Biophysical *Journal*
Article

Biophysical Society

hdANM: a new comprehensive dynamics model for protein hinges

Pranav M. Khade,[1] Domenico Scaramozzino,[2] Ambuj Kumar,[1] Giuseppe Lacidogna,[2] Alberto Carpinteri,[2] and Robert L. Jernigan[1,*]

[1]Roy J. Carver Department of Biochemistry, Biophysics, and Molecular Biology, Iowa State University, Ames, Iowa and [2]Department of Structural, Geotechnical, and Building Engineering, Politecnico di Torino, Torino, Italy

ABSTRACT Hinge motions are essential for many protein functions, and their dynamics are important to understand underlying biological mechanisms. The ways that these motions are represented by various computational methods differ significantly. By focusing on a specific class of motion, we have developed a new hinge-domain anisotropic network model (hdANM) that is based on the prior identification of flexible hinges and rigid domains in the protein structure and the subsequent generation of global hinge motions. This yields a set of motions in which the relative translations and rotations of the rigid domains are modulated and controlled by the deformation of the flexible hinges, leading to a more restricted, specific view of these motions. hdANM is the first model, to our knowledge, that combines information about protein hinges and domains to model the characteristic hinge motions of a protein. The motions predicted with this new elastic network model provide important conceptual advantages for understanding the underlying biological mechanisms. As a matter of fact, the generated hinge movements are found to resemble the expected mechanisms required for the biological functions of diverse proteins. Another advantage of this model is that the domain-level coarse graining makes it significantly more computationally efficient, enabling the generation of hinge motions within even the largest molecular assemblies, such as those from cryo-electron microscopy. hdANM is also comprehensive as it can perform in the same way as the well-known protein dynamics models (anisotropic network model, rotations-translations of blocks, and nonlinear rigid block normal mode analysis), depending on the definition of flexible and rigid parts in the protein structure and on whether the motions are extrapolated in a linear or nonlinear fashion. Furthermore, our results indicate that hdANM produces more realistic motions as compared to the anisotropic network model. hdANM is an open-source software, freely available, and hosted on a user-friendly website.

SIGNIFICANCE Comprehending protein motions is essential for understanding protein structure-function relationships. Protein motions are often regulated by the translation and rotation of rigid domains connected by flexible hinge regions. The hinge-domain anisotropic network model is the first, to our knowledge, computational elastic network model that utilizes explicit hinge and domain structural information to directly simulate protein motions, thereby creating realistic motions that can be informative about the expected mechanism.

INTRODUCTION

Proteins participate in virtually every critical process within cells and are central to nearly all cellular functions. A central paradigm of structural biology relies on the significance of protein structure to understand biological function. Proteins have essential functional dynamics and do not persist in the conformation observed in the static crystal structures, but rather act as dynamic entities having internal motions that relate directly to their essential functions. Therefore, the functional behaviors of proteins are determined by their structures and, in further detail, by the dynamics of their structures. Protein structures exhibit different parts with different levels of cohesion, principally controlled by the packing densities. Densely packed cores have a tendency to show cooperative motions, which are important for protein mechanisms. These cooperative motions of the rigid parts with respect to one another can generate large conformational changes that are essential for protein function. With the recent advances in structure

Submitted May 28, 2021, and accepted for publication October 18, 2021.

*Correspondence: jernigan@iastate.edu

Pranav M. Khade, Domenico Scaramozzino, and Ambuj Kumar contributed equally to this work.

Editor: Alexandr Kornev.

https://doi.org/10.1016/j.bpj.2021.10.017

predictions from AlphaFold2 (1) and RoseTTAFold (2), as well as the ever-growing numbers of experimental x-ray, NMR, and cryo-electron microscopy structures, the study of large-scale protein motions and the associated molecular mechanisms is being pursued ahead.

In their environment within the cell, a protein's motions are affected by the specific interactions it has with other proteins, nucleic acids, small molecules, and solvent as well as by their own activated and inactivated conformations. Enzymes may typically undergo several specific stages such as activation, substrate binding, chemical turnover, and product release, with each of these stages requiring unique, often large, conformational changes. Experimentally, it is difficult to obtain snapshots of all of these states to learn explicitly the detailed mechanisms necessary for each step in a specific biological process. Theoretical models and computational tools are essential for providing the required insights to understand protein mechanisms. For this purpose, various computational approaches, such as atomic molecular dynamics (MD) simulations (3) and elastic network models (ENMs) (4–7), have been pursued. MD simulations have been widely applied to study various molecular motions, such as the local fluctuations within a protein, protein folding, conformational transitions, and ligand transport (3). MD simulations help in exploring the conformational landscape available to a protein by linking its structure, its dynamics, and its energy landscape. Although the advances in MD simulations have significantly improved the ability to study slower protein motions occurring on longer timescales, slow processes are still not accessible for the largest systems.

Normal mode analysis has been an alternative widely applied technique to study protein dynamics (8), drug design (9), and molecular docking (10), as well as to understand large-scale protein motions (11) in general. Numerical studies based on normal mode analysis have also shown that global protein vibrations occur at terahertz frequencies and are correlated with the biologically functional motions, suggesting that resonance phenomena at these frequencies might be involved in some protein actions (12–14). ENMs have been particularly useful in studying the dynamics and the intrinsic motions of proteins and in providing valuable mechanistic insights. Despite modeling the protein structure as a simplified network of springs, these models have been able to provide accurate information about the characteristic protein dynamics in terms of relative amplitudes and directionalities of fluctuations (15). Because they provide this information directly, these models have proven invaluable for understanding mechanisms. Most importantly, they quickly provide access to even the slowest functional motions of a protein structure. After the design of the first ENMs (4,5) and successes in the prediction of slow large-scale protein dynamics, they have been used extensively and have evolved into several different types. Some, such as the Gaussian network model (GNM) and the anisotropic network model (ANM) (7), can be used with either an atomistic or a coarse-grained representation of protein structure; in the latter case, the residues are being represented by their C^{α} atom only, in which the nodes of the network within a defined distance cutoff are connected by elastic springs. The GNM and ANM are the two most commonly used ENMs, and whereas the former yields only the magnitudes of dynamic fluctuations, the latter also provides the directionality of the predicted motions. Other models, such as rotations-translations of blocks (RTB) (16,17), adopt a different coarser-grained strategy for protein structure, modeled as a set of rotating and translating rigid blocks. Recently, the nonlinear rigid block normal mode analysis (NOLB) (18) was formulated, based on the RTB model, providing more realistic motions in relatively short computational times. NOLB starts from the RTB model of protein structure, but it further performs a nonlinear extrapolation of the protein motions by following curvilinear pathways within the large-scale regime—a major advantage for hinges, which are intrinsically curvilinear in nature. This is obtained by exploiting the nonlinear displacements induced by the rotations of the blocks around their centers. These nonlinear motions, due to block rotations, allow it to avoid the unrealistic stretching deformations that are usually found when the ENMs are amplified in a simple linear fashion, as is the case for ANM.

The key to any dynamics simulation being useful depends on its ability to capture a wide range of motion types, from local vibrational fluctuations to the largest global movements (19). Many protein motions can be defined by the translation and rotation of rigid blocks (rigid domains) mediated by the detailed motions at hinge residues (19). According to Gerstein's MolMovDB database (20), hinge motions are the most common type, accounting for ~45% of the total cases. Thus, understanding the fundamental aspects of hinge-bending mechanisms is likely to yield improved comprehension of the relationship between the structure and function for many proteins. Moreover, hinge motions, when combined with domain translation and rotation, have been suggested to play an important role in regulating the global motions of proteins (21,22). Therefore, combining hinge and domain information into a single model can assist in studying the dynamics related to protein mechanism with greater fidelity. The ANM and RTB approaches follow separate routes, with the protein structure being modeled either as a flexible network of springs or as a union of rigid blocks, respectively. Here, we take the features of both the ANM and the RTB approaches and combine the hinge and domain information together, proposing a more general and comprehensive model that we name the hinge-domain anisotropic network model (hdANM, available at https://hdanm.bb.iastate.edu/). This hdANM treats the protein as an elastic body in which both flexible spring networks (hinges) and rigid blocks (domains) are considered. This approach allows localizing the deformations within the flexible hinge regions, which in turn accommodates the large-scale motions

due to the rotations and translations of the rigid domains. Furthermore, because of the presence of rigid domains, we can also extrapolate the hdANM modes to follow curvilinear pathways due to the rotations of the domains that are important for hinge motions in general, thus yielding more realistic estimates of the large-scale protein mechanisms and not leading to the unrealistic stretches of structure observed in the ANM. Our method is a more general ENM, incorporating features from ANM, RTB, and NOLB, including flexible spring networks for the protein hinges as in ANM, rigid blocks for the domains as in the RTB, and nonlinear extrapolations for a more realistic description of the large-scale curvilinear motions as in NOLB. As a result, given an accurate description of domains and hinges, this new, to our knowledge, method retains all the advantages of previous ENMs while providing the option of studying the global protein motions with explicit hinge motions.

Estimating the hinge-domain motions of a protein first requires the identification of hinges and domains within a protein structure. Recently, we developed a novel, to our knowledge, hinge prediction method (19), called PACKMAN (https://packman.bb.iastate.edu/) software, that utilizes packing density to identify the domains and hinges of a protein structure as the most and least densely packed parts, respectively. PACKMAN hinge prediction is based on Delaunay tessellation, clustering, and nonparametric statistics to define the densely packed regions within a protein structure. In this work, we use our previously developed PACKMAN for hinge and domain identifications (based on the packing densities) to define the input hinge-domain partition for the hdANM calculations. As a result, hdANM can be seen to be a generalization of the traditional ANM models based upon additional hinge and domain assignments. One of the most prominent features of the hdANM is its general and comprehensive nature. Users can define all residues as hinge nodes and thus convert the model into a regular ANM or define various blocks as rigid domains to obtain an RTB model. Finally, the user can also select the nonlinear extrapolation option to obtain an NOLB-like model. Apart from using hinge and domain residues predicted by PACKMAN, hdANM also allows the user to provide manually the hinge and domain residue specifications. Therefore, with reliable identifications of hinges and domains, hdANM provides a significant improvement to protein dynamics. Furthermore, hdANM also enables a user to specify domain sizes, which can strongly affect the required computational efficiency because domains are reduced to a single geometric point. Despite this coarse graining, the method still has the ability to generate reliable motions, with any unrealistic stretching of bonds being mitigated. Here, principal components (PCs) are used to assess the performance of hdANM for predicting the known protein conformational motions because PCs of experimental structures have been widely used to characterize protein motions (23–25).

MATERIAL AND METHODS

hdANM

Conceptually, hdANM can be considered as a fundamental generalization of the previously developed ANM and RTB methods. The ANM takes into account the three-dimensional structure of a protein considered as a network of linear springs connecting the nodes. These springs are meant to represent the interactions between different atoms (atomic models) or residues (coarse-grained models), whose force constants play the role of the stiffness of each connection defining the $3N \times 3N$ Hessian of the system $\boldsymbol{H}$, which contains the second derivatives of the potential energy. The details are not so important because it has been shown that the overall shape of a protein is the main determinant of its dynamics. This matrix is then used to obtain information about the fundamental vibrational frequencies and mode shapes of the protein structure by solving the eigenvalue-eigenvector problem to evaluate the intrinsic fluctuations of the system (7). Following a different approach from a modeling perspective, the RTB method starts from the Hessian matrix $\boldsymbol{H}$ and projects it into a smaller subspace, obtained by dividing the molecule into n_b blocks. It considers these blocks as rigid bodies in three-dimensional space, each of them having six degrees of freedom (DOFs), namely three translations and three rotations. Hence, the RTB Hessian $\boldsymbol{H_b}$ of the system is a $6n_b \times 6n_b$ matrix, which is obtained by projecting the initial DOFs of the network ($3N$) into the final ones ($6n_b$) by means of a $3N \times 6n_b$ projection matrix. The hdANM presented here models the protein with both flexible and rigid parts (see the Supporting materials and methods).

Here, we use our previously developed method PACKMAN (19) to identify hinge and domain regions within the protein structure. The flexible regions of the network are the hinges with the domains being treated as rigid bodies. In this way, by using hdANM, the molecule can be modeled as a group of flexible and rigid parts. If no rigid domains were considered, then the hdANM would simply correspond to the standard ANM; at the other extreme, if no hinges were included but only a set of rigid domains were present in the model, then the hdANM would simply become the RTB model. However, the advantage of this new ENM is not simply that it can perform as either the ANM or RTB, but rather, when both hinges (flexible segments) and domains (rigid segments) are considered, the model combines the main features of RTB and ANM to yield more realistic motions of the protein. Moreover, the hdANM also treats rotations of the domains as nonlinear motions, thus obtaining large-scale realistic curvilinear pathways, similarly to the NOLB. The full details for the basis of the hdANM are provided in the Supporting materials and methods.

PCs and overlap analysis

The main comparison with experimental data that can be made for the ENMs is against the apparent motions obtained from a set of experimentally determined x-ray crystal structures, in a similar way to previous studies (23,26). For each protein, a data set of n crystal structures with m C^{α} atoms are collected from the Protein Data Bank (PDB). The position of each atom is specified by its Cartesian coordinates, x, y, and z. We collect all the crystal structures for a single protein and combine them into an $n \times 3m$ matrix M. The $3m \times 3m$ dimensional variance-covariance matrix $\boldsymbol{C}$ can be computed from

$$c_{ij} = \sum_{k=1}^{n} \frac{\left(\xi_{ki} - \overline{\xi_i}\right)\left(\xi_{kj} - \overline{\xi_j}\right)}{n-1}, \quad \forall\ 1 \le i,\ j \le 3m, \qquad (1)$$

where ξ_{ki} represents the position of the ith point in the kth structure in the M matrix and $\overline{\xi_i}$ represents the mean position of the ith point across all the structures, the sum being over all n structures. The covariance matrix $\boldsymbol{C}$ can be decomposed as $\boldsymbol{C} = \mathbb{E}\Delta\mathbb{E}^{\boldsymbol{T}}$, in which $\mathbb{E}$ is the eigenvector matrix,

with each column of $3m$-values. Each column of the $\mathbb{E}$ matrix contains linearly independent, orthogonal vectors for the direction of variances in the M matrix. Eigenvalues are sorted in descending order, each eigenvalue related to the amount of variance captured by that mode from the M matrix. The projection of points of each eigenvector is called the PC. The overlaps between the vector direction of each PC and the vector direction of each normal mode can be calculated as

$$O = \frac{|\boldsymbol{A}\cdot\boldsymbol{B}|}{\|\boldsymbol{A}\cdot\boldsymbol{B}\|}, \tag{2}$$

where $\boldsymbol{A}$ is the normal mode obtained from the dynamics model (ANM, RTB, hdANM, etc.) and $\boldsymbol{B}$ is the PC obtained from the set of crystal structures. Here, we use the initial instantaneous direction of hdANM modes to calculate the overlap values with the experimental PCs. This overlap value provides a numerical measure, between 0 and 1, of the similarity in directions between the computed motion and the experimentally observed ones within the set of crystal structures. Overlaps with the experimental PCs have also been calculated by using the ANM and RTB modes.

Cross correlation calculation

The hdANM eigenvectors ($\boldsymbol{u}$) and eigenvalues (λ) from singular-value decomposition can also be used to calculate the pseudoinverse hdANM Hessian matrix ($\boldsymbol{H}^{-1}$) as

$$\boldsymbol{H}^{-1} = \sum_{i\geq 7}\frac{1}{\lambda_i}\boldsymbol{u}_i\boldsymbol{u}_i^{\boldsymbol{T}} \tag{3}$$

The first six eigenvectors and eigenvalues are not included in the calculation because they correspond to the rigid-body translations and rotations of the whole protein structure. The cross correlation matrix $\boldsymbol{D}$ can be estimated by calculating the correlation between each protein residue pair i and j, using

$$D_{ij} = \frac{tr\left(\boldsymbol{H}_{\boldsymbol{ij}}^{-1}\right)}{\sqrt{tr\left(\boldsymbol{H}_{\boldsymbol{ii}}^{-1}\right)\times tr\left(\boldsymbol{H}_{\boldsymbol{jj}}^{-1}\right)}}, \tag{4}$$

where $\boldsymbol{H}^{-1}{}_{\boldsymbol{ij}}$ represents the 3×3 sub-block of the pseudoinverse Hessian matrix $\boldsymbol{H}^{-1}$ associated with nodes i and j, and the operator $tr()$ represents the trace of the matrix within the brackets.

Contacts

Elastic contacts are defined by using a distance threshold of 10 Å between C^{α} atoms, and, from experience with the GNM and ANM, we would expect the results to be relatively insensitive to the exact value of this cutoff distance. Their evolution throughout the protein motion is analyzed to assess whether the proposed computational model is able to preserve the contacts between the nodes of the network during the motion, thus avoiding unrealistic stretching of certain parts of the protein as well as the occurrence of too many broken bonds. Additional details related to the elastic contact calculation and analysis are provided in the Supporting materials and methods.

RESULTS AND DISCUSSION

The main aim of this work has been to generate an improved novel, to our knowledge, and comprehensive ENM, allowing the computation of normal modes for even the largest macromolecular complexes by combining translations and rotations within the rigid domains with deformations within the hinge residues, as well as to take advantage of our recent work, PACKMAN, that reliably divides structures into hinges and rigid domains. Coarse graining of the protein structures has often been used to reduce the computational cost, although none of the previous models have explicitly considered using the characteristic hinge-domain partitioning of a protein structure to achieve this reduction. hdANM is a reliable way of combining hinge and domain information together in an ENM to study the characteristic hinge motions of a protein. The most commonly used ENMs rely either on ANM or RTB. In ANM, each point in a protein structure is represented by a flexible node connected by linear springs, whereas RTB builds a union of rigid blocks composed of protein atoms that rotate and translate around their centers of mass. ANM has been mainly used to investigate the low-frequency modes of a protein structure that are difficult to obtain from MD. However, because of the lack of rigid domains, the effects of its motions are spread throughout the structure, often resulting in unrealistic stretching of some parts of the protein, especially when the motions are extended to changes. Such unrealistic stretching is avoided by the introduction of curvilinear motions into the RTB approach in the NOLB method. Although the definition of rigid blocks may help with preventing unrealistic stretching, the somewhat arbitrary definition of these blocks in the RTB-NOLB framework may lead to deviations from the actual translations and rotations observed in the protein. Moreover, the absence of flexible regions in the RTB models prevents the observation of the details of motions at hinge sites.

Therefore, the model developed in this work has two gains: not only does it allow the heavy coarse graining of a large molecule, thereby reducing the computational burden, but it also presents the first model, to our knowledge, to combine flexible hinges and rigid domains together for the extraction of global protein motions. Here, multiple protein structures are used to evaluate the ability of hdANM to capture the global motions. Our results indicate that, despite extreme coarse graining, the new, to our knowledge, hdANM method yields good overlap agreement with the PCs from experimental structures (Table 1) while also generating realistic rotational-translational motions of the protein domains in the large-scale regime (see Supporting Videos). It should be noted that a certain loss in the overlap value with the PCs has been observed when compared with ANM and RTB in some cases investigated. This is presumed to arise mainly because of the major coarse graining at the domain levels in hdANM. However, the instantaneous direction of one of the first low-frequency motions predicted by the hdANM has always been found to provide a non-negligible overlap with the first PC of the investigated proteins (overlap values higher than 60%).

TABLE 1 Overlaps between hdANM, ANM, and RTB modes with PCs

Protein (PDB)	PC (% variance)	hdANM[a]	ANM	RTB
Calmodulin (PDB:1EXR)	PC1 (42)	0.73 (mode 3)	0.58 (mode 1)	0.59 (mode 2)
	PC2 (22)	0.68 (mode 2)	0.62 (mode 2)	0.78 (mode 1)
HLA class I histocompatibility antigen, A-2 alpha chain (PDB:3KLA)	PC 1 (33)	0.86 (mode 1)	0.88 (mode 1)	0.80 (mode 1)
	PC 2 (15)	0.51 (mode 2)	0.67 (mode 2)	0.61 (mode 2)
MHC class I antigen (PDB:1ZSD)	PC 1 (40)	0.64 (mode 1)	0.64 (mode 1)	0.73 (mode 1)
	PC 2 (20)	0.36 (mode 3)	0.41 (mode 3)	0.34 (mode 2)
DNA polymerase β (PDB:8ICZ)	PC 1 (90)	0.71 (mode 2)	0.77 (mode 2)	0.67 (mode 1)
	PC 2 (9)	0.59 (mode 4)	0.66 (mode 4)	0.47 (mode 2)
LAO-binding protein (PDB:2LAO)	PC1 (99)	0.61(mode 1)	0.73 (mode 1)	0.95 (mode 1)
	PC2 (1)	0.40 (mode 2)	0.58 (mode 2)	0.65 (mode 2)

Overlap scores from modes that have the highest overlap with PC1 and PC2 are shown here.
[a]The specific domains and hinge residues for the investigated proteins used for this analysis can be found in Table S1.

The different levels of coarse graining between hdANM and the two standard approaches (ANM and RTB) can be seen in Table 2. This table reports the DOFs for the ANM, RTB, and hdANM treatments of the investigated proteins. As reported above and in the Supporting materials and methods, the DOFs of the ANM are $3N$ (N being the number of the protein residues), and the DOFs of the RTB are $6n_b$ (n_b being the total number of defined rigid blocks), whereas the DOFs of the hdANM are $6D + 3n_H$ (D being the total number of rigid domains in the protein structure and n_H being the total number of residues within the flexible hinge regions). As can be seen from Table 2, hdANM leads to a reduction in the DOFs of more than 50% for all cases. Notice that this reduction could be further increased if larger domains are considered and applied in the hinge-domain partition of the structure. Because the number of DOFs is directly related to the dimensions of the matrices to be diagonalized in the eigenvalue decomposition (Hessian and mass matrices), such DOF reduction makes the hdANM significantly more computationally efficient than ANM and RTB. Because eigenvalue decomposition algorithms often present a computational complexity of the order of $O(DOF^3)$, a reduction of 50% in the total number of DOFs might lead to a reduction in the computational complexity of around 75–88%. Nevertheless, the results shown in Table 1 show that, despite this larger amount of coarse graining with respect to ANM and RTB, the first PC is captured by one of the low-frequency hdANM modes with an overlap similarity higher than 60%. Thinking about larger molecular assemblies such as microtubules, these can range from one layer containing 13 tubulin dimers and more than 10,000 residues to a very large fiber reaching a length of 800 nm with more than 1 million residues. It is clear there that the major reduction in the DOFs obtained by modeling the hinges and domains of a microtubule fiber with hdANM is critical for such cases to enable the generation of dynamics. Also, the advent of widespread structure determination by cryo-electron microscopy immediately rapidly increases the need to access the dynamics of ever larger molecular complexes.

Functions of proteins largely depend on their global intrinsic motions. Despite the large number of studies, it remains a challenging task to compute the global motions required for functional and mechanistic insights. ENMs such as ANM and GNM enable the study of intrinsic protein motions (15), but large structural deformations often show unrealistic stretching of springs throughout a structure, making it challenging to understand realistically the underlying biological mechanism. Our results indicate significantly smaller changes in elastically stretched distances for hdANM nonlinear motions in comparison with traditional ANM modes for larger root mean-square deviation (RMSD) values (Fig. 1; Figs. S1–S5). We have also investigated the evolution of the number of contacts throughout the simulated motions

TABLE 2 DOFs considered in the hdANM, ANM, and RTB models of the investigated proteins

	hdANM			ANM		RTB	
Protein (PDB)	D[a]	n_H[a]	DOFs	N	DOFs	n_b ($n_{r,b}$[b])	DOFs
Calmodulin (PDB:1EXR)	2	36	120	146	438	146 (1)	876
HLA class I histocompatibility antigen, A-2 alpha chain (PDB:3KLA)	7	98	336	275	825	138 (2)	828
MHC class I antigen (PDB:1ZSD)	7	101	345	276	828	138 (2)	828
DNA polymerase β (PDB:8ICZ)	7	129	429	327	981	164 (2)	984
LAO-binding protein (PDB:2LAO)	5	92	306	238	714	119 (2)	714

[a]The specific domains and hinge residues for the investigated proteins used for this analysis can be found in Table S1.
[b]$n_{r,b}$ represents the number of residues per block considered for the definition of the RTB model. This value follows from the default settings obtained by the RTB webserver, elNémo (15).

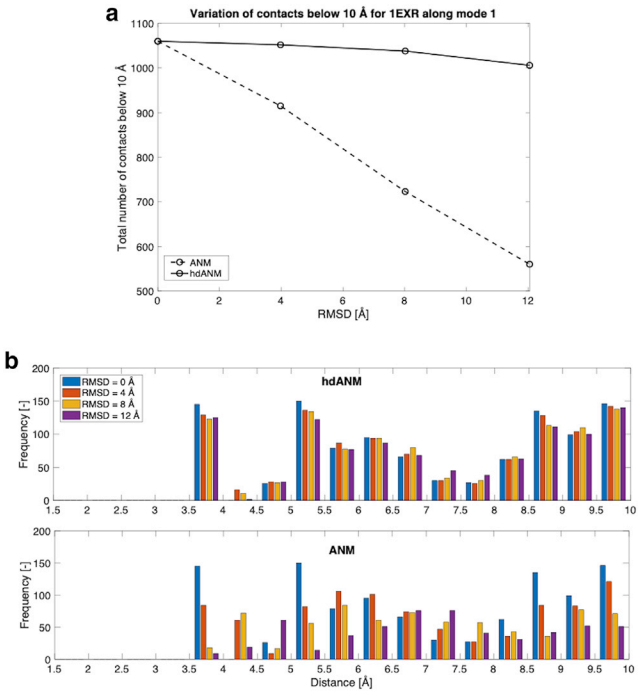

FIGURE 1 Conservation of contacts in calmodulin during the motion of the first normal mode of hdANM in comparison with ANM. (*a*) Number of contacts within 10 Å in calmodulin (PDB: 1EXR). (*b*) Number of contacts within 10 Å in calmodulin (PDB: 1EXR) distributed over different ranges of distance. To see this figure in color, go online.

for different ranges of distances. The number of contacts within each bin was calculated for 0, 4, 8, and 12 Å extents of RMSD deformations for the first ANM mode and the first hdANM nonlinear motion for calmodulin, as shown in Fig. 1. Fig. 1 *a* shows the total variation of elastic contacts throughout the hdANM and ANM motions. As can be seen, the motion evaluated from the first low-frequency hdANM mode leads only to a slight decrease in the number of contacts during the deformation of the structure (1060 contacts in the reference structure vs. 1004 after a 12 Å RMSD deformation, with a loss of 56 contacts, i.e., −5.3%). Conversely, following the first ANM mode to that same extent leads to a major reduction in the total number of contacts (1060 in the reference structure vs. 556 after 12 Å RMSD deformation, with a loss of 504 contacts, i.e., 47.5%). This suggests that because of the presence of rigid domains and the nonlinear extrapolation of motions due to domain rotations, the hdANM motions are better able to preserve the connectivity of the protein network throughout its motions, even for large-scale deformations. Fig. 1 *b* reports the distribution of these elastic contacts based on their length (from 0 to 10 Å, with bin widths of 0.5 Å) and its evolution for increasing RMSD values. Because the changes in the distribution are much smaller for hdANM than for ANM, this suggests that the hdANM motions preserve the local internal distances significantly better than does ANM, indicating that hdANM introduces less distortion in the network of the protein structure even for large-scale deformations (Fig. 1; Figs. S1–S5). As suggested above, this is mainly due to the presence of rigid domains explicitly treated as rigid in the hdANM as well as the extrapolation of the motions to follow curvilinear paths in the large-scale regime. This preserves most of the contacts in the protein network and allows them to maintain their original distance distribution throughout the entire pathway, showing substantially less disruption to the structure (Fig. 1; Figs. S1–S5).

After this, we look at seven specific proteins, in which hdANM is expected to capture their functionally relevant hinge motions—annexin, inorganic pyrophosphatase (PPase), calmodulin, major histocompatibility complex class I (MHC I) protein, DNA polymerase β, lysine/arginine/ornithine-binding periplasmic protein (LAO-binding protein), and human immunodeficiency virus 1 (HIV-1) protease.

Annexins are homologous proteins that bind to membranes in a calcium-dependent manner. The precise physiological roles of annexin are not well defined. They have been shown to have a variety of activities such as facilitating membrane fusion, anticoagulation and antiinflammatory response signals, calcium channel formation, and phospholipase A2 inhibition, as well as engaging in cytoskeleton interactions (27). The annexin structure is composed of a planar array of four homologous domains (I–IV) or repeats, each consisting of five α-II/III repeats arranged in twofold symmetry separated by a groove (28). There are two major hinge motions that are expected to regulate the biological activity of annexin: 1) the angle changes between the two modules, which may regulate the calcium binding and channel formation, and 2) rotation of torsional dihedral angles between the modules (28). Diagrammatic representations of these two hinge motions were shown by Cregut et al. (Fig. 2 in (26)). Based on the definition of the hinges from PACKMAN, which are similar to the experimentally known hinges (28) in the annexin structure as reported in Table S2, hdANM is able to capture the first hinge motion in the second nonrigid body motion and the second hinge motion in the first nonrigid body motion (Videos S1 and S2, respectively). Four rigid domain motions can be clearly observed within the hdANM cross correlation map (Fig. S6). The opening and closing of the channel shown in the second nonrigid body motion of hdANM agrees with the mechanism of calcium channel formation that was proposed for annexin (28).

PPases are essential enzymes that play a significant role in regulating the cellular concentration of inorganic pyrophosphate (29). PPase regulation of cellular phosphate assists in regulating biosynthetic reactions such as nucleic acid and protein synthesis to completion (29). PPases catalyze the hydrolysis of the symmetrical pyrophosphate and thereby regulate the intracellular equilibria of various important physiological reactions. Family II PPases contain N- and C-terminal domains, separated by hinge residues

(30). They contain a conserved Asp-His-His motif in the N-terminal domain that is essential for metal ion binding (29). During catalysis, the C-terminal domain closes over the N-terminal domain by translating and rotating across the hinges (Fig. 2). hdANM captures the translation and rotation involved in its first nonrigid motion without any extensive stretching of bonds within the protein domains (Video S3). The motion cross correlations are shown in Fig. S7.

Calmodulin is a calcium-binding messenger protein that regulates multiple signaling pathways. Calmodulin activity is regulated by the motion of the hinge residues around its central axis. Binding of calcium to calcium-free (apo) calmodulin triggers an initial conformational change that causes the central helix (residues 65–77) to straighten along with a relative rotation of the two globular domains (31,32). When calcium-bound calmodulin binds a peptide, a conformational change is triggered, allowing the two globular domains to wrap around the peptide, pushing the structure toward its closed conformation (31,32). Here, we observe a twisting motion of the globular domains that includes a bending motion at the central hinge, pushing the structure toward its closed conformation (Fig. 3; Video S4). We are able to capture the twisting motion of the globular domains significantly better than ANM. ANM shows extensive stretching of bonds within the domains and hinges when following the twisting motion (Video S5). Moreover, despite the higher level of coarse graining, there is an observed 73% overlap between the first PC and the third lowest-frequency hdANM mode, which is greater than the 58 and 59% overlap between the first PC and the best overlapping ANM and RTB modes, respectively (Table 1). The cross correlations of calmodulin domains and hinge region are shown in Fig. S8.

The antigen-presenting MHC I protein plays an important role in regulating the adaptive immune system by presenting peptides for recognition by T cells (33). Antigen binding is mediated by two domains, originating from a single, heavy α-chain. These two domains are composed of a β-sheet that forms a base for the molecule with two α-helices on top of it (33). These two α-helices (α1 and α2) are sufficiently separated to accommodate the binding of a variety of peptides (Fig. 4). The binding of peptides between these two α-helices is mediated by the formation of hydrogen bonds between the side chains of the MHC I molecules and the peptide backbone (33). Furthermore, the occupancy of the peptide in the pocket is also regulated by the conformational state of the pocket. Therefore, a better understanding of the MHC I conformational changes during the equilibrium fluctuations of the structure is important for making predictions of the affinity for the individual MHC-antigen interactions. Our results show that hdANM produces a rotational-transitional motion of the MHC I domains that is associated with its open-to-closed transition (Fig. 4; Video S7). Obtaining the MHC I open-to-closed conformational change, as shown in Fig. 4, without any accumulation of unrealistically elongated stretched bonds will lead to a better understanding of the peptide binding affinity to the MHC. Residue motion dynamics, shown in the cross correlation map (Fig. S9), represents the correlated motions within the MHC I protein. It should be noted that the hdANM model should yield better estimates of entropy changes than either GNM, ANM, or NOLB.

DNA polymerase β is a repair polymerase that plays a key role in base-excision repair. It has 5′-deoxyribose-5-phosphate lyase activity that removes the 5′ sugar phosphate and acts as a DNA polymerase to add one nucleotide to the 3′ end of a single-nucleotide gap (34). The protein undergoes a stepwise distributive motion to perform "gap filling" during DNA synthesis (35). The polymerase and lyase domains of DNA polymerase β contribute to the DNA repair enzymatic activity (35). Crystal structures of the catalytic intermediate of this molecule have provided mechanistic insight into its role in the genome (36). Understanding the global motion of this structure can yield a deeper molecular insight into the process required for DNA repair (37). Here, we apply hdANM to a DNA polymerase β crystal structure and compute its overlap with the PCs obtained from a set of crystal structures. Our results indicate that hdANM shows a 71% overlap with the first PC, whereas ANM and RTB show a maximal overlap with the first PC of 77 and 67%, respectively (Table 1). This again indicates that, although we are coarse graining the molecule to a very high extent (see Table 2), we are still able to capture much of the conformational fluctuations within the set of crystal structures, with the same level of agreement as for

FIGURE 2 Conformational changes in inorganic PPase family II, moving from open (PDB: 1K23, chain A) to closed form (PDB: 1K20, chain A). Here, the open and closed conformations are from the PDB, and intermediate conformations are generated with hdANM (see Video S3) with the open conformation. The open-to-closed motion is mediated by rotation in the C-terminal domain after a rigid-body motion and hinge bending, leading to occlusion of the active site residues. Similarly, the closed-to-open transition enables the exposure of active site residues, enabling enzyme activation. Secondary structure information has been taken from the open conformation.

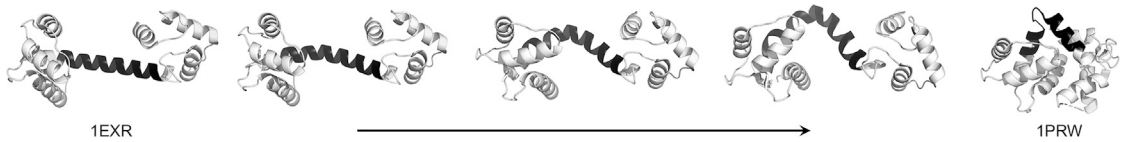

FIGURE 3 Hinge bending motion in calmodulin occurs on the pathway from its open (PDB: 1EXR) to closed form (PDB: 1PRW) (see Video S4 for details). The two EF-hand domains at the opposite ends of the structure in the open form, each containing helix-loop-helix motifs, undergo a mirrored type of motion followed by an orthogonal conformational transition, deforming and breaking the central helix and allowing the structure to move from the open to the closed form. Intermediate conformations are generated by applying hdANM (see Videos S4 and S6) on the open conformation. Secondary structure information has been taken from the open conformation (PDB: 1EXR).

previous coarse-grained dynamics approaches. Furthermore, the first nonrigid motion shows the large-scale opening and closing of the protein likely involved in regulating its enzymatic activity (Video S8). Fig. S10 shows the correlated motion of residues within the protein.

LAO-binding protein is a component of a bacterial periplasmic transport system called the ABC transporter system, which helps in translocating substrate from the periplasm to the cytoplasm (37). LAO-binding protein has a bilobate structure in which the first lobe consists of residues 1–88 and the second lobe consists of residues 93–185. The two lobes are attached by two short peptide segments from residues 89–92 and 186–194 (38). The connecting strands provide a base for the cleft and form a hinge between the two lobes. Our result shows that hdANM captures only 61% overlap with its first PC, whereas, in this case, ANM and RTB are able to capture 73% and, astonishingly, 95% of the first PC. The worse performance seen in the hdANM overlap may be associated with the domain-level coarse graining of the molecule. The difference between the overlap values found with the standard approaches, namely the ANM (73% overlap) and RTB (95% overlap), shows that the motions extracted for this specific protein are very sensitive to the details of the way in which its structure is modeled. However, despite this loss in the PC overlap, with hdANM, we are able to observe a more realistic opening and closing motion of the domains across the hinges in the large-scale regime without accumulating any severe deformations within the domains (Fig. 5; Video S9). The hdANM cross correlation map clearly represents the correlated dynamics of residues within each lobe of LAO-binding protein (Fig. S11).

HIV-1 protease dynamics has been of great interest because of its significant role as a drug target. It functions as a homodimer with each subunit comprised of 99 amino acids. It has a conserved (Asp25-Thr26-Gly27) catalytic triad active site (23). It also contains three domains essential for stabilization, dimerization, catalytic stability, and ligand binding. NMR studies and MD simulations indicated a role of the highly mobile flaps in regulating its substrate and inhibitor binding activity (39,40). These flaps play an important role in regulating the gating mechanism for the binding of substrate. It also helps in orienting the substrate in a conformation suitable for protease catalytic activity. The third nonrigid motion of the HIV-1 protease dimer obtained from hdANM clearly shows opening and closing of the flaps (Fig. 6; Video S10), which is essential for substrate binding and catalysis.

PC overlaps with the motions from hdANM have been computed for a diverse set of proteins to observe the differences between hdANM in comparisons with ANM and RTB. As mentioned above for specific cases, the results obtained indicate that the large coarse graining of domains in hdANM leads to both gains and losses in the overlaps with the PCs (see Table 1). For example, in the case of calmodulin, the PC overlap for hdANM is higher than that obtained with ANM and RTB. However, for other cases, we observe some loss in the PC overlaps with respect to the older approaches (see Table 1). The loss in overlaps with some of the experimental ensembles is most likely attributable to the complete rigidity assumed in the hdANM modeling for the domains, which conversely can have smaller internal flexible regions in the real protein structures, permitting some of the less densely packed intradomain regions to experience a certain amount of deformation. Examples of

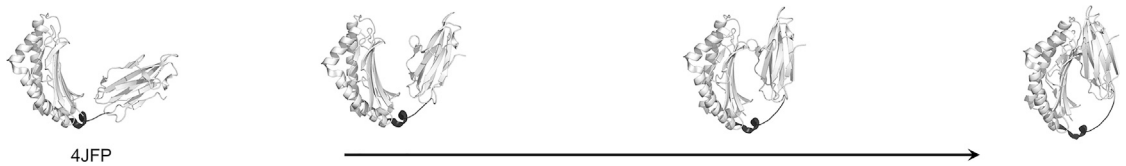

FIGURE 4 Hinge motion in MHC Class I structure. Open (PDB: 4JFP) to closed conformational transition (see Video S7 for details). Here, the open conformation is from PDB, and intermediate and closed conformations are generated by applying hdANM to the open conformation. The MHC Class I peptide selection process depends on the conformational along the pathway from the open-to-closed conformational transition. The conformations sampled during the transition can have a direct effect on the selection of peptides that bind. Secondary structures shown are taken from the open conformation (PDB: 4JFP).

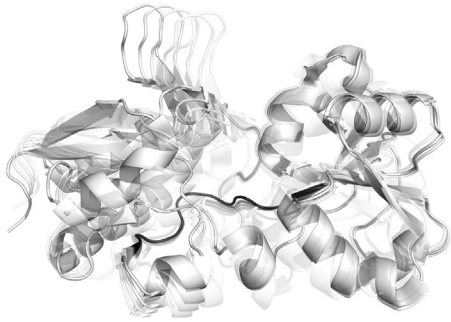

FIGURE 5 Conformations sampled along the pathway for the LAO-binding protein changing from open (6ML0) to closed (6MLP) forms are represented by hdANM in its first nonrigid mode. Here, open and closed conformations are collected from the PDB database, and intermediate conformations are generated by applying hdANM (see Video S9) on the open conformation. Secondary structure information is added from open conformation (PDB: 6ML0).

such motions are seen in our prior studies of the functional motions of the loops in triose phosphate isomerase (41). Nevertheless, there are major computational gains by coarse graining residues within the protein domains, which also provides a major conceptual advantage, hdANM being the first model, to our knowledge, to combine hinge and domain information together to model protein motions. Also, we have already seen how the first PC is often captured with more than 60% agreement by one of the low-frequency hdANM modes.

CONCLUSIONS

hdANM is the first model, to our knowledge, that builds a foundation for combining hinge and domain information within a protein, which is a major conceptual advantage in elastic network modeling for generating the detailed characteristic motions of large proteins with high computational efficiency. Selecting individual hinges and building models has a somewhat variable performance for the extracted motions; therefore, the results with all the possible hinges would likely show improvements if the hinges were selected heuristically. The details of hinge motions differ significantly among various proteins. The hinge in calmodulin is a point hinge that allows rotation as well as bending between the connected domains. Historically, this has been a particularly difficult hinge to understand, but our new, to our knowledge, model leads to a simple yet detailed understanding of its mechanics. In another case, we have seen how annexin contains multiple hinges arranged in parallel, which restricts the DOFs associated with the dynamics of that protein. Exploration and classification of such unique hinge motions will enable us to understand the functional mechanisms of proteins and assist in the design of new proteins.

ENMs have been extensively used to study the global motions and dynamics of many proteins with significant successes. Although the advances in ENMs have enabled us to obtain significant mechanistic insights into a wide variety of protein structures (42–45), the explicit motion of domains and hinge regions has not, to our knowledge, been studied as directly as here. The model developed in this work constitutes a hybrid approach that generates global motions of a protein using prior domain and hinge classification. Moreover, the PC results indicate that the significantly larger coarse graining of the molecule does not cause any severe changes in the representations of the motions when compared to those observable within sets of protein crystal structures; the overlap values obtained with our method are often the same order of magnitude obtained by the standard approaches, such as ANM and RTB. With further advancements in hinge-domain prediction methods and even selecting the hinges heuristically, we can improve hdANM to obtain even more realistic global motions of proteins. Other characteristic types of protein motions, such as shear motion, may also benefit from an approach similar to hdANM. Overall, this approach leads to a view of significantly more efficient and deterministic protein motion with less randomness, as required for proteins to achieve their remarkably high levels of efficiency.

SUPPORTING MATERIAL

Supporting material can be found online at https://doi.org/10.1016/j.bpj.2021.10.017.

AUTHOR CONTRIBUTIONS

Designed research, A.K., P.M.K., and D.S.; performed research, A.K., P.M.K., and D.S.; validation, A.K., P.M.K., D.S., G.L., A.C., and R.L.J.;

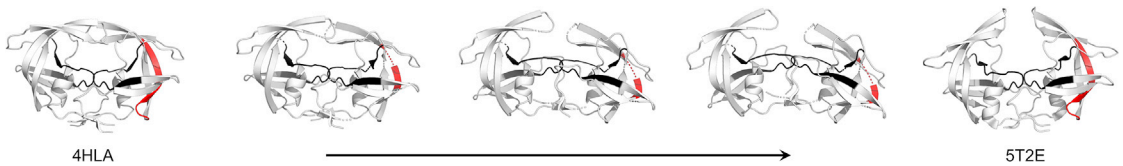

FIGURE 6 Motion of the HIV protease transition from closed (PDB: 4HLA) to open (PDB: 5T2E) form is represented by hdANM in its third nonrigid mode. The parts colored in red and black are the two distinct hinges. Here, open and closed conformations are collected from the PDB, and intermediate conformations are generated by applying hdANM (see Video S10) on the open conformation. Secondary structure information has been taken from the closed conformation. To see this figure in color, go online.

data analysis, A.K., P.M.K., and D.S.; data curation, A.K., P.M.K., and D.S.; manuscript writing, A.K., P.M.K., D.S., and R.L.J.; supervision, G.L., A.C., and R.L.J.

ACKNOWLEDGMENTS

We thank Dr. Guang Song and Dr. Sergei Grudinin for useful insights into hdANM performance evaluation. We also thank Research IT at Iowa State University for helping with many aspects of the computing.

We gratefully acknowledge the support provided by National Institutes of Health grant R01GM127701 and National Science Foundation grant DBI 1661391.

REFERENCES

1. Tunyasuvunakool, K., J. Adler, …, D. Hassabis. 2021. Highly accurate protein structure prediction for the human proteome. *Nature.* 596:590–596.
2. Baek, M., F. DiMaio, …, D. Baker. 2021. Accurate prediction of protein structures and interactions using a three-track neural network. *Science.* 373:871–876.
3. Hospital, A., J. R. Goñi, …, J. L. Gelpí. 2015. Molecular dynamics simulations: advances and applications. *Adv. Appl. Bioinform. Chem.* 8:37–47.
4. Tirion, M. M. 1996. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys. Rev. Lett.* 77:1905–1908.
5. Bahar, I., A. R. Atilgan, and B. Erman. 1997. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Fold. Des.* 2:173–181.
6. Haliloglu, T., I. Bahar, and B. Erman. 1997. Gaussian dynamics of folded proteins. *Phys. Rev. Lett.* 79:3090–3093.
7. Atilgan, A. R., S. R. Durell, …, I. Bahar. 2001. Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys. J.* 80:505–515.
8. Reuveni, S., R. Granek, and J. Klafter. 2008. Proteins: coexistence of stability and flexibility. *Phys. Rev. Lett.* 100:208101.
9. Yan, W., D. Zhang, …, G. Hu. 2018. Recent advances on the network models in target-based drug discovery. *Curr. Top. Med. Chem.* 18:1031–1043.
10. Wang, A., Y. Zhang, …, G. Li. 2020. Higher accuracy achieved for protein-ligand binding pose prediction by elastic network model-based ensemble docking. *J. Chem. Inf. Model.* 60:2939–2950.
11. Yang, L., G. Song, and R. L. Jernigan. 2007. How well can we understand large-scale protein motions using normal modes of elastic network models? *Biophys. J.* 93:920–929.
12. Carpinteri, A., G. Lacidogna, …, A. Bassani. 2017. Terahertz mechanical vibrations in lysozyme: Raman spectroscopy vs modal analysis. *J. Mol. Struct.* 1139:222–230.
13. Carpinteri, A., G. Piana, …, G. Lacidogna. 2019. Terahertz vibration modes in Na/K-ATPase. *J. Biomol. Struct. Dyn.* 37:256–264.
14. Scaramozzino, D., G. Lacidogna, …, A. Carpinteri. 2019. A finite-element-based coarse-grained model for global protein vibration. *Meccanica.* 54:1927–1940.
15. Yang, L., G. Song, and R. L. Jernigan. 2009. Protein elastic network models and the ranges of cooperativity. *Proc. Natl. Acad. Sci. USA.* 106:12347–12352.
16. Tama, F., F. X. Gadea, …, Y. H. Sanejouand. 2000. Building-block approach for determining low-frequency normal modes of macromolecules. *Proteins.* 41:1–7.
17. Suhre, K., and Y. H. Sanejouand. 2004. ElNemo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement. *Nucleic Acids Res.* 32:W610–W614.
18. Hoffmann, A., and S. Grudinin. 2017. NOLB: nonlinear rigid block normal-mode analysis method. *J. Chem. Theory Comput.* 13:2123–2134.
19. Khade, P. M., A. Kumar, and R. L. Jernigan. 2020. Characterizing and predicting protein hinges for mechanistic insight. *J. Mol. Biol.* 432:508–522.
20. Gerstein, M., and W. Krebs. 1998. A database of macromolecular motions. *Nucleic Acids Res.* 26:4280–4290.
21. Chapman, B. K., O. Davulcu, …, M. S. Chapman. 2015. Parsimony in protein conformational change. *Structure.* 23:1190–1198.
22. Magnusson, U., B. N. Chaudhuri, …, S. L. Mowbray. 2002. Hinge-bending motion of D-allose-binding protein from Escherichia coli: three open conformations. *J. Biol. Chem.* 277:14077–14084.
23. Yang, L., G. Song, …, R. L. Jernigan. 2008. Close correspondence between the motions from principal component analysis of multiple HIV-1 protease structures and elastic network modes. *Structure.* 16:321–330.
24. Yang, L. W., E. Eyal, …, A. Kitao. 2009. Principal component analysis of native ensembles of biomolecular structures (PCA_NEST): insights into functional dynamics. *Bioinformatics.* 25:606–614.
25. Sankar, K., S. K. Mishra, and R. L. Jernigan. 2018. Comparisons of protein dynamics from experimental structure ensembles, molecular dynamics ensembles, and coarse-grained elastic network models. *J. Phys. Chem. B.* 122:5409–5417.
26. Bahar, I., T. R. Lezon, …, E. Eyal. 2010. Global dynamics of proteins: bridging between structure and function. *Annu. Rev. Biophys.* 39:23–42.
27. Lizarbe, M. A., J. I. Barrasa, …, J. Turnay. 2013. Annexin-phospholipid interactions. Functional implications. *Int. J. Mol. Sci.* 14:2652–2683.
28. Cregut, D., G. Drin, …, L. Chiche. 1998. Hinge-bending motions in annexins: molecular dynamics and essential dynamics of apo-annexin V and of calcium bound annexin V and I. *Protein Eng.* 11:891–900.
29. Kajander, T., J. Kellosalo, and A. Goldman. 2013. Inorganic pyrophosphatases: one substrate, three mechanisms. *FEBS Lett.* 587:1863–1869.
30. Ahn, S., A. J. Milner, …, S. A. White. 2001. The "open" and "closed" structures of the type-C inorganic pyrophosphatases from Bacillus subtilis and Streptococcus gordonii. *J. Mol. Biol.* 313:797–811.
31. Shamsuddin, R., M. Doktorova, …, K. McMenimen. 2014. Computational prediction of hinge axes in proteins. *BMC Bioinformatics.* 15 (*Suppl 8*):S2.
32. Wriggers, W., E. Mehler, …, K. Schulten. 1998. Structure and dynamics of calmodulin in solution. *Biophys. J.* 74:1622–1639.
33. Wieczorek, M., E. T. Abualrous, …, C. Freund. 2017. Major histocompatibility complex (MHC) class I and MHC class II proteins: conformational plasticity in antigen presentation. *Front. Immunol.* 8:292.
34. Allinson, S. L., I. I. Dianova, and G. L. Dianov. 2001. DNA polymerase beta is the major dRP lyase involved in repair of oxidative base lesions in DNA by mammalian cell extracts. *EMBO J.* 20:6919–6926.
35. Singhal, R. K., R. Prasad, and S. H. Wilson. 1995. DNA polymerase beta conducts the gap-filling step in uracil-initiated base excision repair in a bovine testis nuclear extract. *J. Biol. Chem.* 270:949–957.
36. Beard, W. A., and S. H. Wilson. 2014. Structure and mechanism of DNA polymerase β. *Biochemistry.* 53:2768–2780.
37. Opresko, P. L., R. Shiman, and K. A. Eckert. 2000. Hydrophobic interactions in the hinge domain of DNA polymerase beta are important but not sufficient for maintaining fidelity of DNA synthesis. *Biochemistry.* 39:11399–11407.
38. Oh, B. H., J. Pandit, …, S. H. Kim. 1993. Three-dimensional structures of the periplasmic lysine/arginine/ornithine-binding protein with and without a ligand. *J. Biol. Chem.* 268:11348–11355.
39. Hornak, V., A. Okur, …, C. Simmerling. 2006. HIV-1 protease flaps spontaneously open and reclose in molecular dynamics simulations. *Proc. Natl. Acad. Sci. USA.* 103:915–920.

40. Özer, N., A. Özen, …, T. Haliloğlu. 2015. Drug-resistant HIV-1 protease regains functional dynamics through cleavage site coevolution. *Evol. Appl.* 8:185–198.

41. Katebi, A. R., and R. L. Jernigan. 2014. The critical role of the loops of triosephosphate isomerase for its oligomerization, dynamics, and functionality. *Protein Sci.* 23:213–228.

42. Na, H., and G. Song. 2018. All-atom normal mode dynamics of HIV-1 capsid. *PLoS Comput. Biol.* 14:e1006456.

43. Song, G. 2019. Structure-based insights into the mechanism of nucleotide import by HIV-1 capsid. *J. Struct. Biol.* 207:123–135.

44. Zimmermann, M. T., K. Jia, and R. L. Jernigan. 2016. Ribosome mechanics informs about mechanism. *J. Mol. Biol.* 428:802–810.

45. Liu, J., K. Sankar, …, R. L. Jernigan. 2017. Directional force originating from ATP hydrolysis drives the GroEL conformational change. *Biophys. J.* 112:1561–1570.