PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Density-cluster NMA: A new protein decomposition technique for coarse-grained normal mode analysis

Omar N. A. Demerdash[1,2] and Julie C. Mitchell[3,4]*

[1] Medical Scientist Training Program, University of Wisconsin-Madison, Madison, Wisconsin
[2] Biophysics Program, University of Wisconsin-Madison, Madison, Wisconsin
[3] Department of Biochemistry, University of Wisconsin-Madison, Madison, Wisconsin
[4] Department of Mathematics, University of Wisconsin-Madison, Madison, Wisconsin

ABSTRACT

Normal mode analysis has emerged as a useful technique for investigating protein motions on long time scales. This is largely due to the advent of coarse-graining techniques, particularly Hooke's Law-based potentials and the rotational–translational blocking (RTB) method for reducing the size of the force-constant matrix, the Hessian. Here we present a new method for domain decomposition for use in RTB that is based on hierarchical clustering of atomic density gradients, which we call Density-Cluster RTB (DCRTB). The method reduces the number of degrees of freedom by 85–90% compared with the standard blocking approaches. We compared the normal modes from DCRTB against standard RTB using 1–4 residues in sequence in a single block, with good agreement between the two methods. We also show that Density-Cluster RTB and standard RTB perform well in capturing the experimentally determined direction of conformational change. Significantly, we report superior correlation of DCRTB with *B*-factors compared with 1–4 residue per block RTB. Finally, we show significant reduction in computational cost for Density-Cluster RTB that is nearly 100-fold for many examples.

Proteins 2012; 80:1766–1779.

Key words: multiscale; protein motion; dimension reduction; domain decomposition.

INTRODUCTION

Proteins motions encompass all time scales, but the most functionally relevant ones often occur at scales of micro- to millisecond. Such motions are highly collective in nature, involving many degrees of freedom; moreover, they are subject to solvent damping, rendering the motions anharmonic. Normal mode analysis (NMA) has emerged as a very popular technique for studying these motions, despite being a *harmonic* technique. It is harmonic in that it involves diagonalizing a force-constant matrix, or Hessian, H, into a matrix of eigenvectors (U) describing the directions of the harmonic motion and a diagonal matrix of eigenvalues (Λ), the square root of which are the frequencies associated *with the corresponding eigenvectors.*

$$H = U\Lambda U^{\mathrm{T}} \tag{1}$$

Elastic network models (ENM), like the anisotropic network model[1,2] and Gaussian network model (GNM),[3] have simplified NMA by reducing the potential to a single-parameter Hooke's law potential energy. The Gaussian network model is useful for predicting fluctuations but not directions of motions, as the three NMA degrees of freedom for a given atom are equal. ANM, however, is able to predict directions of motion and fluctuations. ENM has the key advantage of having a force-constant matrix that is easy to calculate and does not require a stringently minimized structure. The minimization required within the framework of MM is often a nontrivial task and has a tendency to distort the protein. In ENM, the original structure is automatically at a minimum with respect to the harmonic potential. Recently, a

Additional Supporting Information may be found in the online version of this article.

Grant sponsor: US Department of Energy Genomics: GTL and SciDAC Programs (DE-FG02-04ER25627)

*Correspondence to: Julie C. Mitchell, Department of Biochemistry, University of Wisconsin-Madison, 433 Babcock Drive, Madison, WI. E-mail: jcmitchell@wisc.edu.

Received 1 December 2011; Revised 13 February 2012; Accepted 12 March 2012
Published online 20 March 2012 in Wiley Online Library (wileyonlinelibrary.com).
DOI: 10.1002/prot.24072

promising method known as the minimalist network model was developed that allows MM force field-based NMA to be performed without prior minimization by modifying the original Hessian to be positive definite.[4]

Another advance in coarse-grained NMA is rotational-translational blocking (RTB).[5,6] In this technique, the original Hessian matrix is projected onto a smaller subspace defined by rigid groups of atoms. The resulting eigenvectors represent combined rigid motions on the blocks. Reduction of the Hessian to be diagonalized saves substantial computational effort. In many RTB methods, rigid blocks are defined largely as a single residue or short polypeptide. In their original report of the RTB method, Tama *et al.* did attempt coarser partitioning by dividing the system into blocks according to secondary structure. This blocking performed reasonably well, though not as good as a finer, 1–5 residue-per-block partitioning.

It has been shown that both the RTB and the ENM yield normal modes that are comparable to traditional NMA derived from molecular mechanics force fields but come at a greatly reduced computational cost.[7,8] Of clear importance is that the normal modes (usually some combination of low-frequency modes) adequately approximate the experimentally determined direction of motion. Krebs *et al.*[8] showed that the three slowest modes can often approximate biologically relevant conformational change.

Others have undertaken further efforts to develop coarse-grained NMA approaches that either reduce computation time or improve accuracy. Ahmed and Gohlke[9] extended the RTB method by partitioning the protein into blocks after performing rigidity analysis with FIRST.[10] In this approach, a group of residues determined to be rigid according to FIRST, and not necessarily contiguous in the primary sequence, are grouped into a single block. An important motivation for this approach is the recognition that nonlocal interactions are important determinants of a protein's functionally relevant deformations, in addition to the local interactions underlying secondary structure.[6] Tama *et al.* also performed blocking by taking the same block lengths produced by the secondary structure blocking but instead randomly distributing these stretches over the protein without regard for the actual secondary structure. When this was done, the resulting normal modes performed equivalently in describing the protein's conformational change compared to secondary structure blocking, suggesting that blocking performed linearly along the sequence is more dependent on block length rather than the specific secondary structure.

Schuyler and Chirikjian also developed a method for defining rigid blocks, but did not implement their partitioning within the RTB framework.[11] Rather than projecting an all-atom Hessian into a smaller subspace defined by the blocks, they construct the stiffness matrix describing the interactions among blocks *after* defining the blocks. As the emphasis of their report was the theory of their mechanical model, they did not develop a blocking method based on any physical parameters. Rather, their clustering approach is ad hoc, either grouping helices together and partitioning the rest of the protein uniformly in the original application[11] or, in a later application,[12] simply assigning the blocks according to chain or, in one example, according to obvious geometric boundaries that conveniently coincided with boundaries of functional domains. Nonetheless, they generate physically meaningful motions at small computational cost with their technique.

Efforts have also been undertaken to model flexibility of the individual blocks, which are rigid under standard RTB. To this end, vibrational subsystem analysis (VSA) was developed.[13–15] In the application of VSA to large systems,[13] a given block was divided into subsystem and environment, where the atoms in the environment fluctuate under the influence of the atoms of the subsystem. Significantly, VSA performed significantly better than standard RTB, as demonstrated by the greater agreement of VSA eigenvectors and eigenvalues with all-atom NMA compared with RTB, but was consistently slower and was not assessed in terms of its ability to reproduce experimentally determined flexibility.

The application of domain decomposition techniques is of value beyond normal mode analysis. Many previously reported domain identification schemes perform flexibility analysis on multiple experimentally determined structures of a given protein or on fluctuation models calculated from NMA or molecular dynamics (MD). Several techniques rely on clustering of regions of the protein based on similarity of RTB rotational vectors[16,17] or similarity of rotational and translation vectors combined with prior deformation analysis.[18] Other methods rely on clustering of motional covariance from NMA or MD[19–22] or examination of the directions of motion from a single GNM-NMA eigenvector.[23] Still others rely on fragmentation and similarity in Cα distances among multiple known structures.[24–26]

A few newer methods are able to decompose a protein into domains based on a single structure. Keating *et al.*[27] developed a method known as StoneHinge that decomposes a protein based on an extension of the aforementioned rigidity analysis program FIRST. In contrast to these techniques, which are based on kinematic metrics, FlexOracle partitions the protein according to an energetic analysis of cleaved fragments, where cleavage that is energetically favorable is hypothesized to correspond to hinge regions, while cleavage of a putative domain would result in a comparatively less favorable energy.

In the current work, we develop a method for protein decomposition based on a single structure that requires no prior computation of motion from NMA or MD trajectories for implementation in RTB NMA. The method is computationally efficient, is able to group residues that

are not contiguous in primary sequence, and is based on as few parameters as possible, while at the same time yielding accurate low-frequency normal modes. Furthermore, the computational methods underlying our desired approach are not geared toward a specific type of inter-domain motion, such as rotation about a hinge, but should rather be able to identify a range of motions, such as shear, flexible-loop. Moreover, the method is valid across scales from single chain to complex, multi-chain assemblies.

We define "Density-Cluster RTB" by grouping residues according to similarities in their atomic density differences. This approach is based on the hypothesis that similarity in an atomic density "gradient" among a set of residues would reflect a similar propensity for motion and similar deformability. We apply a hierarchical clustering of atomic density differences calculated between atoms within a pre-specified cut-off radius, which are rapidly calculated with the Fast Atomic Density Evaluator (FADE).[28]

We implemented this blocking scheme in the open-source program DIAGRTB,[5,6,29,30] based on an anisotropic model and all-heavy atom NMA. Thus, full local contact information was used to derive standard RTB results for comparison, in line with the density-based rationale of our method. Normal modes were calculated across a benchmark taken from the Database of Macromolecular Movements.[31] These are proteins that undergo significant conformational change in performing their functions, representing motions at different length scales (sub-domain, domain, and inter-domain) and of different types (hinge, shear, refolding of structure, allosteric, etc.). We also included a set allosteric proteins obtained from published benchmark data sets[32,33] along with several very large, well-studied complexes for which large-scale motion is critical for their function, namely, GroEL-GroES, F1-ATPase, and the 70S ribosome. We compared Density-Cluster RTB with standard 1 residue/block, or 2–4 residues per block in terms of their ability to capture functional motions with low-order modes as well as speed of calculation. We did not compare with MM-based NMA or ANM without blocking, as 1 residue/block RTB has already been extensively validated against these methods.[5,6,34]

Low-frequency normal modes calculated with the Density-Cluster RTB (DCRTB) are in good agreement with low-frequency modes calculated with 1 residue/block RTB. Moreover, we show that DCRTB modes are in good agreement with the direction of motion associated with the conformational change, and even show better correlation with *B*-factors than 1–4 residue/block partitioning for a number of examples. Finally, the method is computationally cheap, with dramatic reductions in computer time demonstrated for the systems at all scales, but especially for the largest. For example, calculation of the all-atom DCRTB normal modes for the 70S ribosome took about 30 min on a desktop computer (an iMac containing a 3.06 GHz Intel Core 2 Duo CPU with 4 GB of memory), compared with over 35 h for 4 residue/block RTB on the same iMac.

MATERIALS AND METHODS

Domain decomposition

The atomic density of the protein is calculated using the Fast Atomic Density Evaluator (FADE). This algorithm rests on the power law relationship between atomic density N and distance r from a reference point x:[28]

$$N(x, r) = r^{\lambda_i} \tag{2}$$

FADE uses a fast Fourier transform (FFT) algorithm to calculate the density exponent λ at grid points. A density value for each residue is assigned by estimating the value of λ at each grid point near an atom and then averaging over local λ values. A matrix M of density differences is calculated as follows:

$$\begin{aligned} M_{ij} &= \bar{\lambda}_i - \bar{\lambda}_j \qquad \text{if} \quad d_{ij} \leq 20 \\ M_{ij} &= 0 \qquad\qquad\quad \text{if} \quad d_{ij} > 20 \end{aligned} \tag{3}$$

where d_{ij} is the inter-residue distance taken to be the shortest distance between any atom of residue i and any atom of residue j. Density differences are calculated only if the inter-residue distance falls within a cut-off of 20 Å. (For aspartate transcarbamoylase, GroEL, GroEL-GroES, and the ribosome a 40-Å cut-off was used due to the larger size of these systems.) This ensures that the matrix M reflects only local correlations, as density similarities between distant protein regions are unlikely to be physically meaningful.

Using the Pearson correlation between the calculated density differences as distance metric, residues can be clustered into groups in which there is high correlation in the density profile. Clustering was done using maximal linkage, the most stringent criterion for successive cluster merging. We also performed hierarchical clustering with average linkage, but we found that complete linkage performed better for Density-Cluster RTB, in particular the ability of the modes to capture the experimentally measured conformational change. Complete linkage also decomposed the protein into fewer blocks than average linkage, thus giving added efficiency.

Following hierarchical clustering, one has a tree structure that represents a hierarchy of data similarity. In hierarchical clustering, elements are successively merged into clusters based on the distance criteria, here following the regime of complete linkage defined according to the maximum distance within a cluster. A merging of groups occurs at a node. Here, the lowest nodes correspond to the residues and the highest node corresponds to the entire protein. Branches represent distances between successive merging of nodes. The length of a branch is deter-

mined by the change in maximum intracluster distance when one cluster is split into two, hence long branches occur when splitting a cluster results in two clusters with significantly smaller diameters than the original cluster. The branches of the dendrogram are ranked by the distance between upper and lower nodes. The residues at the termini of the highest 10% of branches are determined, and for each of these top branches, its set of terminal branches, representing individual residues, are combined into a single rigid block. Due to the hierarchical aspect of the clustering, some tall branches may be daughters of other tall branches, leading to a situation where some initially determined clusters may be a subset of a larger cluster. In this case, the lower branch defines its own rigid block, with the remaining elements of the larger cluster defining a separate rigid block.

Anisotropic network model

Interactions between all heavy atoms are calculated using a harmonic potential evaluated for atoms within a cut-off distance as

$$U = \sum_{i<j} C_0 \left(r_{ij} - r_{ij}^0 \right)^2 \cdot h\left(r_c - r_{ij}^0 \right) \tag{4}$$

where r_{ij} and r_{ij}^0 are the instantaneous and equilibrium distances between atoms i and j, respectively, r_c is the cut-off distance (6 Å) for inter-atomic interactions, C_0 is an arbitrary force constant set to 1, and h enforces the cut-off, so that

$$\begin{aligned} h(r_c - r_{ij}^0) &= 1 \qquad \text{if} \quad r_{ij} \le r_c \\ h(r_c - r_{ij}^0) &= 0 \qquad \text{if} \quad r_{ij} > r_c \end{aligned} \tag{5}$$

Rotational–translational blocking

In RTB, the original Hessian is projected into a subspace of rotational–translational eigenvectors of the rigid blocks. The original *3N* X *3N* Hessian H (where N is the number of atoms) is reduced to a *6n* X *6n* Hessian H_{proj} (where n is the number of blocks), described as

$$H_{proj} = P^T H P \tag{6}$$

where P is an orthonormal *3N* X *6n* matrix of translational/rotational eigenvectors of each rigid block:[5,6,9]

$$\begin{aligned} P^{\mu}_{J,j_\nu} &= \sqrt{\frac{m_j}{M_j}} \delta_{\mu\nu} && \mu = 1,2,3 \\ P^{\mu}_{J,j_\nu} &= \sum_{\alpha,\beta} (I_J)^{-1/2}_{(\mu-3),\alpha} \sqrt{m_j} \left(r_j - r_J^0 \right)_\beta \epsilon_{\alpha\beta\nu} && \mu = 4,5,6 \end{aligned} \tag{7}$$

J and j subscripts label blocks and atoms, respectively. μ = 1,2,3 and μ = 4,5,6 label translation and rotation, respectively. m_j and r_j are the mass and Cartesian coordinates of atom j, respectively. M_J, I_J, and r_j^0 are the total mass, moment of inertia and center of mass of block J. δ is the Kronecker delta, and ϵ is the permutation over α, β, and ν = [1,2,3].

The resulting projected Hessian H_{proj} is diagonalized to give eigenvectors in the subspace defined by the rigid blocks, U_{proj}. Eigenvectors in the original space can be obtained by multiplying U_{proj} by the projection matrix:

$$U = P U_{proj} \tag{8}$$

where P is the projection matrix defined above. These calculations were carried out using a modified version of DIAGRTB.[5,6,29,30]

Normal mode evaluation metrics

Density-Cluster RTB normal modes and 1 residue/block normal modes (2–4 residues per block for aspartate transcarbamoylase and GroEL-GroES) were compared using the following measures, which assess agreement between blocking methods and with experimentally observed conformational change.

Overlap. The overlap O_{ij} of a given normal mode U_i of one method with a mode U_j of another method is given by

$$O_{ij} = \mathrm{abs}\left(U_i^T \cdot U_j \right) \tag{9}$$

This is simply the absolute value of the inner product of the normal modes. Repeating the calculation for all mode combinations results in a matrix of overlap values.

Spanning coefficient. The spanning coefficient, S, measures how well a normal mode from one method is represented by the space of normal modes calculated using another method. Here, we wish to know how well each normal mode of the 1 residue/block RTB NMA is approximated by the space of normal modes calculated by Density-Cluster RTB. For the kth 1 residue/block mode, the spanning coefficient S_k is calculated as follows:

$$S_k = \sum_i \left(U_{i;\mathrm{clust}}^T \cdot U_{k;1\mathrm{res}} \right)^2 \tag{10}$$

where $U_{i;\mathrm{clust}}$ is the ith normal mode calculated with Density-Cluster RTB and $U_{k;1\mathrm{res}}$ is the kth normal mode calculated with 1 residue/block NMA.

Root mean-squared fluctuation. Root mean-squared fluctuation (RMSF) is a standard measure for calculating atomic flexibility from normal modes given by

$$\mathrm{RMSF}_j = \sqrt{\sum_i \frac{U_{ij}^2}{\lambda_i}} \tag{11}$$

where RMSF_j is the root mean-squared fluctuation of the jth degree of freedom ($x/y/z$ component of motion), U_{ij} is the jth component of the ith eigenvector, and λ_I is the eigenvalue associated with the ith eigenvector.

Cumulative involvement coefficient. The cumulative involvement coefficient (CI) measures how well a set of normal modes U_k capture an observed protein motion:

$$\mathrm{CI} = \sum_k \left(\frac{\Delta R}{\|\Delta R\|} \cdot U_k \right)^2 \tag{12}$$

ΔR is the conformational change represented as a difference between the Cartesian coordinates of the protein in the two states. The argument inside the parentheses is the inner product between the kth normal mode and the normalized conformational change vector.

Correlation between RMSFs and B-factors. The correlation between B-factors and RMSFs calculated by normal modes was computed as follows:

$$C = \sum_i \frac{\mathrm{RMSF}_i \cdot \sqrt{B_i}}{\sqrt{\sum_j \mathrm{MSF}_j} \cdot \sqrt{\sum_j B_j}} \tag{13}$$

where RMSF_i is the root mean-squared fluctuation of the ith atom and B_i its B-factor.

RESULTS AND DISCUSSION

Description of dataset and domain decomposition

Density-Cluster RTB and 1 residue/block RTB were assessed using examples from the Database of Macromolecular Movements.[31] This is a well-annotated database of functionally relevant large-scale motions categorized according to the size of the mobile regions and the type of motion. The size categories include fragments, entire domains, and subunits/assemblies larger than a domain. Types of motions include hinge and shear motions, partial refolding of tertiary structure, and more complex motions that may be allosteric. The proteins studied here are a representative sampling of all the different scales and motion types in the database (Supporting Information Table S1). In addition to testing whether Density-Cluster RTB is capable of modeling different motion types, we assessed its performance for several large allosteric systems and molecular machines, for which more detailed calculations are prohibitively expensive (Supporting Information Table S1).

Our strategy rests on the hypothesis that regions of similar atomic density gradients within a local spatial neighborhood will undergo similar motions. First, the protein's atomic density is calculated using the fast Fourier transform-based Fast Atomic Density Evaluator or FADE.[28] Average atomic densities are assigned to each residue according to densities of the nearest grid points. Differences in average densities for residues within a 20Å cut-off were calculated, or, for larger systems such as the 70S ribosome, GroEL-GroES, and aspartate transcarbamoylase, a 40 Å cut-off was used to further coarsen the resolution. We also tried a cut-off of 10 Å on the smaller systems, which resulted in similar quality normal modes to those obtained using the 20 Å cut-off. However, since this resulted in a larger number of rigid blocks, it was not as efficient.

Each matrix of atomic density differences was subjected to hierarchical clustering, using the Pearson correlation between rows as the similarity measure. The resulting tree structure was used to obtain the clusters. Each branch was ranked by the distance between its associated upper and lower nodes, and the 10% of branches with the longest inter-nodal distances (or 5% in the case of especially large systems like the ribosome) are taken to represent the clusters of residues used for the blocking. The significance of long branches is that the addition of the lower node to the cluster considerably increases the largest distance/dissimilarity among elements of the cluster. Supporting Information Table S2 gives the results of the cluster decomposition along with the number of residues for each protein studied. In most cases, the number of blocks is a 12–15% fraction of the number of residues. For the NMA calculations, the number of residues and the number of blocks represent the number of rigid bodies for 1 residue/block RTB and cluster RTB, respectively.

Finally, we examined the spatial distribution of the blocks when mapped onto the protein structure to determine whether the blocks coincided with known functional domains and were sensible from a physical standpoint. We also compared our decomposition with that of rigidity analysis as calculated using FIRST,[10] which demonstrated success when applied to proteins for use with RTB.[9] For the examples that are discussed below, FIRST tends to form a small number of large blocks and a large number of very small blocks, which divide the protein into constrained regions connected by flexible ones. Whole secondary structure components are often assigned to single blocks using FIRST. In contrast, our method tends to divide the protein into a smaller number of rigid blocks, as seen in Supporting Information Table S3.

Although the partitions generated by Density Cluster and FIRST are not the same, there are clear similarities. Figure 1 and Supporting Information Figure S1 show adenylate kinase and several other proteins colored according to the identity of the blocks generated using the Density Cluster approach. Supporting Information Figure S2 shows the decomposition of these same proteins using FIRST. Adenylate kinase is a monomeric enzyme that catalyzes the transfer of a phosphoryl group from ATP to

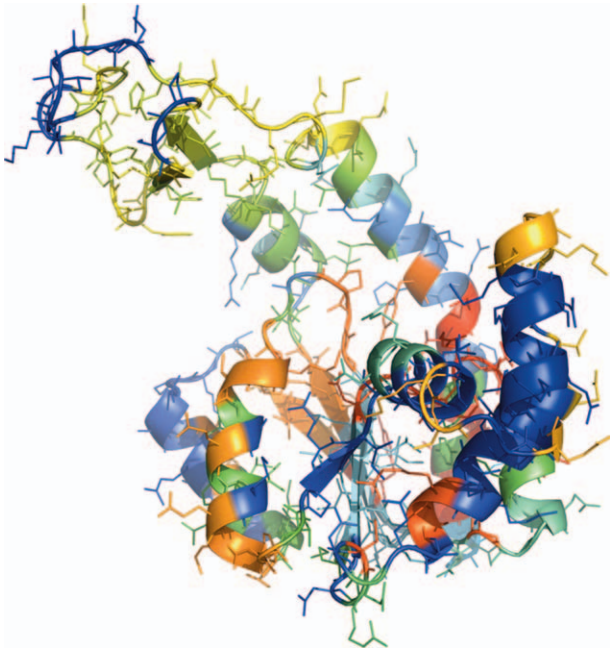

Figure 1

Adenylate kinase (PDB ID: 4ake) with blocks mapped onto the structure by color code. Residues in a given region colored the same way belong to the same block. There is no physical significance of the colors themselves; they are only labels.

AMP[35,36] and has been well-studied by NMA.[6,37,38] ATP and AMP are each bound by a domain, the Lid and NMPbind domains, respectively. These domains are responsible for the conformational change, "closing down" on their respective bound molecules prior to catalysis. Each of these domains appears to be composed of one main block, colored light green for the Lid and NMPbind in dark blue (Fig. 1). In Supporting Information Figure S2A, we see that secondary structure elements tend to be grouped together using FIRST, and that there are a few large blocks (in color) surrounded by many smaller blocks. As with the Density-Cluster decomposition, the Lid and NMPbind domains appear to each be composed of a dominant block colored chartreuse and turquoise, corresponding roughly to the light green block of the Lid and the dark blue block of the NMPbind.

Myosin II is one of the main proteins involved in muscle contraction. Under the control of ATP binding and hydrolysis, the motor domain of myosin II undergoes successive cycles of ratcheting motions relative to the long lever arm, along with binding and unbinding of the motor domain and actin.[39–42] The ratcheting motion involves a large conformational change of the converter domain,[43,44] the link between the motor domain and the lever arm (note that this domain is absent in the crystal structure depicted in Supporting Information Fig. S1B.) The converter is comprised primarily of a single green block, suggesting that the important conformational changes are not internal but rather are relative to other structural elements. Near the converter domain, the relay helix is another structure that undergoes substantial conformational change and plays an important role in coupling the active site conformation to that of the converter/lever arm.[45] It appears divided into four regions that are qualitatively consistent with its functional role. As seen in Supporting Information Figure S1B, the region proximal to the converter is part of a shared rigid block. Adjacent to this block, but oriented away from the converter, are blocks in light and dark blue, as well as a more distant large orange block. Although difficult to see in the displayed view, the relay helix ends in a section comprising a red and a green block. The red block is shared by part of Switch II, an important functional motif whose conformation determines myosin's ability to bind ATP. Finally, a large portion of the upper 50 K domain is assigned to a single navy blue block, consistent with the fact that this domain exhibits significant mobility with respect the regions adjacent to it. The FIRST decomposition also shows a dominant block comprising a large portion of the upper 50 K domain (the green block in Supporting Information Fig. S2B). However, FIRST assigns the bulk of the protein to a single very large block (dark blue in Supporting Information Fig. S2B), which contrasts the Density-Cluster decomposition.

Human immunodeficiency virus type 1 (HIV-1) reverse transcriptase catalyzes the synthesis of DNA from RNA,[46–49] making it an important drug target. It is a heterodimer, with one chain (p51) comprising a DNA polymerase domain consisting of fingers, palm, and thumb motifs, and the other (p66) composed of a RNaseH domain and an additional DNA polymerase domain. DNA is bound in a cleft formed by the subdomains of p66, which exhibit a great degree of flexibility[50–52] consistent with their role in regulating DNA binding. In Supporting Information Figure S1C, one can see that the fingers and thumbs of p66 are, respectively, comprised of 1 (in green) or 2 (in yellow and blue) main blocks. A previously reported method[18] yielded a similar domain decomposition for these two subunits. Supporting Information Figure S2C shows that FIRST also divides these functionally relevant domains into several blocks, where the fingers are composed of a dominant block (olive green) and the thumb is composed of a sky blue and a brick red block.

GroEL-GroES is a large, multimeric, protein-folding chaperone that sequesters proteins in its barrel-shaped structure and catalyzes their correct folding through coupling to ATP hydrolysis.[53–59] The barrel structure of GroEL consists of two rings, each composed of seven chains. These chains are each composed of three domains: an equatorial domain adjacent to the interface between the rings, an intermediate domain, and an apical domain. These domains exhibit an interesting switching

behavior associated with ATP binding to the equatorial domain. When ATP is bound, the motions of the equatorial and intermediate domains become coupled. In the ATP-unbound state, the motion of the apical and intermediate domains are coupled.[60] We see in Supporting Information Figure S1D that a large portion of the equatorial domain is composed of two blocks (in green and red), and that the "core" of the apical domain is composed of a single large block (in sky blue). In contrast, the intermediate domain is a mosaic of blocks, including shared blocks with the other two domains, consistent with its role as a conformational coupling switch. The FIRST decomposition of GroEL-GroES represented in Supporting Information Figure S2D shows a decomposition that has little obvious similarity with Density-Cluster, which is in contrast with the previous examples.

Upon examining the structures in Figure 1 and Supporting Information Figure S1, it appears that some helices are transected longitudinally by the blocking. This was initially disconcerting, as helices are known to be stable structures. However, we observed similar divisions in the deformation energy-based decomposition method of Hinsen and co-workers.[18] Indeed, deformability, and even local unfolding, are hypothesized to be an important in determining a protein's function.[61–66] Moreover, if the helical axis acts as a hinge, this type of decomposition is sensible.

Comparison of density-cluster RTB with 1 residue/block RTB

We compared the normal modes calculated with the Density-Cluster RTB method with 1 residue/block RTB and with N residue/block RTB, where N is the ratio of the number of residues to the number of Density Cluster blocks, which varies depending on the example. In the case of N residue/block RTB, N residues that are contiguous along the protein's primary sequence are grouped together into a single block. This was performed as a control to determine whether the same reduction in degrees of freedom would achieve similar achieve the same accuracy as Density-Cluster RTB.

We also compared Density-Cluster RTB with RTB using FIRST's domain decomposition, henceforth referred to as Rigidity-Cluster RTB. We attempted to compare the methods across a range of systems, consisting of examples in our data set that performed well in the previous study by Ahmed and Gohlke[9] as well as the examples that were discussed in the previous section. Unfortunately, FIRST decomposed the protein into blocks that were too large to be dealt with by the DIAGRTB code. However, we did examine the results of FIRST for lysine, arginine, ornithine-binding protein (2lao) and thymidylate synthase, as well as HIV-1 reverse transcriptase.

In all cases, normal modes were based on the anisotropic model applied to heavy atoms, and eigenvalue calculations were performed using a modified version of DIAGRTB.[5,6,29,30] For some larger systems, 1 residue/block NMA could not be performed due to the size of the resulting Hessian, so comparisons are based on blocks of 2-3 consecutive residues.

The agreement between modes generated by different methods was assessed by means of the overlap measure, which is simply the inner product of the mode vectors. There is high overlap for the 2–3 lowest frequency nontrivial normal modes in almost all cases, and this was true across all types of motion classes and length scales [Fig. 2(A) and Supporting Information Fig. S3]. Larger proteins, whose motions tend to be more collective, often had good agreement for up to 20–30 low order modes. Rigidity-Cluster RTB also yielded high overlap in these lowest frequency modes.

The spanning coefficient is another measure that was used to evaluate the agreement between normal modes calculated with the two blocking schemes. The spanning coefficient is a measure of how well each normal mode calculated with the reference 1-residue/block method is approximated using the space of normal modes calculated with Density-Cluster RTB. The spanning coefficient calculation yields a similar result as the overlap measure [Fig. 2(B) and Supporting Information Fig. S4], which is intuitive, as the spanning coefficient is essentially a cumulative overlap (see Methods). The 1-residue block normal modes are well spanned (spanning coefficient of 0.5 or greater) by Density-Cluster RTB for the first 50–60 modes for smaller proteins and 100 or more modes for larger systems. The span is similar over the lowest order modes to that obtained with the control case of N residue/block modes. However, the span decays rapidly for a number of examples in the case of N residue/block modes compared with Density-Cluster modes [Fig. 2(B) and Supporting Information Fig. S4]. In those cases, higher frequency 1 residue/block modes are better spanned by Density-Cluster modes.

The ability of the methods to describe protein flexibility was also measured using root mean-squared fluctuation (RMSF). The methods perform equivalently in modeling the flexibility as assessed by this measure [Fig. 2(C) and Supporting Information Fig. S5]. The methods differed somewhat in the calculated magnitude of the RMSFs, with Density-Cluster RTB, N residue/block RTB, and Rigidity-Cluster RTB yielding smaller magnitude RMSFs than 1 residue/block RTB. This is expected given that Density-Cluster RTB, Rigidity-Cluster RTB, and N residue/block RTB impose greater rigidity than 1-residue/block RTB. For the low frequency modes used to calculate RMSF, we verified that in all cases, correlation of the eigenvalues of Rigidity-Cluster RTB and N residue/block RTB with 1 residue/block RTB was 97–99% for the first 200 lowest eigenvalues (Data not shown.).

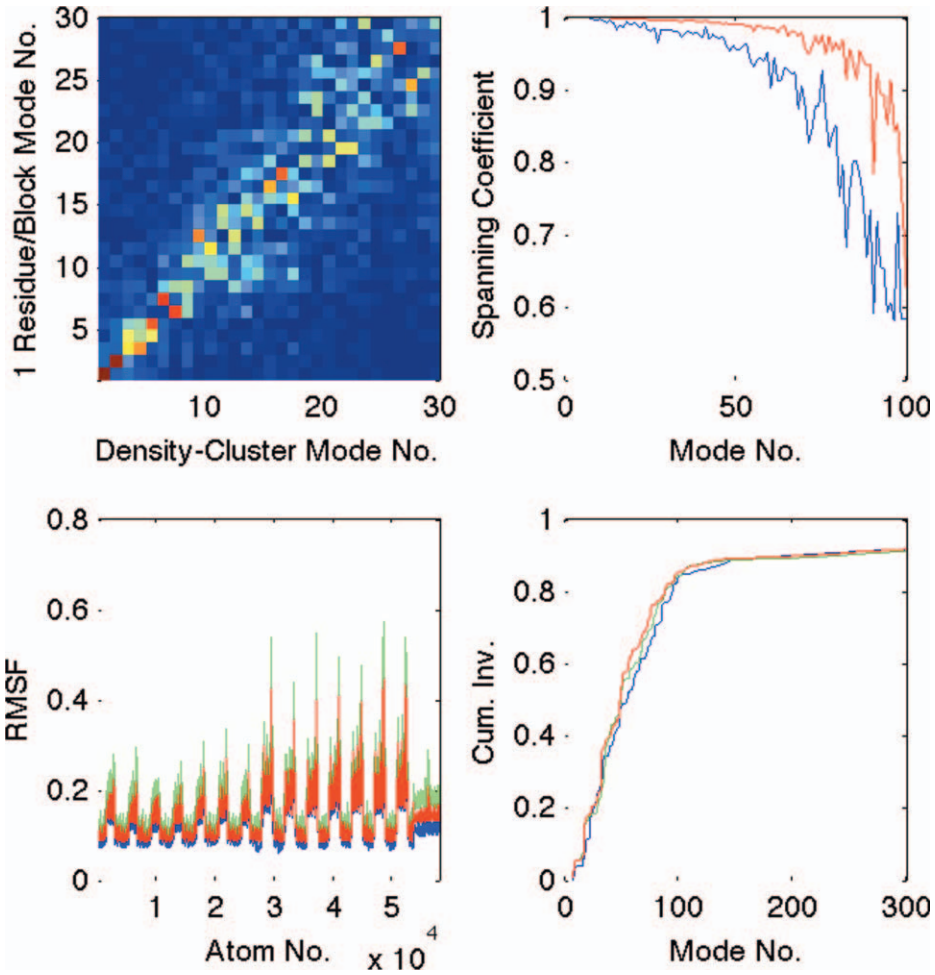

Figure 2

Performance of Density-Cluster RTB modes compared with standard 1 residue/block modes for GroEL-GroES (PDB ID: 2c7c). Other examples in our dataset performed similarly and are given in Supporting Information. (**A**) Overlap matrix describing agreement of low-frequency Density-Cluster RTB modes (bottom axis) with 1 residue/block modes (left axis). The color scale is a heat map that runs from blue (overlap of zero) to maroon red (overlap of 1). (**B**) Spanning coefficient describing how well each 1 residue/block mode is represented by the entire space of density-cluster modes (blue) and N residue/block modes (red). The mode index starts with 1, although mode 7 is the first nontrivial mode. (**C**) Root mean-squared fluctuations calculated with Density-Cluster RTB (blue), 1 residue/block RTB (green), and N residue/block RTB (red). (**D**) Cumulative Involvement Coefficient, a measure of how well the normal modes capture the experimentally known direction of conformational change, calculated with Density-Cluster RTB (blue), 1 residue/block RTB (green), and N residue/block RTB (red) plotted versus mode index. The mode index starts with 1, although mode 7 is the first nontrivial mode.

Agreement with experimentally determined conformational changes and *B*-factors

In addition to comparing Density-Cluster RTB against other methods, its ability to describe known conformational changes was also assessed. In particular, normal modes were calculated using the unbound/inactive state, and their ability to describe the conformational transition from the unbound/inactive to the bound/active state was assessed using the cumulative involvement coefficient, which is defined similarly to the spanning coefficient. However, in this case, we measure how well the vector of positional differences between two experimentally determined structures is spanned by the normal modes calculated with either the reference 1 residue/block or Density-Cluster RTB methods. It can be seen [Fig. 2(D) and Supporting Information Fig. S6] that small protein motions are not as well described by Density-Cluster RTB as by 1 residue/block RTB, although the conformational changes of four proteins less than 300 residues were well described with Density-Cluster RTB: ran GTPase (PDB ID: 1byu), CheY (3chy), LAO-binding protein (2lao), and adenylate kinase (4ake). With the exception of CheY, these four proteins represent hinge motions, which are probably the simplest type of motion represented by our dataset. However, larger motions are well described by the method, and, in a few cases (ran

GTPase [1byu]; glycogen phosphorylase [1gpb]; Bam H1 restriction endonuclease [1bam]; PcrA DNA helicase [1pjr]; aspartate transcarbamoylase [1rac]), Density-Cluster RTB has higher cumulative involvement than 1 residue/block RTB. With the exception of ran GTPase, these proteins represent complex motions that are not simply hinge or shear. Since NMA is primarily used to describe complex, large-scale motions on time scales inaccessible by MD, good performance for such large cases is most desirable. We note that the control calculation of N residue/block RTB yields cumulative involvement that is largely in agreement with 1 residue/block RTB and Density-Cluster RTB. We also focused on the cumulative involvement in the lowest frequency nontrivial modes to determine whether functional significance is more concentrated in this range. We note that 14 examples have higher cumulative involvement at mode 20, and 8 of these (CheY, GroEL (1gr5), hemoglobin, aspartate transcarbamoylase, F1-ATPase, rhoA, glutamate dehydrogenase, and phosphofructokinase) are allosteric proteins, which, in general, undergo a more complicated conformational change than proteins that bind a single substrate (Supporting Information Table S4). For the three examples that we were able to perform using Rigid-Cluster RTB, two of these, lysine, arginine, ornithine-binding protein (2lao), and thymidylate synthase (3tms), both showed significantly higher cumulative involvement with Density-Cluster than Rigidity-Cluster RTB in the low-frequency range. Furthermore, cumulative involvement calculated with Density Cluster NMA are in much better agreement with 1 residue/block RTB (Supporting Information Fig. S6E). In addition, the method has an important advantage over 1 residue/block RTB, namely computational efficiency, which will be discussed in the next section.

In addition to the experimentally determined direction of conformational change, we also compared Density-Cluster RTB with 1 residue/block RTB and N residue/block RTB in their ability to describe experimentally determined magnitudes of fluctuation, the B-factors. We see that in 20 of the 32 cases the correlation between calculated RMSF and *B*-factors in Density-Cluster RTB is higher than that obtained with 1-residue/block RTB. Significantly, these 20 proteins encompass the entire range of length scale and type of conformational change (Table I). For the three proteins for which we were able to perform Rigidity Cluster RTB, Density-Cluster RTB achieved a higher correlation with *B*-factors. The poorer performance of 1 residue/block RTB is likely related to ANM's tendency to exaggerate the motions of floppy or loosely packed regions, the so-called tip effect. Indeed, introducing additional appropriate stiffness into the system, as we have done with our method, has been shown to ameliorate the tip effect and improve agreement with experimentally determined magnitudes of motion.[67–69] That Density-Cluster RTB results in greater rigidity is also demonstrated by the smaller magnitude RMSFs compared with 1 residue/block RTB [Fig. 2(C) and Supporting Information Fig. S5].

Computational cost

Because Density-Cluster RTB and other coarse-grained methods are aimed at the largest systems with a view toward reduced computational expense, we compared the calculation time of 1 residue/block RTB with Density-Cluster RTB for the largest systems in our study, along with some smaller allosteric systems, to determine whether computational savings occur on all length scales. Of these, the 70S ribosome, GroEL, GroEL-GroES, F1-ATPase, aspartate transcarbamoylase, and glutamate dehydrogenase are large multimeric protein assemblies with between 2000 and 8000 residues (Table II). For these massive systems, 1 residue/block RTB could not be performed, as the resulting Hessian was too large. Hence, Density-Cluster RTB was compared with 2–4 residues/block for these examples. Table II shows dramatic computational savings across all length scales, but particularly for the largest examples. Compute times are at least 83% shorter for Density-Cluster RTB compared with 1–4 residue/block RTB. Especially striking was the computational saving for the 70S ribosome, where Density-Cluster RTB resulted in a nearly 2-fold order of magnitude reduction in the computing time (Table II). For most cases, the setup and calculation of the density clusters was instantaneous, except for GroEL and GroEL-GroES, which each took around 10 min. Overall, we observed faster calculations for Density-Cluster RTB compared with Rigidity Cluster RTB, with especially significant reduction in the computational cost for HIV-1 reverse transcriptase (Table III).

Examination of low-frequency modes

In addition to validating Density-Cluster RTB by comparison with the established single-residue blocking and numerical comparison with experimental conformational change, we also examined the animations of Density-Cluster RTB normal modes to ensure that they yield physically reasonable kinematics and are in agreement with previously published NMA calculations. In general, the low-frequency modes that we examined are in good agreement with previous calculations and with experimentally inferred motions. Depicted in Figure 3 and Supporting Information Figure S6 are four well-studied proteins that are each perturbed along the first nontrivial mode, mode 7.

Myosin II is the molecular motor that causes muscle contraction by cycles of binding and unbinding to actin filaments that are coupled to ATP binding. It consists of a motor domain that contains the structural motifs responsible for actin and ATP binding and a lever arm. Consistent with its role in linking the motor domain to

Table I
Correlation of Density-Cluster RTB, 1–4 Residue/Block RTB, N-Residue/Block RTB, and Rigid Cluster RTB with B-*Factors*

	Proteins from database of macromolecular movements				
PDB ID	Protein name	Cluster	1 res./block	N res./block	Rigid clust.
3tms	Thymidylate synthase	0.90	0.86	0.94	0.70
4dfr	Dihydrofolate reductase	0.94	0.88	0.98	
1byu	GDP-bound Ran	0.75	0.76	0.95	
3chy	CheY (inactive state)	0.92	0.94	0.98	
1j74	Ubiquitin-conjugating enzyme Mms2	0.80	0.89	0.98	
1i69	OxyR transcription factor	0.94	0.92	0.98	
6tim	Triosephosphate isomerase	0.92	0.91	0.94	
2lao	Lysine, arginine, ornithine-binding protein	0.97	0.96	0.98	0.88
1g7s	Translation-initiation factor IF2/EIF5B	0.93	0.93	0.93	
1d6m	DNA topoisomerase III	0.97	0.93	0.96	
2hmi	HIV-1 reverse transcriptase	0.92	0.88	0.91	0.89
4ake	adenylate kinase	0.96	0.94	0.95	
8adh	APO-liver alcohol dehydrogenase	0.96	0.90	0.97	
9aat	Aspartate amino transferase	0.95	0.95	0.95	
1lih	Aspartate receptor	0.88	0.84	0.90	
1su4	Sarcoplasmic/endoplasmic reticulum calcium ATPase 1	0.91	0.92	0.93	
1pjr	PcrA DNA helicase	0.94	0.94	0.96	
1bam	Endonuclease BamH I	0.97	0.97	0.98	
4hhb	Hemoglobin	0.97	0.96	0.97	

	Allosteric proteins, multimeric assemblies, and large molecular machines			
PDB ID	Protein name	Cluster	1 res./block	N res./block
1rac	aspartate transcarbamoylase	0.93	0.92	0.93
1an0	cdc-42	0.88	0.84	0.91
1dbq	purR repressor	0.87	0.86	0.88
1EYJ	fructose-1,6-bisphosphatase	0.95	0.92	0.96
1ftn	RhoA	0.95	0.88	0.97
1gpb	glycogen phosphorylase	0.83	0.80	0.90
1nr7	glutamate dehydrogenase	0.78	0.76	0.78
1tlf	lac repressor	0.85	0.85	0.85
1v4t	Glucokinase	0.95	0.68	0.77
1vom	myosin II	0.91	0.83	0.87
6pfk	Phosphofructokinase	0.96	0.96	0.97
1bmf	F1-atpase	0.93	0.89	0.92

Table II
Computational Timings for Allosteric Proteins and Large Molecular Machines

Protein name	No. of res.	No. of atoms	Cluster	1 res./block	N res./block	Fractional reduction in computation time for cluster RTB
ATCase**	2778	21,666	0:09:24	4:06:56	0:06:53	0.036
GroEL T/T***	7238	52,668	1:32:49	22:32:29	1:15:39	0.069
GroEL R/T***	7350	53,970	1:40:04	24:00:41	1:14:57	0.07
GroEL R′/T-GroES****	7966	57,946	1:57:40	12:52:56	2:43:17	0.153
GroEL(R″/T)-GroES****	7923	58,674	1:59:53	13:06:49	1:46:44	0.152
cdc-42	189	1477	0:00:08	0:00:48	0:00:04	0.167
purR repressor	552	4344	0:00:45	0:15:59	0:00:25	0.0469
FBPase	1308	9984	0:06:56	3:53:37	0:02:54	0.03
RhoA	117	1405	0:00:06	0:00:42	0:00:04	0.143
glyc. phosphorylase	1646	13,382	0:06:53	7:33:11	0:03:06	0.015
glut. dehydrogenase**	2976	23,244	0:24:05	5:07:45	0:12:01	0.078
lac repressor	1184	10,700	0:03:40	2:34:33	0:01:52	0.024
Glucokinase	424	3325	0:00:26	0:07:26	0:00:17	0.058
myosin II	730	5750	0:01:15	0:37:30	0:00:38	0.034
phosphofructokinase	1276	9395	0:03:24	3:17:53	0:01:40	0.017
F1-atpase**	2987	22,722	0:17:35	5:59:13	0:08:02	0.047
70S Ribosome****	9788	13,816	0:32:41	35:32:49	0:24:10	0.015

Two residues/block RTB (**) was done due to insufficient memory. Three residues/block RTB (***) and four residues/block RTB (****) were done for the same reason for the indicated systems.

Table III
Computational Times for Examples for which Rigidity Analysis was Performed

Protein name	No. of residues	No. of atoms	Cluster RTB	Rigidity cluster
Thymidylate synthase	264	2150	0:00:17	0:01:06
Lysine, arginine, ornithine-binding protein	238	1822	0:00:13	0:00:30
HIV-1 reverse transcriptase	988	7630	0:03:40	0:36:00

the lever arm, the converter exhibits large motions in the lowest frequency mode [Fig. 3(A) and Supporting Information Fig. S7A]; this is also in agreement with X-ray crystal structures of myosin II in different states of ATP hydrolysis[43,44] and a previous NMA study.[34] The depicted normal mode also shows simultaneous flexing of the upper 50-K domain, consistent with the prior work of Li and Cui.[34] Other low frequency modes are similarly dominated by motions of the converter and upper 50-K domains. Mode 8 captures significant motions of the relay helix, a large structural element thought to be important for coupling ATP-binding to pivoting at the converter.

GroEL-GroES is a 21-chain multimer that functions as a folding chaperone, sequestering misfolded or partially folded proteins and catalyzing their folding via coupling to ATP hydrolysis. Each chain of GroEL comprises an equatorial domain, intermediate domain, and apical domain. The chains of GroEL exhibit an important allosteric switching behavior linked to ATP binding.[56] When ATP is bound to the equatorial domain, the equatorial domain and intermediate domain are correlated and both rigid, while the apical domain remains flexible. In the absence of ATP, the behavior of the intermediate domain switches to become correlated in its motion with the apical domain, in which both are flexible in the ATP-

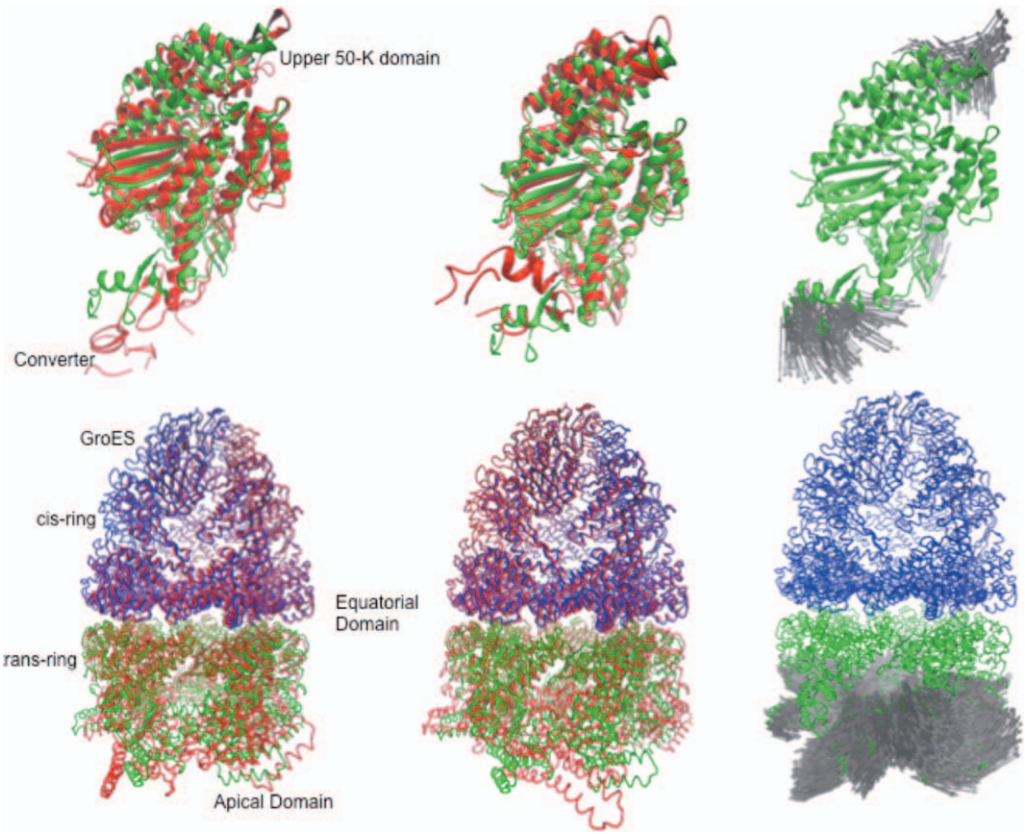

Figure 3
Protein structures deformed by the first nontrivial Density-Cluster RTB normal mode. In all cases, the deformed structure is in red, and the static structure is in blue or green. (**A**) Myosin II. The motion comprises a ratcheting of the converter domain and a simultaneous bending of the upper 50 K domain. (**B**) GroEL-GroES. Here a torsional motion of the two rings about the equator is seen with large amplitude motions of the apical domains of the individual subunits of the *trans*-ring accompanied by much smaller amplitude motions of the same domains of the *cis*-ring, illustrating the allosteric switching induced by nucleotide binding to the *cis*-ring.

unbound state. This switching behavior is observed in the first nontrivial mode of GroEL-GroES [Fig. 3(B) and Supporting Information Fig. S7B], where we observe that the apical and intermediate domain motions are large and correlated in the ATP-unbound *trans*-ring, while the equatorial and intermediate domains are much more rigid in the ATP-bound *cis*-ring. In modes 7–9, we also observe a twisting of the two rings in opposite directions with respect to each other, consistent with a key prior NMA study.[60]

F_0F_1-ATP synthase is the complex that links transmembrane proton pumping to ATP synthesis.[43,70–72] F_1-ATPase is the soluble portion of this complex and consists of a γ subunit whose motion is coupled to conformational change in the nucleotide-binding subunits. In particular, the large-amplitude motion of the γ subunit was found to be strongly coupled to the flexible β_E subunit,[73] and results from our method are consistent with this finding (Supporting Information Fig. S7C). Also in line with this prior study is the observation that the other β subunits remain relatively rigid, while one of the α subunits, α_{TP}, is highly flexible.

The 70S ribosome is a large assembly comprised of RNA and protein chains that translate mRNA to a polypeptide chain, and a large conformational change occurs during protein synthesis.[74–76] Most notably, the ribosomal 30S and 50S subunits are observed to undergo a ratchet-like rotation relative to each other[75] that is thought to affect the translocation of mRNA during translation. Indeed, we observed this motion in all four of the lowest frequency modes (Supporting Information Fig. S7D), and this is consistent with previous NMA studies.[77,78] Also in line with these studies are observed large-amplitude motions of the L1 stalk, which has a postulated role in binding tRNA and removing it from the E-site.

CONCLUSION

In this work, we have developed a new method for domain decomposition for RTB normal mode calculations, which is based on hierarchical clustering of local atomic density differences. Our method shows a clear computational advantage for large systems, yielding high-quality modes with a 10–100-fold reduction in computational expense when compared with 1–4 residues/block. The method is more computationally efficient than NMA with rigidity analysis, which was slower or could not be performed at all due to the size of the largest blocks. The modes calculated using this technique were shown to be in good agreement with normal modes calculated with standard 1–4 residues/block RTB. When comparing the agreement of Density-Cluster NMA with 1 residue/block or N residues/block, it is notable that our method is better able to span the higher frequency modes than N residue/block NMA and performs similarly to 1 residue/block over a large range of modes.

Most importantly, we have demonstrated that Density Cluster modes are able to capture experimentally observed conformational changes for a diverse set of proteins. The magnitudes of fluctuation calculated using Density-Cluster RTB correlate better with crystallographic *B*-factors in most cases than 1 residue/block RTB, and this was especially true for larger protein systems. The good performance on large systems is especially noteworthy, because the examples studied here have motions that are highly complex, involve many length scales, and cannot be described by simple mechanical terms as hinge-bending or shear motions. However, for smaller systems, the standard 1 residue/block method tended to give better agreement with the conformational change. This is not necessarily unexpected, as it is reasonable to postulate that smaller proteins must be described with more-detailed models, as their conformational changes may tend to involve finely tuned rearrangements of small motifs. In contrast, larger systems, whose motions are more collective and may involve large protein regions, may be more suited to a domain decomposition approach.

Domain decomposition is not a new technique, and it has been applied toward normal mode analysis and in other contexts. However, most existing domain decomposition techniques require multiple structures or motion data calculated for a single structure. FIRST is the only systematic method that we are aware of for domain decomposition based on a single structure.[9] The Density-Cluster RTB method developed here offers several advantages. Unlike FIRST, the initial decomposition does not require the use of any force field, thus reducing the number of parameters to which our results may be sensitive. Moreover, compared with the domain partitioning obtained with FIRST,[9] the Density Cluster method provides a greater reduction in the number of degrees of freedom (~85–90% reduction in number of degrees of freedom versus 60–70% reduction in the systems studied by Ahmed and Gohlke). In contrast to the clustering scheme of Schuyler and Chirikjian,[11,12] our method is fully automated and does not require human intervention. Finally, the use of atomic density is intuitively satisfying, as atomic density is closely correlated with burial and solvent accessibility, which in turn are correlated with flexibility. Density-Cluster decomposition thus provides a computationally efficient and accurate means of probing a protein's dynamic properties based on a single structure.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Gary Wesenberg for critical reading of the manuscript and assistance with FORTRAN coding, as well as Dr. George N. Phillips, Jr.

and Qiang Cui for providing helpful suggestions throughout the course of this work.

REFERENCES

1. Atilgan AR, Durell SR, Jernigan RL, Demirel MC, Keskin O, Bahar I. Anisotropy of fluctuation dynamics of proteins with an elastic network model. Biophys J 2001;80:505–515.
2. Tirion MM. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. Phys Rev Lett 1996;77:1905–1908.
3. Bahar I, Atilgan AR, Erman B. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. Fold Des 1997;2:173–181.
4. Lu M, Ma J. A minimalist network model for coarse-grained normal mode analysis and its application to biomolecular x-ray crystallography. Proc Natl Acad Sci U S A 2008;105:15358–15363.
5. Li G, Cui Q. A coarse-grained normal mode approach for macromolecules: an efficient implementation and application to Ca(2+)-ATPase. Biophys J 2002;83:2457–2474.
6. Tama F, Gadea FX, Marques O, Sanejouand YH. Building-block approach for determining low-frequency normal modes of macromolecules. Proteins 2000;41:1–7.
7. Hinsen K. Analysis of domain motions by approximate normal mode calculations. Proteins 1998;33:417–429.
8. Krebs WG, Alexandrov V, Wilson CA, Echols N, Yu H, M.Gerstein . Normal mode analysis of macromolecular motions in a database framework: developing mode concentration as a useful classifying statistic. Proteins 2002;48:682–695.
9. Ahmed A, Gohlke H. Multiscale modeling of macromolecular conformational changes combining concepts from rigidity and elastic network theory. Proteins 2006;63:1038–1051.
10. Jacobs DJ, Rader AJ, Kuhn LA, Thorpe MF. Protein flexibility predictions using graph theory. Proteins 2001;44:150–165.
11. Schuyler AD, Chirikjian GS. Normal mode analysis of proteins: a comparison of rigid cluster modes with Cα coarse graining. J Mol Graph Model 2004;22:183–193.
12. Schuyler AD, Chirikjian GS. Efficient determination of low-frequency normal modes of large protein structures by cluster-NMA. J Mol Graph Model 2005;24:46–58.
13. Hafner J, Zheng W. Approximate normal mode analysis based on vibrational subsystem analysis with high accuracy and efficiency. J Chem Phys 2009;130:194111.
14. Woodcock HL, Zheng W, Ghysels A, Shao Y, Kong J, Brooks BR. Vibrational subsystem analysis: a method for probing free energies and correlations in the harmonic limit. J Chem Phys 2008;129:214109.
15. Zheng W, Brooks BR. Probing the local dynamics of nucleotide-binding pocket coupled to the global dynamics: myosin versus kinesin. Biophys J 2005;89:167–178.
16. Hayward S, Berendsen HJ. Systematic analysis of domain motions in proteins from conformational change: new results on citrate synthase and T4 lysozyme. Proteins 1998;30:144–154.
17. Hayward S, Kitao A, Berendsen HJ. Model-free methods of analyzing domain motions in proteins from simulation: a comparison of normal mode analysis and molecular dynamics simulation of lysozyme. Proteins 1997;27:425–437.
18. Hinsen K, Thomas A, Field MJ. Analysis of domain motions in large proteins. Proteins 1999;34:369–382.
19. Flores SC, Keating KS, Painter J, Morcos F, Nguyen K, Merritt EA, Kuhn LA, Gerstein MB. HingeMaster: normal mode hinge prediction approach and integration of complementary predictors. Proteins 2008;73:299–319.
20. Emekli U, Schneidman-Duhovny D, Wolfson HJ, Nussinov R, T.Haliloglu . HingeProt: automated prediction of hinges in protein structures. Proteins 2008;70:1219–1227.
21. Yesylevskyy SO, Kharkyanen VN, Demchenko AP. Hierarchical clustering of the correlation patterns: new method of domain identification in proteins. Biophys Chem 2006;119:84–93.
22. Yesylevskyy SO, Kharkyanen VN, Demchenko AP. Dynamic protein domains: identification, interdependence, and stability. Biophys J 2006;91:670–685.
23. Kundu S, Sorensen DC, Phillips GN, Jr. Automatic domain decomposition of proteins by a Gaussian network model. Proteins 2004;57:725–733.
24. Shatsky M, Nussinov R, Wolfson HJ. Flexible protein alignment and hinge detection. Proteins 2002;48:242–256.
25. Wriggers W, Schulten K. Protein domain movements: detection of rigid domains and visualization of hinges in comparisons of atomic coordinates. Proteins 1997;29:1–14.
26. Abyzov A, Bjornson R, Felipe M, M.Gerstein . RigidFinder: a fast and sensitive method to detect rigid blocks in large macromolecular complexes. Proteins 2010;78:309–324.
27. Keating KS, Flores SC, Gerstein MB, Kuhn LA. StoneHinge: hinge prediction by network analysis of individual protein structures. Protein Sci 2009;18:359–371.
28. Mitchell JC, Kerr R, Ten Eyck LF. Rapid atomic density methods for molecular shape characterization. J Mol Graph Model 2001;19:325–330.
29. Durand P, Trinquier G, Sanejouand YH. A new approach for determining low-frequency normal modes in macromolecules. Biopolymers 1994;34:759–771.
30. Suhre K, Sanejouand YH. ElNemo: a normal mode server for protein movement analysis and the generation of templates for molecular replacement. Nucleic Acids Res 2004;32:W610–W614.
31. Gerstein M, Krebs W. A database of macromolecular motions. Nucleic Acids Res 1998;26:4280–4290.
32. Demerdash ON, Daily MD, Mitchell JC. Structure-based predictive models for allosteric hot spots. PLoS Comput Biol 2009;5:e1000531.
33. Daily MD, Gray JJ. Local motions in a benchmark of allosteric proteins. Proteins 2007;67:385–399.
34. Li G, Cui Q. Analysis of functional motions in Brownian molecular machines with an efficient block normal mode approach: myosin-II and Ca2+ -ATPase. Biophys J 2004;86:743–763.
35. Muller CW, Schlauderer GJ, Reinstein J, Schulz GE. Adenylate kinase motions during catalysis: an energetic counterweight balancing substrate binding. Structure 1996;4:147–156.
36. Muller CW, Schulz GE. Structure of the complex between adenylate kinase from Escherichia coli and the inhibitor Ap5A refined at 1.9 A resolution. A model for a catalytic transition state. J Mol Biol 1992;224:159–177.
37. Maragakis P, Karplus M. Large amplitude conformational change in proteins explored with a plastic network model: adenylate kinase. J Mol Biol 2005;352:807–822.
38. Miyashita O, Onuchic JN, Wolynes PG. Nonlinear elasticity, proteinquakes, and the energy landscapes of functional transitions in proteins. Proc Natl Acad Sci U S A 2003;100:12570–12575.
39. Geeves MA, Holmes KC. The molecular mechanism of muscle contraction. Adv Protein Chem 2005;71:161–193.
40. Geeves MA, Holmes KC. Structural mechanism of muscle contraction. Annu Rev Biochem 1999;68:687–728.
41. Trentham DR, Eccleston JF, Bagshaw CR. Kinetic analysis of ATPase mechanisms. Q Rev Biophys 1976;9:217–281.
42. Lymn RW, Taylor EW. Mechanism of adenosine triphosphate hydrolysis by actomyosin. Biochemistry 1971;10:4617–4624.
43. Bauer CB, Holden HM, Thoden JB, Smith R, I.Rayment . X-ray structures of the apo and MgATP-bound states of Dictyostelium discoideum myosin motor domain. J Biol Chem 2000;275:38494–38499.
44. Smith CA, Rayment I. X-ray structure of the magnesium(II). ADP. vanadate complex of the Dictyostelium discoideum myosin motor domain to 1.9 A resolution. Biochemistry 1996;35:5404–5417.

45. Yu H, Ma L, Yang Y, Cui Q. Mechanochemical coupling in the myosin motor domain. I. Insights from equilibrium active-site simulations. PLoS Comput Biol 2007;3:e21.
46. Ding J, Das K, Hsiou Y, Sarafianos SG, Clark AD, Jr., Jacobo-Molina A, Tantillo C, Hughes SH, Arnold E. Structure and functional implications of the polymerase active site region in a complex of HIV-1 RT with a double-stranded DNA template-primer and an antibody Fab fragment at 2.8 A resolution. J Mol Biol 1998;284:1095–1111.
47. Arts EJ, Le Grice SF. Interaction of retroviral reverse transcriptase with template-primer duplexes during replication. Prog Nucleic Acid Res Mol Biol 1998;58:339–393.
48. Whitcomb JM, Hughes SH. Retroviral reverse transcription and integration: progress and problems. Annu Rev Cell Biol 1992;8:275–306.
49. Jacobo-Molina A, Arnold E. HIV reverse transcriptase structure-function relationships. Biochemistry 1991;30:6351–6356.
50. Seckler JM, Howard KJ, Barkley MD, Wintrode PL. Solution structural dynamics of HIV-1 reverse transcriptase heterodimer. Biochemistry 2009;48:7646–7655.
51. Madrid M, Lukin JA, Madura JD, Ding J, Arnold E. Molecular dynamics of HIV-1 reverse transcriptase indicates increased flexibility upon DNA binding. Proteins 2001;45:176–182.
52. Bahar I, Erman B, Jernigan RL, Atilgan AR, Covell DG. Collective motions in HIV-1 reverse transcriptase: examination of flexibility and enzyme function. J Mol Biol 1999;285:1023–1037.
53. Thirumalai D, Lorimer GH. Chaperonin-mediated protein folding. Annu Rev Biophys Biomol Struct 2001;30:245–269.
54. Sigler PB, Xu Z, Rye HS, Burston SG, Fenton WA, Horwich AL. Structure and function in GroEL-mediated protein folding. Annu Rev Biochem 1998;67:581–608.
55. Horovitz A. Structural aspects of GroEL function. Curr Opin Struct Biol 1998;8:93–100.
56. Xu Z, Horwich AL, Sigler PB. The crystal structure of the asymmetric GroEL-GroES-(ADP)7 chaperonin complex. Nature 1997;388:741–750.
57. Weissman JS, Rye HS, Fenton WA, Beechem JM, Horwich AL. Characterization of the active intermediate of a GroEL-GroES-mediated protein folding reaction. Cell 1996;84:481–490.
58. Corrales FJ, Fersht AR. Toward a mechanism for GroEL.GroES chaperone activity: an ATPase-gated and -pulsed folding and annealing cage. Proc Natl Acad Sci U S A 1996;93:4509–4512.
59. Clarke AR. Molecular chaperones in protein folding and translocation. Curr Opin Struct Biol 1996;6:43–50.
60. Keskin O, Bahar I, Flatow D, Covell DG, Jernigan RL. Molecular mechanisms of chaperonin GroEL-GroES function. Biochemistry 2002;41:491–501.
61. Terada TP, Sasai M, Yomo T. Conformational change of the actomyosin complex drives the multiple stepping movement. Proc Natl Acad Sci U S A 2002;99:9202–9206.
62. Luque I, Leavitt SA, Freire E. The linkage between protein folding and functional cooperativity: two sides of the same coin? Annu Rev Biophys Biomol Struct 2002;31:235–256.
63. Rief M, Gautel M, Oesterhelt F, Fernandez JM, Gaub HE. Reversible unfolding of individual titin immunoglobulin domains by AFM. Science 1997;276:1109–1112.
64. Kellermayer MS, Smith SB, Granzier HL, Bustamante C. Folding-unfolding transitions in single titin molecules characterized with laser tweezers. Science 1997;276:1112–1116.
65. Hilser VJ, Freire E. Predicting the equilibrium protein folding pathway: structure-based analysis of staphylococcal nuclease. Proteins 1997;27:171–183.
66. Ansari A, Berendzen J, Bowne SF, Frauenfelder H, Iben IE, Sauke TB, Shyamsunder E, Young RD. Protein states and proteinquakes. Proc Natl Acad Sci U S A 1985;82:5000–5004.
67. Lu M, Poon B, Ma J. A new method for coarse-grained elastic normal-mode analysis. J Chem Theory Comput 2006;2:464–471.
68. Gohlke H, Thorpe MF. A natural coarse graining for simulating large biomolecular motion. Biophys J 2006;91:2115–2120.
69. Lu M, Ma J. Normal mode analysis with molecular geometry restraints: bridging molecular mechanics and elastic models. Arch Biochem Biophys 2011;508:64–71.
70. Weber J, Senior AE. ATP synthase: what we know about ATP hydrolysis and what we do not know about ATP synthesis. Biochim Biophys Acta 2000;1458:300–309.
71. Ren H, Allison WS. On what makes the gamma subunit spin during ATP hydrolysis by F(1). Biochim Biophys Acta 2000;1458:221–233.
72. Boyer PD. The ATP synthase–a splendid molecular machine. Annu Rev Biochem 1997;66:717–749.
73. Cui Q, Li G, Ma J, Karplus M. A normal mode analysis of structural plasticity in the biomolecular motor F(1)-ATPase. J Mol Biol 2004;340:345–372.
74. Stark H, Rodnina MV, Wieden HJ, van Heel M, Wintermeyer W. Large-scale movement of elongation factor G and extensive conformational change of the ribosome during translocation. Cell 2000;100:301–309.
75. Frank J, Agrawal RK. A ratchet-like inter-subunit reorganization of the ribosome during translocation. Nature 2000;406:318–322.
76. Agrawal RK, Heagle AB, Penczek P, Grassucci RA, Frank J. EF-G-dependent GTP hydrolysis induces translocation accompanied by large conformational changes in the 70S ribosome. Nat Struct Biol 1999;6:643–647.
77. Wang Y, Rader AJ, Bahar I, Jernigan RL. Global ribosome motions revealed with elastic network model. J Struct Biol 2004;147:302–314.
78. Tama F, Valle M, Frank J, Brooks CL, 3rd. Dynamic reorganization of the functionally active ribosome explored by normal mode analysis and cryo-electron microscopy. Proc Natl Acad Sci U S A 2003;100:9319–9323.