. We also thank Edward Lyman and Ji Young Lee for their generous donation of the A2A receptor simulations, the University of Rochester Center for Integrated Research Computing for the resources to run the opsin simulation, and Michael Pitman and the IBM T. J. Watson Research Center for the resources required to run the simulations of β_2AR, CB2, and rhodopsin. This work was supported by NIH project number 1R01GM095496, and N.L. is supported by the Institutional Ruth L. Kirschstein National Research Service Award, grant number GM068411.

■ REFERENCES

(1) Shaw, D. E.; Dror, R. O.; Salmon, J. K.; Grossman, J. P.; Mackenzie, K. M.; Bank, J. A.; Young, C.; Deneroff, M. M.; Batson, B.; Bowers, K. J.; Chow, E.; Eastwood, M. P.; Ierardi, D. J.; Klepeis, J. L.; Kuskin, J. S.; Larson, R. H.; Lindorff-Larsen, K.; Maragakis, P.; Moraes, M. A.; Piana, S.; Shan, Y.; Towles, B. Millisecond-scale molecular dynamics simulations on Anton. *Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis*; Association for Computing Machinery: New York, 2009.

(2) Allen, F.; Almasi, G.; Andreoni, W.; Beece, D.; Berne, B. J.; Bright, A.; Brunheroto, J.; Cascaval, C.; Castanos, J.; Coteus, P.; Crumley, P.; Curioni, A.; Denneau, M.; Donath, W.; Eleftheriou, M.; Fitch, B.; Fleisher, B.; Georgiou, C. J.; Germain, R.; Giampapa, M.; Gresh, D.; Gupta, M.; Haring, R.; Ho, H.; Hochschild, P.; Hummel, S.; Jonas, T.; Lieber, D.; Martyna, G.; Maturu, K.; Moreira, J.; Newns, D.; Newton, M.; Philhower, R.; Picunko, T.; Pitera, J.; Pitman, M.; Rand, R.; Royyuru, A.; Salapura, V.; Sanomiya, A.; Shah, R.; Sham, Y.; Singh, S.; Snir, M.; Suits, F.; Swetz, R.; Swope, W. C.; Vishnumurthy, N.; Ward, T. J. C.; Warren, H.; Zhou, R. *IBM Syst. J.* **2001**, *40*, 310.

(3) Shirts, M.; Pande, V. S. *Science* **2000**, *290*, 1903–1904.

(4) Bahar, I.; Lezon, T. R.; Bakan, A.; Shrivastava, I. H. *Chem. Rev.* **2010**, *110*, 1463–1497.

(5) Jernigan, R. L.; Demirel, M. C.; Bahar, I. *Int. J. Quantum Chem.* **1999**, *75*, 301–312.

(6) Tama, F.; Valle, M.; Frank, J.; Brooks, C. L. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 9319–9323.

(7) Wang, Y.; Rader, A. J.; Bahar, I.; Jernigan, R. L. *J. Struct. Biol.* **2004**, *147*, 302–314.

(8) Zheng, W.; Brooks, B. R. *Biophys. J.* **2005**, *89*, 167–178.

(9) Isin, B.; Rader, A. J.; Dhiman, H. K.; Klein-Seetharaman, J.; Bahar, I. *Proteins: Struct., Funct., Bioinf.* **2006**, *65*, 970–983.

(10) Yang, L.; Song, G.; Jernigan, R. L. *Biophys. J.* **2007**, *93*, 920–929.

(11) Zheng, W.; Brooks, B. R.; Thirumalai, D. *Biophys. J.* **2007**, *93*, 2289–2299.

(12) Niv, M. Y.; Filizola, M. *Proteins: Struct., Funct., Bioinf.* **2008**, *71*, 575–586.

(13) Ahmed, A.; Villinger, S.; Gohlke, H. *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 3341–3352.

(14) Chennubhotla, C.; Rader, A. J.; Yang, L.-W.; Bahar, I. *Phys. Biol.* **2005**, *2*, S173–S180.

(15) Rueda, M.; Chacóshon, P.; Orozco, M. *Structure* **2007**, *15*, 565–575.

(16) Yang, L.-W.; Eyal, E.; Chennubhotla, C.; Jee, J.; Gronenborn, A. M.; Bahar, I. *Structure* **2007**, *15*, 741–749.

(17) Delarue, M.; Dumas, P. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 6957–6962.

(18) Tama, F.; Miyashita, O.; Brooks, C. L. J. *Mol. Biol.* **2004**, *337*, 985−999.
(19) Zheng, W. *Biophys. J.* **2011**, *100*, 478−488.
(20) Grossfield, A.; Zuckerman, D. M. *Annu. Rep. Comput. Chem.* **2009**, *5*, 23−48.
(21) Woodcock, H. L.; Zheng, W.; Ghysels, A.; Shao, Y.; Kong, J.; Brooks, B. R. *J. Chem. Phys.* **2008**, *129*, 214109.
(22) Ma, J. *Structure* **2005**, *13*, 373−380.
(23) Atilgan, A. R.; Durell, S. R.; Jernigan, R. L.; Demirel, M. C.; Keskin, O.; Bahar, I. *Biophys. J.* **2001**, *80*, 505−515.
(24) Hinsen, K.; Petrescu, A.-J.; Dellerue, S.; Bellissent-Funel, M.-C.; Kneller, G. R. *Chem. Phys.* **2000**, *261*, 25−37.
(25) Kovacs, J. A.; Chacón, P.; Abagyan, R. *Proteins: Struct., Funct., Bioinf.* **2004**, *56*, 661−668.
(26) Micheletti, C.; Carloni, P.; Maritan, A. *Proteins: Struct., Funct., Bioinf.* **2004**, *55*, 635−645.
(27) Jeong, J. I.; Jang, Y.; Kim, M. K. *J. Mol. Graphics Modell.* **2006**, *24*, 296−306.
(28) Kondrashov, D. A.; Cui, Q.; Phillips, G. N. *Biophys. J.* **2006**, *91*, 2760−2767.
(29) Song, G.; Jernigan, R. L. *Proteins: Struct., Funct., Bioinf.* **2006**, *63*, 197−209.
(30) Moritsugu, K.; Smith, J. C. *Biophys. J.* **2007**, *93*, 3460−3469.
(31) Lyman, E.; Pfaendtner, J.; Voth, G. A. *Biophys. J.* **2008**, *95*, 4183−4192.
(32) Yang, L.; Song, G.; Jernigan, R. L. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 12347−12352.
(33) Orellana, L.; Rueda, M.; Ferrer-Costa, C.; Lopez-Blanco, J. R.; Chacón, P.; Orozco, M. *J. Chem. Theory Comput.* **2010**, *6*, 2910−2923.
(34) Yang, L.-W. *Biophys. J.* **2011**, *100*, 1784−1793.
(35) Romo, T. D.; Grossfield, A. *Proteins: Struct., Funct., Bioinf.* **2011**, *79*, 23−34.
(36) Doruker, P.; Atilgan, A. R.; Bahar, I. *Proteins: Struct., Funct., Bioinf.* **2000**, *40*, 512−524.
(37) Brooks, B. R.; Janěžič, D.; Karplus, M. *J. Comput. Chem.* **1995**, *16*, 1522−1542.
(38) Hess, B. *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.* **2002**, *65*, 031910.
(39) Lyman, E.; Zuckerman, D. M. *J. Phys. Chem. B* **2007**, *111*, 12876−12882.
(40) Romo, T. D.; Grossfield, A. *J. Chem. Theory Comput.* **2011**, *7*, 2464−2472.
(41) Liu, L.; Gronenborn, A. M.; Bahar, I. *Proteins: Struct., Funct., Bioinf.* **2011**, *80*, 616−625.
(42) Tirion. *Phys. Rev. Lett.* **1996**, *77*, 1905−1908.
(43) Bahar, I.; Atilgan, A. R.; Erman, B. *Folding Des.* **1997**, *2*, 173−181.
(44) Tama, F.; Gadea, F. X.; Marques, O.; Sanejouand, Y. H. *Proteins: Struct., Funct., Bioinf.* **2000**, *41*, 1−7.
(45) Ming, D.; Wall, M. E. *Phys. Rev. Lett.* **2005**, *95*, 198103.
(46) Riccardi, D.; Cui, Q.; Phillips, G. N. *Biophys. J.* **2010**, *99*, 2616−2625.
(47) Grossfield, A.; Feller, S. E.; Pitman, M. C. *Proteins: Struct., Funct., Bioinf.* **2007**, *67*, 31−40.
(48) Hinsen, K. *Proteins: Struct., Funct., Bioinf.* **1998**, *33*, 417−429.
(49) Eyal, E.; Yang, L.-W.; Bahar, I. *Bioinformatics* **2006**, *22*, 2619−2627.
(50) Romo, T. D.; Grossfield, A. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* **2009**, *2009*, 2332−2335.
(51) Romo, T. D.; Grossfield, A. *LOOS: Lightweight Object Oriented Structure analysis; Grossfield Lab*; University of Rochester Medical School: Rochester, NY. http://loos.sourceforge.net (accessed Mar 5, 2012).
(52) Bahar, I.; Jernigan, R. L. *J. Mol. Biol.* **1997**, *266*, 195−214.
(53) Lau, P.-W.; Grossfield, A.; Feller, S. E.; Pitman, M. C.; Brown, M. F. *J. Mol. Biol.* **2007**, *372*, 906−917.
(54) Grossfield, A.; Pitman, M. C.; Feller, S. E.; Soubias, O.; Gawrisch, K. *J. Mol. Biol.* **2008**, *381*, 478−486.
(55) Romo, T. D.; Grossfield, A.; Pitman, M. C. *Biophys. J.* **2010**, *98*, 76−84.
(56) Hurst, D. P.; Grossfield, A.; Lynch, D. L.; Feller, S.; Romo, T. D.; Gawrisch, K.; Pitman, M. C.; Reggio, P. H. *J. Biol. Chem.* **2010**, *285*, 17954−17964.
(57) Feller, S. E.; Gawrisch, K.; MacKerell, A. D. *J. Am. Chem. Soc.* **2002**, *124*, 318−326.
(58) Klauda, J. B.; Brooks, B. R.; MacKerell, A. D.; Venable, R. M.; Pastor, R. W. *J. Phys. Chem. B* **2005**, *109*, 5300−5311.
(59) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R., Jr.; Evanseck, J.; Field, M.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D.; Prodhom, B.; Reiher, W., III; Roux, B.; Schlenkrich, M.; Smith, J.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(60) Lee, J. Y.; Lyman, E. *Biophys. J.* **2012**, *102*, 2114−2120.
(61) Faraldo-Gómez, J. D.; Forrest, L. R.; Baaden, M.; Bond, P. J.; Domene, C.; Patargias, G.; Cuthbertson, J.; Sansom, M. S. P. *Proteins: Struct., Funct., Bioinf.* **2004**, *57*, 783−791.
(62) Nelder JA, M. R. *Comput. J.* **1964**, *7*, 308−313.
(63) Rouse, P. E. *J. Chem. Phys.* **1953**, *21*, 1272−1280.
(64) Romo, T. D.; Clarage, J. B.; Sorensen, D. C.; Phillips, G. N. *Proteins: Struct., Funct., Bioinf.* **1995**, *22*, 311−321.
(65) Tozzini, V. Q. *Rev. Biophys.* **2010**, *43*, 333−371.
(66) Zuckerman, D. M. *J. Phys. Chem. B* **2004**, *108*, 5127−5137.