Exploring conformational space with thermal fluctuations obtained by normal mode analysis

Tadeo E. Saldaño[1], Victor M. Freixas[1], Silvio C. E. Tosatto[2], Gustavo Parisi[1], and Sebastian Fernandez-Alberti[1*]

[1]Universidad Nacional de Quilmes/CONICET, Roque Saenz Peña 352, B1876BXD Bernal, Argentina.

[2]Department of Biomedical Sciences, University of Padova, Viale G. Colombo 3, 5131 Padova, Italy.

[*]corresponding autor, sfalberti@gmail.com

ORCID iDs

Tadeo E. Saldaño, 0000-0003-0017-4660
Victor M. Freixas, 0000-0003-1733-4827
Silvio C. E. Tosatto, 0000-0003-4525-7793
Gustavo Parisi, 0000-0001-7444-1624
Sebastian Fernandez-Alberti, 0000-0002-0916-5069

ABSTRACT

Proteins in their native states can be represented as ensembles of conformers in dynamical equilibrium. Thermal fluctuations are responsible for transitions between these conformers. Normal modes analysis (NMA) using elastic network models (ENM) provides an efficient procedure to explore global dynamics of proteins commonly associated to conformational transitions. In the present work, we present an iterative approach to explore protein conformational spaces by introducing structural distortions according to their equilibrium dynamics at room temperature. The approach can be used either to perform unbiased explorations of conformational space or to explore guided pathways connecting two different conformations, e.g., apo and holo forms. In order to test its performance, four proteins with different magnitude of structural distortions upon ligand binding have been tested. In all cases, the conformational selection model has been confirmed and the conformational space between apo and holo forms has been encompassed. Different strategies have been tested that impact either on the efficiency to achieve a desired conformational change or to achieve a balanced exploration of the protein conformational multiplicity.

Keywords: protein conformational space, normal modes.

I. INTRODUCTION

The native state of proteins can be represented as an ensemble of conformers in dynamical equilibrium[1]. That is, conformers are separated by energy barriers that can be overcome by thermal fluctuations. Conformational transitions can mean a wide range of structural changes ranging from spatially localized distortions (e.g. side-chain reorientation that can act as gates for ligand entry, small loop or helical hairpin fluctuations or changes in the relative orientation of residues in the active site[2]) to collective global rearrangements (e.g. changes in the relative orientation between domains, or fluctuations of large loop or intrinsically disordered regions[3,4].

The existence of an ensemble of conformers in dynamical equilibrium validates the generalized conformational selection model[5–10] originally formulated by Monod, Wyman and Changeux to explain corporativism and allosterim[11–13]. According to this model, the pre-existing conformational multiplicity of a protein is required so that it can fulfil its function. Currently, many aspects of the protein function, such as enzyme catalysis[14], signal transduction[15], protein-protein interaction[16], promiscuity[17] and allosterism[1,13] are explained making use of its conformational diversity. Therefore, the exploration of the conformational space of a protein is a valuable task that contributes to understand its function.

Conformational diversity of proteins is also employed to improve the performance of different bioinformatics tools like molecular docking[18], protein-protein interaction prediction[19], evaluation of protein structural models[20], prediction of observed substitution patterns of sequence divergence during evolution[21], and coevolutionary measurements between residues[22]. In order to give rise to this demand, different databases of conformational diversity in the native state of proteins (CoDNaS[23], PDBFlex[24]) have been developed.

Ligand binding can be characterized in terms of structural differences between apo and holo conformations of a protein[25,26]. The pre-existence of both conformers within the conformational space of a protein in the absence of ligand has been extensively verified by multiple experimental techniques like X-ray and cryo-electron microscope images, kinetic studies, single molecule fluorescence and NMR[27–30]. Therefore, the conformational change between apo and holo conformations should be achieved by their intramolecular vibrational dynamics. The energy barriers that separate these conformers are commonly overcome by thermal fluctuations. Therefore, protein vibrations under thermal equilibrium conditions modulate the conformational diversity and the relative population of each conformer on the whole ensemble.

Numerous computational methods have been developed to explore the conformational space of a protein[31]. On one hand, Molecular Dynamics (MD) techniques suffer from the inefficiency in reaching certain high energy barriers between conformers. The height and the collective character of a conformational change impact on the time-scale for it to occur. In order to overcome this issue, several accelerated conformational sampling techniques have been developed, e.g., metadynamics[32], replica-exchange MD (REMD)[33,34], self-guided MD[35], and targeted MD[36,37] among others.

Coarse-grained techniques have been developed not only to reduce the MD computational costs[38–43] but also to contribute in protein design, modelling of protein energy landscape, and conformational transition pathways[44–47]. Methods based on Normal modes

analysis (NMA) using elastic network models (ENM)[48,49] have shown to be largely efficient to explore global dynamics (i.e. low frequency modes) of proteins commonly associated to conformational transitions[50–53]. Moreover, it has been proved that only a few low-frequency normal modes are necessary to achieve a good description of most of the conformational changes associated to the biological function of proteins[54]. Since ENM is limited to second order approximation of the potential energy surfaces, pathway generation requires iterative deformations to explore the conformational space and/or pathways of conformational transitions of proteins[50,51,53,55–66]. Most of the methods propose iterative structural distortions of an original protein structure either performing unbiased explorations of conformational space or exploring guided pathways connecting two different conformations, e.g., apo and holo forms. NMA performed on each new structures obtained at each iteration step proves successful in exploring the conformational space through effective anharmonic paths. The continuously updating of normal modes allows to accomplish the desired conformational transitions. Methods like MAP (MinActionPath)[67], mENM[68], iENM[69], aANM[70], Climber[71], NMSim[72], MDdMD[73], GodMD[74], ANMPathway[75], iMODS[76] and coMD[56], among others, represent conformational transition modelling methods that allow to define transition paths connecting two protein conformations.

Herein, we present an alternative procedure to efficiently explore the conformational space of proteins. The method attempts to efficiently provide realistic ensembles of protein conformations that can be subsequently used in further studies involving large number of homologous proteins. Therefore, it represents an efficient and faster alternative respect to other accurate atomistic methods for sampling of conformational space[60]. Different biophysics and bioinformatic studies require of initial confident ensemble of protein conformations, i.e., flexible docking and drug design approaches, studies of the dynamism of homologous proteins, improvements of homology modelling techniques[77], effect of mutations on changes in the relative population of conformers of the native state[78], coarse-grained MD simulations to explore protein energy landscapes, structurally constrained evolution methods[79] and validation of NMR ensembles[80] among others.

The method is presented in its unbiased and targeted versions. Inspired on previous works[61], the method is a three-step approach involving: (a) identification of the residue interaction network (RIN); (b) NMA based on this RIN; (c) generation, selection and optimization of new protein structures displaced in the direction of selected NMA modes. These three steps are repeated iteratively and random structural distortions are introduced in step (c) according to the protein equilibrium dynamics at a given temperature. That is, the exploration of conformational space is performed fulfilling the equilibrium probability distribution of each collective normal mode at that temperature. Different strategies to generate the new set of structural distortions at each iteration step are discussed.

The paper is organized as follows. In Section II we briefly describe our iterative procedure to protein conformational spaces, either in its unbiased or targeted version. In Section III we present and discuss their efficiency on explore the conformational sampling of four proteins with different magnitude of structural distortions upon ligand binding. Conclusions are given in Section IV.

II. METHODS

A. Dataset

Four proteins have been selected to test the performance of our procedure to explore the conformational space of proteins. They have been selected to represent conformational changes with different magnitude of structural distortions upon ligand binding: Calmodulin (CaM) (PDB ID 1CLL and 1CDL for apo and holo conformations respectively) with RMSD (root-mean-square deviation) = 15.2 Å between both conformations, Adenylate Kinase (AK) (PDB ID 4AKE and 2ECK for apo and holo conformations respectively) with RMSD = 6.9 Å, Lysine-Arginine-Ornithine (LAO) binding protein (PDB ID 2LAO and 1LST for apo and holo conformations respectively) with RMSD = 4.7 Å, Maltodextrin binding protein (MBP) (PDB ID 1OMP and 1ANF for apo and holo conformations respectively) with RMSD = 3.8 Å.

B. Normal Mode Analysis (NMA)

Normal mode analysis has been performed using the Elastic Network Model (ENM)[48,81] that considers the protein as an elastic network with nodes linked by springs within a cutoff distance r_c. Herein, the atoms of protein backbone, C_β and the center of mass of side chains are taken as nodes. The value of r_c is varied from 7Å to 20Å in order to optimize the correlation between theoretical and experimental B-factors.

The interaction potential between the N nodes is then defined as[48,82,83]

$$E(r_i, r_j) = \frac{1}{2} k_{ij} \left(|r_{ij}| - |r_{ij}^0| \right)^2 \tag{1}$$

where $r_{ij} \equiv r_i - r_j$ is the vector connecting nodes i and j, and the zero superscript indicates the equilibrium position that corresponds to the coordinates of the nodes in the evaluated structure. The value of the force constant k_{ij} depends on the type of interaction between nodes i and j[52,84], that have been defined using the RING program (Residue Interaction Network Generator)[85,86].

Normal modes, obtained as a set of eigenvectors $\{\mathbf{Q}_k\}_{k=1,\,3N}$ of the Hessian matrix **H** of the interaction potential energy, are 3N vectors whose elements $\{Q_k^j\}_{j=1,3N}$ represent the relative displacements of Cartesian coordinates of each j^{th} residue. Each normal mode $\mathbf{Q}_k$ has its corresponding frequency $w_k = \sqrt{\lambda_k}$, where λ_k is the k^{th} eigenvalue of **H**. Therefore, their evolution in time can be expressed as

$$Q_k = A_k \cos(\omega t) \tag{2}$$

with A_k being the amplitude of the normal-mode motion. A_k can be calculated considering that the average potential energy in the direction of $\mathbf{Q}_k$ can be written as

$$\langle V_k \rangle = \frac{1}{2} \lambda_k \langle Q_k^2 \rangle = \frac{1}{4} \lambda_k A_k^2 \tag{3}$$

taking into account that, according to the equipartition theorem $\langle V_k \rangle = \frac{1}{2} k_B T$, we obtain

$$A_k = \left(\frac{2 k_B T}{\lambda_k} \right)^{1/2} \tag{4}.$$

Any vector **v** can be expressed in terms of normal modes as

$$v = \sum_{k=1}^{3N-6} (v \cdot Q_k) Q_k = \sum_{k=1}^{3N-6} c_k Q_k \tag{5}$$

with

$$c_k = \sum_{j=1}^{3N} (v \cdot Q_k) \tag{6}$$

the minimum number of modes required to represent vector **v** is calculated using the mode participation number[52,87] as

$$P = \left(\sum_{k=1}^{3N-6} (c_k)^4 \right)^{-1} \tag{7}$$

the P modes with largest values of $(c_k)^2$ represent the minimum set of modes that describe **v**.

The values of $\{\lambda_k\}_{k=1,3N}$ are scaled as $\lambda_k^* = \gamma \lambda_k$ with the scaling constant γ calculated in order to best fit the experimental temperature factors $B_i^{\exp}$ with the theoretical mean square residue fluctuations $\langle \Delta r_i^2 \rangle = \langle (r_i - r_i^0)^2 \rangle$ from its equilibrium position[52]. This is done by calculating the theoretical temperature factors B_i^{theor} as

$$B_i^{theor} = \frac{8\pi}{3} \langle \Delta r_i^2 \rangle \tag{8}$$

where $\langle \Delta r_i^2 \rangle$ can be expressed as[88]

$$\langle \Delta r_i^2 \rangle = 3 k_B T \sum_{k=1}^{3N-6} \left[\lambda_k^{-1} Q_k Q_k^T \right]_{ii} \tag{9}$$

being k_B is the Boltzmann constant, T is the absolute temperature. Then, the scaling constant γ is obtained as

$$\gamma = 8\pi^2 k_B T \frac{\sum_i^{3N} B_i^{theor}}{\sum_i^{N} B_i^{\exp}} \qquad (10).$$

C. Distribution function for a harmonic oscillator in thermal equilibrium

Each normal mode $\mathbf{Q}_k$ represents a harmonic oscillator whose total energy can be calculated as

$$E_k(Q_k, P_k) = \frac{P_k^2}{2} + \frac{1}{2} w_k^2 Q_k^2 \qquad (11)$$

with Q_k and P_k being the normal mode displacement and momentum respectively. The corresponding partition function is written as:

$$Z_k = \frac{1}{h} \int_{-\infty}^{\infty} dQ_k \int_{-\infty}^{\infty} dP_k \exp\left(-\frac{P_k^2}{2k_B T}\right) \exp\left(-\frac{w_k^2 Q_k^2}{2k_B T}\right) = \frac{k_B T}{\hbar w_k} \qquad (12)$$

being $\hbar = h/2\pi$ with h the Plank constant and T the desired temperature.

Then, the distribution function of displacements for each normal mode in thermal equilibrium $\rho_k(Q_k)$is calculated as

$$\rho_k(Q_k, P_k) = \frac{\exp\left(-\frac{E_k(Q_k, P_k)}{k_B} T\right)}{Z_k} \qquad (13)$$

using eq. (11) and integrating on P_k we obtain

$$\rho_k(Q_k) = \int_{-\infty}^{\infty} \hbar w_k \frac{\exp\left(-\frac{E_k(Q_k, P_k)}{k_B T}\right)}{k_B T} dP_k = \frac{\hbar w_k}{k_B T} \sqrt{\frac{\pi}{2k_B T}} \exp\left(-\frac{w_k^2 Q_k^2}{2k_B T}\right) \qquad (14)$$

for which we have used eq. (13).

D. Iterative procedure for conformational sampling

A flowchart of our iterative procedure is displayed in **Figure 1**. We start by performing a NMA analysis on the structure of the initial conformation, commonly chosen as the apo

conformation as it has been described in section II. B. Then, a set of one hundred of new structures are generated randomly in the direction of each of the *X* lowest frequency normal modes within the range $\left[-A_k : A_k\right]$ (see eq. 4) and the distribution function of displacements in thermal equilibrium $\rho_k\left(Q_k\right)$ (eq.14). In the present paper, we used *X*=10 except in cases that are explicitly indicated. Thereafter, two different strategies are followed to select one structure from the set as the new reference structure. We call them the unbiased and targeted versions of the procedure. In the unbiased version, the structure with the largest RMSD respect to the initial one is selected as the new reference structure to generate a new set of structures in the next iteration. In the targeted version, the structure with the smaller RMSD respect to the target structure, that is, the structure corresponding to the holo conformation is selected. As it has been pointed out previously (**section II-B**), NMA is performed considering as nodes only the atoms of protein backbone, C_β and the center of mass of side chains. Therefore, once the new structure has been chosen as the new reference structure, its all-atom representations is reconstructed using the SCWRL program[89] to take into account changes in side-chain conformations. After that, a new iteration is started by performing NMA on it.

III. RESULTS AND DISCUSSION

We have applied our procedure, either in its unbiased and targeted versions, on four proteins with different magnitudes of structural distortions upon ligand binding: Calmodulin (CaM), Adenylate Kinase (AK), Lysine-Arginine-Ornithine (LAO) binding protein and Maltodextrin binding protein (MBP). The Cα root-mean-square-fluctuations (RMSFs) obtained from the unbiased ensembles were compared with conformational variations obtained from experimental apo and holo conformations (**Figure 2**). The Pearson correlation coefficient ρ between them was 0.70, 0.86, 0.62, and 0.92 for CaM, AK, LAO and MBP respectively. These results validate our exploration of the conformational space between the apo and holo structures. The low values of ρ obtained for CaM and LAO indicate that the ten lowest frequency normal modes are not adequate to explore the fraction of the conformational space that connects both conformations. In order to improve the performance of the method on these two proteins, we have modified the criteria to select the modes to distort the reference structure. Instead of using the 10 lowest normal modes, we use the *P* modes with largest values $\left(c_k\right)^2$ of corresponding to the vector difference, obtained by superposing the reference structure and the holo structure and expressed in terms of normal modes of the reference structure (see eq. 5-7). The distribution of values of *P*, calculated at each iteration during the unbiased simulations of CaM and LAO, are shown in **Figure 3(a-b)**. We can observe that *P* ranges between ↔ 5 to 30 modes. That is, the apo-holo conformational change of CaM and LAO involves other modes than the 10 lowest frequency ones. The Pearson correlation coefficient ρ between the calculated and experimental RMSFs obtained from unbiased simulations using the *P* modes was 0.87, and 0.93 for CaM, and LAO respectively (**Figure 3c-d**). That is, a significant improve of the method is achieved by selecting the set of normal modes in a more specific way. Actually, normal modes change from one iteration to the other. This can be seen in **Figure 4** where the overlaps between normal modes of the initial apo structure and the corresponding ones obtained at each iteration step are displayed. We can see that most of normal modes lose their original identity in ↔ 15-20 iterations. The lowest normal modes seem to be more robust with respect to structural distortions introduced at each iteration step. This is consistent with the previously

reported robustness of the lowest and most collective motions with respect to mutational perturbations[90,91]. Particularly, the identity of the lowest-frequency MBP modes seems to persist during the whole unbiased simulations. MBP is the example with the least structural change between apo and holo structures (RMSD = 3.77 Å). Therefore, MBP does not experience significant structural distortions during our simulations and, therefore, the similarity of normal modes is expected to be more preserved than in the other examples.

A. Calmodulin

CaM belongs to the family of calcium binding proteins containing EF-hands or helix-loop-helix motifs. It participates in a large number of biological functions modulating the activities of other proteins, like protein kinases, NAD kinase, phospho-diesterase, and calcium pumps among others. The conformational change between apo (extended)[92] and holo (collapsed)[93] structures involves a large structural distortion (RMSD = 15.2 Å). **Figure 5** shows the results of our targeted simulations. The apo 1CLL crystallographic structure is chosen as the initial reference structure. After ~50 iterations, it was possible to achieve a random structure with RMSD = 1.6 Å respect to the holo (1CLD) structure (see **Table I**). The main structural change concern a hinge bending motion connecting the N- and C- lobes (**Figures 5 b** and **c** and **Figure S1**), leading to a change of the end-to-end linker distance (defined between Cαs of residues 69 and 91)[94] of 23.2 Å. This value is close to the one calculated at the holo structure (21.7 Å) and smaller than the one corresponding to the apo conformation (33.7 Å). In order to further validate our conformational exploration, the distance between the N- and C- lobes center of masses and the dihedral angle ϴ connecting them were also measured[94]. The values of N- and C- lobes center of mass distances, calculated on apo- and holo- structures, are 41.9 Å and 23.8 Å respectively. The corresponding value, calculated on the structure that best match with the holo-structure was 24.5 Å. Besides, the dihedral angle ϴ was -81°, 121°, and 116° calculated on apo, holo, and the structure that best match with the holo-structure respectively. This value were in excellent agreement with previous experimental and theoretical (molecular dynamics) results[94].

As it has been shown in **Figure 3(a)**, the conformational change is not limited to the participation of the ten lowest frequency normal modes. Actually few middle-frequency modes associated to local rearrangements seem also to be involved[61]. Therefore, an unbiased sampling based on a conformational space exploration using a few lowest normal modes is not expected to be able to efficiently reproduce it[61]. Our unbiased simulations were only able to achieve a structure with of 5.1 Å respect to the holo (1CLD) structure (see **Table I**). In order to improve this performance, we perform fifteen unbiased simulations using the *P* modes with largest values of $(c_k)^2$ associated to the vector difference between apo and holo structures (see eq. 5-7). These simulations were performed staring from the same apo structure but using different random-seed. The results are shown in **Figure 6**. While some simulations explore a conformational space far away from the region of the holo structure, others manage to reach it. The best match between a random generated structure and holo structure was 2.8 Å, confirming that an adequate selection of the set of modes to be used to explore the conformational space can significantly improve the performance of the method. The end-to-end linker distance results 23.8 Å, very close to the value of 23.2 Å obtained in the targeted simulations. The

corresponding values of distance between the N- and C- lobes center of masses and the dihedral angle ❶ were 27.3 Å and 131° respectively.

In order to further validate our findings, we have compared the structures obtained during unbiased simulations with the 160 models determined by NMR for apo CaM[95]. Values of RMSD respect to apo and holo structures for these NMR models are pointed out in **Figure 6(a).** Most of NMR models are within the subspace of structures generated during our set of unbiased simulations, confirming the reliability of our method to generate conformational ensembles.

At this point it is interesting to stress that our results represent a relative improvement respect to previous reported results obtained by unbiased CaM conformational space explorations[61]. CaM conformational changes are subjected to local rearrangements with significant impacts on the N- and C- lobes exposure surfaces[96]. These local rearrangements cannot be taken into account by low-frequency normal modes[61,97]. Besides, the apo-to-holo conformational change of CaM involves a substantial change in its residue interaction network. Nevertheless, several features of our unbiased procedure for conformational space exploration contribute to overcome these issue. Firstly, structural distortions introduced at each iteration step are within the limits provided by the calculated thermal fluctuation amplitudes of each normal mode (see **Section II.C**). This significantly reduces the extent of deformations increasing the reliability of the resulted structures and allowing the incorporation of structural changes in a smoother manner than other methods. This can be seen in **Figure S2** where statistical parameters concerning the set of new structures generated randomly at each iteration step, are displayed. Maximum values of RMSD are typically lower than 0.7 Å and standard deviations below 0.1 Å. Structural distortions seems to depend on the extent of apo-to-holo conformational change and tend to decrease with the # of iteration steps. Secondly, middle-high frequency normal modes present smaller amplitudes than low-frequency ones, as it is indicated in equation 4. Therefore the reliability of the new structures is not restricted to the use of a reduced number of low-frequency modes and the method is flexible to consider any previously selected combination of low-middle-high frequency normal modes. Besides, required changes in side-chain orientations, that should involve middle-high frequency modes, are optimized at each iteration step (see **Section II.D**). Finally, our NMA is performed according to interaction contacts defined at each iteration step using re-evaluated residue interaction networks (see **Section II.B**).

B. Adenylate Kinase

AK is a phosphotransferase that contributes to maintain the ATP/ADP balance in cells by catalyzing their interconversion. It represents example of a large functional conformational transition[98] involving a hinge motion connecting the open (apo) and closed (holo) conformations. The conformational change implies a RMSD = 6.9 Å between both conformations. The hinge motion induces displacements of the so-called LID and NMP domains relative to a third domain (CORE) to exclude water from the active site during the catalysis[99]. The conformational selection model has been previously confirm for AK, either experimentally[14,100] and theorietically[101]. This has been also confirm by our targeted (**Figure 7**) and unbiased (**Figure 8**) simulations, from which we managed to reach structures with RMSD equal to 1.1 Å and 2.4 Å respectively (see **Table I**). The structural changes observed in the apo-to-holo path are shown in **Figure S3**. The main conformational change involves the loop of the LID domain, particularly the fragments R124-V132 and P140-R156. Our final structures present a relative displacement of the α-helix composed by residues K47-A55 respect to the

holo structure. Therefore, this final rearrangement seems to involve other modes than the ten lowest-frequency ones.

The relative positions of LID, CORE, and NMP domains can be described by the LID-CORE and NMP-CORE angles[102]. While the initial apo structure presents values of 71° and 148° for LID-CORE and NMP-CORE angles respectively, the values at holo structure are 16° and 25° respectively. These values are well reproduced in our two types of simulations: 16° and 25°, calculated on the structure that best fit with the holo structure for targeted simulations, and 20° and 27° for the unbiased simulations respectively.

Targeted simulations on AK (**Figure 7**) reach the holo structure in less iterations than CaM (**Figure 5**). This is expected due to the smaller apo-to-holo conformational change involved. In the case of the unbiased simulations (**Figure 8**), we observe that some of them explore regions of the conformational space situated far from those corresponding to the apo and holo structures. These structures reach LID-CORE and NMP-CORE angles of ⇽80° and 210° respectively. On the contrary, a few simulations seem to experience iterative structural distortions that drive the protein, in a sequential way, to the holo conformation. Since all unbiased simulations started from the same initial apo structure, it seems that directions selected during the first iterations have a significant influence on the structural distortions introduced in the subsequent iterations. That is, explorations of the conformational space perform on each simulation are not exhaustive and they seem not to be reversible in all cases. In this way several simulations starting with different random seeds are recommended.

C. Lysine-Arginine-Ornithine (LAO) binding protein

LAO is a periplasmic substrate-binding protein (permease) that is part of the bacterial periplasmic transport system in charge of the transport of different kind of substrates like amino acids, peptides, sugars, vitamins, and inorganic ions[103,104]. It has a bi-lobal structure with domains 1 and 2 that are connected by two connecting segments. The structure presents a deep cleft between both domains. Each domain is composed of five β-strands and four α-helices. It undergoes a conformational change which is presumably an integral aspect of its mechanism. Our targeted simulations were able to reproduce this conformational change (see **Figure S4**), reaching a random structure with RMSD = 0.6 Å respect to the holo conformation (see **Table I**).

As it was the case with CaM, our unbiased simulations were not able to explore the complete apo-to-holo conformational change and the best match of a random structure with the holo conformation was RMDS=2.4Å. This was also seen in previous conformational space explorations performed by Gohlke et al.[61]. Therefore, we have extended the vibrational space to consider the *P* modes involved in the vector difference between apo and holo structures (eq. 5-7). In this way, we were able to achieve a RMSD = 0.7 respect to the holo structure. **Figure 9** shows the result of these simulations. As in the case of AK, we can observe that a few simulations follow conformational pathways that move the structure away from the apo and holo forms. As it was the case for AK, this behaviour can be related to the fact that the apo-to-holo conformational change involves global and hinge-bending motions of domains 1 and 2[105].

D. Maltodextrin binding protein (MBP)

MBP is a periplasmic-binding protein that belongs to the maltose transport system, an ATP-binding cassette (ABC) transporter that accumulates linear maltooligosaccharides and

results essential for the survival of Gram-negative bacterial cells. Its structure consists of two globular domains with Rossman fold or α/β motif. These two domains are separated by a ligand-binding site groove and the ligand-binding conformational change consists on a hinge-bending between these domains[106–108].

MBP has been chosen as an prototype example for which previous pioneers studies have shown that the collective lowest-frequency modes closely map the apo form onto the holo form[48]. Actually, an overlap >0.8 between one collective mode and the vector difference between apo and holo forms have been reported[109]. Therefore, it is a suitable example to test our procedure in all its versions. As it was expected, either targeted and unbiased simulations have allowed to achieve the apo to holo conformational change obtaining random structures with RMSD equal to 0.9 Å and 1.2 Å respectively (see **Table I**). Results concerning the unbiased simulations are shown in **Figure S5**.

E. Other selections of the reference structure

At this point it is important to mention that other options to select the reference structure at each iteration step for the unbiased simulations have been tested. Particularly, we have tested two other options instead of the structure with the largest RMSD respect to the initial one. On one hand, we have performed simulations considering selecting the structure with the largest RMSD respect to the structure used to introduce the structural distortions at each iteration step (see **Section II. D**) at room temperature. On the other hand, we have selected the structure with the smaller RMSD respect to it. In none of these variants we manage to explore the conformational space that connects apo and holo structures. Besides, tests considering the holo structure as the initial one have also been performed. As it has been observed in previous works[109], the compactness of holo conformations makes its vibrations less suitable to be reproduced by low-frequency normal modes using ENM. Our simulations using the holo structure as the initial one were not able to explore the conformational space involving the apo structure.

IV. CONCLUSIONS

We have presented a new iterative procedure to explore conformational spaces of proteins based on their equilibrium dynamics at room temperature. It makes use of vibrations obtained by normal mode analysis according to protein elastic network models. In that sense, the method does not use further ad hoc prescriptions and structural distortions are introduced randomly within the range of protein thermal fluctuations.

The method has been tested on four proteins with different magnitudes of structural distortions upon ligand binding have been tested. Two different strategies have been tested. On one hand, the targeted strategy allows to find a conformational path that connects an initial chosen structure (e.g. the apo state) and a final target one (e.g. the holo state).On the other hand, the unbiased strategy explores the conformational space available for the initial chosen structure according to its thermal fluctuations, without any further guided direction. We found that the method results sufficiently efficient using both strategies. Furthermore, the two reported strategies can be interpreted as limited cases of a large variety of situations in which only partial targeted structural information is provided to guide the conformational sampling exploration, e.g., certain specific final contacts or relative orientations between residues.

The performance of the method is subjected to the efficiency with which the selected set of normal modes introduces structural distortions that allow a broad exploration of the conformational space. We think that the simplicity of the method, together with its flexibility in terms of either the number and type of normal modes selected, and the information provided as a guide for conformational space exploration, makes it a suitable way to explore the conformational multiplicity of proteins. Furthermore, the method introduces structural distortions according to protein thermal fluctuations while it optimizes side-chain orientations. Because of that, it can result useful in providing the starting points for more detailed atomistic or coarse grained MD simulations without subjecting the molecular system to sudden, relatively localized structural stresses that can lead to subsequent unstable MD simulations. Besides, the method results efficient enough to be applied in biophysics and bioinformatics approaches that require fast generation of ensembles of protein conformations, i.e., flexible docking and drug design approaches, and different kind of comparative studies of the dynamism among homologous proteins.

SUPPORTING INFORMATION AVAILABLE

Figure S1. Superimposed putty-style cartoon representation of the holo structure with the random structure that best match with at different steps during the targeted simulation of CaM. The variable-width and colour (from blue red in increasing order of RMSD) correspond to values of RMSD per site.

Figure S2. Maximum values of RMSD (a, b) and standard deviations (c, d) for the random structures respect to the corresponding reference structure obtained at each iteration step during (a, c) targeted and (b, d) unbiased simulations.

Figure S3. Superimposed putty-style cartoon representation of the holo structure with the random structure that best match with at different steps during the targeted simulation of AK. The variable-width and colour (from blue red in increasing order of RMSD) correspond to values of RMSD per site.

Figure S4. (a) Values of RMSD for the random structures respect to apo and holo structures during the targeted simulation of LAO. Iterations are identified with different colours; (b) superposition of apo (green) and holo (blue) structures of CaM (RMSD = 4.7 Å); (c) superposition of holo (blue) structure with the random structure that best match with it (i.e. lower RMSD = 0.6 Å) (cyan).

Figure S5. (a) Values of RMSD for the random structures respect to apo and holo structures during the unbiased simulation of MBP. Different simulations are identified with different colours; superimposed putty-style cartoon representation of the (b) apo and (c) holo structures with the random structure that best match with the holo structure. The variable-width and colour (from blue red in increasing order of RMSD) correspond to values of RMSD per site.

ACKNOWLEDGEMENTS

GP and SFA are researchers of CONICET and TS is a Postdoctoral fellow of CONICET. This work was supported by Universidad Nacional de Quilmes (PUNQ 1004/11) and from the European Union's Horizon 2020 research and innovation program under grant agreement No 778247.

FIGURE CAPTIONS

Figure 1. Workflow for exploration of protein conformational space.

Figure 2. Comparison of Cα root-mean-square-fluctuations (RMSFs) obtained from the unbiased ensembles (red) and obtained from experimental apo and holo conformations (blue) for (a) CaM, (b) AK, (c) LAO and (d) MBP.

Figure 3. Distribution of values of *P* throughout the unbiased simulations for (a) CaM, and (b) LAO. Comparison of Cα root-mean-square-fluctuations (RMSFs) obtained from the unbiased ensembles (red) and obtained from experimental apo and holo conformations (blue) for (c) CaM, and (d) LAO using the *P* modes that overlap the most with the vector difference that connect apo and holo structures.

Figure 4. Overlaps between normal modes of the initial apo structure and the corresponding ones obtained at each iteration step during targeted simulations for (a) CaM, (b) AK, (c) LAO and (d) MBP.

Figure 5. (a) Values of RMSD for the random structures respect to apo and holo structures during the targeted simulation of CaM. Iterations are identified with different colours; (b) superposition of apo (green) and holo (blue) structures of CaM (RMSD = 15.2 Å); (c) superposition of holo (blue) structure with the random structure that best match with it (i.e. lower RMSD = 1.6 Å) (cyan).

Figure 6. (a) Values of RMSD for the random structures respect to apo and holo structures during the unbiased simulation of CaM. Different simulations are identified with different colours; values of RMSD respect to apo and holo structures for NMR models (obtained from the PDB file 2K0E) are indicated as black dots. Superimposed putty-style cartoon representation of the (b) apo and (c) holo structures with the random structure that best match with the holo structure. The variable-width and colour (from blue red in increasing order of RMSD) correspond to values of RMSD per site.

Figure 7. (a) Values of RMSD for the random structures respect to apo and holo structures during the targeted simulation of AK. Iterations are identified with different colours; (b) superposition of apo (green) and holo (blue) structures of AKM (RMSD = 6.9 Å); (c) superposition of holo (blue) structure with the random structure that best match with it (i.e. lower RMSD = 2.4 Å) (orange).

Figure 8. (a) Values of RMSD for the random structures respect to apo and holo structures during the unbiased simulation of AK. Different simulations are identified with different colours; superimposed putty-style cartoon representation of the (b) apo and (c) holo structures

with the random structure that best match with the holo structure. The variable-width and colour (from blue red in increasing order of RMSD) correspond to values of RMSD per site.

Figure 9. (a) Values of RMSD for the random structures respect to apo and holo structures during the unbiased simulation of LAO. Different simulations are identified with different colours; superimposed putty-style cartoon representation of the (b) apo and (c) holo structures with the random structure that best match with the holo structure. The variable-width and colour (from blue red in increasing order of RMSD) correspond to values of RMSD per site.

BIBLIOGRAPHY

(1) Gunasekaran, K.; Ma, B.; Nussinov, R. Is Allostery an Intrinsic Property of All Dynamic Proteins? *Proteins Struct. Funct. Bioinforma.* **2004**, *57* (3), 433–443. https://doi.org/10.1002/prot.20232.

(2) Gora, A.; Brezovsky, J.; Damborsky, J. Gates of Enzymes. *Chem. Rev.* **2013**, *113* (8), 5871. https://doi.org/10.1021/CR300384W.

(3) Tobi, D.; Bahar, I. Structural Changes Involved in Protein Binding Correlate with Intrinsic Motions of Proteins in the Unbound State. *Proc. Natl. Acad. Sci.* **2005**, *102* (52), 18908–18913. https://doi.org/10.1073/pnas.0507603102.

(4) Goh, C.-S.; Milburn, D.; Gerstein, M. Conformational Changes Associated with Protein–Protein Interactions. *Curr. Opin. Struct. Biol.* **2004**, *14* (1), 104–109. https://doi.org/10.1016/j.sbi.2004.01.005.

(5) Zhuravlev, P. I.; Papoian, G. A. Protein Functional Landscapes, Dynamics, Allostery: A Tortuous Path towards a Universal Theoretical Framework. *Q. Rev. Biophys.* **2010**, *43* (3), 295–332. https://doi.org/10.1017/S0033583510000119.

(6) Ruvinsky, A. M.; Kirys, T.; Tuzikov, A. V.; Vakser, I. A. Ensemble-Based Characterization of Unbound and Bound States on Protein Energy Landscape. *Protein Sci.* **2013**, *22* (6), 734–744. https://doi.org/10.1002/pro.2256.

(7) Kumar, S.; Ma, B.; Tsai, C.-J.; Sinha, N.; Nussinov, R. Folding and Binding Cascades: Dynamic Landscapes and Population Shifts. *Protein Sci.* **2008**, *9* (1), 10–19. https://doi.org/10.1110/ps.9.1.10.

(8) Csermely, P.; Palotai, R.; Nussinov, R. Induced Fit, Conformational Selection and Independent Dynamic Segments: An Extended View of Binding Events. *Trends Biochem. Sci.* **2010**, *35* (10), 539–546. https://doi.org/10.1016/j.tibs.2010.04.009.

(9) Boehr, D. D.; Nussinov, R.; Wright, P. E. The Role of Dynamic Conformational Ensembles in Biomolecular Recognition. *Nat. Chem. Biol.* **2009**, *5* (11), 789–796. https://doi.org/10.1038/nchembio.232.

(10) Okazaki, K.; Takada, S. Dynamic Energy Landscape View of Coupled Binding and Protein Conformational Change: Induced-Fit versus Population-Shift Mechanisms. *Proc. Natl. Acad. Sci.* **2008**, *105* (32), 11182–11187. https://doi.org/10.1073/PNAS.0802524105.

(11) Monod, J.; Wyman, J.; Changeux, J.-P. On the Nature of Allosteric Transitions: A Plausible Model. *J. Mol. Biol.* **1965**, *12* (1), 88–118. https://doi.org/10.1016/S0022-2836(65)80285-6.

(12) James, L. C.; Tawfik, D. S. Conformational Diversity and Protein Evolution – a 60-Year-Old Hypothesis Revisited. *Trends Biochem. Sci.* **2003**, *28* (7), 361–368. https://doi.org/10.1016/S0968-0004(03)00135-X.

(13) Changeux, J.-P. Allostery and the Monod-Wyman-Changeux Model After 50 Years. *Annu. Rev. Biophys.* **2012**, *41* (1), 103–133. https://doi.org/10.1146/annurev-biophys-050511-102222.

(14) 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.; et al. Intrinsic Motions along an Enzymatic Reaction Trajectory. *Nature* **2007**, *450* (7171), 838–844. https://doi.org/10.1038/nature06410.

(15) Nussinov, R.; Ma, B. Protein Dynamics and Conformational Selection in Bidirectional Signal Transduction. *BMC Biol.* **2012**, *10* (1), 2. https://doi.org/10.1186/1741-7007-10-2.

(16) Chen, R.; Mintseris, J.; Janin, J.; Weng, Z. A Protein-Protein Docking Benchmark. *Proteins Struct. Funct. Genet.* **2003**, *52* (1), 88–91. https://doi.org/10.1002/prot.10390.

(17) KHERSONSKY, O.; ROODVELDT, C.; TAWFIK, D. Enzyme Promiscuity: Evolutionary and Mechanistic Aspects. *Curr. Opin. Chem. Biol.* **2006**, *10* (5), 498–508. https://doi.org/10.1016/j.cbpa.2006.08.011.

(18) Kuzu, G.; Gursoy, A.; Nussinov, R.; Keskin, O. Exploiting Conformational Ensembles in Modeling Protein–Protein Interactions on the Proteome Scale. *J. Proteome Res.* **2013**, *12* (6), 2641–2653. https://doi.org/10.1021/pr400006k.

(19) Osguthorpe, D. J.; Sherman, W.; Hagler, A. T. Generation of Receptor Structural Ensembles for Virtual Screening Using Binding Site Shape Analysis and Clustering. *Chem. Biol. Drug Des.* **2012**, *80* (2), 182–193. https://doi.org/10.1111/j.1747-0285.2012.01396.x.

(20) Palopoli, N.; Lanzarotti, E.; Parisi, G. BeEP Server: Using Evolutionary Information for Quality Assessment of Protein Structure Models. *Nucleic Acids Res.* **2013**, *41* (W1), W398–W405. https://doi.org/10.1093/nar/gkt453.

(21) Juritz, E.; Palopoli, N.; Fornasari, M. S.; Fernandez-Alberti, S.; Parisi, G. Protein Conformational Diversity Modulates Sequence Divergence. *Mol. Biol. Evol.* **2013**, *30* (1), 79–87. https://doi.org/10.1093/molbev/mss080.

(22) Parisi, G.; Zea, D. J.; Monzon, A. M.; Marino-Buslje, C. Conformational Diversity and the Emergence of Sequence Signatures during Evolution. *Curr. Opin. Struct. Biol.* **2015**, *32*, 58–65. https://doi.org/10.1016/j.sbi.2015.02.005.

(23) Monzon, A. M.; Rohr, C. O.; Fornasari, M. S.; Parisi, G. CoDNaS 2.0: A Comprehensive Database of Protein Conformational Diversity in the Native State. *Database* **2016**, *2016*, 1–8. https://doi.org/10.1093/database/baw038.

(24) Hrabe, T.; Li, Z.; Sedova, M.; Rotkiewicz, P.; Jaroszewski, L.; Godzik, A. PDBFlex: Exploring Flexibility in Protein Structures. *Nucleic Acids Res.* **2016**, *44* (D1), D423–D428. https://doi.org/10.1093/nar/gkv1316.

(25) Gunasekaran, K.; Nussinov, R. How Different Are Structurally Flexible and Rigid Binding Sites? Sequence and Structural Features Discriminating Proteins That Do and Do Not Undergo Conformational Change upon Ligand Binding. *J. Mol. Biol.* **2007**, *365* (1), 257–273. https://doi.org/10.1016/j.jmb.2006.09.062.

(26) Gutteridge, A.; Thornton, J. Conformational Changes Observed in Enzyme Crystal Structures upon Substrate Binding. *J. Mol. Biol.* **2005**, *346* (1), 21–28. https://doi.org/10.1016/j.jmb.2004.11.013.

(27) Lange, O. F.; Lakomek, N.-A.; Fares, C.; Schroder, G. F.; Walter, K. F. A.; Becker, S.; Meiler, J.; Grubmuller, H.; Griesinger, C.; de Groot, B. L. Recognition Dynamics Up to Microseconds Revealed from an RDC-Derived Ubiquitin Ensemble in Solution. *Science (80-.).* **2008**, *320* (5882), 1471–1475. https://doi.org/10.1126/science.1157092.

(28) Wu, Z.; Elgart, V.; Qian, H.; Xing, J. Amplification and Detection of Single-Molecule Conformational Fluctuation through a Protein Interaction Network with Bimodal Distributions. *J. Phys. Chem. B* **2009**, *113* (36), 12375–12381. https://doi.org/10.1021/jp903548d.

(29) Weikl, T. R.; von Deuster, C. Selected-Fit versus Induced-Fit Protein Binding: Kinetic Differences and Mutational Analysis. *Proteins Struct. Funct. Bioinforma.* **2009**, *75* (1), 104–110. https://doi.org/10.1002/prot.22223.

(30) Tzeng, S.-R.; Kalodimos, C. G. Protein Dynamics and Allostery: An NMR View. *Curr. Opin. Struct. Biol.* **2011**, *21* (1), 62–67. https://doi.org/10.1016/j.sbi.2010.10.007.

(31) Orellana, L. Large-Scale Conformational Changes and Protein Function: Breaking the in Silico Barrier. *Front. Mol. Biosci.* **2019**, *6* (November). https://doi.org/10.3389/fmolb.2019.00117.

(32) Bussi, G.; Laio, A.; Parrinello, M. Equilibrium Free Energies from Nonequilibrium Metadynamics. *Phys. Rev. Lett.* **2006**, *96* (9), 090601. https://doi.org/10.1103/PhysRevLett.96.090601.

(33) Zhou, R. Replica Exchange Molecular Dynamics Method for Protein Folding Simulation. In *Protein Folding Protocols*; Humana Press: New Jersey, 2007; Vol. 350, pp 205–224. https://doi.org/10.1385/1-59745-189-4:205.

(34) Sugita, Y.; Okamoto, Y. Replica-Exchange Molecular Dynamics Method for Protein Folding. *Chem. Phys. Lett.* **1999**, *314* (1–2), 141–151. https://doi.org/10.1016/S0009-2614(99)01123-9.

(35) Wu, X. and; Wang*, S. Self-Guided Molecular Dynamics Simulation for Efficient Conformational Search. **1998**. https://doi.org/10.1021/JP9817372.

(36) 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. https://doi.org/10.1063/1.1861885.

(37) Schlitter, J.; Engels, M.; Krüger, P. Targeted Molecular Dynamics: A New Approach for Searching Pathways of Conformational Transitions. *J. Mol. Graph.* **1994**, *12* (2), 84–89.

(38) Zhang, Z. Systematic Methods for Defining Coarse-Grained Maps in Large Biomolecules. In *Advances in experimental medicine and biology*; 2015; Vol. 827, pp 33–48. https://doi.org/10.1007/978-94-017-9245-5_4.

(39) Voth, G. A. *Coarse-Graining of Condensed Phase and Biomolecular Systems*; CRC Press, 2009.

(40) Sieradzan, A. K.; Niadzvedtski, A.; Scheraga, H. A.; Liwo, A. Revised Backbone-Virtual-Bond-Angle Potentials to Treat the l - and d -Amino Acid Residues in the Coarse-Grained United Residue (UNRES) Force Field. *J. Chem. Theory Comput.* **2014**, *10* (5), 2194–2203. https://doi.org/10.1021/ct500119r.

(41) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *J. Phys. Chem. B* **2007**, *111* (27), 7812–7824. https://doi.org/10.1021/jp071097f.

(42) Liwo, A.; He, Y.; Scheraga, H. A. Coarse-Grained Force Field: General Folding Theory. *Phys. Chem. Chem. Phys.* **2011**, *13* (38), 16890–16901. https://doi.org/10.1039/c1cp20752k.

(43) Darré, L.; Machado, M. R.; Brandner, A. F.; González, H. C.; Ferreira, S.; Pantano, S. SIRAH: A Structurally Unbiased Coarse-Grained Force Field for Proteins with Aqueous

Solvation and Long-Range Electrostatics. *J. Chem. Theory Comput.* **2015**, *11* (2), 723–739. https://doi.org/10.1021/ct5007746.

(44) Zheng, W.; Wen, H. A Survey of Coarse-Grained Methods for Modeling Protein Conformational Transitions. *Curr. Opin. Struct. Biol.* **2017**, *42*, 24–30. https://doi.org/10.1016/J.SBI.2016.10.008.

(45) Kmiecik, S.; Gront, D.; Kolinski, M.; Wieteska, L.; Dawid, A. E.; Kolinski, A. Coarse-Grained Protein Models and Their Applications. *Chemical Reviews*. American Chemical Society July 27, 2016, pp 7898–7936. https://doi.org/10.1021/acs.chemrev.6b00163.

(46) Kmiecik, S.; Kouza, M.; Badaczewska-Dawid, A.; Kloczkowski, A.; Kolinski, A. Modeling of Protein Structural Flexibility and Large-Scale Dynamics: Coarse-Grained Simulations and Elastic Network Models. *Int. J. Mol. Sci.* **2018**, *19* (11), 3496. https://doi.org/10.3390/ijms19113496.

(47) Dawid, A. E.; Gront, D.; Kolinski, A. SURPASS Low-Resolution Coarse-Grained Protein Modeling. *J. Chem. Theory Comput.* **2017**, *13* (11), 5766–5779. https://doi.org/10.1021/acs.jctc.7b00642.

(48) Tirion, M. M. Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis. *Phys. Rev. Lett.* **1996**, *77* (9), 1905–1908. https://doi.org/10.1103/PhysRevLett.77.1905.

(49) Emperador, A.; Carrillo, O.; Rueda, M.; Orozco, M. Exploring the Suitability of Coarse-Grained Techniques for the Representation of Protein Dynamics. *Biophys. J.* **2008**, *95* (5), 2127–2138. https://doi.org/10.1529/BIOPHYSJ.107.119115.

(50) Kirillova, S.; Cortés, J.; Stefaniu, A.; Siméon, T. An NMA-Guided Path Planning Approach for Computing Large-Amplitude Conformational Changes in Proteins. *Proteins Struct. Funct. Bioinforma.* **2007**, *70* (1), 131–143. https://doi.org/10.1002/prot.21570.

(51) Schuyler, A. D.; Jernigan, R. L.; Qasba, P. K.; Ramakrishnan, B.; Chirikjian, G. S. Iterative Cluster-NMA: A Tool for Generating Conformational Transitions in Proteins. *Proteins* **2009**, *74* (3), 760–776. https://doi.org/10.1002/prot.22200.

(52) Saldaño, T. E.; Monzon, A. M.; Parisi, G.; Fernandez-Alberti, S. Evolutionary Conserved Positions Define Protein Conformational Diversity. *PLoS Comput. Biol.* **2016**, *12* (3), e1004775. https://doi.org/10.1371/journal.pcbi.1004775.

(53) Mahajan, S.; Sanejouand, Y. H. Jumping between Protein Conformers Using Normal Modes. *J. Comput. Chem.* **2017**, *38* (18), 1622–1630. https://doi.org/10.1002/jcc.24803.

(54) Tama, F.; Sanejouand, Y.-H. Conformational Change of Proteins Arising from Normal Mode Calculations. *Protein Eng. Des. Sel.* **2001**, *14* (1), 1–6. https://doi.org/10.1093/protein/14.1.1.

(55) Uyar, A.; Kantarci-Carsibasi, N.; Haliloglu, T.; Doruker, P. Features of Large Hinge-Bending Conformational Transitions. Prediction of Closed Structure from Open State. *Biophys. J.* **2014**, *106* (12), 2656–2666. https://doi.org/10.1016/j.bpj.2014.05.017.

(56) Gur, M.; Madura, J. D.; Bahar, I. Global Transitions of Proteins Explored by a Multiscale Hybrid Methodology: Application to Adenylate Kinase. *Biophys. J.* **2013**, *105* (7), 1643–1652. https://doi.org/10.1016/j.bpj.2013.07.058.

(57) Das, A.; Gur, M.; Cheng, M. H.; Jo, S.; Bahar, I.; Roux, B. Exploring the Conformational Transitions of Biomolecular Systems Using a Simple Two-State Anisotropic Network Model. *PLoS Comput. Biol.* **2014**, *10* (4), e1003521. https://doi.org/10.1371/journal.pcbi.1003521.

(58) Seeliger, D.; de Groot, B. L. Conformational Transitions upon Ligand Binding: Holo-Structure Prediction from Apo Conformations. *PLoS Comput. Biol.* **2010**, *6* (1), e1000634. https://doi.org/10.1371/journal.pcbi.1000634.

(59) Seo, S.; Jang, Y.; Qian, P.; Liu, W. K.; Choi, J.-B.; Lim, B. S.; Kim, M. K. Efficient Prediction of Protein Conformational Pathways Based on the Hybrid Elastic Network Model. *J. Mol. Graph. Model.* **2014**, *47*, 25–36. https://doi.org/10.1016/j.jmgm.2013.10.009.

(60) Kurkcuoglu, Z.; Bahar, I.; Doruker, P. ClustENM: ENM-Based Sampling of Essential Conformational Space at Full Atomic Resolution. *J. Chem. Theory Comput.* **2016**, *12* (9), 4549–4562. https://doi.org/10.1021/acs.jctc.6b00319.

(61) Ahmed, A.; Rippmann, F.; Barnickel, G.; Gohlke, H. A Normal Mode-Based Geometric Simulation Approach for Exploring Biologically Relevant Conformational Transitions in Proteins. *J. Chem. Inf. Model.* **2011**, *51* (7), 1604–1622. https://doi.org/10.1021/ci100461k.

(62) 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. https://doi.org/10.1016/S0006-3495(02)73931-3.

(63) Zheng, W.; Brooks, B. R. Modeling Protein Conformational Changes by Iterative Fitting of Distance Constraints Using Reoriented Normal Modes. *Biophys. J.* **2006**, *90* (12), 4327–4336. https://doi.org/10.1529/biophysj.105.076836.

(64) Zheng, W.; Brooks, B. R. Normal-Modes-Based Prediction of Protein Conformational Changes Guided by Distance Constraints. *Biophys. J.* **2005**, *88* (5), 3109–3117. https://doi.org/10.1529/biophysj.104.058453.

(65) 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. https://doi.org/10.1016/j.jmb.2005.07.031.

(66) 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. https://doi.org/10.1529/biophysj.107.128447.

(67) Franklin, J.; Koehl, P.; Doniach, S.; Delarue, M. MinActionPath: Maximum Likelihood Trajectory for Large-Scale Structural Transitions in a Coarse-Grained Locally Harmonic Energy Landscape. *Nucleic Acids Res.* **2007**, *35* (suppl_2), W477–W482. https://doi.org/10.1093/nar/gkm342.

(68) Zheng, W.; Brooks, B. R.; Hummer, G. Protein Conformational Transitions Explored by Mixed Elastic Network Models. *Proteins Struct. Funct. Genet.* **2007**, *69* (1), 43–57. https://doi.org/10.1002/prot.21465.

(69) Tekpinar, M.; Zheng, W. Predicting Order of Conformational Changes during Protein Conformational Transitions Using an Interpolated Elastic Network Model. *Proteins Struct. Funct. Bioinforma.* **2010**, *78* (11), 2469–2481. https://doi.org/10.1002/prot.22755.

(70) Yang, Z.; Májek, P.; Bahar, I. Allosteric Transitions of Supramolecular Systems Explored by Network Models: Application to Chaperonin GroEL. *PLoS Comput. Biol.* **2009**, *5* (4). https://doi.org/10.1371/journal.pcbi.1000360.

(71) Weiss, D. R.; Levitt, M. Can Morphing Methods Predict Intermediate Structures? *J. Mol. Biol.* **2009**, *385* (2), 665–674. https://doi.org/10.1016/j.jmb.2008.10.064.

(72) Krüger, D. M.; Ahmed, A.; Gohlke, H. NMSim Web Server: Integrated Approach for Normal Mode-Based Geometric Simulations of Biologically Relevant Conformational Transitions in Proteins. *Nucleic Acids Res.* **2012**, *40* (W1). https://doi.org/10.1093/nar/gks478.

(73) Sfriso, P.; Emperador, A.; Orellana, L.; Hospital, A.; Gelpí, J. L.; Orozco, M. Finding Conformational Transition Pathways from Discrete Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2012**, *8* (11), 4707–4718. https://doi.org/10.1021/ct300494q.

(74) Sfriso, P.; Hospital, A.; Emperador, A.; Orozco, M. Exploration of Conformational Transition Pathways from Coarse-Grained Simulations. *Bioinformatics* **2013**, *29* (16), 1980–1986. https://doi.org/10.1093/bioinformatics/btt324.

(75) Das, A.; Gur, M.; Cheng, M. H.; Jo, S.; Bahar, I.; Roux, B. Exploring the Conformational Transitions of Biomolecular Systems Using a Simple Two-State Anisotropic Network Model. *PLoS Comput. Biol.* **2014**, *10* (4). https://doi.org/10.1371/journal.pcbi.1003521.

(76) López-Blanco, J. R.; Aliaga, J. I.; Quintana-Ortí, E. S.; Chacón, P. IMODS: Internal Coordinates Normal Mode Analysis Server. *Nucleic Acids Res.* **2014**, *42* (W1). https://doi.org/10.1093/nar/gku339.

(77) Monzon, A. M.; Zea, D. J.; Marino-Buslje, C.; Parisi, G. Homology Modeling in a Dynamical World. *Protein Sci.* **2017**, *26* (11), 2195–2206. https://doi.org/10.1002/pro.3274.

(78) Juritz, E.; Fornasari, M.; Martelli, P.; Fariselli, P.; Casadio, R.; Parisi, G. On the Effect of Protein Conformation Diversity in Discriminating among Neutral and Disease Related Single Amino Acid Substitutions. *BMC Genomics* **2012**, *13* (Suppl 4), S5. https://doi.org/10.1186/1471-2164-13-S4-S5.

(79) Parisi, G.; Echave, J. Structural Constraints and Emergence of Sequence Patterns in Protein Evolution. *Mol. Biol. Evol.* **2001**, *18* (5), 750–756. https://doi.org/10.1093/oxfordjournals.molbev.a003857.

(80) Wei, G.; Xi, W.; Nussinov, R.; Ma, B. Protein Ensembles: How Does Nature Harness Thermodynamic Fluctuations for Life? The Diverse Functional Roles of Conformational Ensembles in the Cell. *Chemical Reviews*. American Chemical Society June 8, 2016, pp 6516–6551. https://doi.org/10.1021/acs.chemrev.5b00562.

(81) Bahar, I.; Erman, B.; Jernigan, R. L.; Atilgan, A. R.; Covell, D. G. Collective Motions in HIV-1 Reverse Transcriptase: Examination of Flexibility and Enzyme Function. *J. Mol. Biol.* **1999**, *285* (3), 1023–1037. https://doi.org/10.1006/jmbi.1998.2371.

(82) Hinsen, K. Analysis of Domain Motions by Approximate Normal Mode Calculations. *Proteins Struct. Funct. Genet.* **1998**, *33* (3), 417–429. https://doi.org/10.1002/(SICI)1097-0134(19981115)33:3<417::AID-PROT10>3.0.CO;2-8.

(83) 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. https://doi.org/10.1016/S0006-3495(01)76033-X.

(84) Jeong, J. I.; Jang, Y.; Kim, M. K. A Connection Rule for -Carbon Coarse-Grained Elastic Network Models Using Chemical Bond Information. *J. Mol. Graph. Model.* **2006**, *24* (4), 296–306. https://doi.org/10.1016/j.jmgm.2005.09.006.

(85) Martin, A. J. M.; Vidotto, M.; Boscariol, F.; Di Domenico, T.; Walsh, I.; Tosatto, S. C. E. RING: Networking Interacting Residues, Evolutionary Information and Energetics in

Protein Structures. *Bioinformatics* **2011**, *27* (14), 2003–2005. https://doi.org/10.1093/bioinformatics/btr191.

(86) Piovesan, D.; Minervini, G.; Tosatto, S. C. E. The RING 2.0 Web Server for High Quality Residue Interaction Networks. *Nucleic Acids Res.* **2016**, gkw315. https://doi.org/10.1093/nar/gkw315.

(87) Bell, R. J.; Dean, P.; Hibbins-Butler, D. C. Localization of Normal Modes in Vitreous Silica, Germania and Beryllium Fluoride. *J. Phys. C Solid State Phys.* **1970**, *3* (10), 2111–2118. https://doi.org/10.1088/0022-3719/3/10/013.

(88) Tama, F.; Gadea, F. X.; Marques, O.; Sanejouand, Y.-H. Building-Block Approach for Determining Low-Frequency Normal Modes of Macromolecules. *Proteins Struct. Funct. Genet.* **2000**, *41* (1), 1–7. https://doi.org/10.1002/1097-0134(20001001)41:1<1::AID-PROT10>3.0.CO;2-P.

(89) Krivov, G. G.; Shapovalov, M. V.; Dunbrack, R. L. Improved Prediction of Protein Side-Chain Conformations with SCWRL4. *Proteins Struct. Funct. Bioinforma.* **2009**, *77* (4), 778–795. https://doi.org/10.1002/prot.22488.

(90) Maguid, S.; Fernandez-Alberti, S.; Ferrelli, L.; Echave, J. Exploring the Common Dynamics of Homologous Proteins. Application to the Globin Family. *Biophys. J.* **2005**, *89* (1), 3–13. https://doi.org/10.1529/biophysj.104.053041.

(91) Maguid, S.; Fernandez-Alberti, S.; Echave, J. Evolutionary Conservation of Protein Vibrational Dynamics. *Gene* **2008**, *422* (1–2), 7–13. https://doi.org/10.1016/j.gene.2008.06.002.

(92) Chattopadhyaya, R.; Meador, W. E.; Means, A. R.; Quiocho, F. A. Calmodulin Structure Refined at 1.7 A Resolution. *J. Mol. Biol.* **1992**, *228* (4), 1177–1192. https://doi.org/10.1016/0022-2836(92)90324-d.

(93) Meador, W.; Means, A.; Quiocho, F. Target Enzyme Recognition by Calmodulin: 2.4 A Structure of a Calmodulin-Peptide Complex. *Science (80-.).* **1992**, *257* (5074), 1251–1255. https://doi.org/10.1126/science.1519061.

(94) Aykut, A. O.; Atilgan, A. R.; Atilgan, C. Designing Molecular Dynamics Simulations to Shift Populations of the Conformational States of Calmodulin. *PLoS Comput. Biol.* **2013**, *9* (12), e1003366. https://doi.org/10.1371/journal.pcbi.1003366.

(95) Gsponer, J.; Christodoulou, J.; Cavalli, A.; Bui, J. M.; Richter, B.; Dobson, C. M.; Vendruscolo, M. A Coupled Equilibrium Shift Mechanism in Calmodulin-Mediated Signal Transduction. *Structure* **2008**, *16* (5), 736–746. https://doi.org/10.1016/j.str.2008.02.017.

(96) Tanaka, T.; Hidaka, H. Hydrophobic Regions Function in Calmodulin-Enzyme(s) Interactions. *J. Biol. Chem.* **1980**, *255* (23), 11078–11080.

(97) 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. https://doi.org/10.1529/biophysj.106.095927.

(98) Yan, H.; Tsai, M. D. Nucleoside Monophosphate Kinases: Structure, Mechanism, and Substrate Specificity. *Adv. Enzymol. Relat. Areas Mol. Biol.* **1999**, *73*, 103–134, x.

(99) 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. https://doi.org/10.1016/s0969-2126(01)00181-2.

(100) Hanson, J. A.; Duderstadt, K.; Watkins, L. P.; Bhattacharyya, S.; Brokaw, J.; Chu, J.-W.; Yang, H. Illuminating the Mechanistic Roles of Enzyme Conformational Dynamics. *Proc. Natl. Acad. Sci.* **2007**, *104* (46), 18055–18060. https://doi.org/10.1073/pnas.0708600104.

(101) Seyler, S. L.; Beckstein, O. Sampling Large Conformational Transitions: Adenylate Kinase as a Testing Ground. *Mol. Simul.* **2014**, *40* (10–11), 855–877. https://doi.org/10.1080/08927022.2014.919497.

(102) Beckstein, O.; Denning, E. J.; Perilla, J. R.; Woolf, T. B. Zipping and Unzipping of Adenylate Kinase: Atomistic Insights into the Ensemble of Open↔Closed Transitions. *J. Mol. Biol.* **2009**, *394* (1), 160–176. https://doi.org/10.1016/j.jmb.2009.09.009.

(103) Oh, B. H.; Pandit, J.; Kang, C. H.; Nikaido, K.; Gokcen, S.; Ames, G. F. L.; Kim, S. H. Three-Dimensional Structures of the Periplasmic Lysine/Arginine/Ornithine- Binding Protein with and without a Ligand. *J. Biol. Chem.* **1993**, *268* (15), 11348–11355.

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

(105) Echols, N.; Milburn, D.; Gerstein, M. MolMovDB: Analysis and Visualization of Conformational Change and Structural Flexibility. *Nucleic Acids Res.* **2003**, *31* (1), 478–482. https://doi.org/10.1093/nar/gkg104.

(106) Quiocho, F. A.; Spurlino, J. C.; Rodseth, L. E. Extensive Features of Tight Oligosaccharide Binding Revealed in High-Resolution Structures of the Maltodextrin Transport/Chemosensory Receptor. *Structure* **1997**, *5* (8), 997–1015. https://doi.org/10.1016/S0969-2126(97)00253-0.

(107) Sharff, A. J.; Rodseth, L. E.; Spurlino, J. C.; Quiocho, F. A. Crystallographic Evidence of a Large Ligand-Induced Hinge-Twist Motion between the Two Domains of the Maltodextrin Binding Protein Involved in Active Transport and Chemotaxis. *Biochemistry* **1992**, *31* (44), 10657–10663. https://doi.org/10.1021/bi00159a003.

(108) Stockner, T.; Vogel, H. J.; Tieleman, D. P. A Salt-Bridge Motif Involved in Ligand Binding and Large-Scale Domain Motions of the Maltose-Binding Protein. *Biophys. J.* **2005**, *89* (5), 3362–3371. https://doi.org/10.1529/biophysj.105.069443.

(109) Tama, F.; Sanejouand, Y. H. Conformational Change of Proteins Arising from Normal Mode Calculations. *Protein Eng.* **2001**, *14* (1), 1–6.https://doi.org/10.1093/protein/14.1.1.

Table 1. Results of iterative procedures for conformational sampling

Protein	apo[a]	targeted[b]	unbiased[b]	unbiased with selected modes[b]
CaM	15.2	1.6	5.1	2.8
AK	6.9	1.1	2.4	-
LAO	4.7	0.6	2.4	0.7
MBP	3,8	0.9	1.2	-

[a]Backbone RMSD with respect to the experimental holo structure. [b]RMSD of the conformation most similar to the holo form.

Figure 1

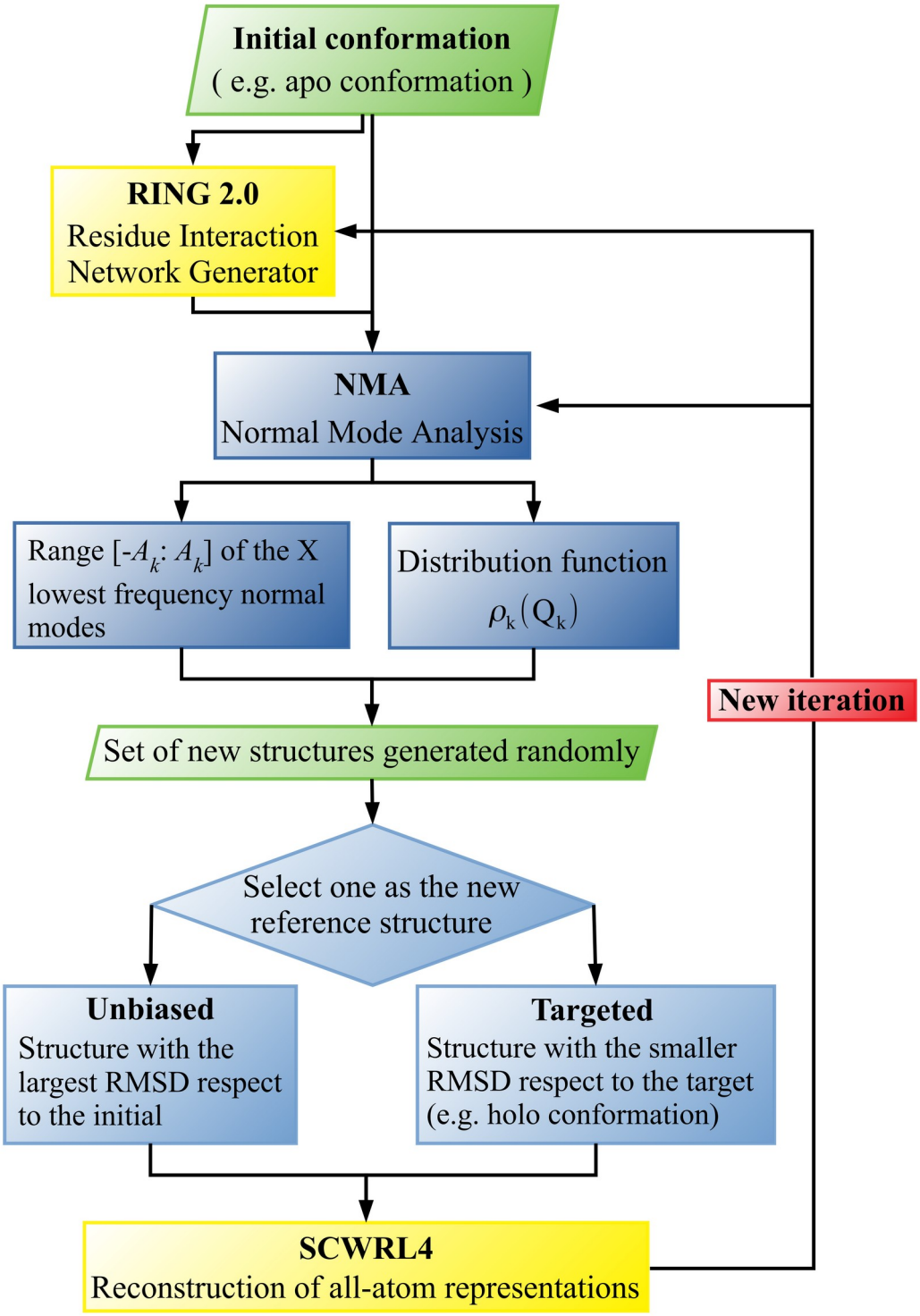

Initial conformation
(e.g. apo conformation)
RING 2.0
Residue Interaction Network Generator
NMA
Normal Mode Analysis
Range [-A_k: A_k] of the X lowest frequency normal modes
Distribution function $\rho_k(Q_k)$
New iteration
Set of new structures generated randomly
Select one as the new reference structure
Unbiased
Structure with the largest RMSD respect to the initial
Targeted
Structure with the smaller RMSD respect to the target (e.g. holo conformation)
SCWRL4
Reconstruction of all-atom representations

Figure 2

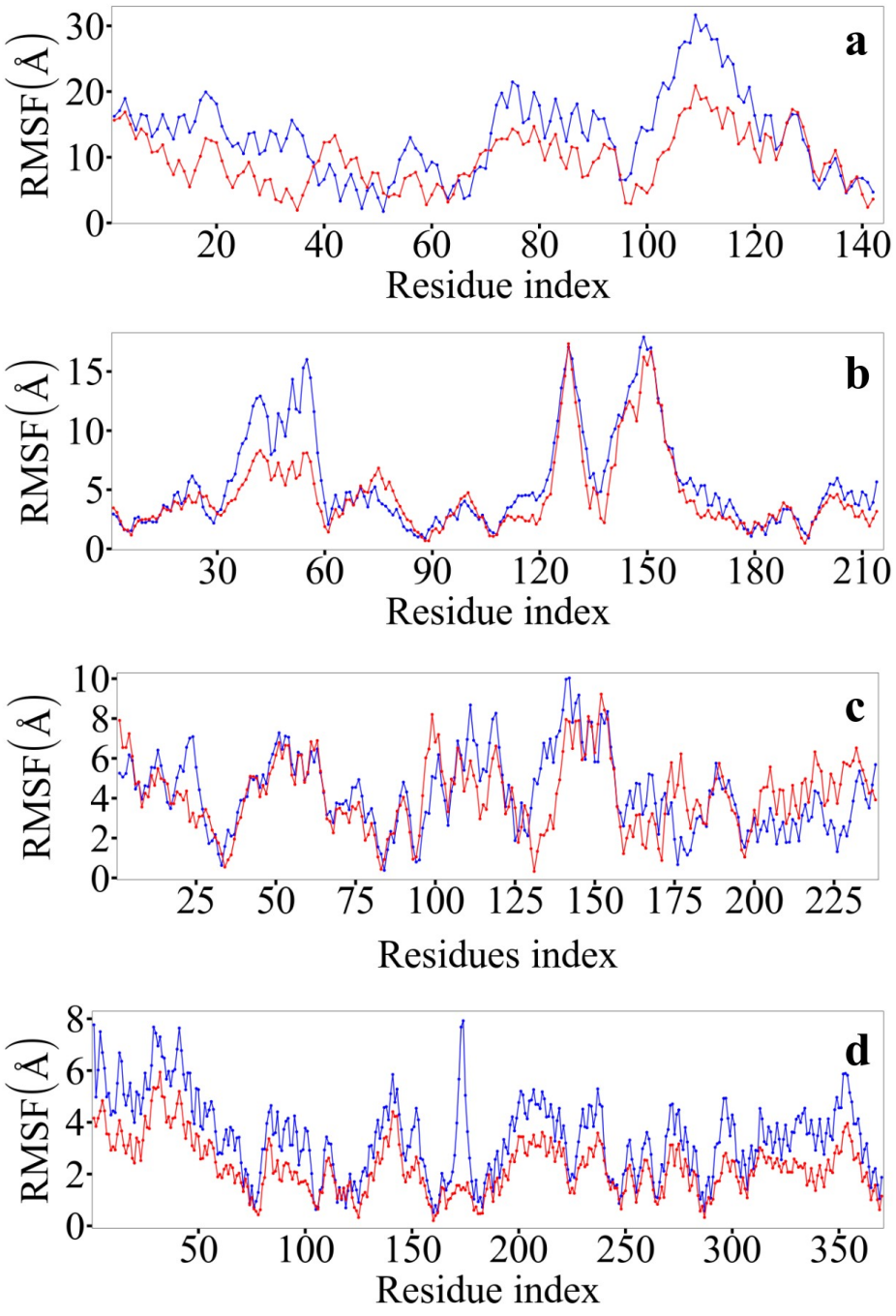

a
RMSF(Å)
Residue index
b
RMSF(Å)
Residue index
c
RMSF(Å)
Residues index
d
RMSF(Å)
Residue index

Figure 3

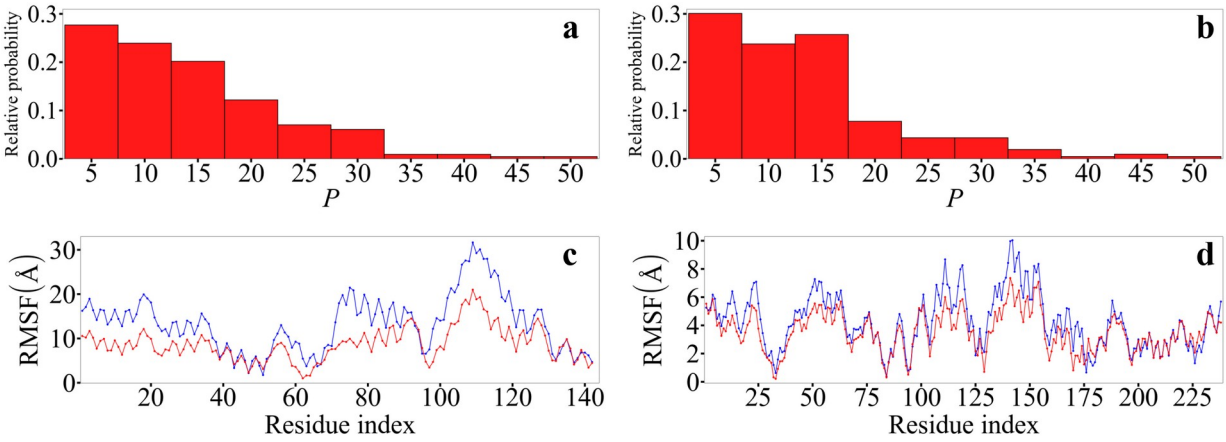


Figure 4

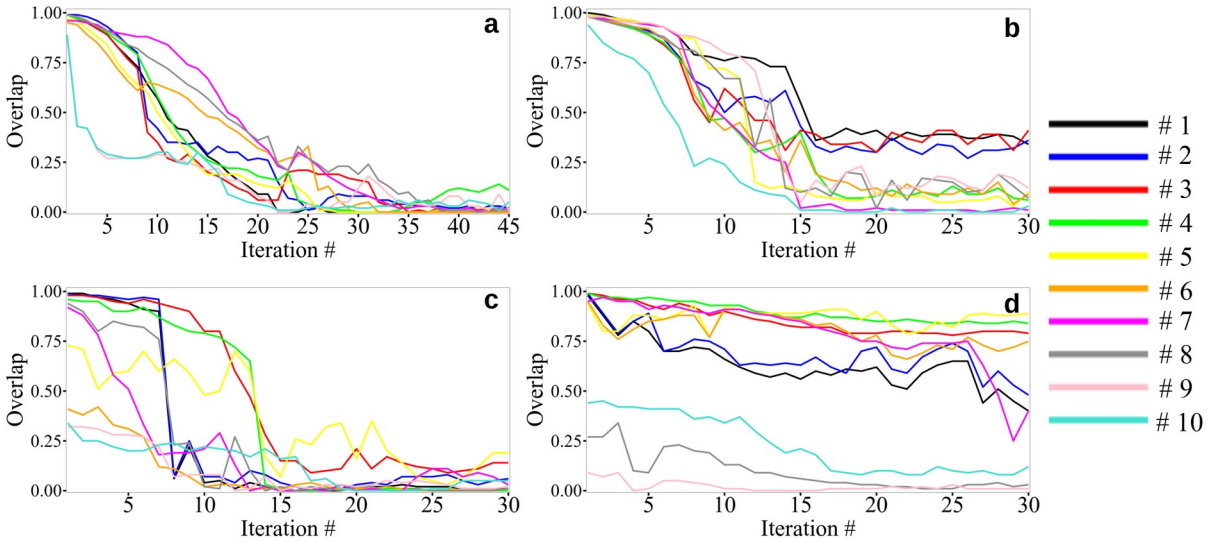

a
b
c
d
Overlap
Iteration #
1.00
0.75
0.50
0.25
0.00
5
10
15
20
25
30
35
40
45
1
2
3
4
5
6
7
8
9
10

Figure 5

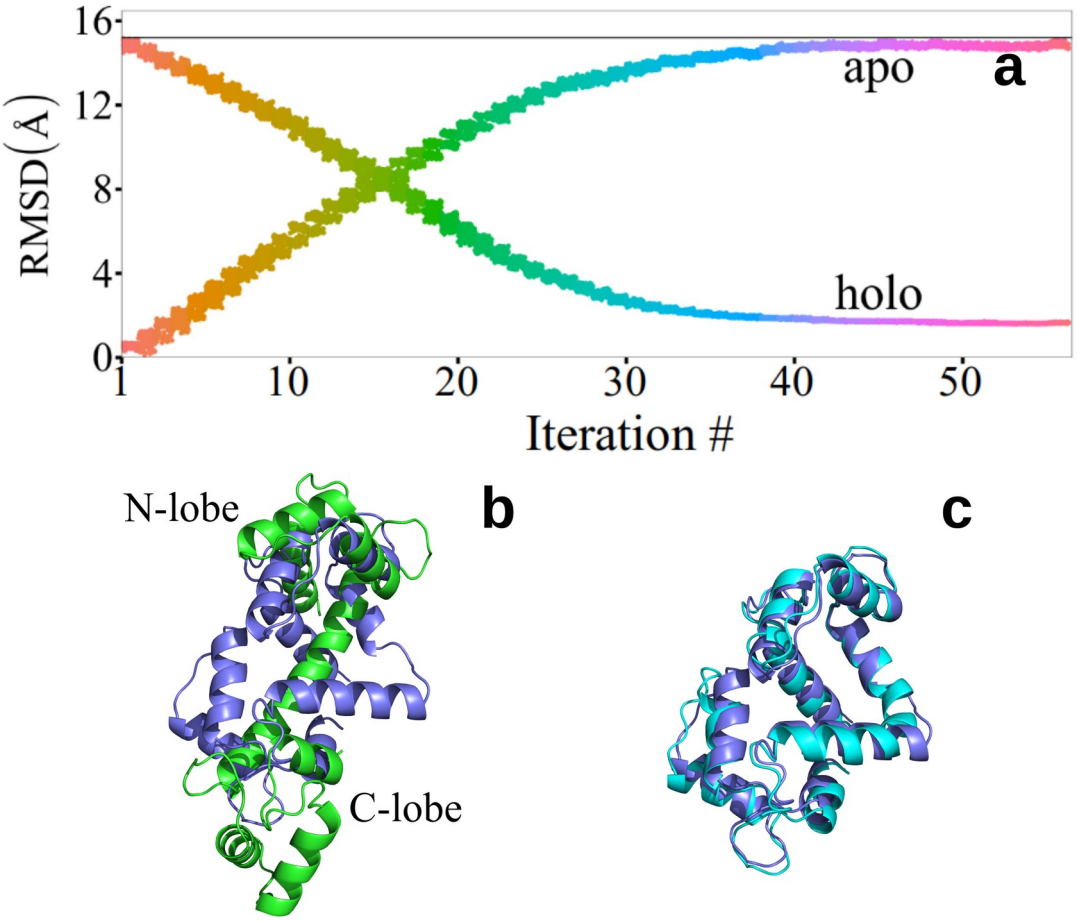

16
12
8
4
0
RMSD(Å)
1
10
20
30
40
50
Iteration #
apo
holo
a
N-lobe
C-lobe
b
c

Figure 6

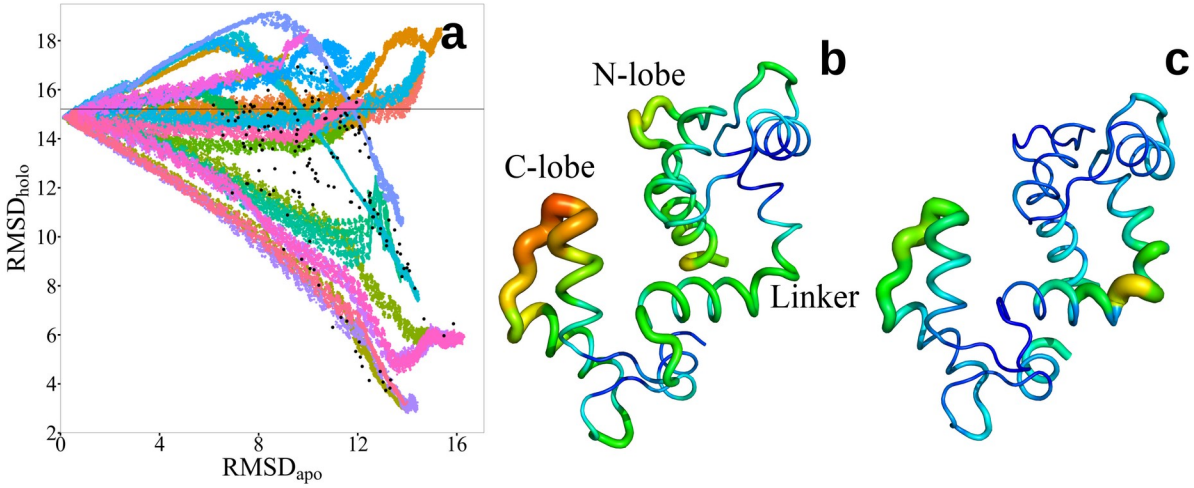

a
b
c
N-lobe
C-lobe
Linker
RMSD$_{holo}$
RMSD$_{apo}$
2
4
6
8
10
12
14
16
18
0
4
8
12
16

Figure 7

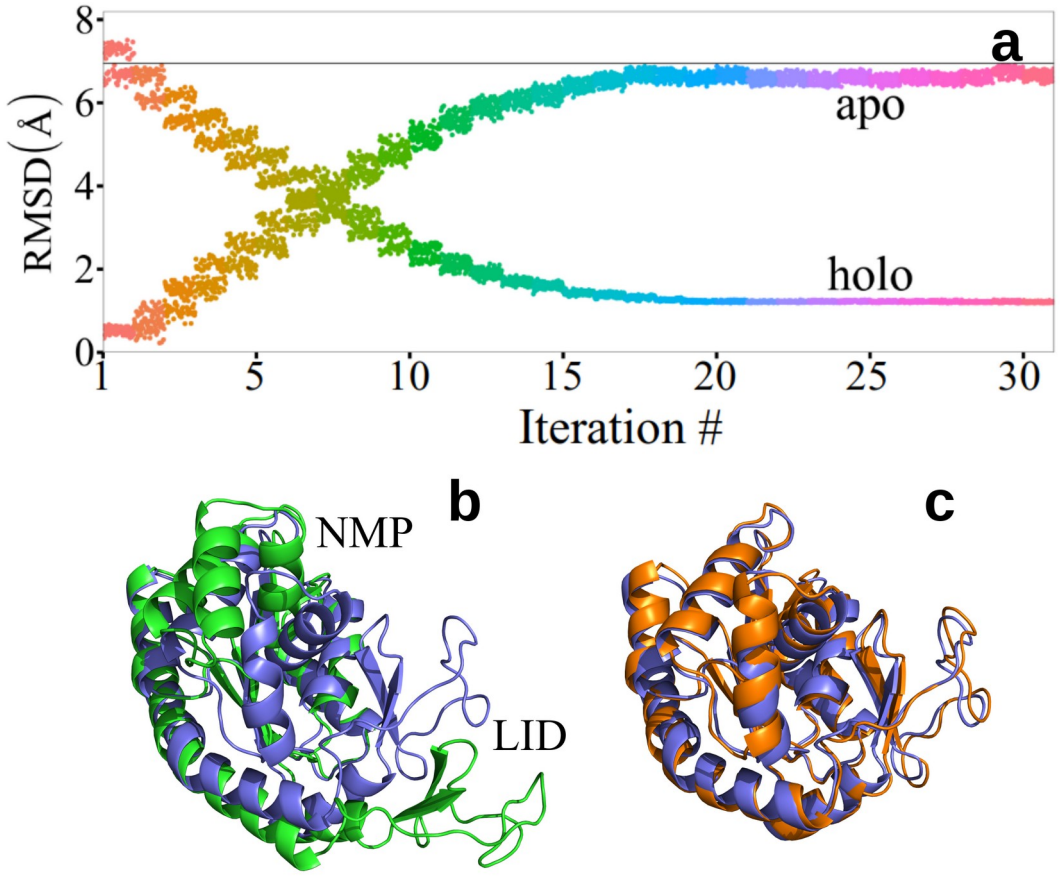


Figure 8

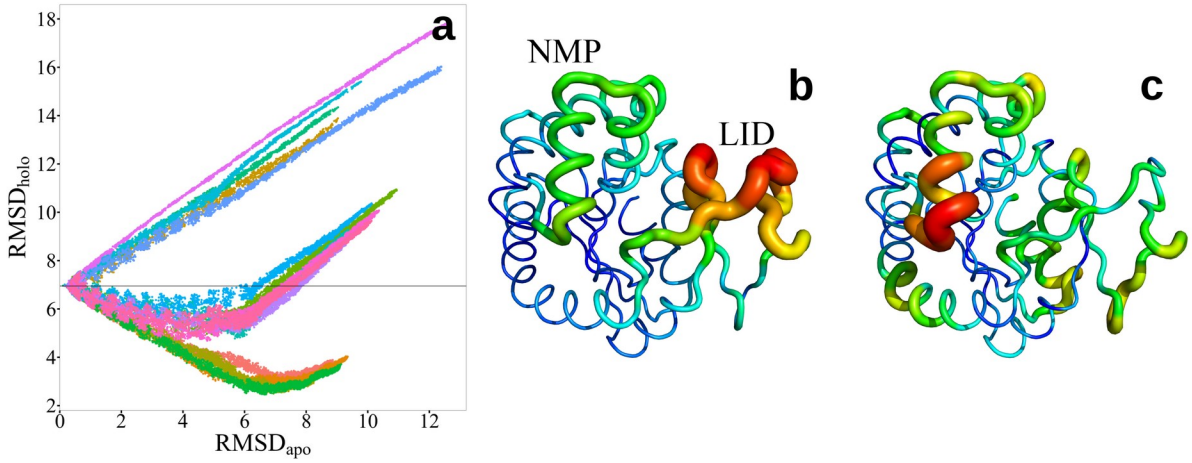
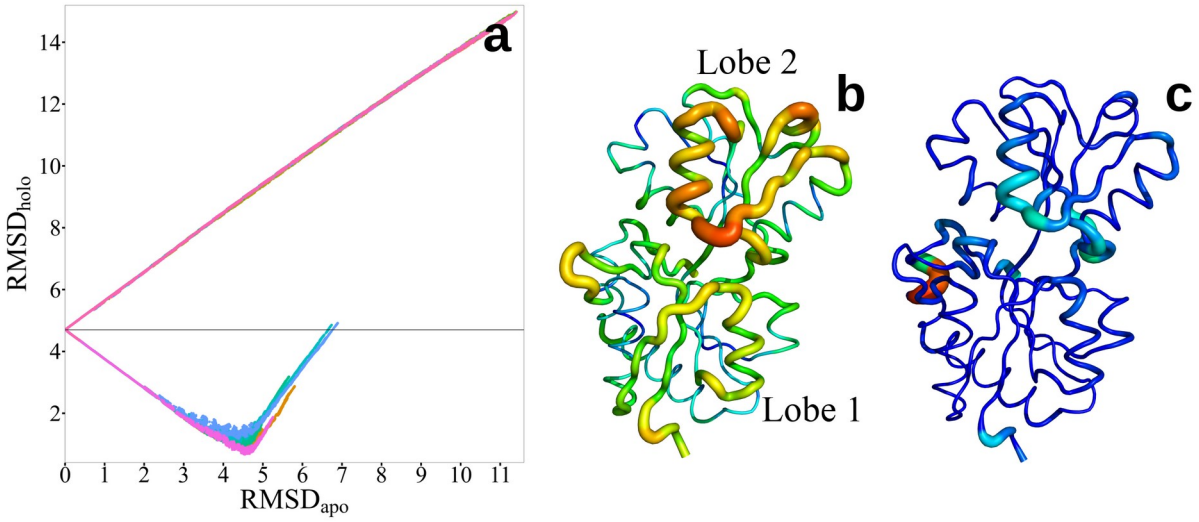


Figure 9


a
b
c
Lobe 2
Lobe 1
RMSD$_{holo}$
RMSD$_{apo}$
0 1 2 3 4 5 6 7 8 9 10 11
2 4 6 8 10 12 14