JOURNAL OF CHEMICAL INFORMATION AND MODELING

ARTICLE

pubs.acs.org/jcim

A Normal Mode-Based Geometric Simulation Approach for Exploring Biologically Relevant Conformational Transitions in Proteins

Aqeel Ahmed,[†,∥] Friedrich Rippmann,[‡] Gerhard Barnickel,[‡] and Holger Gohlke*[,†,§]

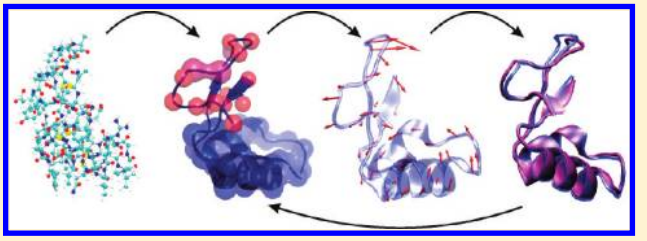
[†]Department of Biological Sciences, Molecular Bioinformatics Group, Goethe University, Frankfurt, Germany

[‡]Bio- and Chemoinformatics, Merck Serono, Merck KGaA, Darmstadt, Germany

[§]Department of Mathematics and Natural Sciences, Heinrich Heine University, Düsseldorf, Germany

Supporting Information

ABSTRACT: A three-step approach for multiscale modeling of protein conformational changes is presented that incorporates information about preferred directions of protein motions into a geometric simulation algorithm. The first two steps are based on a rigid cluster normal-mode analysis (RCNMA). Low-frequency normal modes are used in the third step (NMSim) to extend the recently introduced idea of constrained geometric simulations of diffusive motions in proteins by biasing backbone motions of the protein, whereas side-chain motions are biased toward favorable rotamer states. The generated structures are iteratively corrected regarding steric clashes and stereochemical constraint violations. The approach allows performing three simulation types: unbiased exploration of conformational space; pathway generation by a targeted simulation; and radius of gyration-guided simulation. When applied to a data set of proteins with experimentally observed conformational changes, conformational variabilities are reproduced very well for 4 out of 5 proteins that show domain motions, with correlation coefficients $r > 0.70$ and as high as $r = 0.92$ in the case of adenylate kinase. In 7 out of 8 cases, NMSim simulations starting from unbound structures are able to sample conformations that are similar (root-mean-square deviation = 1.0−3.1 Å) to ligand bound conformations. An NMSim generated pathway of conformational change of adenylate kinase correctly describes the sequence of domain closing. The NMSim approach is a computationally efficient alternative to molecular dynamics simulations for conformational sampling of proteins. The generated conformations and pathways of conformational transitions can serve as input to docking approaches or as starting points for more sophisticated sampling techniques.



INTRODUCTION

Biological macromolecules have an intrinsic ability to switch between conformationally distinct states under native conditions. Conformational transitions occur on a wide range of scales, both in time and space. As a prominent example, adenylate kinase undergoes large conformational changes of its domains during a catalytic cycle.[1,2] These movements are coupled to small amplitude fluctuations on the picosecond time scale of backbone atoms.[3] The ability to undergo conformational transitions becomes particularly pronounced in the case of ligand binding to several pharmacologically important proteins, e.g., HIV-1 protease,[4] aldose reductase,[5] FK506 binding protein,[6] renin,[7] and DHFR.[6] The mutual conformational adaptation of binding partners is referred to as plasticity. Receptor plasticity is also a hallmark of DNA[8] or RNA[9,10] targets. The examples of receptor plasticity demonstrate that the "rigid receptor hypothesis",[11] which has served as an underlying principle in structure-based ligand design (SBLD), is no longer tenable. Instead, the ability to understand and predict receptor plasticity becomes central for a more in-depth understanding of molecular recognition processes[12] and success in SBLD.[13−16]

Motion (or mobility) of a biomacromolecule is a prerequisite for plasticity. Knowledge about biomacromolecular mobility can be obtained from different sources. As for experimental approaches, X-ray crystallography allows to deduce information about protein mobility from B-factor values or by analyzing structures crystallized in different conformational states. In general, however, only a restricted description of the available conformational space is obtained.[17] NMR spectroscopy provides information about protein dynamics in a more direct manner, e.g., in terms of order parameters, relaxation rates, and conformational variabilities within the structure ensemble. However, despite many advances, the technique is still restricted to proteins of a limited size.[18] As for computational approaches, molecular dynamics (MD)[19−21] simulation is one of the most widely applied and accurate computational techniques currently being used in the field of biomacromolecular computation. Several efforts have been made to overcome the problem of restricted sampling in MD due to slow barrier crossing on the rugged energy landscape of biomacromolecules.[22,23] These accelerated conformational search techniques include conformational flooding,[24] replica-exchange MD (REMD),[25,26] self-guided MD,[27] and targeted MD (TMD).[28,29] Despite these improvements and increased computational power, MD simulations are still computationally expensive.[30,31]

Received: November 25, 2010
Published: June 04, 2011

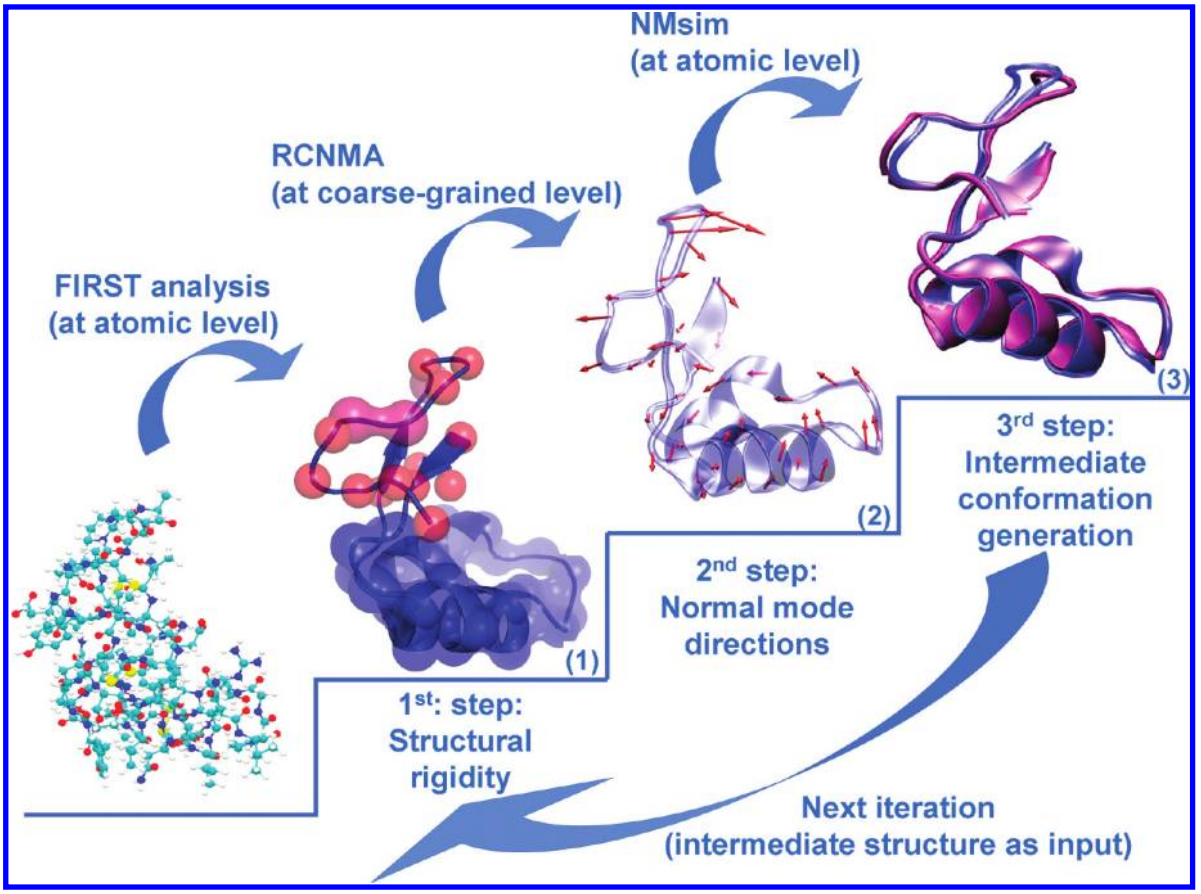


Figure 1. Overview of the RCNMA/NMSim approach. In the first step, a rigid cluster decomposition (RCD) is obtained by FIRST analysis. In the second step, the RCD is utilized by RCNMA for the calculation of normal modes. In the third step, the mode directions are used by NMSim to generate stereochemically allowed protein conformations. In order to generate a NMSim trajectory, step two and three are repeated using the previously generated structure as input.

Hence, there is a need to develop computational approaches that are computationally more efficiently in exploring the conformational space. The distance geometry-based approach CONCOORD generates multiple protein conformations by satisfying distance constraints derived from a starting structure.[32,33] FRODA, a constrained geometric simulation-based approach, generates conformations by simulating diffusive motions of flexible regions and rigid clusters of proteins.[34] Both these approaches have been shown to be promising in SBLD.[35,36] However, neither one of the approaches uses any directional guidance for sampling biologically relevant conformations despite the fact that conformational changes upon ligand binding occur preferentially along lowest frequency (energy) normal modes of the unbound protein. These normal modes involve large-amplitude and correlated motions.[37−40]

Normal modes can be efficiently and accurately predicted by coarse-grained normal mode (CGNM) approaches, such as the elastic network model (ENM)[38,41−43] and the rigid cluster normal-mode analysis (RCNMA).[44] For example, the main directions of conformational changes in tyrosine phosphatase and adenylate kinase upon ligand binding overlap with low-frequency modes calculated by RCNMA for the corresponding unbound conformations.[44] A large-scale study on a data set of 335 proteins has recently shown that the low-frequency modes from ENM/RCNMA agree well with essential dynamics (ED) modes from MD simulations, both in terms of directions and relative amplitudes of motions.[45] The calculation of modes by CGNM approaches only takes seconds for these proteins and, therefore, can be applied also to larger biomacromolecules as well as in an iterative manner. Consequently, several approaches have already made use of such directional information, e.g., for steering MD simulations,[46−48] incorporating receptor flexibility in docking approaches,[49−51] flexible fitting of molecular structures,[52−55] and efficient generation of pathways of conformational changes.[56−58]

In this study, a novel three-step approach, termed NMSim, for the multiscale modeling of protein conformational changes is presented (Figure 1). The first two steps are based on previous work from our lab.[44] Initially, static properties are determined from an all-atom representation of the protein by decomposing the molecule into rigid clusters and flexible regions using the graph theoretical approach FIRST.[59] In a second step, dynamical properties of the molecule are revealed by the rotations−translations of blocks approach (RTB)[60] using an ENM representation of the coarse-grained protein, as implemented in the RCNMA approach.[44] In this step, only rigid body motions are allowed for rigid clusters, while links between them are treated as fully flexible. In the final step, termed NMSim, the recently introduced idea of constrained geometric simulations of diffusive motions in proteins[34] is extended. New protein conformations are generated in that backbone motions are biased toward directions that lie in the subspace spanned by low-frequency normal modes, and side chain motions are biased toward attractive basins derived from

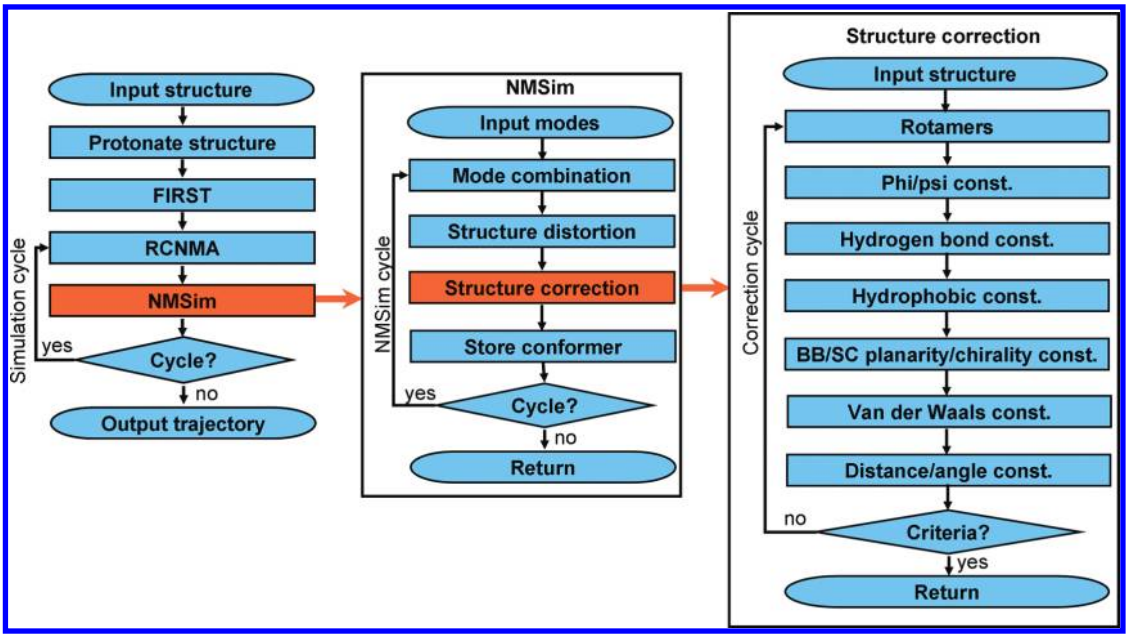


Figure 2. Program flow and modules of the NMSim approach. Modules colored in orange are further expanded on their right (BB: backbone; SC: side-chain).

experimental rotamer information. The generated structures are then iteratively corrected regarding steric clashes and constraint violations. Particular attention has been paid to the stereochemical accuracy of the generated conformations. This requires that backbone torsion angles are constrained to favorable regions, in addition to covalent and noncovalent bonds, like hydrogen bonds and hydrophobic interactions. In total, when applied repetitively over all three steps, the procedure efficiently generates a series of conformations that lie preferentially in the subspace spanned by low-frequency normal modes. Results of the NMSim approach are initially compared with those from state-of-the-art MD simulations on hen egg white lysozyme (HEWL). The NMSim approach is then applied to a data set of eight proteins where conformational changes in terms of domain and loops motions have been observed experimentally. In 7 out of 8 cases, NMSim simulations starting from unbound structures are able to sample conformations that are similar [root-mean-square deviation (RMSD) <3.1 Å] to ligand bound conformations. Biasing the search toward structures with lower radius of gyration considerably improves the sampling of ligand-bound conformations in NMSim. The results show that incorporating directional information about collective motions into a constrained geometric simulation-based approach allows for a thorough sampling of the biologically relevant conformational space and provides a computationally efficient alternative to MD simulations for conformation generation.

NMSim Module. A flowchart of the NMSim approach is shown in Figure 2. The procedure starts with the structural rigidity analysis of the input protein structure, represented as a bond-bending network containing covalent and noncovalent bonds. This yields a decomposition of the protein into rigid clusters and flexible regions in between. Subsequently, normal modes are calculated for the input structure by the RCNMA module. Finally, the NMSim module, first, distorts the structure along directions of low-frequency normal modes and, second, generates a stereochemically allowed conformation from the distorted structure.

The RCNMA and NMSim modules are repetitively called during the simulation. In each RCNMA call, a new set of normal modes is calculated using the structure previously generated by NMSim as input. In each NMSim call, multiple new conformations are generated from linear combinations of these normal modes. The RCNMA approach has been introduced in ref 44, and a detailed description is given in the Appendix. In the following, the NMSim module is described in more detail.

Directions for Structure Distortion. RCNMA yields information about amplitudes and directions of motions of C_α atoms in terms of normal modes $\vec{C}$ and eigenvalues Λ. However, in order to efficiently sample the conformational space of the protein, all non-C_α atoms of the structure must also be displaced in the structure distortion step. For this, the direction of the displacement of atom i of residue j $\vec{P}^k_{i(j)}$ (eq 1) is obtained by adding a random direction component $\vec{E}^k_{i(j)}$ (eq 2) to the k^{th} normal mode direction $\vec{C}^k_{\alpha(j)}$ of the respective C_α atom:

$$\vec{P}^k_{i(j)} = F_{i(j)} * \vec{E}^k_{i(j)} + (1 - F_{i(j)}) * \vec{C}^k_{\alpha(j)} \quad (1)$$

We note that, while the $\vec{C}^k_{\alpha(j)}$ vectors are mutually orthogonal, the $\vec{P}^k_{i(j)}$ vectors are not due to the random direction component added. $\vec{E}^k_{i(j)}$ is obtained from

$$\vec{E}^k_{i(j)} = \vec{R}^k_{i(j)} * [(rand * RANDSCALING) + |\vec{C}^k_{\alpha(j)}|] \quad (2)$$

where $\vec{R}^k_{i(j)}$ is a random unit vector, which is scaled by the magnitude of $\vec{C}^k_{\alpha(j)}$ plus a random component formed by the product of a uniformly distributed random number $rand \in [0,1]$ and the user-defined parameter RANDSCALING. By default, RANDSCALING = 0.3 Å. We note that, while the magnitude of the motion due to $\vec{C}^k_{\alpha(j)}$ depends on the protein size (on average, each C_α atom moves by $1/\sqrt{n}$ Å, with n being the number of C_α atoms), the magnitude of the random component is independent of the protein size. Thus, for proteins larger than the ones investigated here, it may be advantageous to decrease the value of RANDSCALING in order to reduce the influence of the random component on the displacement of atom i.

$F_{i(j)}$ (eq 3) is a weighting factor that is determined by the distance of atom i from the respective C_α atom. This follows the idea that non-C_α atoms that are close to the C_α atom should have a large $\vec{C}^k_{\alpha(j)}$ direction component (for C_α atoms, $\vec{P}^k_{i(j)} = \vec{C}^k_{\alpha(j)}$ in eq 1), whereas atoms in the tail region of a side chain should have a large random component. In the overall NMSim procedure, this ensures that side chain conformations will be explored predominantly in a random manner in stereochemically allowed space with a bias toward attractive basins derived from experimental rotamer information (see Appendix), while backbone conformations will be explored predominantly in the given low-frequency normal-mode space:

$$F_{i(j)} = \frac{D_{i(j)}}{D_{max(j)}} \tag{3}$$

Here, $D_{i(j)}$ is the distance between atom i and the C_α atom of j, and $D_{\max(j)}$ is the maximum distance found in residue j.

Linear Combination of Modes in Unbiased NMSim Runs. For unbiased NMSim runs, the vectors $\vec{P}^k_{i(j)}$ are then linearly combined (eq 4). A coefficient in the linear combination is defined as the ratio of a uniformly distributed random number $O^k \in [-1,1]$ and a factor $\omega^k = (\Lambda^k)^{1/2}$. Here, Λ^k is the eigenvalue of normal mode k.

$$\vec{V}_i = \sum_{k=7}^{m} \frac{O^k}{\omega^k} \vec{P}^k_{i(j)} \tag{4}$$

All $\vec{P}^k_{i(j)}$ vectors belonging to the m low-frequency normal modes (except for the first six zero-frequency modes) are considered in the linear combination. By default, $m = 56$ unless stated otherwise. The above definition of coefficients ensures that a $\vec{P}^k_{i(j)}$ vector can enter the linear combination either with a positive or a negative phase. Furthermore, the influence of a $\vec{P}^k_{i(j)}$ vector is weighted according to the eigenvalue of the normal mode k, giving more weight to the low-frequency, i.e., large amplitude, modes.

Linear Combination of Modes in Targeted NMSim Runs. In targeted NMSim runs, a conformational change vector $\Delta\vec{r} = \vec{r}_c - \vec{r}_o$ is used to guide the trajectory toward a target structure $\vec{r}_c$, starting from a structure $\vec{r}_o$. The vectors $\vec{r}_c$ and $\vec{r}_o$ are the C_α atomic coordinates of the two conformations. A coefficient Z^k (eq 5) is now obtained as the projection of the conformational change vector $\Delta\vec{r}$ onto the normal mode vector $\vec{C}^k$.

$$Z^k = \Delta\vec{r} \cdot \vec{C}^k \tag{5}$$

Z^k is either used to select the normal mode that overlaps best with the direction of conformational change or to bias the linear combination of displacement vectors of atom i of residue j according to eq 6:

$$\vec{V}_i = \sum_{k=7}^{m} Z^k \vec{P}^k_{i(j)} \tag{6}$$

Generating a Distorted Structure. Finally, $\vec{V} = [\vec{V}_1, \vec{V}_2, \vec{V}_3, ..., \vec{V}_N]$ is used to displace atoms of an input structure in an NMSim cycle. For this, the magnitude of $\vec{V}$ is adjusted such as to account for a predefined step size along the NMSim trajectory. Here, the RMSD between the current structure and the next distorted structure is used to define the step size, specified by the parameter RMSDSTEPSIZE. This results in

$$\vec{Q} = RMSDSTEPSIZE * N^{1/2} * \frac{\vec{V}}{|\vec{V}|} \tag{7}$$

where $\vec{Q}$ is a displacement vector that causes the distortion of the structure, and N is the number of atoms in the structure. Equation 7 ensures that the overall distortion of the structure is constant in each NMSim step. By default the RMSDSTEPSIZE = 0.5 Å.

Structure Correction. Next, a distorted structure is efficiently corrected using a geometry-based constraints correction approach. For this, the network of constraints from the rigidity analysis is used where covalent and noncovalent bonds are considered as constraints. In addition, constraints for φ/ψ backbone torsion angles are introduced, which are derived from favorable regions on a Ramachandran map. For χ angles of side chains, a knowledge-based approach is applied by forcing side chains to move into the closest favorable rotamer state during the structure correction. Finally, chirality and planarity constraints are considered for backbone and side chains, and steric clashes between atoms are removed. A detailed description of the structure correction module is given in the Appendix.

Radius of Gyration-Guided NMSim Runs. The search for a ligand-bound conformation of a protein can be drastically improved if structural characteristics of the complex are incorporated in order to guide the search. In the case of large-scale conformational changes, like domain closures in proteins upon ligand binding, it is well-known that the compactness of the protein structure increases upon binding.[61,62] The radius of gyration R_g (eq 8) is an appropriate measure to describe the compactness of a protein:

$$R_g^2 = \frac{1}{n} \sum_{i=1}^{n} (\vec{r}_i - \vec{R}_c)^2 \tag{8}$$

where $\vec{R}_c$ is the center of geometry of n C_α atoms and $\vec{r}_i$ is the atomic position of atom i. Here, only C_α atoms are considered.

In a radius of gyration-guided NMSim simulation, the trajectory is tailored toward the bound structure by selecting the pathway that leads to a decrease in R_g. We note that R_g is not used for enforcing the generation of a more compact protein structure by influencing the structure distortion step: Conformations are still generated by structure distortion along directions of random linear combinations of low-frequency normal modes and subsequent structure correction. Rather, the guiding occurs because out of several (by default, three) such conformations generated by repeated NMSim cycles, the one with the lowest R_g is selected as a starting point for further trajectory exploration in the next simulation cycle. That way, the pathway of structural change still lies in the subspace spanned by low-frequency normal modes. We also note that no experimentally determined target value of the radius of gyration of the bound structure is needed here.

MATERIALS AND METHODS

Comparison with MD Simulation. For comparison with NMSim generated conformations, MD generated conformations of HEWL were taken from a recent study by Koller et al.[63] Here, a MD simulation of 100 ns length of HEWL (PDB code 1hel)[64] was performed with Amber9 under periodic boundary conditions in the NVT ensemble at $T = 300$ K. The force field ff99SB was used together with the TIP3P water model.[63] Unbiased NMSim runs were performed on the same starting structure with the default parameter set described in Table S1, Supporting Information (see also the Ensemble Generation with NMSim Section). In total, four different NMSim simulations were performed by varying the number of low-frequency modes

Table 1. Protein Data Set Used in the Present Study

protein	open[a]	closed[a]	no. of residues[b]	structural regions[c]
		Domain Motions		
adenylate kinase (ADK)	4ake (A)	1ake	214	core: 1−28, 80−112, 173−214; ATP: 119−156; NMP: 31−72
aspartate aminotransferase (AST)	9aat (A)	1ama	388	large: 42−322; small:15−33, 330−356, 362−410
calmodulin (CLM)	1cfd (A)	1ckk	148	C-term. domain: 82−146; N-term. domain: 5−75
citrate synthase (CTS)	5csc (A,B)	6csc	860	large: 3−55, 66−272, 330−335, 382−433; small: 56−63, 284−327, 338−378
LAO binding protein (LAO)	2lao (A)	1lst	238	lobe I: 1−88 and 195−238; lobe II: 93−185
		Loop Motions		
tyrosine phosphatase (TYP)	1ypt (A)	1yts	278	$\beta7$-$\alpha4$ loop: 350−360
triosephosphate isomerase (TIM)	8tim (B)	1tph	245	loop 6: 166−176
CAMP-dependent protein kinase (CAPK)	1jlu (E)	1fmo	336	glycine-rich loop: 50−55

[a] PDB codes for unbound (open) and ligand-bound (closed) structures. The PDB chain that was used as a starting structure is given in parentheses. [b] 13 residues (Ser3-Asp15) in AST were removed as they were found to be highly fluctuating. Two and three residues were removed, respectively, from the open TIM and CAPK structures in order to generate open and closed structures with an identical set of residues. [c] Residue numbers for domain[74,93,94] and important loop regions.

(i.e., m = 11, 16, 31, and 56 in eq 4). In each of these NMSim runs, 10 000 conformations were generated using a simulation cycle of 1000 and an NMSim cycle of 10 steps. Every 10th structure was then taken for the analysis, resulting in an ensemble of 1000 conformations. In order to see the effect of the length of the simulation, 3 ensemble sets were created using the first 100, 500, and 1000 conformations, respectively, henceforth referred to as "one-tenth", "one-half" and "complete" sets.

In order to compare the MD generated ensemble to the NMSim generated ones, the mass-weighted averages of root-mean-square fluctuations (RMSF) of heavy atoms were calculated for each residue. Furthermore, essential dynamics (ED)[65,66] analysis was performed in order to determine the extent of conformational sampling during the NMSim simulations. To do so, ED calculations were performed on the MD ensemble, and the NMSim ensemble sets were then projected onto the plane spanned by the first two ED modes with the highest eigenvalues. Ptraj of the Amber8 package[67] was used for the fluctuation and the ED calculations.

The stereochemical quality of NMSim generated structures (using m = 56) was analyzed by the Procheck[68] program and compared to the MD generated structures. For this, 100 equally spaced structures were taken from the MD and NMSim trajectories. Additionally, these results were compared to the stereochemical quality of 130 structures of HEWL obtained from the Protein Data Bank; these structures were selected by requiring a 100% sequence identity with the structure of PDB code 1hel. This set will be referred to as "EXP" hereafter. Furthermore, 100 high-resolution (0.8−1.7 Å) and nonhomologous crystal structures available from the Richardsons' lab[69] were also subjected to stereochemical analysis. This set will be referred to as "EXPTOP" hereafter.

Finally, the effective conformational energy of the MD and NMSim generated structures as well as the EXP structures was calculated. The effective energy is the sum of the gas-phase energy, determined with the force field ff99SB, and a solvation free energy, determined with the generalized Born model by Onufriev, Bashford, and Case[70−72] (i.e., igb = 5 in Amber9) plus a nonpolar contribution proportional to the solvent-accessible surface, using γ = 0.0072 kcal mol^{-1} Å^{-2} as a surface tension. As structures obtained from MD, NMSim, or experiment do not necessarily reside in local minima of the effective energy surface, all structures were minimized by 100 steps of conjugate gradient minimization with respect to the effective energy, thereby tethering all atoms with harmonic restraints to the starting coordinates using a force constant of 2 kcal mol^{-1} Å^{-2}. That way any gross conformational change of a structure is prevented; on average, the all-atom RMSD between the starting and relaxed structure is <0.2 Å.

Data Set of Proteins. The NMSim approach was then applied to a data set of eight proteins for which important conformational changes have been observed upon ligand binding and for which crystal structures of an unbound (open) and ligand-bound (closed) conformation were available. In order to analyze the usefulness and the limitations of the NMSim approach, the data set is subdivided into two categories, domain and loop motions, based on the types of conformational changes observed upon ligand binding. The data set is listed in Table 1. The proteins in the data set have been used previously in other normal-mode studies[38,44,49,73] and show both "induced fit" and "conformational selection" types of conformational changes.[74,75]

Ensemble Generation with NMSim. In order to explore the extent up to which experimentally observed conformational

changes can be simulated by NMSim, the method was applied to the open conformation of proteins in the data set (Table 1). Three types of simulations were performed: unbiased, radius of gyration guided, and targeted.

Unbiased NMSim runs are performed without any information about the target conformation, and hence, random linear combinations of low-frequency normal modes are used to generate a trajectory (eq 4). In general, unbiased NMSim is run with the default parameter set described in Table S1, Supporting Information. By default, 5000 conformations are generated per run using 500 simulation cycles and 10 NMSim cycles. Every 10th conformation is selected for analysis. For the domain motion data set, only the first five normal modes and a smaller side-chain randomization parameter (i.e., RANDSCALING = 0.05 Å) are used in order to focus more on the exploration of large-scale backbone conformations. Furthermore, 10 independent unbiased trajectories are generated for each protein in the domain motion data set, generating 500 conformations per run using 500 simulation cycles and one NMSim cycle. These settings proved particularly promising for a large-scale exploration of the conformational space of a protein. For generating 5000 conformations of adenylate kinase with 214 residues, NMSim takes about 2 days of computational time on a single core of a current desktop computer.

Radius of gyration-guided NMSim runs are performed with the assumption that the closed structure has a smaller R_g than the open structure. Consequently, protein movements that lead to lower R_g values are favored throughout the run by generating three conformations at each step and then selecting the one with the lowest R_g for further trajectory exploration. Thus, out of 1500 conformations generated per run by default, using 500 simulation cycles and one NMSim cycle, 500 conformations represent the pathway of motion at the end.

Targeted NMSim runs are performed incorporating information about the closed conformation in that the best combination of normal modes as defined by eq 6 is used at each structure distortion step. By default, 500 conformations are generated using 500 simulation cycles and one NMSim cycle.

Analysis of Simulation Results. Intrinsic fluctuations of a protein near its equilibrium state in the open conformation correlate with the conformational change of the protein upon complex formation.[3,76] In order to verify if this argument holds for NMSim generated structures, C_α root-mean-square fluctuations (RMSF) derived from the 5000 structures generated by unbiased NMSim simulations are compared with conformational variabilities between the respective open and closed conformations. These RMSF were calculated using ptraj of the Amber 8 package.[67] In order to observe to what extent a closed structure is approached during NMSim simulations started from the open conformation, the backbone RMSD was calculated between the generated structures and the closed conformation. For proteins with loop motions, the backbone RMSD of only the moving loop region (Table 1) is calculated after aligning the rest of the protein backbone. Furthermore, essential dynamics (ED)[65,66] calculations were performed for adenylate kinase (ADK) in order to compare essential motions observed in unbiased NMSim simulations with those derived from an ensemble of experimental structures. For this, a principal component analysis was performed using ptraj of the Amber8 package[67] on C_α atoms of an ensemble of eleven crystal structures of ADK[77−80] [PDB code (chain): 1ake (A), 1ank (A), 4ake (A), 4ake (B), 1dvr (A), 1dvr (B), 1e4v (A), 1e4y (A), 1e4y (B), 2eck (A), 2eck (B)]. Subsequently, all experimental and NMSim generated structures were projected onto the plane described by the first two ED modes. Finally, calculations of effective conformational energies were performed for subsets of ADK conformations generated by unbiased NMSim simulations, as described above.

Pathway Generation with NMSim. Pathways of conformational changes from the open to the closed structures were generated for ADK using two types of simulation: targeted NMSim and radius of gyration-guided NMSim simulations. In general, default parameters were used for both types of simulations (Table S1, Supporting Information). However, each intermediate conformation was generated using the single "best" mode instead of a linear combination of modes. For targeted NMSim simulations, out of the first 50 modes, the mode that best overlaps with the direction of conformational change (eq 5) was used for this in each NMSim cycle. For radius of gyration-guided NMSim simulations, 10 structures were generated by following either direction of the first five modes, and the structure with the lowest radius of gyration was selected for further exploration of the pathway in each NMSim cycle.

In order to analyze the order of the domain closure in ADK, the reaction coordinates described by Whitford et al.[81] were used. The reaction coordinate $R^{\mathrm{LID-CORE}}$ ($R^{\mathrm{NMP-CORE}}$) is defined as the distance between the centers of mass of the LID (NMP) and CORE domains, respectively. In order to further verify the NMSim pathway, intermediate structures generated by NMSim simulations were compared with 11 crystal structures of ADK in terms of C_α RMSD.[82] For this, a crystal structure was assigned to each intermediate structure based on the lowest C_α RMSD identified between the set of 11 crystal structures and the respective intermediate structures, similar to a study by Maragakis and Karplus.[82]

RESULTS AND DISCUSSION

Comparison with MD Simulation: HEWL As a Test Case. Initially, results of NMSim simulations were compared with those from state-of-the-art MD simulations using HEWL as a test case. HEWL is a well-studied[40,83−86] protein comprised of 129 residues, which has repeatedly been used for the evaluation of force fields.[63,87−90] The MD simulation of HEWL[63] took approximately 4 months on 4 CPUs on a state-of-the-art workstation, whereas the NMSim simulation took 30 h on a single CPU on the same workstation.

RMSFs obtained from different NMSim runs agree well with fluctuations derived from the MD ensemble with correlation coefficients between 0.72 to 0.79 for the complete NMSim simulations (Table S3 and Figure S9, Supporting Information). These results vary little if the number of low-frequency modes is varied between 5 and 50. As for the influence of the length of the simulation, using only the first half of the NMSim trajectory yields essentially the same good agreement between NMSim and MD derived residue fluctuations as obtained for the complete trajectory. If only one-tenth of the NMSim trajectory is used, the agreement between the fluctuations deteriorates, however.

In order to determine the extent of conformational sampling during the NMSim simulations, essential dynamics (ED) calculations[65,66] were performed. In general, NMSim conformations are found to be well distributed along the first two principal directions of motions observed during the MD simulation in the case of the complete simulation as well as the one-half simulation (Figure S10, Supporting Information). In contrast, if only one-tenth of the NMSim trajectory is considered the conformational

sampling becomes much more restricted. The results also show that, regardless of the simulation length, the principle directions of motions derived by MD are better captured by NMSim when 5 or 10 low-frequency modes are used instead of 25 or 50 modes. This finding can be understood in that fewer modes lead to a more pronounced sampling in the principal directions of motions. In contrast, using more modes allows a more detailed sampling but leads to a more restricted sampling in the principal directions. Thus, for proteins with pronounced conformational changes, reducing m is expected to result in a better sampling of the dominating motions. Following this observation, we have used the first 5 low-frequency modes for proteins that involve domain motions and the first 50 modes for proteins with all other types of motions.

The quality of NMSim generated conformations were analyzed using Procheck[68] and was compared with MD generated conformations and crystallographic structures (Table S4, Supporting Information). The Procheck results show that the NMSim generated conformations are of a higher stereochemical quality than the MD derived conformations. Specifically modeling φ/ψ constraints in NMSim resulted in the highest population of the core region (92%) of the Ramachandran map, which is similar to the value obtained for the high-resolution crystallographic structures of the EXPTOP set (91%). In turn, the population of the generously allowed or disallowed regions of the Ramachandran map is zero in the case of NMSim generated conformations. The Procheck G-factor provides a measure of how normal a given stereochemical property is. A low G-factor indicates that the property corresponds to a low-probability conformation; ideally, the G-factor value should be above -0.5. On average, the overall G-factor value for NMSim generated conformations is around -0.3, whereas this value is considerably lower (-0.86) for MD structures. Similarly, aromatic, carbonyl-containing, and guanidino moieties of side chains are planar in NMSim generated conformations; this proportion is higher than the one found in crystallographic structures. Only 56% of these moieties are planar in MD generated conformations.

In terms of the effective conformational energies, on average, the NMSim generated structures are almost similar (-5078 ± 5.0 kcal mol^{-1}; average $\pm$ standard error in the mean; Table S5, Supporting Information) to the EXP structures (-5075 ± 24.4 kcal mol^{-1}), whereas the MD structures show an effective conformational energy that is lower (-5383 ± 2.4 kcal mol^{-1}). We note, however, that the MD structures were generated with the same force field that was also used for the effective energy calculations, whereas the NMSim and EXP structures were only allowed to relax by means of a restrained minimization in order to prevent gross conformational changes. As a result, particularly dihedral, van der Waals, and electrostatic energy components are larger in the case of NMSim generated and EXP structures than in the case of MD generated structures (Table S5, Supporting Information). In turn, the generalized Born and nonpolar components are lower for NMSim generated and EXP structures. In summary, these results demonstrate that NMSim simulations generate stereochemically and energetically favorable structures.

Data Set of Proteins with Pronounced Conformational Changes. Next, the NMSim approach was applied to proteins for which important conformational changes have been observed upon ligand binding. The domain motion data set (Table 1) contains five proteins that are diverse in terms of their structures, sizes, and motions. Adenylate kinase (ADK) is a monomeric enzyme that contains a main domain (CORE), an ATP-binding domain

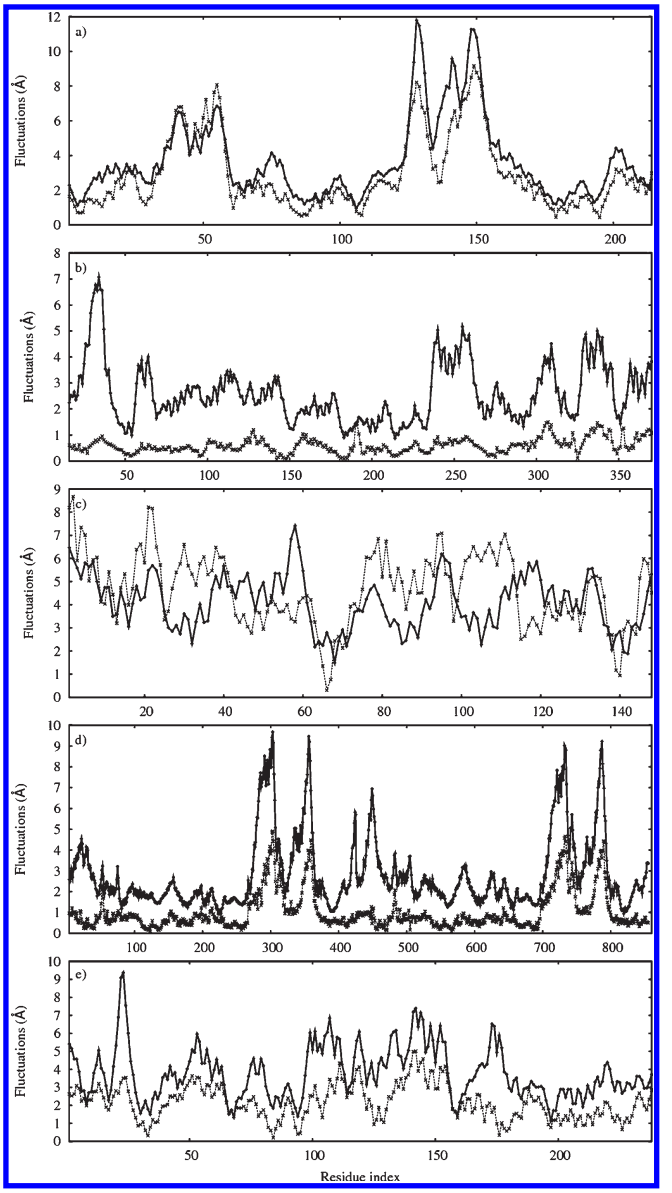


Figure 3. C_α atom fluctuations of proteins with domain movements obtained from ensembles generated by unbiased NMSim simulations (solid lines; +) and structural deviations between the open and closed structures (dotted lines; ×). (a) Adenylate kinase; (b) aspartate aminotransferase; (c) calmodulin; (d) citrate synthase; and (e) LAO binding protein.

(LID), and a NMP-binding domain (NMPbind).[77] Aspartate aminotransferase (AST) is a homodimeric enzyme,[91] as is citrate synthase (CTS), the largest protein in the data set with 860 residues.[92] Calmodulin (CLM), the smallest protein in the data set with 148 residues, consists of two globular domains separated by a flexible linker.[93] The structure of the lysine/arginine/ornithine-binding protein (LAO) is bilobate, and the two lobes are held together by two connecting segments.[94] ADK and LAO show global and hinge-bending motions of domains,[95] AST and CTS show localized motions of small domains and sheer motions,[95] and CLM shows a large-scale bend and twist motion of two domains.[96]

In addition to domain motions, three functionally important loop motions are investigated in this study (Table 1). For

Table 2. Results of NMSim Simulations

protein	RMSD[a] open[b]	RMSD[a] unbiased[c]	RMSD[a] targeted[c]	RMSD[a] radius of gyration guided[c]	correlation NMSim[d]	correlation ENM[e]
			Domain			
adenylate kinase	7.15	3.06 (1.0, 3.0)	0.93	2.36 (62.8, 71.8)	0.92	0.90
aspartate aminotransferase	1.55	0.98 (2.1, 8.7)	0.60	1.21 (46.4, 61.8)	0.71	0.62
calmodulin	9.80	6.71 (1.0, 3.1)	2.95	5.32 (40.2, 73.6)	0.32	0.22
citrate synthase	2.70	1.55 (0.6, 7.0)	0.91	1.37 (48.4, 87.6)	0.86	0.80
LAO binding protein	4.67	2.31 (1.0, 3.8)	0.59	1.75 (51.2, 82.4)	0.70	0.53
			Loop			
tyrosine phosphatase	3.18	1.86 (21.6, 44.6)	0.95	1.58 (48.8, 79.0)	0.43 (0.64)	0.42
triosephosphate isomerase	4.50	2.01 (7.0, 16.8)	0.90	2.24 (23.6, 66.4)	0.39 (0.06)	0.37
CAMP-dependent protein kinase	1.68	1.14 (7.4, 20.2)	0.79	0.67 (20.0, 31.2)	0.28 (0.34)	0.41

[a] Backbone RMSD with respect to the closed structure, in Å. For proteins with loop motions, only the backbone RMSD of the moving loop region is calculated after aligning the rest of the protein. [b] RMSD between open and closed structures. [c] RMSD of the conformation most similar to the closed structure. In parentheses, the percentages of generated structures that are similar to the closed structure are given, considering RMSD thresholds <0.5 and <1.0 Å with respect to the conformation being itself most similar to the closed structure. The conformations were generated by the denoted NMSim simulation variants. [d] Correlation coefficient between C_α atom fluctuations obtained for an ensemble generated by unbiased NMSim simulations and structural deviations determined from the open and closed structures. In parentheses, the correlation coefficient between C_α atom fluctuations obtained for the NMSim generated ensemble and fluctuations derived from B-factor values of the open structure are given. [e] Correlation coefficient between C_α atom fluctuations obtained for an ENM representation of the open structure using the ElNémo web server (with default settings)[106] and conformational variabilities determined from the open and closed structures.

tyrosine phosphatases (TYP), a ligand-induced conformational change has been observed, which moves Asp356 on the $\beta7-\alpha4$ loop into the active site where it can function as a general acid.[97] In triosephosphate isomerase (TIM), the functionally important loop 6 undergoes a remarkable conformational change upon ligand binding, and the tip of the loop moves about 7 Å to form a lid over the bound ligand.[98,99] NMR results show that the loop also closes as a natural motion of the enzyme in the absence of the substrate.[100] For both TYP and TIM, loop motions have already been shown to be governed by low-frequency modes.[44,73] In contrast, it has been suggested[49] that higher frequency modes are involved in the midscale loop rearrangements of the catalytic subunit of cAMP-dependent protein kinase (CAPK). Thus, CAPK is a valuable test case to investigate to what extent observed movements can still be modeled when using a combination of only the first 50 modes but without considering any a priori information about the closed loop conformation.

Computed RMS Atomic Fluctuations vs Conformational Variabilities from Experiment. C_α RMSFs observed in ensembles generated by unbiased NMSim simulations were compared with conformational variabilities derived from the respective open and closed conformations (Figure 3). For the domain motion data set, the overall trends of experimental conformational variabilities are reproduced very well by the NMSim simulations, with correlation coefficients $r > 0.70$ for all cases except CLM and as high as $r = 0.92$ in the case of ADK. Here, the NMSim simulation captures the hinge-bending movements of the LID and NMPbind domains very well. Likewise, mobile regions are well recognized in the case of CTS, which shows a sheer motion upon ligand binding,[95] resulting in high conformational variabilities of two regions in either monomer. The low correlation coefficient observed for CLM ($r = 0.32$) can be attributed to local rearrangements within the two domains upon Ca^{2+} binding.[93] These rearrangements lead to the exposure of large hydrophobic surfaces in both domains of CLM.[101] There are two limitations of the present NMSim approach that prevent a successful modeling of such a conformational change. First, the local rearrangements are not well described by low-frequency modes,[102] especially as the overall intrinsic motions of CLM are dominated by large-scale movements of the domains. Second, the conformational change requires a substantial change in the constraint network in that hydrophobic constraints that are present in the starting structure need to be broken, which is not permitted in the current NMSim version. Regarding the magnitude of fluctuations, a good agreement is observed in the cases of ADK, CLM, and LAO. This result is remarkable, as no scaling of the computed values is applied here, in contrast to what is usually done if results are reported for ENM approaches.[103] However, in the case of AST and CTS, RMSF derived from NMSim simulations exceed observed movements in the experimental structures. This observation indicates that the constraint network underlying the geometric simulations might be under constrained in those cases.

For the loop motion data set, C_α RMSFs derived from NMSim generated conformations are compared with structural deviations derived from the respective open and closed structures as well as C_α RMSFs computed from B-factor values of the open structures (Table 2). In contrast to the domain motion data set, in general weak correlations are found with correlation coefficients between ~0.3 and ~0.4 when fluctuations from NMSim generated ensembles are compared to structural deviations (Table 2). On the one hand, in the case of TYP and TIM, large fluctuations in the $\beta7$-$\alpha4$ loop and loop 6 regions are observed in NMSim generated structures that match perfectly with the observed conformational changes upon ligand binding (Figure S1, Supporting Information). On the other hand, large fluctuations in some parts of the proteins, e.g. residues 335−343 in TYP and residues 65−78 in TIM, do not correlate with observed conformational changes. Still, the fluctuations predicted by NMSim do correlate with those derived from B-factor values ($r = 0.64$), which demonstrates that highly fluctuating regions in the NMSim ensemble of TYP agree with regions that have an

intrinsic ability to move. Apparently, the two crystal structures of TYP considered here do not capture these movements. In contrast to the TIM and TYP cases, the glycine-rich loop in CAPK does not show large fluctuations in the NMSim ensemble, although this loop has been previously reported to be mobile.[104,105] Instead, the region containing the F-to-G helix loop and the G helix (residues 238−250) is found to be overly mobile in NMSim.

These results demonstrate that the NMSim approach effectively captures the information available in low-frequency normal modes and translates it into mobility predictions, in particular for large-scale movements. This result is notable in that the reported fluctuations are derived from NMSim simulations and do not consider any experimental information about the closed structure. Previous studies have also shown a good agreement between fluctuations obtained from a (collection of) normal mode(s) and experimentally observed conformational changes.[37,38,43,44,76] When compared to RMSF obtained for an ENM representation of the open structures by the ElNémo server,[106] in four out of eight cases NMSim outperforms ENM in reproducing experimental conformational variabilities as judged by differences in the correlation coefficients of at least 0.05 and as high as 0.17 (Table 2); in turn, ENM outperforms NMSim in one case.

Closed Conformations Observed in Unbiased NMSim Simulations Started from an Open Structure. The above findings provided the incentive for us to search for closed (bound) conformations in the ensembles generated by NMSim simulations, which were started from open structures. For the domain motion data set, Figure S2, Supporting Information shows the results of comparing NMSim generated conformations with respective closed conformations in terms of backbone RMSD for all 5 proteins and 10 independent NMSim trajectories in each case. In CTS—dominated by sheer motions—most of the trajectories move initially toward the closed structure before they start to drift away. In contrast, for systems dominated by hinge bending motions, like ADK or LAO, most of the trajectories move preferentially either toward or away from the closed conformation. Still, some trajectories show a reversal of the preferred direction of motion, such that both domain closing and opening movements are explored during one simulation run.

As for RMSD between experimental structures, open and closed conformations deviate by 1.55 (AST) to 9.80 Å (CLM) (Table 2). In view of these numbers, it is reassuring that for all cases structural deviations could be reduced by unbiased NMSim simulations, with a maximal reduction of 4.1 Å RMSD in the case of ADK. When considering minimal RMSD between simulated conformations and respective bound conformations, values as low as 1.0 (AST) to 3.0 Å (ADK) are found for all cases except CLM. In the case of ADK, the conformation generated by NMSim is thus more similar to the closed structure than the one reported previously in a study using tCONCOORD (RMSD = 3.3 Å).[33] As for CLM, no conformation could be generated that is satisfyingly similar to the closed structure despite movements of 3 Å RMSD toward the closed structure. As discussed above, this is due to local rearrangements within the two domains of CLM,[93] which are not well described by low-frequency modes.[107] In addition to reporting the minimal RMSD value found for an NMSim generated conformation with respect to the closed structure, the ensemble population density can be analyzed with respect to the RMSD value. Considering RMSD values up to 0.5 (1.0) Å larger than the minimal RMSD, on average, 1 (5)% of the generated structures is similar to the NMSim generated conformation that itself is most similar to the closed structure (Table 2). This demonstrates that conformations with a minimal RMSD are not isolated incidences in RMSD space but are surrounded by structurally similar conformations.

For the loop motion data set, backbone RMSD of the selected loop regions with respect to the closed structures is determined for each NMSim generated conformation along the trajectory (Table 2 and Figure S3, Supporting Information). In the case of TYP and TIM, the unbiased NMSim simulations show opening and closing movements of the $\beta7-\alpha4$ loop and loop 6, respectively. The structural deviations between open and closed structures can be reduced by up to 2.5 Å RMSD (TIM), and the generated conformations most similar to the experimentally determined closed structures show deviations of 1.2 (CAPK), 1.9 (TYP), and 2.0 (TIM) Å (Table 2). As for the ensemble population density, on average, 12 (27)% of the generated loop conformations are similar to the NMSim generated conformation that itself is most similar to the closed structure, again considering RMSD values up to 0.5 (1.0) Å RMSD larger than the minimal RMSD (Table 2). These results confirm that even loops are engaged in intrinsic motions of a protein toward a closed conformation, as has been shown both experimentally[100] and theoretically[73,108] for the loop 6 closure in TIM. Consequently, these motions can be exploited for the generation of conformations that resemble bound protein states.

The above results on proteins with pronounced conformational changes show that NMSim simulations can be successfully applied for conformational sampling in the case of both hinge and sheer motions, as long as the motion is not dominated by a local rearrangement of secondary structure. However, as the approach relies on the directional guiding of low-frequency normal modes, it might fail for systems where normal mode approaches have general limitations, e.g., in the case of closed-to-open transitions. Finally, NMSim was shown to work also for functionally important loop motions as long as these motions are intrinsic and not ligand induced.

Pathway Generation by Targeted NMSim Simulations. If information about a closed conformation is available, then this information can be used for biasing coefficients of the linear combination of normal modes that is used for the distortion of the protein structure (eq 6). This leads to a targeted NMSim simulation, which allows for the generation of a nonlinear pathway between the open and closed conformations. Such a pathway provides a valuable means for visualizing even complex protein movements, which helps in understanding the relationship between protein motion and function (see also below).[34,109] Moreover, it can serve as input to more sophisticated techniques such a nudged elastic band,[110] transition path sampling,[111] or umbrella sampling.[112,113]

When performing targeted NMSim simulations starting from open structures of the domain motion data set, deviations between the open and closed structures could be reduced by more than 6.0 Å RMSD. For all systems but CLM, a final RMSD < 1.0 Å with respect to the closed conformation is obtained (Table 2). There are two reasons for the residual deviation from the closed structure. First, only the first 50 low-frequency normal modes are applied here in the structure distortion step such that small-scale motions required for the final approach to the closed structure are not accessible in the spanned subspace.[107] Along these lines, a related study combining normal mode calculations with Monte Carlo simulation techniques reached the closed

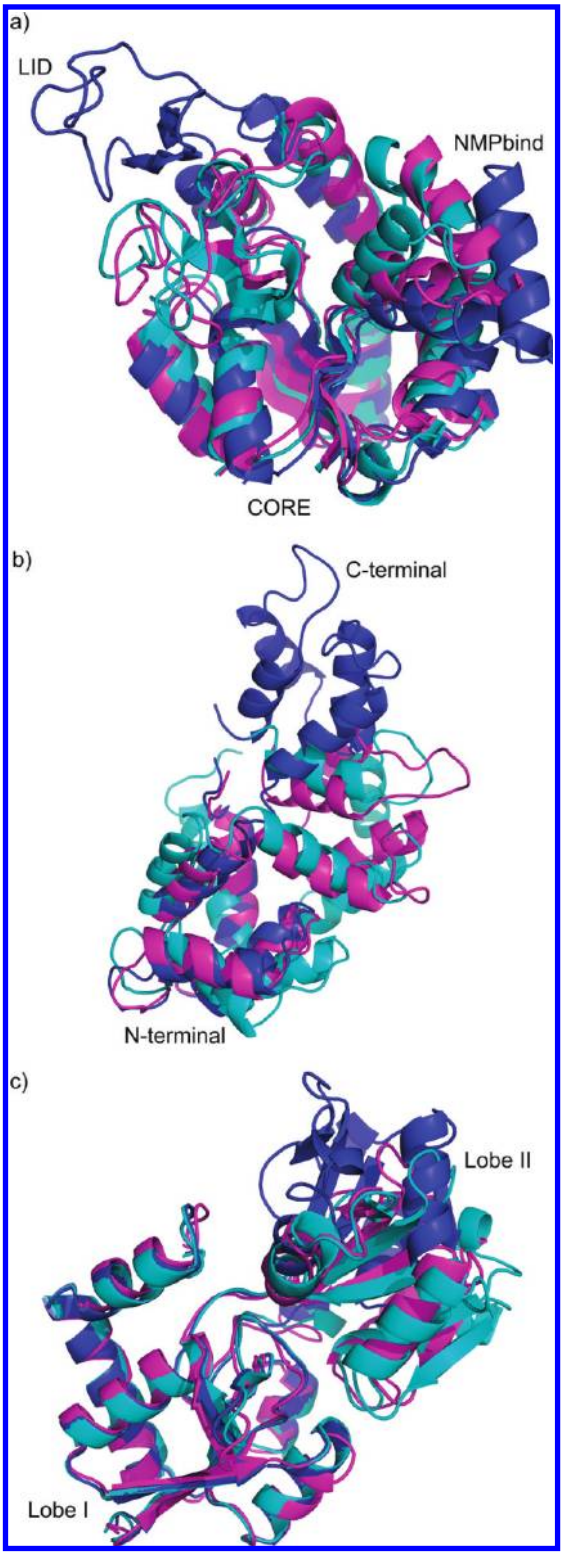


Figure 4. Superimposition of the open (blue), the closed (cyan), and the NMSim generated conformation most similar to the closed structure (magenta) using radius of gyration-guided NMSim simulations. (a) Adenylate kinase; b) calmodulin; and c) LAO binding protein.

structure of ADK only up to 2.3 Å RMSD using 10 low-frequency modes.[114] Second, details of the constraint network may differ between the open and closed structures, so that rigid or overconstrained regions determined in the rigid cluster decomposition (RCD) step with the open structure may not be appropriate anymore for conformations similar to the closed structure. The latter may be overcome by approaches for breaking and reforming constraints during a simulation, which is an area of active research in our group. In the case of the loop motion data set, conformations generated by targeted NMSim simulations approach the closed structure up to ~0.9 Å RMSD for all three proteins (Table 2). These values could even be further decreased if higher frequency modes are used in addition, as shown previously.[49]

Improved Sampling of Closed Conformations by Radius of Gyration-Guided NMSim Simulations. Results from the targeted NMSim simulations provide a lower bound up to which the closed structure can be approached. For all investigated systems, conformations that are more similar to the closed structures have been obtained compared to if unbiased NMSim simulations had been performed. This finding prompted us to develop a simulation variant where a driving force fosters the transition to the closed structure but without requiring any a priori knowledge about the closed structure. This led to radius of gyration-guided NMSim runs (eq 8), which are based on the assumption that ligand binding usually results in a more compact protein conformation due to domain or loop closure.[61,62] The way this NMSim variant was implemented ensures that the structure generation is not influenced by a driving force. Thus, trajectories are still allowed to proceed through the subspace spanned by low-frequency normal modes. Instead, the radius of gyration criterion is only used for selecting more compact conformations along a trajectory. Furthermore, no experimental target value of the radius of gyration is required.

The comparison between the unbiased and the radius of gyration-guided NMSim simulations reveals for the domain motion data set that the latter method generates conformations that are more similar to a closed structure for four out of five cases (Table 2). This improvement can be as large as 1.4 Å in the case of CLM, although in no case a conformation as similar to a closed structure as obtained by targeted NMSim simulations is generated. The improvement is, in general, more obvious for hinge bending motions (ADK, LAO, and CLM) than sheer motions (CTS). This reflects that proteins that show hinge-bending motions are subject to a larger compaction upon ligand binding than are proteins with sheer motions. In terms of efficiency, the improvement is achieved with about 3.3 times lower computational costs: Instead of performing 10 independent unbiased NMSim simulations each generating 5000 conformations, here a single radius of gyration-guided simulation is performed that generates 500 conformations. Computing a single trajectory is sufficient because it was found in initial tests that trajectories generated by radius of gyration-guided NMSim simulations do not differ significantly (data not shown).

Figures 4 and S4, Supporting Information illustrate for proteins of the domain motion data set the extent by which radius of gyration-guided NMSim simulations were successful in reaching a closed structure. The generated conformation that is most similar to the closed structure is shown along with the respective open and closed structures. In the ADK case (Figure 4a), the large-scale conformational change of the LID domain is well described by the radius of gyration-guided simulation despite the fact that no a piori information about the closed structure was used. However, the NMPbind domain moves only halfway toward the closed structure, in agreement with the suggestion

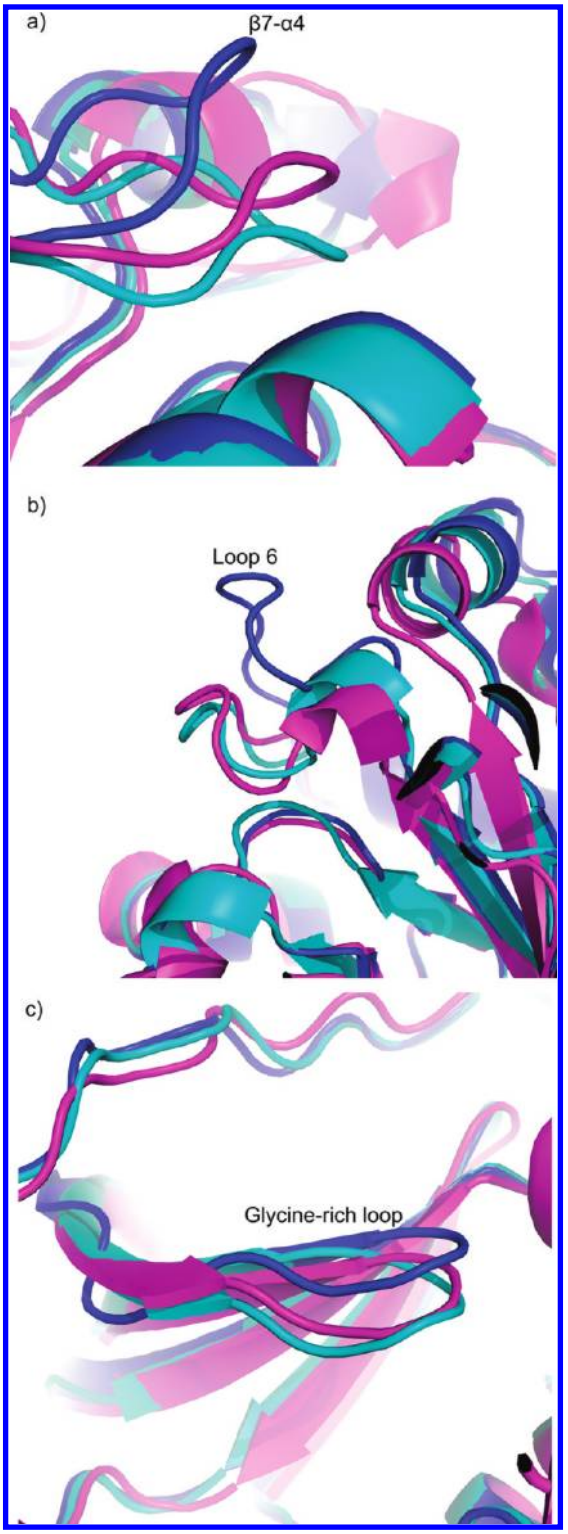


Figure 5. Superimposition of the open (blue), the closed (cyan), and the NMSim generated conformation most similar to the closed structure (magenta) using radius of gyration-guided NMSim simulations. (a) Tyrosine phosphatase; (b) triosephosphate isomerase; and (c) cAMP-dependent protein kinase.

that its closing follows a ligand-induced mechanism.[81] Still, a conformation as similar as 2.4 Å to the closed structure results. In the LAO case (Figure 4c), again a large hinge-bending motion of 3 Å RMSD toward the closed conformation is obvious, and the closed conformation is approached up to 1.7 Å RMSD. Finally, the radius of gyration-guided NMSim simulation captures the large-scale movement of CLM, too, (Figure 4b) but fails to reproduce the local rearrangements within the two domains due to Ca^{2+}-binding,[93] resulting in an RMSD with respect to the closed structure of 5.3 Å.

For the loop motion data set, fostering a lower R_g while moving in the subspace spanned by low-frequency normal modes again guides the trajectories toward the experimentally observed closed structures in all three cases (Figure S3, Supporting Information): For TYP, TIM, and CAPK, conformations as similar as 1.58, 2.23, and 0.66 Å RMSD are generated, respectively (Figure 5, Table 2). Thereby, lower RMSD values than in the case of unbiased NMSim simulations are obtained for TYP and CAPK. Furthermore, in the TYP and TIM cases, the loops fluctuate around the closed conformation, which means that fostering lower R_g values does not deteriorate the loop region in the course of the trajectory. However, this is not always the case because for CAPK the loop starts to move away from a state similar to the closed structure.

These results are encouraging in view of protein−ligand docking. Here, a drop in docking accuracy compared to redocking was often found to be mirrored by the degree to which a protein moves upon ligand binding,[15,115] so that docking to an apo form usually shows the largest deterioration.[13] In this regard, being able to move closer to a bound conformation by up to 4.8 Å RMSD and to come as close as 0.7 Å RMSD to a bound conformation without requiring any a priori knowledge about this conformation is a promising achievement.

Furthermore, when analyzing the population density of ensembles generated by radius of gyration-guided NMSim runs, it becomes obvious that in 6 out of 8 cases more than 40% of the ensemble is very similar to that conformation being itself most similar to the closed structure (again using an RMSD threshold up to 0.5 Å) (Table 2). This illustrates the potential of radius of gyration-guided NMSim runs for generating a focused set of candidate structures for docking approaches when starting from an apo structure. The number of candidate structures can be further reduced by structurally clustering the ensemble and by only considering cluster representatives. Evaluating the generated structures energetically appears to be difficult because the conformational variability of a receptor leads to a disfavorable reorganization energy that can be large and vary strongly, even for relatively well preorganized binding sites.[116] As an alternative, a scoring function that evaluates the "ligandability" of a generated conformation can be helpful, as has been shown with an even simpler approach that worked well in a related scenario.[117]

Essential Dynamics and Sequence of Domain Movements: ADK As a Test Case. In order to further determine the extent of conformational sampling during NMSim simulations, essential dynamics (ED) calculations[65,66] were performed on ensembles of NMSim generated conformations and experimentally determined structures. ADK was chosen for this because ADK is a well-studied protein in terms of catalytic mechanism and conformational flexibility and has been used as a test case in other theoretical studies.[81,82,118] The ensemble of experimentally determined structures consists of 11 crystal structures of ADK.[77−80] These structures can be subdivided into three groups with respect to the observed ADK conformations: (i) structures near the open conformation [PDB code (chain): 4ake (A) and 4ake (B)];

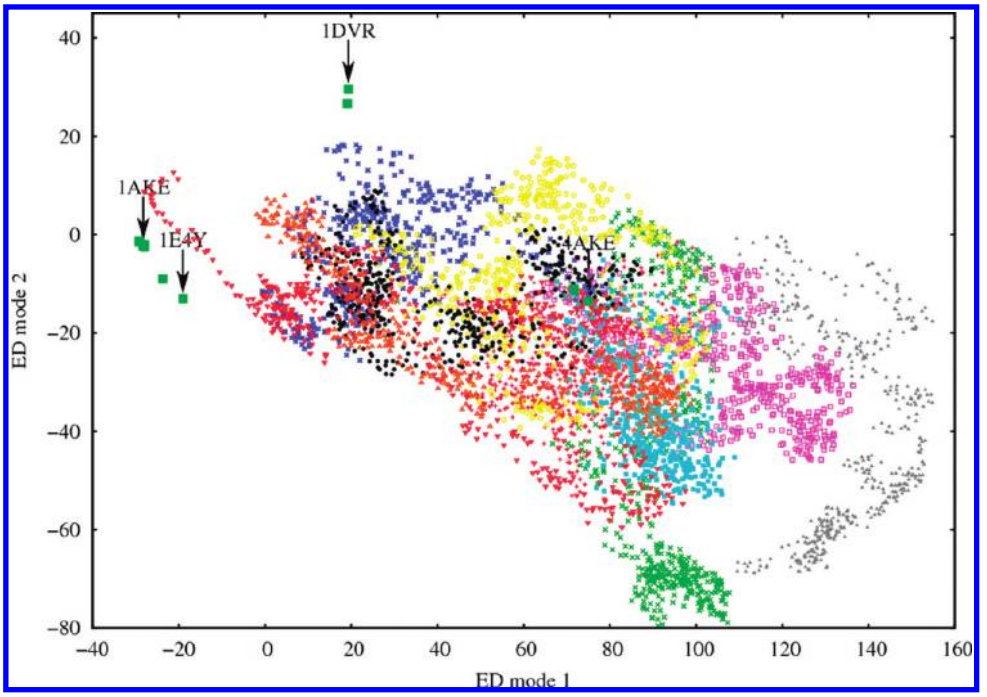


Figure 6. Projection of conformations from 10 independent trajectories generated by unbiased NMSim simulations (each colored separately) and 11 experimentally determined structures (large green squares) of adenylate kinase onto the plane described by the first 2 ED modes derived from the 11 experimentally determined structures.

(ii) intermediate structures in between the open and closed conformations, where the LID domain is completely closed and the NMPbind domain is still open [1dvr (A) and 1dvr (B)]; (iii) structures near the closed conformation [1ake (A), 1ank (A), 1e4v (A), 1e4y (A), 1e4y (B), 2eck (A), and 2eck (B)]. The ensemble of NMSim conformations was obtained by pooling 10 trajectories generated by unbiased NMSim runs, which had been started from the open form of ADK (4ake (A)).

The projection of crystal and NMSim generated ADK structures onto the plane spanned by the first two ED modes obtained from the ensemble of crystal structures is shown in Figure 6. The first ED mode represents to a large extent the movement of the LID domain. The projection map shows that NMSim generated conformations come close to experimental structures in ED space. This is particularly pronounced in the case of the closed ADK structure, where in 1 out of 10 NMSim trajectories conformations were generated that show a closing of the LID and NMP domains to an extent seen in the closed experimental structures (red triangles in Figure 6). In contrast, 2 out of 10 NMSim trajectories sample conformations that show a further opening of the LID domain (gray triangles and pink squares in Figure 6). So far, no structure from experiment has been described that shows such an opening. Considering those 100 conformations with the largest projection values along the first ED mode, an average effective conformational energy of −7398 ± 5.5 kcal mol^{-1} is computed. For comparison, the average effective conformational energy over 500 conformations of the combined ensemble (obtained by extracting every 10th structure of each trajectory) is −7412 ± 4.56 kcal mol^{-1}. Although these numbers indicate that the further opened conformations are less favorable, they do not rule out that these conformations can exist given the statistical uncertainties in the effective conformational energy calculations and the general difficulty of reliably estimating reorganization energies.[116]

The projection of structures generated from the targeted and the radius of gyration-guided NMSim simulations onto the plane spanned by the first two ED modes obtained from the ensemble of crystal structures is shown in Figure S5, Supporting Information. Due to the biased sampling, movements of the protein toward conformations that are further opened than the starting structure, as observed in the unbiased case, do not occur. In contrast, the sampling is more focused on generating conformations that lie in between the opened and closed conformations, as expected. In the case of the radius of gyration-guided NMSim simulation, the generated conformations cover an area on the projection plane close to the intermediate structures [1dvr (A) and (B)] that also shows a large density of projected conformations in the case of the unbiased simulations. This suggests that, although the radius of gyration-guided NMSim simulation focuses the sampling, the guidance is apparently gentle enough so that intermediate structures can still be approached as in the unbiased simulation (see also below).

Next, we investigated whether pathways generated by targeted or radius of gyration-guided NMSim simulations of ADK provide a realistic representation of the sequence of conformational changes that occur during the transition from the open to the closed structure. For analyzing the order of closing of domains, reaction coordinates described by Whitford et al.[81] were used (see Materials and Methods Section). As depicted in Figure 7, in the targeted NMSim simulation, the closing of the LID domain precedes the closing of the NMPbind domain. This is in agreement with previous studies,[58,81] which have suggested that a sequential domain closure has evolved such as to prevent nonproductive substrate binding.[81] Out of 50 normal modes available for the targeted NMSim simulation (eq 6), the first 5 lowest frequency modes are applied throughout most of the transition (Figure 7), suggesting that the closing of the domains follows a low-energy pathway. In fact, the initial closing of the

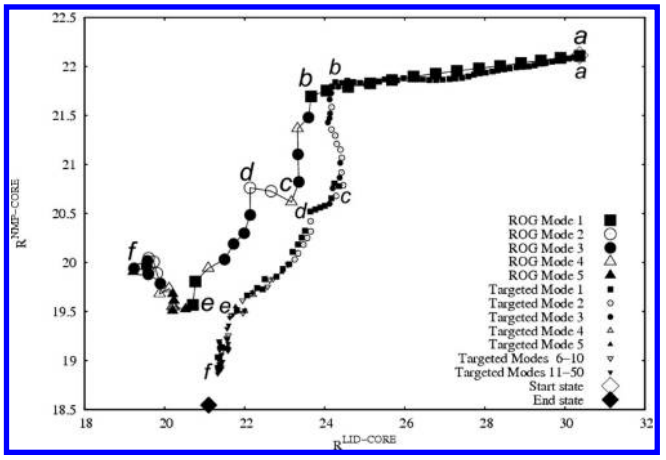


Figure 7. Pathway of conformational transition of ADK generated by targeted (small symbols) and radius of gyration-guided (large symbols) NMSim simulations, which were started from the open conformation. The abscissa denotes the distance between the LID domain and the CORE domain centers of mass; the ordinate denotes the distance between the NMPbind domain and the CORE domain centers of mass. Symbols denote the normal modes used to generate an intermediate conformation from a previous one. The start (unfilled diamond) and end (filled diamond) states are represented by PDB codes 4ake and 1ake, respectively. States along the trajectory are marked by a−f; see text for details.

LID domain (state a−b) is completely dominated by the first normal mode, whereas the partial closing of the NMPbind domain (state b−c) mainly occurs along the directions of the second and third normal modes. Higher frequency modes become involved from state e−f only, i.e., during the final approach of the NMPbind domain to the closed conformation. Notably, a very similar picture emerges in the radius of gyration-guided NMSim simulation of ADK, although here no a priori knowledge about a closed ADK conformation is used. The obtained sequence of domain closure remarkably resembles the one obtained by targeted NMSim simulations between states a−e. Both pathways differ only in the last part (state e−f), where the LID domain moves closer to the CORE domain in the radius of gyration-guided NMSim simulation compared to the targeted NMSim simulation (Figure 7). As for the NMPbind domain, a similar level of domain closure is observed between states b−e in both simulations.

In order to further verify the generated pathways, 1 of the 11 crystal structures of ADK was assigned to each intermediate pathway structure, using the respective lowest C_α RMSD as a criterion similar to a study by Maragakis and Karplus.[82] Plots of the C_α RMSD between intermediate and crystal structures are shown in Figure S6, Supporting Information. For the targeted NMSim simulations, the assigned crystal structures observed along the generated pathway [from open to closed: 4ake (A), 4ake (B), 1dvr (A), 1e4y (B), 1e4y (A), 1ank (A), 2eck (B), 1ake (A)] are in good agreement with the previously suggested sequence of structures.[82] Similarly, crystal structures 4ake (A), 4ake (B), 1dvr (A), 1dvr (B), and 1ake (A) are assigned to the pathway generated by radius of gyration-guided NMSim simulations, which again agrees with the previously reported sequence of crystal structures.[82] This finding is particularly encouraging in the case of 1dvr (A) and 1dvr (B), which lie in between the open and closed ADK conformations, yet are approached up to 2.5−3.0 Å RMSD during the course of both simulations.

CONCLUSION

Efficiently predicting conformational changes of biomacromolecules will be important for understanding biological function and will be valuable for modeling (macro)molecular complex formation and in structure-based drug design. Here, a novel three-step approach, termed NMSim, for multiscale modeling of protein conformational changes has been developed that incorporates information about preferred directions of protein motions into a geometric simulation algorithm. The first two steps are based on previous work in our lab.[44] Initially, static properties of the protein are determined by decomposing the protein into rigid clusters using an all-atom representation of the protein. In a second step, dynamic properties of the molecule are revealed using an elastic network model representation of the coarse-grained protein. The resulting rigid cluster normal mode analysis provides directions of intrinsic motions in terms of normal modes,[43,44] which can be viewed as possible deformations of proteins along low-energy paths. It has been shown by us[44] and others[38,43,102,119] that conformational changes of proteins upon ligand binding occur preferentially along the directions of a few normal modes derived from an unbound protein structure. In the third step, the recently introduced idea of constrained geometric simulations of diffusive motions in proteins[34] is extended, aiming at an efficient sampling of conformational space. For this, backbone motions of the protein are guided by low-frequency normal modes, whereas side-chains perform diffusive motions biased toward energetically favorable rotamer states. The generated structures are iteratively corrected regarding steric clashes and constraint violations. In total, when applied repetitively over all three steps, the procedure efficiently generates a series of stereochemically allowed conformations that lie preferentially in the subspace spanned by low-frequency normal modes. The developed approach allows to perform three simulation types: (i) unbiased exploration of the conformational space; (ii) pathway generation by a targeted simulation; and (iii) radius of gyration-guided simulation to foster the compaction of a protein structure upon ligand binding. We note that in the latter case no experimental radius of gyration is required as input.

Incorporating directional information distinguishes the NMSim approach from other widely used geometry-based simulation approaches, FRODA[34] and CONCOORD.[32,33] While FRODA and NMSim share a natural way of coarse graining the protein structure,[120] they differ at the simulation level. In contrast to NMSim, FRODA simulates only diffusive motions of flexible regions and rigid clusters without considering any information about preferred directions of motion. This limits the sampling of the conformational space by FRODA,[121] particularly in those cases where proteins are rather flexible. The CONCOORD approach iteratively satisfies interatomic distance constraints to generate conformations starting from randomized atomic coordinates. While this allows for a thorough sampling of the conformational space,[121] it precludes the generation of a trajectory of consecutive protein conformations. Thus, from an ensemble of CONCOORD generated conformations, no sequential ordering of conformational transitions could be deduced. Still, we note that none of the geometry-based simulation approaches discussed here samples from a thermodynamic ensemble and, hence, none provides a quantitative description of the distribution of the generated conformations.

When applied to hen egg white lysozyme, NMSim simulations compare favorably with state-of-the-art MD simulations in terms

of root-mean-square fluctuations, extent of sampling, and stereochemical quality as well as computational effort. When applied to a data set of proteins where conformational changes have been observed experimentally, either as domain motions or motions of functionally important loops, experimental conformational variabilities are reproduced very well by the NMSim simulations for 4 out of 5 domain moving proteins with correlation coefficients $r > 0.70$ and as high as $r = 0.92$ in the case of ADK. In 7 out of 8 cases, NMSim simulations starting from unbound structures are able to sample conformations that are similar (RMSD < 3.1 Å) to ligand bound conformations: Minimal RMSD between simulated conformations and respective bound conformations for 4 out 5 proteins with domain motions are between 1.0 (AST) and 3.1 Å (ADK) and for proteins with loop motions are between 1.1 (CAPK) and 2.0 Å (TIM). Biasing the search toward structures with lower radius of gyration considerably improves the sampling of ligand-bound conformations in NMSim, without any a priori information of the bound structure. This is valuable in light of fully flexible docking approaches that require a pregenerated conformational ensemble of the receptor.

The NMSim generated pathway of conformational change from the unbound structure to the ligand bound structure of ADK is validated by a comparison to experimental structures that reflect different conformational states along the transition as proposed by previous studies.[81,82,122] Furthermore, the generated pathway describes the correct sequence of the domain closing in that the closing of the LID domain precedes the closing of the NMPbind domain.

The above results show that incorporating directional information about collective motions into a constrained geometric simulation-based approach allows for the thorough sampling of biologically relevant conformational space. The NMSim approach may thus be a computationally efficient alternative to MD simulations for conformation generation. The generated conformations and pathways of conformational transitions can serve as input to docking approaches, including receptor flexibility,[123] or to more sophisticated techniques, such as nudged elastic band simulations,[110] transition path sampling,[111] or umbrella sampling.[112,113]

APPENDIX

Rigid Cluster Normal Mode Analysis. RCNMA consists of two steps. In the first step, a rigid cluster decomposition of the protein is obtained by FIRST analysis,[59] using an all-atom representation. In the second step, a RTB analysis[60] is performed based on a coarse-grained ENM representation of the protein consisting of rigid clusters connected by flexible links.

Rigid Cluster Decomposition. FIRST identifies and counts the bond rotational degrees of freedom in a molecular framework, whose vertices represent protein atoms and whose edges represent covalent and noncovalent (hydrogen bond and hydrophobic) constraints within the protein.[59,124,125] Flexibility in this network results from dihedral rotations of bonds that are not locked in by other bonds. Each bond is assigned by FIRST to be part of either a rigid cluster or a flexible region. A rigid cluster forms a collection of interlocked bonds for which only rigid-body motions (translation and rotation) are allowed. Under-constrained regions in the network are typically flexible links between rigid clusters.

The molecular framework that represents the protein is completely defined by bond constraints between atoms and next-nearest neighbor constraints that define coordination angles between bonded atoms. Biologically important motions are in many cases characterized by low-frequency, large-amplitude structural fluctuations. By including constraints into the network that represent strong forces, high-frequency motions can be effectively quenched, thereby reducing the complexity of the energy landscape. Here, covalent and hydrogen bonds, salt bridges, and hydrophobic interactions are considered to be strong forces. The configuration of double and partial double bonds (peptide bonds) is restricted by additional constraints.[124] The noncovalent interactions are modeled as described in previous FIRST studies.[124,126,127]

Elastic network model. Based on a simplified representation of the potential energy,[103,128,129] the protein is described as a 3D elastic network. Each amino acid is reduced to a single "particle" (the C_α atom), which acts as a junction in the network. Interactions between these particles are modeled by Hookean springs based on a harmonic pairwise potential,[128] which results in a total potential energy of the system given by eq 9:

$$V = \frac{\gamma}{2}\sum_i \sum_j \theta(r_c - r_{ij}^0)(r_{ij} - r_{ij}^0)^2 \quad (9)$$

where $r_c = 10$ Å is the cutoff up to which interactions between the C_α atoms are taken into account; r_{ij} and r_{ij}^0 are the instantaneous and equilibrium distances between atoms i and j, respectively; $\theta(x)$ is the heaviside step function that accounts for the cutoff effect of the interaction (it is 1 if $x > 0$ and 0 otherwise); and γ is a phenomenological force constant assumed to be the same for all pairwise interactions (it is set to 1 kcal mol^{-1} Å^{-2}).

According to the ENM,[43] the elements of a $3N \times 3N$ Hessian matrix H (where N is the number of C_α atoms) are then obtained from the second derivatives of V with respect to the Cartesian coordinates of atoms i and j. H is diagonalized to obtain the normal modes.

Coarse Graining in RCNMA. The FIRST analysis adds another level of coarse graining of the protein structures. Each rigid cluster forms a block in the RTB approach,[37,60] and flexible regions are modeled on a one-residue-per-block basis (in which case only translational motions of the "block" are considered). Interactions between these blocks are modeled as described for the ENM (eq 9). The $3N \times 3N$ matrix H is therefore reduced to a $6n \times 6n$ dimensional matrix H_{sub} by projecting H into the subspace spanned by translation/rotation basis vectors of n blocks according to eq 10:

$$H_{sub} = P^t H P \quad (10)$$

with P being an orthogonal $3N \times 6n$ projection matrix of the infinitesimal translation/rotation eigenvectors of each block. This leads to a reduction of the memory requirement proportional to $(N/n)^2$ and computational time proportional to $(N/n)^3$, respectively. Diagonalization of the resulting matrix H_{sub} yields the normal modes U_{sub} and eigenvalues Λ (eq 11):

$$H_{sub} U_{sub} = U_{sub} \Lambda \quad (11)$$

Finally, atomic displacements can be obtained by expanding back the eigenvectors U_{sub} from the subspace spanned by translation/rotation basis vectors of the blocks to the Cartesian space (U) (eq 12):

$$U = P U_{sub} \quad (12)$$

The $3N \times 6n$ dimensional matrix U contains $6n$ normal modes $\vec{C}^k$. The k^{th} normal mode direction for the j^{th} C_α atom is then given by $\vec{C}_j^k = [U_{x,k}, U_{y,k}, U_{z,k}]$, with $x = 3^*j - 2$, $y = 3^*j - 1$, and $z = 3^*j$.

Structure Correction

Covalent Bonds. All covalent bonds are modeled by equality distance constraints between the bonded atoms. Additionally, all bond angles in the covalent bond network are modeled as equality distance constraints between the two nonvertex atoms of the angle. Reference distances for these constraints are taken from the input structure.

Noncovalent Bonds. Noncovalent bonds are modeled explicitly and include hydrogen bonds, salt bridges, and hydrophobic interactions. These are recognized from the input starting structure using the FIRST approach[59] and kept throughout the simulation. For example, noncovalent bonds are neither broken nor formed during the NMSim simulation. Each hydrogen bond and salt bridge is modeled by three equality distance constraints between the donor and acceptor atoms, a neighbor atom of the acceptor and the donor atom, and a neighbor atom of the donor and the acceptor atom. That way, hydrogen atoms need not be considered in the NMSim simulation, which increases the efficiency of the calculations. A neighbor atom is selected randomly if the donor/acceptor atom is covalently bonded to multiple atoms. Hydrophobic interactions are also recognized from the input starting structure using the FIRST approach.[59] Each carbon−carbon, carbon−sulfur, or sulfur−sulfur atom pair is recognized as a hydrophobic interaction if the atoms in the pair are within the sum of their van der Waals radii (vdW) plus a cutoff. By default, the cutoff is set to 0.35 Å. Each hydrophobic interaction is then modeled as an upper bound distance constraint between the interacting atoms, i.e., two hydrophobic atoms can only be separated up to a maximum distance, which allows the atoms to slide with respect to each other yet not pull apart.

Steric Clashes. To avoid steric clashes, lower bound distance constraints are defined between atoms that are neither involved in covalent nor in noncovalent interactions (except hydrophobic interactions). The minimum distance of the constraint is defined as the sum of the atomic vdW radii. The latter are taken from Tsai et al.[130] and consider the hybridization states of heavy atoms. This allows to implicitly model the presence of hydrogen atoms. A tolerance factor, defined as the fraction of the sum of vdW radii, is applied to allow a certain steric overlap of atoms (Table S2, Supporting Information). The tolerance factor is higher for atoms involved in 1−4 distance constraints, i.e., atom pairs that are separated by three covalent bonds (Table S2, Supporting Information).

Backbone Torsion Angles φ/ψ. In order to generate protein conformations with favorable secondary structure regions, three basins of attraction for core regions of α_L, α_R, and β backbone conformations are defined using the Ramachandran map described by Morris et al.[131] During the structure correction, φ/ψ angles that lie in allowed or generously allowed regions of α_L, α_R, and β feel an "attraction" toward the center of the core regions. The attraction is generated by adjusting distance constraints between atoms defined below such that if these distance constraints are fully satisfied, φ/ψ angles lie in the center of one of the core regions. In turn, disallowed regions are avoided by corrections of steric clashes.

Except for Gly residues, φ/ψ angles of residue r are modeled by four distance constraints: For modeling φ angles, distances between C_{r-1} and C_r as well as between C_{r-1} and $C_{\beta,r}$ atoms are considered. For modeling ψ angles, distances between N_r and N_{r+1} as well as between N_{r+1} and $C_{\beta,r}$ atoms are considered (Figure S7, Supporting Information). Reference distances for these constraints are set based on the selected basin of attraction. As these constraints are used to bias φ/ψ angles toward favorable secondary structure regions, the constraints are only weakly adjusted during a structure correction cycle. This is achieved by using a small adjustment factor (Table S2, Supporting Information), which has been determined by empirical testing to ensure a limited biasing.

Rotamer Modeling. After displacing side chain atoms in the structure distortion step, the side chain is attracted toward a basin derived from a rotamer library in the structure correction step.[132] This biasing is done as follows: First, the nearest rotamer state is selected for each residue r by searching a rotamer list, which is specific for each amino acid type, for candidates such that the candidate rotamers have all χ angles within a limit CHIDEV_SELLIMIT (Table S1, Supporting Information) of the corresponding χ angles of residue r. The nearest rotamer is then selected from the set of candidates based on the smallest RMSD from residue r. Second, in every structure correction iteration, the χ angles of residue r are adjusted toward the corresponding χ angles of the selected rotamer. The adjustment depends on the χ angle deviation from the selected rotamer χ angle and an adjustment factor (Table S2, Supporting Information). Again, a small adjustment factor is used, which ensures a limited biasing and structural stability. This follows observations according to which the rotamericity of side chains is not perfect in an ensemble of proteins. A recent study shows that between 5 and 30% of the side chains do not correspond to any rotameric state.[133]

Backbone and Side Chain Planarity and Chirality. To acquire side chain planarity in the structure correction step, a superimposition method is used. For this, perfect planar side chain groups are least-squares fit onto their respective distorted counterparts in the structure. Then, the atoms in the distorted planar groups are moved to the coordinate positions of the superimposed planar groups. Finally, distorted covalent and noncovalent constraints are corrected. This procedure is repeated until convergence. To acquire backbone planarity, the same procedure as for side chain planarity is used, but with a relaxed adjustment factor for moving atoms of the distorted planar group toward the superimposed plane (Table S2, Supporting Information). Chirality is another important property to conserve. Changes in the configuration of backbone C_α atoms as well as C_β atoms of threonine and isoleucine side chains are primarily avoided by performing only small structure distortion steps (Table S1, Supporting Information). However, if a change in a configuration is identified, then the generated structure is rejected.

Constraint Adjustment. An iterative approach is applied to satisfy the constraint network consisting of the above-described covalent and noncovalent bonds and the stereochemical criteria. The different types of constraints are satisfied in the sequence shown in the structure correction module in Figure 2. In every structure correction cycle, each constraint is adjusted using respective adjustment factors until the difference between the reference constraint property and the actual property falls below

a respective tolerance value (see Table S2, Supporting Information for adjustment and tolerance values) or until the number of maximum iterations is exceeded (Table S1, Supporting Information).

A schematic diagram for a distance constraint (i.e., constraints for covalent, noncovalent, steric clashes, and backbone torsion angles) correction is shown in Figure S8, Supporting Information. Here two atoms i and j, connected by a distance constraint with a reference distance d_{ij}, are moved in normal mode directions to new positions $\vec{a}$ and $\vec{b}$, respectively, leading to an actual distance d'_{ij} between them. The constraint is corrected by adding vectors $\vec{G}_{ij} = -\vec{G}_{ji}$, respectively, to the current position vectors $\vec{a}$ and $\vec{b}$ to get new coordinate position vectors $\vec{i'}$ and $\vec{j'}$, respectively. The correction vector $\vec{G}_{ij}$ is calculated by

$$\vec{G}_{ij} = \frac{\vec{u}}{|\vec{u}|} * \Delta d_{ij} * AdjustFactor \qquad (13)$$

where $\vec{u} = \vec{a} - \vec{b}$ and $\Delta d_{ij} = d_{ij} - d_{ij}'$. The AdjustFactor for a distance constraint can have a maximum value of 0.5, which means that the constraint will be satisfied by moving both connected atoms midway along the line joining the two atoms.

■ ASSOCIATED CONTENT

Supporting Information. Tables with constraints and simulation parameters used in NMSim, and tables comparing different sets of HEWL structures in terms of RMSF, structural quality, and effective conformational energies. Graphical representations of: (i) RMSF of proteins with loop movements; (ii) RMSD of conformations generated by unbiased NMSim simulations for proteins with domain movements; (iii) RMSD of conformations generated by unbiased or radius of gyration-guided NMSim simulations for proteins with loop movements; (iv) RMSD of conformations generated by radius of gyration-guided NMSim simulations for proteins with domain movements; (v) projection of ensembles generated by targeted and radius of gyration-guided NMSim simulations onto the principal directions obtained from experimental structures; (vi) RMSD of conformations generated by targeted and radius of gyration-guided NMSim simulations; (vii) distance constraints used for modeling φ and ψ torsion angles; (viii) distance constraint correction; (ix) RMSF of HEWL ensembles; (x) projections of HEWL ensembles onto MD essential space. This information is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: gohlke@uni-duesseldorf.de. Telephone: (+49) 211 81-13662.

Present Addresses

[||]Department of Biochemistry and Molecular Biophysics, The University of Arizona, Tucson, Arizona, United States.

■ ACKNOWLEDGMENT

We are grateful to Alrun Koller, Heinrich-Heine-University, Düsseldorf, for providing the MD trajectory of HEWL and Merck KGaA, Darmstadt, for funding this project.

■ REFERENCES

(1) Vonrhein, C.; Schlauderer, G. J.; Schulz, G. E. Movie of the structural changes during a catalytic cycle of nucleoside monophosphate kinases. *Structure* **1995**, *3* (5), 483–490.

(2) Wolf-Watz, M.; Thai, V.; Henzler-Wildman, K.; Hadjipavlou, G.; Eisenmesser, E. Z.; Kern, D. Linkage between dynamics and catalysis in a thermophilic-mesophilic enzyme pair. *Nat. Struct. Mol. Biol.* **2004**, *11* (10), 945–949.

(3) Henzler-Wildman, K. A.; Thai, V.; Lei, M.; Ott, M.; Wolf-Watz, M.; Fenn, T.; Pozharski, E.; Wilson, M. A.; Petsko, G. A.; Karplus, M.; Hubner, C. G.; Kern, D. Intrinsic motions along an enzymatic reaction trajectory. *Nature* **2007**.

(4) Wlodawer, A.; Vondrasek, J. Inhibitors of HIV-1 protease: A major success of structure-assisted drug design. *Annu. Rev. Biophys. Biomol. Struct.* **1998**, *27*, 249–284.

(5) Wilson, D. K.; Tarle, I.; Petrash, J. M.; Quiocho, F. A. Refined 1.8 A structure of human aldose reductase complexed with the potent inhibitor zopolrestat. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90* (21), 9847–9851.

(6) Lin, J. H.; Perryman, A. L.; Schames, J. R.; McCammon, J. A. Computational drug design accommodating receptor flexibility: the relaxed complex scheme. *J. Am. Chem. Soc.* **2002**, *124* (20), 5632–5633.

(7) Märki, H. P.; Fischli, W.; Binggeli, A.; Breu, V.; Bur, D.; Clozel, J. P.; D'Arcy, A.; Grüninger, F.; Güller, R.; Hirth, G.; Lave, T.; Mathews, S.; Müller, M.; Oefner, C.; Stadler, H.; Vieira, E.; Wilhelm, M.; Wostl, W. In Proceedings of 9th RSC-SCI Medicinal Chemistry Symposium, Cambridge, U.K., September 7, 1997; RSC's Biological & Medicinal Chemistry Sector and SCI's Fine Chemicals Group: Cambridge, U.K., 1997.

(8) Allemann, R. K.; Egli, M. DNA recognition and bending. *Chem. Biol.* **1997**, *4* (9), 643–650.

(9) Fourmy, D.; Yoshizawa, S.; Puglisi, J. D. Paromomycin binding induces a local conformational change in the A-site of 16 S rRNA. *J. Mol. Biol.* **1998**, *277* (2), 333–345.

(10) Fulle, S.; Gohlke, H. Constraint counting on RNA structures: linking flexibility and function. *Methods* **2009**, *49* (2), 181–188.

(11) Davis, A. M.; Teague, S. J. Hydrogen Bonding, Hydrophobic Interactions, and Failure of the Rigid Receptor Hypothesis. *Angew. Chem., Int. Ed. Engl.* **1999**, *38* (6), 736–749.

(12) Gohlke, H.; Klebe, G. Approaches to the Description and Prediction of the Binding Affinity of Small-Molecule Ligands to Macromolecular Receptors. *Angew. Chem.* **2002**, *41*, 2644–2676.

(13) Erickson, J. A.; Jalaie, M.; Robertson, D. H.; Lewis, R. A.; Vieth, M. Lessons in molecular recognition: the effects of ligand and protein flexibility on molecular docking accuracy. *J. Med. Chem.* **2004**, *47* (1), 45–55.

(14) Murray, C. W.; Baxter, C. A.; Frenkel, A. D. The sensitivity of the results of molecular docking to induced fit effects: application to thrombin, thermolysin and neuraminidase. *J. Comput.-Aided Mol. Des.* **1999**, *13* (6), 547–562.

(15) Verdonk, M. L.; Mortenson, P. N.; Hall, R. J.; Hartshorn, M. J.; Murray, C. W. Protein-ligand docking against non-native protein conformers. *J. Chem. Inf. Model.* **2008**, *48* (11), 2214–2225.

(16) Cavasotto, C. N.; Abagyan, R. A. Protein flexibility in ligand docking and virtual screening to protein kinases. *J. Mol. Biol.* **2004**, *337* (1), 209–225.

(17) Wall, M. E.; Gallagher, S. C.; Trewhella, J. Large-scale shape changes in proteins and macromolecular complexes. *Annu. Rev. Phys. Chem.* **2000**, *51*, 355–380.

(18) Kay, L. E. Protein dynamics from NMR. *Nat. Struct. Biol.* **1998**, *5* (Suppl), 513–517.

(19) McCammon, J. A.; Gelin, B. R.; Karplus, M. Dynamics of folded proteins. *Nature* **1977**, *267* (5612), 585–590.

(20) Hansson, T.; Oostenbrink, C.; van Gunsteren, W. Molecular dynamics simulations. *Curr. Opin. Struct. Biol.* **2002**, *12* (2), 190–196.

(21) Adcock, S. A.; McCammon, J. A. Molecular dynamics: survey of methods for simulating the activity of proteins. *Chem. Rev.* **2006**, *106* (5), 1589–1615.

(22) Berne, B. J.; Straub, J. E. Novel methods of sampling phase space in the simulation of biological systems. *Curr. Opin. Struct. Biol.* **1997**, *7* (2), 181–189.
(23) Lei, H.; Duan, Y. Improved sampling methods for molecular simulation. *Curr. Opin. Struct. Biol.* **2007**, *17* (2), 187–191.
(24) Grubmuller, H. Predicting slow structural transitions in macromolecular systems: Conformational flooding. *Phys. Rev. E* **1995**, *52* (3), 2893–2906.
(25) Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **1999**, *314* (1−2), 141–151.
(26) Zhou, R. Replica exchange molecular dynamics method for protein folding simulation. *Methods Mol. Biol.* **2007**, *350*, 205–223.
(27) Wu, X.; Wang, S. Self-Guided Molecular Dynamics Simulation for Efficient Conformational Search. *J. Phys. Chem. B* **1998**, *102* (37), 7238–7250.
(28) van der Vaart, A.; Karplus, M. Simulation of conformational transitions by the restricted perturbation-targeted molecular dynamics method. *J. Chem. Phys.* **2005**, *122* (11), 114903.
(29) Schlitter, J.; Engels, M.; Kruger, P. Targeted molecular dynamics: a new approach for searching pathways of conformational transitions. *J. Mol. Graphics* **1994**, *12* (2), 84–89.
(30) Cheatham, T. E.; Young, M. A. Molecular dynamics simulation of nucleic acids: successes, limitations, and promise. *Biopolymers* **2000**, *56* (4), 232–256.
(31) Moraitakis, G.; Purkiss, A.; Goodfellow, J. Simulated dynamics and biological macromolecules. *Rep. Prog. Phys.* **2003**, *66* (3), 383–406.
(32) de Groot, B. L.; van Aalten, D. M. F.; Scheek, R. M.; Amadei, A.; Vriend, G.; Berendsen, H. J. C. Prediction of protein conformational freedom from distance constraints. *Proteins: Struct., Funct., Bioinf.* **1997**, *29* (2), 240–251.
(33) Seeliger, D.; Haas, J.; de Groot, B. Geometry-Based Sampling of Conformational Transitions in Proteins. *Structure* **2007**, *15* (11), 1482–1492.
(34) Wells, S.; Menor, S.; Hespenheide, B.; Thorpe, M. F. Constrained geometric simulation of diffusive motion in proteins. *Phys. Biol.* **2005**, *2* (4), S127–136.
(35) Mustard, D.; Ritchie, D. Docking essential dynamics eigenstructures. *Proteins: Struct., Funct., Bioinf.* **2005**, *60* (2), 269–274.
(36) Jolley, C. C.; Wells, S. A.; Hespenheide, B. M.; Thorpe, M. F.; Fromme, P. Docking of photosystem I subunit C using a constrained geometric simulation. *J. Am. Chem. Soc.* **2006**, *128* (27), 8803–8812.
(37) Tama, F.; Gadea, F. X.; Marques, O.; Sanejouand, Y. H. Building-block approach for determining low-frequency normal modes of macromolecules. *Proteins: Struct., Funct., Bioinf.* **2000**, *41* (1), 1–7.
(38) Tama, F.; Sanejouand, Y. H. Conformational change of proteins arising from normal mode calculations. *Protein Eng.* **2001**, *14* (1), 1–6.
(39) Go, N.; Noguti, T.; Nishikawa, T. Dynamics of a small globular protein in terms of low-frequency vibrational modes. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80* (12), 3696–3700.
(40) Brooks, B.; Karplus, M. Normal modes for specific motions of macromolecules: application to the hinge-bending mode of lysozyme. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82* (15), 4995–4999.
(41) Hinsen, K. Analysis of domain motions by approximate normal mode calculations. *Proteins: Struct., Funct., Bioinf.* **1999**, *33* (3), 417–429.
(42) Hinsen, K.; Thomas, A.; Field, M. J. Analysis of domain motions in large proteins. *Proteins: Struct., Funct., Bioinf.* **1999**, *34* (3), 369–382.
(43) Atilgan, A. R.; Durell, S. R.; Jernigan, R. L.; Demirel, M. C.; Keskin, O.; Bahar, I. Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys. J.* **2001**, *80* (1), 505–515.
(44) Ahmed, A.; Gohlke, H. Multiscale modeling of macromolecular conformational changes combining concepts from rigidity and elastic network theory. *Proteins: Struct., Funct., Bioinf.* **2006**, *63* (4), 1038–1051.
(45) Ahmed, A.; Villinger, S.; Gohlke, H. Large-scale comparison of protein essential dynamics from molecular dynamics simulations and coarse-grained normal mode analyses. *Proteins: Struct., Funct., Bioinf.* **2010**, *78* (16), 3341–3352.
(46) Zhang, Z.; Shi, Y.; Liu, H. Molecular dynamics simulations of peptides and proteins with amplified collective motions. *Biophys. J.* **2003**, *84* (6), 3583–3593.
(47) Tatsumi, R.; Fukunishi, Y.; Nakamura, H. A hybrid method of molecular dynamics and harmonic dynamics for docking of flexible ligand to flexible receptor. *J. Comput. Chem.* **2004**, *25* (16), 1995–2005.
(48) He, J.; Zhang, Z.; Shi, Y.; Liu, H. Efficiently explore the energy landscape of proteins in molecular dynamics simulations by amplifying collective motions. *J. Chem. Phys.* **2003**, *119* (7), 4005–4017.
(49) Cavasotto, C. N.; Kovacs, J. A.; Abagyan, R. A. Representing receptor flexibility in ligand docking through relevant normal modes. *J. Am. Chem. Soc.* **2005**, *127* (26), 9632–9640.
(50) May, A.; Zacharias, M. Protein-protein docking in CAPRI using ATTRACT to account for global and local flexibility. *Proteins: Struct., Funct., Bioinf.* **2007**, *69* (4), 774–780.
(51) May, A.; Zacharias, M. Protein-ligand docking accounting for receptor side chain and global flexibility in normal modes: evaluation on kinase inhibitor cross docking. *J. Med. Chem.* **2008**, *51* (12), 3499–3506.
(52) Delarue, M.; Dumas, P. On the use of low-frequency normal modes to enforce collective movements in refining macromolecular structural models. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101* (18), 6957–6962.
(53) Hinsen, K.; Reuter, N.; Navaza, J.; Stokes, D. L.; Lacapere, J. J. Normal mode-based fitting of atomic structure into electron density maps: application to sarcoplasmic reticulum Ca-ATPase. *Biophys. J.* **2005**, *88* (2), 818–827.
(54) Tama, F.; Miyashita, O.; Brooks, C. L. Flexible multi-scale fitting of atomic structures into low-resolution electron density maps with elastic network normal mode analysis. *J. Mol. Biol.* **2004**, *337* (4), 985–999.
(55) Tama, F.; Miyashita, O.; Brooks, C. L. Normal mode based flexible fitting of high-resolution structure into low-resolution experimental data from cryo-EM. *J. Struct. Biol.* **2004**, *147* (3), 315–326.
(56) Kim, M. K.; Jernigan, R. L.; Chirikjian, G. S. Efficient generation of feasible pathways for protein conformational transitions. *Biophys. J.* **2002**, *83* (3), 1620–1630.
(57) Miyashita, O.; Onuchic, J. N.; Wolynes, P. G. Nonlinear elasticity, proteinquakes, and the energy landscapes of functional transitions in proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100* (22), 12570–12575.
(58) Miyashita, O.; Wolynes, P. G.; Onuchic, J. N. Simple energy landscape model for the kinetics of functional transitions in proteins. *J. Phys. Chem. B* **2005**, *109*, 1959–1969.
(59) Jacobs, D.; Rader, A. J.; Kuhn, L.; Thorpe, M. F. Protein flexibility predictions using graph theory. *Proteins: Struct., Funct., Bioinf.* **2001**, *44* (2), 150–165.
(60) Durand, P.; Trinquier, G.; Sanejouand, Y.-H. A new approach for determining low-frequency normal modes in macromolecules. *Biopolymers* **1994**, *34*, 759–771.
(61) Olah, G. A.; Trakhanov, S.; Trewhella, J.; Quiocho, F. A. Leucine Isoleucine Valine-Binding Protein Contracts Upon Binding of Ligand. *J. Biol. Chem.* **1993**, *268* (22), 16241–16247.
(62) Egea, P. F.; Rochel, N.; Birck, C.; Vachette, P.; Timmins, P. A.; Moras, D. Effects of ligand binding on the association properties and conformation in solution of retinoic acid receptors RXR and RAR. *J. Mol. Biol.* **2001**, *307* (2), 557–576.
(63) Koller, A. N.; Schwalbe, H.; Gohlke, H. Starting structure dependence of NMR order parameters derived from MD simulations: Implications for judging force-field quality. *Biophys. J.* **2008**, *95* (1), L4–L6.
(64) Wilson, K. P.; Malcolm, B. A.; Matthews, B. W. Structural and Thermodynamic Analysis of Compensating Mutations within the Core of Chicken Egg-White Lysozyme. *J. Biol. Chem.* **1992**, *267* (15), 10842–10849.
(65) Amadei, A.; Linssen, A. B.; Berendsen, H. J. Essential dynamics of proteins. *Proteins: Struct., Funct., Bioinf.* **1993**, *17* (4), 412–425.
(66) van Aalten, D. M.; Amadei, A.; Linssen, A. B.; Eijsink, V. G.; Vriend, G.; Berendsen, H. J. The essential dynamics of thermolysin:

confirmation of the hinge-bending motion and comparison of simulations in vacuum and water. *Proteins: Struct., Funct., Bioinf.* **1995**, *22* (1), 45–54.

(67) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber biomolecular simulation programs. *J. Comput. Chem.* **2005**, *26* (16), 1668–1688.

(68) Laskowski, R. A.; Macarthur, M. W.; Moss, D. S.; Thornton, J. M. Procheck - a Program to Check the Stereochemical Quality of Protein Structures. *J. Appl. Crystallogr.* **1993**, *26*, 283–291.

(69) Word, J. M.; Lovell, S. C.; LaBean, T. H.; Taylor, H. C.; Zalis, M. E.; Presley, B. K.; Richardson, J. S.; Richardson, D. C. Visualizing and quantifying molecular goodness-of-fit: Small-probe contact dots with explicit hydrogen atoms. *J. Mol. Biol.* **1999**, *285* (4), 1711–1733.

(70) Onufriev, A.; Bashford, D.; Case, D. A. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins: Struct., Funct., Bioinf.* **2004**, *55* (2), 383–394.

(71) Onufriev, A.; Bashford, D.; Case, D. A. Modification of the Generalized Born Model Suitable for Macromolecules. *J. Phys. Chem. B* **2000**, *104*, 3712–3720.

(72) Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. The GB/SA Continuum Model for Solvation. A Fast Analytical Method for the Calculation of Approximate Born Radii. *J. Phys. Chem. A* **1997**, *101* (16), 3005–3014.

(73) Kurkcuoglu, O.; Jernigan, R. L.; Doruker, P. Loop Motions of Triosephosphate Isomerase Observed with Elastic Networks. *Biochemistry* **2006**, *45* (4), 1173–1182.

(74) Hayward, S. Identification of Specific Interactions that Drive Ligand-induced Closure in Five Enzymes with Classic Domain Movements. *J. Mol. Biol.* **2004**, *339* (4), 1001–1021.

(75) Okazaki, K. I.; Takada, S. Dynamic energy landscape view of coupled binding and protein conformational change: Induced-fit versus population-shift mechanisms. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105* (32), 11182–11187.

(76) Tobi, D.; Bahar, I. Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state. *Proc. Natl. Acad. Sci. U.S.A.* **2005**.

(77) Muller, C. W.; Schulz, G. E. Structure of the Complex between Adenylate Kinase from Escherichia-Coli and the Inhibitor Ap5a Refined at 1.9 Å Resolution - a Model for a Catalytic Transition-State. *J. Mol. Biol.* **1992**, *224* (1), 159–177.

(78) Muller, C. W.; Schlauderer, G. J.; Reinstein, J.; Schulz, G. E. Adenylate kinase motions during catalysis: an energetic counterweight balancing substrate binding. *Structure* **1996**, *4* (2), 147–156.

(79) Schlauderer, G. J.; Schulz, G. E. The structure of bovine mitochondrial adenylate kinase: comparison with isoenzymes in other compartments. *Protein Sci.* **1996**, *5* (3), 434–441.

(80) Schlauderer, G. J.; Proba, K.; Schulz, G. E. Structure of a mutant adenylate kinase ligated with an ATP-analogue showing domain closure over ATP. *J. Mol. Biol.* **1996**, *256* (2), 223–227.

(81) Whitford, P. C.; Gosavi, S.; Onuchic, J. N. Conformational transitions in adenylate kinase - Allosteric communication reduces misligation. *J. Biol. Chem.* **2008**, *283* (4), 2042–2048.

(82) Maragakis, P.; Karplus, M. Large Amplitude Conformational Change in Proteins Explored with a Plastic Network Model: Adenylate Kinase. *J. Mol. Biol.* **2005**, *352* (4), 807–822.

(83) McCammon, J. A.; Gelin, B. R.; Karplus, M.; Wolynes, P. G. The hinge-bending mode in lysozyme. *Nature* **1976**, *262* (5566), 325–326.

(84) Blake, C. C. F.; Cassels, R.; Dobson, C. M.; Poulsen, F. M.; Williams, R. J. P.; Wilson, K. S. Structure and binding properties of hen lysozyme modified at tryptophan 62. *J. Mol. Biol.* **1981**, *147*, 73–95.

(85) Levitt, M.; Sander, C.; Stern, P. S. Protein normal-mode dynamics: trypsin inhibitor, crambin, ribonuclease and lysozyme. *J. Mol. Biol.* **1985**, *181* (3), 423–447.

(86) Bruccoleri, R.; Karplus, M.; McCammon, A. The hinge-bending mode of a lysozyme-inhibitor complex. *Biopolymers* **1986**, *25* (9), 1767–1802.

(87) Buck, M.; Bouguet-Bonnet, S.; Pastor, R. W.; MacKerell, A. D. Importance of the CMAP correction to the CHARMM22 protein force field: Dynamics of hen lysozyme. *Biophys. J.* **2006**, *90* (4), L36–L38.

(88) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins: Struct., Funct., Bioinf.* **2006**, *65* (3), 712–725.

(89) Soares, T. A.; Daura, X.; Oostenbrink, C.; Smith, L. J.; van Gunsteren, W. F. Validation of the GROMOS force-field parameter set 45A3 against nuclear magnetic resonance data of hen egg lysozyme. *J. Biomol. NMR* **2004**, *30* (4), 407–422.

(90) Stocker, U.; van Gunsteren, W. F. Molecular dynamics simulation of hen egg white lysozyme: A test of the GROMOS96 force field against nuclear magnetic resonance data. *Proteins: Struct., Funct., Bioinf.* **2000**, *40* (1), 145–153.

(91) Mcphalen, C. A.; Vincent, M. G.; Jansonius, J. N. X-Ray Structure Refinement and Comparison of 3 Forms of Mitochondrial Aspartate-Aminotransferase. *J. Mol. Biol.* **1992**, *225* (2), 495–517.

(92) Liao, D. I.; Karpusas, M.; Remington, S. J. Crystal structure of an open conformation of citrate synthase from chicken heart at 2.8-A resolution. *Biochemistry* **1991**, *30* (24), 6031–6036.

(93) Kuboniwa, H.; Tjandra, N.; Grzesiek, S.; Ren, H.; Klee, C. B.; Bax, A. Solution structure of calcium-free calmodulin. *Nat. Struct. Biol.* **1995**, *2* (9), 768–776.

(94) Kang, C. H.; Shin, W. C.; Yamagata, Y.; Gokcen, S.; Ames, G. F.; Kim, S. H. Crystal-Structure of the Lysine-Binding, Arginine-Binding, Ornithine-Binding Protein (Lao) from Salmonella-Typhimurium at 2.7 Å Resolution. *J. Biol. Chem.* **1991**, *266* (35), 23893–23899.

(95) Echols, N.; Milburn, D.; Gerstein, M. MolMovDB: analysis and visualization of conformational change and structural flexibility. *Nucleic Acids Res.* **2003**, *31* (1), 478–482.

(96) Vandonselaar, M.; Hickie, R. A.; Quail, J. W.; Delbaere, L. T. J. Trifluoperazine-induced conformational change in Ca(2+)-calmodulin. *Nat. Struct. Biol.* **1994**, *1*, 795–801.

(97) Schubert, H. L.; Fauman, E. B.; Stuckey, J. A.; Dixon, J. E.; Saper, M. A. A ligand-induced conformational change in the Yersinia protein tyrosine phosphatase. *Protein Sci.* **1995**, *4* (9), 1904–1913.

(98) Alber, T.; Banner, D. W.; Bloomer, A. C.; Petsko, G. A.; Phillips, D.; Rivers, P. S.; Wilson, I. A. On the three-dimensional structure and catalytic mechanism of triose phosphate isomerase. *Phil. Trans. R. Soc. Lond. B.* **1981**, *293* (1063), 159–171.

(99) Wierenga, R. K.; Noble, M. E.; Postma, J. P.; Groendijk, H.; Kalk, K. H.; Hol, W. G.; Opperdoes, F. R. The crystal structure of the "open" and the "closed" conformation of the flexible loop of trypanosomal triosephosphate isomerase. *Proteins: Struct., Funct., Bioinf.* **1991**, *10* (1), 33–49.

(100) Williams, J. C.; McDermott, A. E. Dynamics of the flexible loop of triosephosphate isomerase: the loop motion is not ligand gated. *Biochemistry* **1995**, *34* (26), 8309–8319.

(101) Tanaka, T.; Hidaka, H. Hydrophobic regions function in calmodulin-enzyme(s) interactions. *J. Biol. Chem.* **1980**, *255* (23), 11078–11080.

(102) Yang, L.; Song, G.; Jernigan, R. L. How well can we understand large-scale protein motions using normal modes of elastic network models?. *Biophys. J.* **2007**, *93* (3), 920–929.

(103) Bahar, I.; Atilgan, A. R.; Erman, B. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Folding Des.* **1997**, *2* (3), 173–181.

(104) Narayana, N.; Cox, S.; Shaltiel, S.; Taylor, S. S.; Xuong, N. Crystal structure of a polyhistidine-tagged recombinant catalytic subunit of cAMP-dependent protein kinase complexed with the peptide inhibitor PKI(5−24) and adenosine. *Biochemistry* **1997**, *36* (15), 4438–4448.

(105) Madhusudan; Trafny, E. A.; Xuong, N. H.; Adams, J. A.; Ten Eyck, L. F.; Taylor, S. S.; Sowadski, J. M. cAMP-dependent protein kinase: crystallographic insights into substrate recognition and phosphotransfer. *Protein Sci.* **1994**, *3* (2), 176–187.

(106) Suhre, K.; Sanejouand, Y. H. ElNemo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement. *Nucleic Acids Res.* **2004**, *32* (Web Server issue), W610–614.

(107) Petrone, P.; Pande, V. Can conformational change be described by only a few normal modes?. *Biophys. J.* **2006**, *90* (5), 1583–1593.

(108) Cansu, S.; Doruker, P. Dimerization affects collective dynamics of triosephosphate isomerase. *Biochemistry* **2008**, *47* (5), 1358–1368.

(109) Krebs, W. G.; Gerstein, M. The morph server: A standardized system for analyzing and visualizing macromolecular motions in a database framework. *Nucleic Acids Res.* **2000**, *28* (8), 1665–1675.

(110) Mathews, D. H.; Case, D. A. Nudged elastic band calculation of minimal energy paths for the conformational change of a GG non-canonical pair. *J. Mol. Biol.* **2006**, *357* (5), 1683–1693.

(111) Rogal, J.; Bolhuis, P. G. Multiple state transition path sampling. *J. Chem. Phys.* **2008**, *129* (22), 224107.

(112) Torrie, G. M.; Valleau, J. P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.* **1977**, *23* (2), 187–199.

(113) Mamonova, T.; Hespenheide, B.; Straub, R.; Thorpe, M. F.; Kurnikova, M. Protein flexibility using constraints from molecular dynamics simulations. *Phys. Biol.* **2005**, *2* (4), S137–147.

(114) Kantarci-Carsibasi, N.; Haliloglu, T.; Doruker, P. Conformational transition pathways explored by Monte Carlo simulation integrated with collective modes. *Biophys. J.* **2008**, *95* (12), 5862–5873.

(115) Ferrara, P.; Gohlke, H.; Price, D. J.; Klebe, G.; Brooks, C. L. Assessing scoring functions for protein-ligand interactions. *J. Med. Chem.* **2004**, *47*, 3032–3047.

(116) Lazaridis, T.; Masunov, A.; Gandolfo, F. Contributions to the binding free energy of ligands to avidin and streptavidin. *Proteins: Struct., Funct., Bioinf.* **2002**, *47* (2), 194–208.

(117) Eyrisch, S.; Helms, V. Transient pockets on protein surfaces involved in protein-protein interaction. *J. Med. Chem.* **2007**, *50* (15), 3457–3464.

(118) Temiz, N. A.; Meirovitch, E.; Bahar, I. Escherichia coli adenylate kinase dynamics: comparison of elastic network model modes with mode-coupling (15)N-NMR relaxation data. *Proteins: Struct., Funct., Bioinf.* **2004**, *57* (3), 468–480.

(119) Trakhanov, S.; Vyas, N. K.; Luecke, H.; Kristensen, D. M.; Ma, J.; Quiocho, F. A. Ligand-free and -bound structures of the binding protein (LivJ) of the Escherichia coli ABC leucine/isoleucine/valine transport system: trajectory and dynamics of the interdomain rotation and ligand specificity. *Biochemistry* **2005**, *44* (17), 6597–6608.

(120) Gohlke, H.; Thorpe, M. A Natural Coarse Graining for Simulating Large Biomolecular Motion. *Biophys. J.* **2006**, *91* (6), 2115–2120.

(121) Ahmed, A.; Gohlke, H. Multiscale modeling of macromolecular conformational changes. In Proceedings of the *1st International Conference on Computational & Mathematical Biomedical Engineering (CMBE09)*; Nithiarasu, P., Löhner, R., Eds.; CMBE: Swansea, 2009, pp 219−222.

(122) Kirillova, S.; Cortes, J.; Stefaniu, A.; Simeon, T. An NMA-guided path planning approach for computing large-amplitude conformational changes in proteins. *Proteins: Struct., Funct., Bioinf.* **2008**, *70* (1), 131–143.

(123) Ahmed, A.; Kazemi, S.; Gohlke, H. Protein Flexibility and Mobility in Structure-Based Drug Design. *Front. Drug Des. Discovery* **2007**, *3*, 455–476.

(124) Rader, A. J.; Hespenheide, B. M.; Kuhn, L. A.; Thorpe, M. F. Protein unfolding: Rigidity lost. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99* (6), 3540–3545.

(125) Jacobs, D. J.; Kuhn, L. A.; Thorpe, M. F. Flexible and rigid regions in proteins. In *Rigidity Theory and Applications*; Thorpe, M. F., Duxbury, P. M., Eds.; Kluwer Academic/Plenum Publishers: New York, 1999; pp 357−384.

(126) Hespenheide, B. M.; Rader, A. J.; Thorpe, M. F.; Kuhn, L. A. Identifying protein folding cores from the evolution of flexible regions during unfolding. *J. Mol. Graphics Modell.* **2002**, *21*, 195–207.

(127) Gohlke, H.; Kuhn, L. A.; Case, D. A. Change in Protein Flexibility upon Complex Formation: Analysis of Ras-Raf using Molecular Dynamics and a Molecular Framework Approach. *Proteins: Struct., Funct., Bioinf.* **2004**, *56*, 322–337.

(128) Tirion, M. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys. Rev. Lett.* **1996**, *77* (9), 1905.

(129) Hinsen, K. Analysis of domain motions by approximate normal mode calculations. *Proteins: Struct., Funct., Bioinf.* **1998**, *33* (3), 417–429.

(130) Tsai, J.; Taylor, R.; Chothia, C.; Gerstein, M. The packing density in proteins: Standard radii and volumes. *J. Mol. Biol.* **1999**, *290* (1), 253–266.

(131) Morris, A. L.; Macarthur, M. W.; Hutchinson, E. G.; Thornton, J. M. Stereochemical quality of protein-structure coordinates. *Proteins: Struct., Funct., Bioinf.* **1992**, *12* (4), 345–364.

(132) Lovell, S. C.; Word, J. M.; Richardson, J. S.; Richardson, D. C. The penultimate rotamer library. *Proteins: Struct., Funct., Bioinf.* **2000**, *40* (3), 389–408.

(133) Schrauber, H.; Eisenhaber, F.; Argos, P. Rotamers: to be or not to be? An analysis of amino acid side-chain conformations in globular proteins. *J. Mol. Biol.* **1993**, *230* (2), 592–612.