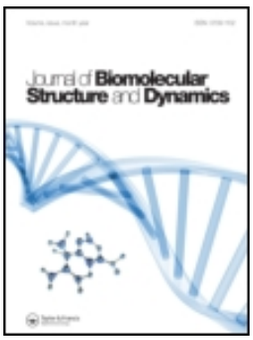

Journal of Biomolecular Structure and Dynamics

ISSN: 0739-1102 (Print) 1538-0254 (Online) Journal homepage: https://www.tandfonline.com/loi/tbsd20

A Docking Study Using Atomistic Conformers Generated via Elastic Network Model for Cyclosporin A/Cyclophilin A Complex

E. Demet Akten , Sertan Cansu & Pemra Doruker

To cite this article: E. Demet Akten , Sertan Cansu & Pemra Doruker (2009) A Docking Study Using Atomistic Conformers Generated via Elastic Network Model for Cyclosporin A/Cyclophilin A Complex, Journal of Biomolecular Structure and Dynamics, 27:1, 13-25, DOI: 10.1080/07391102.2009.10507292

To link to this article: https://doi.org/10.1080/07391102.2009.10507292

Published online: 15 May 2012.

Journal of Biomolecular Structure & Dynamics, ISSN 0739-1102
Volume 27, Issue Number 1, (2009)

A Docking Study Using Atomistic Conformers Generated via Elastic Network Model for Cyclosporin A/Cyclophilin A Complex

http://www.jbsdonline.com

E. Demet Akten[1,3,]**
Sertan Cansu[1,2]
Pemra Doruker[1,2,*]

[1]Polymer Research Center
[2]Department of Chemical Engineering
Bogazici University
Bebek 34342, Istanbul, Turkey
[3]Department of Information Technologies
Kadir Has University
Cibali 34083, Istanbul, Turkey

Abstract

Anisotropic network model is used to generate a set of distinct conformations for cylophilin A (CypA). The native structure is deformed to different extents along each of the lowest-frequency modes (first 7 modes) both in negative and positive directions. Each node of the elastic network represents either a single atom in the high-resolution model or a single residue in the low-resolution model. Realistic conformations with energies close to or lower than the crystal structure and with satisfactory internal geometry are recovered by energy minimization using implicit solvation model. These conformations are then used for ensemble docking to the ligand cyclosporin A for both a further test of accuracy of generated conformers and exploration of different binding modes. Higher number of correctly docked ligands are obtained for conformations with low deformation factors as a result of lower root mean square distances with respect to crystal structure. Yet, surprisingly, the lowest binding energy is obtained for one of the highly deformed conformations as a result of its special contact with arginine side chain oriented towards binding site. Considering the fact that the cyclic ligand's backbone and protein's side chains are held rigid during docking, the conformers generated by high- and low-resolution elastic network models are almost equally successful in providing the correct binding mode. The shape of the binding pocket that incorporates crucial interaction sites for hydrogen bond formation is found to be another important determining factor for the success of the dock. Also, the small backbone variations of a few Ångstroms in magnitude at the loop regions surrounding the binding pocket can cause amino acids' side chains to be displaced by magnitudes of up to 10 Å and therefore have a strong influence on the efficiency of the conformational search during docking.

Key words: Elastic network model; Protein flexibility; Structure-based drug design; Cyclophilin A; Cyclosporin A; Coarse-grained model.

Introduction

Protein's conformational dynamics underlies its biological activity, such as catalysis and signal transduction. Binding of ligands to proteins is commonly accompanied by special conformational changes, such as those between the open/apo and closed/bound states (1, 2). The preexisting equilibrium/conformational selection hypothesis (3-5) has emerged as an alternative to induced-fit mechanism (1). The model states that the native state of a protein is described by an ensemble of closely-related conformations in equilibrium. Binding of a ligand will only shift the equilibrium toward a specific conformation, thus allowing a specific signal to propagate. The dynamic 'personality' of a protein is an intrinsic property and is encoded in its three-dimensional structure. In other words, the ligand does not trigger the formation of a new structure but, instead, selects a pre-existing structure. Recent experimental studies have shown that the bound conformational states are also accessible to the apo form of the protein in solution (6, 7).

[*]Phone: +90-212-3597365
Fax: +90-212-2575032
Email: doruker@boun.edu.tr
[**]Phone: +90-212-3597002
Email: demetakten@gmail.com

Protein's full flexibility is a determining factor in the interaction of proteins with ligands and other macromolecules. Thus, both main- and side-chain flexibility needs to be taken into account for improved accuracy in docking studies (8-11). Flexibility of the protein, especially of its main chain, has often been discarded in docking studies due to its high cost of computation (12-14). Sometimes, only small side-chain and backbone variations in protein's loop regions around the binding site are allowed during docking (15, 16). However, a more accurate description of the conformational variations describing the protein's intrinsic plasticity is essential for molecular recognition studies. This problem is largely solved by 'ensemble docking' algorithms that provide a feasible set of conformations for the bound, unbound, and intermediate states of the protein. Different methodologies exist to generate conformational ensembles for flexible docking, such as normal mode calculations (17), molecular dynamics (18), and Monte Carlo simulations (19-21). Recent studies have used multicanonical molecular dynamics simulations (22-23) and provided successful molecular recognitions.

However, for most MD studies, the computational cost is a major drawback. In this respect, coarse-grained normal mode analysis (NMA) has become one of the standard techniques in studying the conformational dynamics of proteins and their complexes. Elastic network models, which provide analytical solutions to the equilibrium dynamics of biomolecules (24, 25), are low-resolution models for normal mode analysis. These models replace detailed inter-atomic potentials with uniform harmonic potentials between close-neighboring atoms (26) or residue pairs (24, 25). As elastic network models provide high computational efficiency and accuracy in the prediction of functionally important collective dynamics (24, 27, 28), they have been implemented recently in various macromolecular docking algorithms (17, 29-30).

In this study, we employ a so-called reverse-mapping technique, where elastic network models at different resolutions are used to generate realistic flexible conformations with full atomistic details. Alternative conformations are generated by deforming the native x-ray structure of a protein along the low-frequency collective normal modes of the elastic network. The new conformers are then brought back to local minima by means of energy minimization using implicit solvation model, which removes steric clashes and unrealistic geometries due to forcing the native structure along harmonic directions. Here, each conformation is generated by deformation along a single mode, which does not guarantee a complete search of the conformational space and identification of all binding modes. However, it is known that one of the low frequency modes often correlates with the direction of conformational transition, specifically for proteins showing large conformational changes upon binding (31, 32), which justifies our approach. Moreover, it is straightforward to modify our reverse-mapping technique to perform simultaneous deformation along several modes for conformer generation.

Elastic network models can be employed at different resolutions, namely atomistic (33), residue-level (using C^{α} atoms as nodes) or even lower resolutions (34). Here, we will assess the atomistic and residue-based networks in terms of ensemble generation and their subsequent application in ligand docking. We are especially interested in the capability of the residue-based, low-resolution model to generate a representative set of conformers to be used in docking, compared to atomistic, high-resolution model. This study will confirm the accuracy of our model in generating realistic conformers that can be used in detailed molecular recognition studies such as docking.

The protein chosen for validation of our elastic network models and reverse-mapping technique is cyclophilin A (CypA) with 165 residues, which catalyzes the *cis-trans* isomerization in proteins. Independent from its catalytic activity, it is a target for the immunosuppressive drug cyclosporine A (CsA), a cyclic undecapeptide. The tightly bound complex formation of CsA/CypA, which targets the protein phosphatase calcineurin A (35), is required for immunosuppressive activity.

Here, we will employ the ensemble docking approach in order to test the validity of our conformations and to bring out new and more favorable binding modes. Zavodsky *et al*. (36) also studied the same protein-ligand complex and generated multiple receptor conformations via a graph-theoretical approach called FIRST (37) by identifying rigid and flexible regions based on the protein bond network and considering the flexibility of each loop region, independently. The major difference of our methodology lies in the fact that loop motions and resulting conformations are correlated with each other and the rest of the protein, as they are generated based on the collective normal modes.

Materials and Methods

Reverse-mapping Procedure

New protein conformations are generated using the low-frequency normal modes from the anisotropic network model (ANM) (24). Here we employ both coarse-grained and atomistic ANM, where the positions of the nodes are defined by the coordinates of α-Carbons and heavy atoms (excluding H atoms), respectively, in the folded three-dimensional (3D) structure of the protein. The node pairs that fall within a specific cutoff distance are connected by identical harmonic springs representative of the bonded and nonbonded interactions between the node pairs. The total potential energy of the structure is expressed as a Gaussian form

$$V = (1/2)\Delta R^T H \Delta R \qquad [1]$$

where $\Delta\mathbf{R}$ is a 3*N*-dimensional vector of the positional fluctuations of the residues, $\Delta\mathbf{R}^T$ is its transpose, and $\mathbf{H}$ is the Hessian matrix. By performing an orthogonal transformation of the Hessian matrix, the overall motion can be expressed as a sum over the (3*N*-6) normal modes. The mean-square fluctuation of the ith node based on the kth non-zero normal mode is given by

$$\langle \Delta \mathbf{R}_i^2 \rangle_k = \left(\frac{k_B T}{\gamma} \right) \frac{(\mathbf{u}_{ik})^2}{\lambda_k} \qquad [2]$$

with the corresponding eigenvalue λ_k and the displacement vector $\mathbf{u}_{ik}$ (the ith element of the kth eigenvector). Here k_B is the Boltzmann constant, T is the absolute temperature, and γ is the uniform spring constant.

Harmonic vibrational analysis of the protein network is performed around an equilibrium/native structure of the protein in the bound form (PDB code: 1bck.pdb). It should be noted that our choice of the X-ray structure in the bound form cannot produce biased results, since the bound and unbound (PDB code: 2cpl.pdb) forms of the protein are almost indistinguishable with a root mean square distance (RMSD) of 0.22 Å with respect to each other. New conformations are generated by deforming the native structure along both positive and negative directions of the selected low-frequency eigenvectors representing collective motions.

$$\mathbf{R}_i^{new,k} = (\pm DF) \frac{\mathbf{u}_{ik}}{\sqrt{\lambda_k}} + \mathbf{R}_i^{ref} \qquad [2]$$

Here, $\mathbf{R}_i^{ref}$ and $\mathbf{R}_i^{new,k}$ refer to the coordinates of the ith node in the reference x-ray structure and two new conformers generated along the kth mode (in positive and negative directions). Deformation factor (*DF*) can be adjusted to produce distinct conformations along each mode by overcoming energy barriers. We will employ the same deformation factor along each mode, which leads to higher deformations along lower frequency modes. In the case of coarse-grained ANM, atomistic conformers can be generated by rigidly moving all the atoms in a residue with the corresponding coarse-grained C$^{\alpha}$ node, analogous to building block approach of Tama *et al*. (38).

The deformed conformations have high initial energies due to close contacts between atoms and distorted local geometries. In order to relax these structures into realistic conformations, energy minimization is performed using the program AMBER (39, 40) with the ff03 forcefield parameters (41). For each conformer, 500 cycles of steepest descent algorithm are performed, followed by conjugate gradients until the energy gradient per atom reaches 0.01 kcal·mol^{-1}·Å^{-1}. Pairwise generalized Born model (42, 43) is used to mimic the effect of water molecules during the minimizations. The cutoff for non-bonded interactions is fixed at 16 Å. Salt concentration for 1-1 mobile counter ions in the solution is set to 0.1 M according to generalized Born theory based on the Debye-Hückel limiting law for ion screening of interactions (44).

Ensemble Docking

The latest version of the program Autodock v.4 (45), is used for docking cyclosporin with flexible side chains against the low energy conformers of cyclophilin A (CypA) obtained by reverse-mapping the ANM conformations, as described above. A pre-calculated three-dimensional energy grid of equally spaced discrete points is generated prior to docking, for a rapid energy evaluation, using the program AutoGrid (46). The grid box with dimensions of 27.75 Å × 18.75 Å × 22.5 Å covers the entire binding site and its neighboring residues. The distance between two grid points was set to 0.375 Å.

The ligand is allowed to be flexible around 17 rotatable bonds on its side chains only. The backbone of the cyclic conformation of the ligand is retained as in the crystal structure of the complex (PDB code: 1bck) with all peptide bonds in trans form (see Fig. 1). Docking is computationally challenging due to large number of rotatable bonds. As a search method, Lamarckian genetic algorithm (LGA) (45), a hybrid of global and local search, is employed. The ligand conformation extracted from the crystal structure is used as a template to create a population of 200 distinct ligand conformers with randomly assigned torsion angles to rotatable bonds, position of the center of the ligand, and quaternion for its overall orientation. For each protein conformation, a total of 100 independent docking runs, each starting with a different ligand population, are performed. Each run is composed of 5×10^7 energy evaluations, which makes a total of 5×10^9 energy evaluations for one protein conformer only. Considering the large size of the ligand and docking conditions given above, each docking run takes an average of 180 minutes on a 1.5 GHz/6MB L3 Cache Intel® Itanium2 processor. Same docking procedure is applied to a total of 34 protein conformers, 22 of them generated by atomistic (high-resolution) ANM and 12 of them by coarse-grained (low-resolution) ANM.

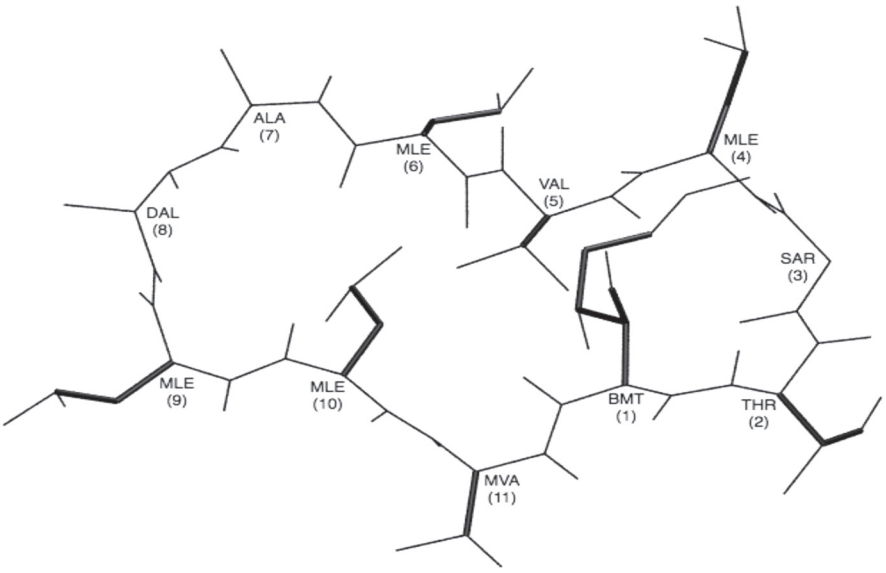

Figure 1: Ligand *2-Thr* Cyclosporin A with rotatable bonds on its side chain indicated with thick lines.

For a correctly docked ligand, it is necessary to maintain some critical non-bonded contact distances between the protein and its ligand within acceptable ranges. Table SI given as ***Supplementary Materials*** lists the most conserved interaction sites and their average non-bonded contact distances between cyclosporin A and cyclophilin A determined from X-ray crystallographic analysis (47). The first five contacts are strongly conserved hydrogen bonds. The remaining ones on the list comprise two C…N contacts, six C…O contacts, and six C…C contacts, mostly dominated by MeVal11 of cyclosporin A, which is known to interact at six different recognition sites on cyclophilin A. In this study, the criteria for a correct dock is to maintain a maximum distance of 4 Å for at least two hydrogen bonds, and a maximum distance of 5 Å for at least eight contacts where C atom is involved (no. 6 to 19 on the list).

Free energy scoring function used in Autodock v.4 is an empirical binding free energy based on AMBER force field (48). It is composed of intermolecular interactions (van der Waals, hydrogen bond, desolvation, and electrostatics) between ligand and target protein, ligand's internal energy, ligand's torsional free energy, and unbound ligand's energy. Here, the ligand's torsional free energy is always +4.12 kcal/mol, since the same number of allowed rotatable bonds is used for every run, and also the energy of its unbound state is statistically predetermined prior to docking only once and does not change for every run. Therefore, the values of the ligand's torsional free energy and the unbound ligand's energy are not incorporated into our scoring function during analysis.

Results And Discussion

Validation of Reverse-mapped Atomistic CypA Conformers

Slowest seven modes are used for reverse-mapping by energy minimization. Same deformation is applied along positive and negative directions of each eigenvector, leading to *a* and *b* conformers, respectively. Six modes [1, 2, 3, 4, 5, and 7] are later chosen for ensemble docking, and their corresponding reverse-mapped and energy minimized conformers are listed in Tables I and II. The second and third columns list final energies after minimization and final RMSD values calculated with respect to crystal structure based on backbone atoms only. The energy of the crystal structure indicated as x-ray in both tables is also minimized using the same procedure.

Two different deformation factors, denoted as low and high DF (L and H in Table I), are used to generate two distinct sets of conformations using atomistic ANM. The average RMSDs of these sets (before energy minimization) are 0.7 ± 0.1 Å and 1.7 ± 0.4 Å from the crystal structure, respectively. For example H-1a refers to the conformer generated along the positive direction of first eigenvector using the high DF. Similarly, the modes listed in Table II are obtained from coarse-grained ANM, where a single

Table I
Docking results for cyclophilin A conformers generated from atomistic ANM.

conf #	Energy[a] (kcal/mol)	RMSD (Å)[b]	# of correct docks	Scoring Energy Average / Min.[c] (kcal/mol)	# of bonds[d]	*d* (Å)[e]
x-ray	-4750	0	9	-14.5 / -15.9	4.1 / 12.7	14.9
L-1a	-4737	0.6 (1.0)	11	-14.8 / -15.5	4.2 / 12.3	15.8
L-1b	-4755	0.7 (1.1)	6	-13.9 / -14.6	3.2 / 13.2	14.2
L-2a	-4753	0.6 (0.8)	2	-12.9 / -13.0	3.5 / 10.5	14.5
L-2b	-4746	0.5 (0.7)	5	-15.2 / -15.5	4 / 13.2	14.9
L-3a	-4751	0.3 (0.4)	4	-14.0 / -14.3	4 / 12.7	14.8
L-3b	-4750	0.4 (0.5)	10	-15.5 / -16.2	4.9 / 12.9	14.9
L-4a	-4,740	0.5 (0.3)	N/A[f]			
L-4b	-4,759	0.4 (0.4)	N/A[f]			
L-5a	-4761	0.5 (0.7)	9	-14.6 / -15.4	3.9 / 12.8	15.2
L-5b	-4741	0.4 (0.7)	1	-12.3 / -12.3	3 / 11	14.4
L-6a	-4,757	0.2 (0.3)	N/A[f]			
L-6b	-4,740	0.3 (0.5)	N/A[f]			
L-7a	-4765	0.4 (0.6)	8	-15.1 / -15.3	4.5 / 13.2	14.6
L-7b	-4733	0.3 (0.3)	13	-14.5 / -15.7	3.7 / 12	14.9
H-1a	-4712	1.2 (2.2)	4	-12.6 / -13.2	3.5 / 8.2	16.2
H-1b	-4746	1.2 (2.0)	0	-	-	13.7
H-2a	-4623	1.0 (1.4)	0	-	-	14.2
H-2b	-4725	1.1 (2.0)	5	-13.4 / -14.2	2.6 / 11.8	15.2
H-3a	-4757	0.6 (0.9)	2	-13.8 / -13.8	3 / 10.5	14.5
H-3b	-4739	0.8 (1.1)	3	-14.6 / -15.1	4.7 / 13	14.7
H-4a	-4720	0.8 (0.6)	0	-	-	14.6
H-4b	-4745	0.8 (0.8)	13	-14.7 / -15.8	3.8 / 12.6	14.9
H-5a	-4,729	1.1 (2.0)	9	-14.1 / -14.9	3.5 / 12.4	15.5
H-5b	-4718	1.0 (1.9)	0	-	-	13.9
H-6a	-4,745	0.5 (0.8)	N/A[f]			
H-6b	-4,751	0.6 (0.9)	N/A[f]			
H-7a	-4739	0.9 (1.3)	4	-15.6 / -16.6	4.2 / 13	14.7
H-7b	-4722	0.8 (1.7)	0	-	-	14.8

[a] Energy of the ligand-free protein conformation after energy minimization. The first residue Ala has been deleted from the x-ray structure due to its extremely mobility in ANM slow modes. [b] RMSD with respect to energy-minimized crystal structure (1bck.pdb). The RMSD value given in parentheses is for the loop regions located between residues 65-76, 99-108, and 143-155. [c] Scoring energy is the sum of intermolecular energy and internal energy of the ligand (see text for more details). Both average and minimum values are given. [d] The number of nonbonded interactions between ligand and protein averaged over all correctly docked ligands. First entry is for hydrogen bonds, and the second entry is for hydrophobic contacts. [e] The distance *d* is between two sets of residues (Arg54, Gln62) and (Asn101, Ala102, His125) (see text for more details). [f] N/A: Modes that are excluded from docking calculations due to lower RMSDs in the loop regions, as explained in the text in more detail.

Table II
Docking results for cyclophilin A conformers generated from coarse-grained ANM.

conf #[f]	Energy[a] (kcal/mol)	RMSD (Å)[b]	# of correct docks	Scoring Energy Average / Min.[c] (kcal/mol)	# of bonds[d]	*d* (Å)[e]
x-ray	-4750	0	9	-14.5 / -15.9	4.1 / 12.7	14.9
1a (4a)	-4734	0.9 (1.7)	3	-12.9 / -13.3	2.3 / 9.3	14.1
1b (4b)	-4587	0.8 (1.5)	9	-14.6 / -15.3	4. / 12.3	15.9
2a (1a)	-4726	0.8 (1.1)	2	-13.8 / -14.9	3.5 / 10.	15.9
2b (1b)	-4739	1.2 (1.7)	0	-	-	13.9
3a (2b)	-4706	0.7 (0.9)	9	-15.3 / -15.7	3.2 / 11.9	15.3
3b (2a)	-4737	0.9 (1.1)	1	-12.6	4. / 10.	14.2
4a (3a)	-4747	0.6 (0.6)	2	-13.4 / -14.5	3.5 / 10.5	14.8
4b (3b)	-4731	0.6 (0.7)	0	-	-	14.6
5a (5a, 6a)	-4733	0.9 (1.5)	3	-11.9 / -12.1	2.3 / 10.	14.7
5b (5b, 6b)	-4723	0.9 (1.4)	0	-	-	14.4
6a (7a, 8a)	-4753	0.7 (0.9)	N/A[g]			
6b (7b, 8b)	-4741	0.6 (0.6)	N/A[g]			
7a (7a)	-4740	0.5 (0.7)	1	-12.9	4. / 12.	14.7
7b (7b)	-4731	0.4 (0.5)	7	-13.6 / -14.8	3.6 / 12.3	14.9

[a-e] Same explanations as in Table I. [f] The coarse-grained ANM modes are listed with their corresponding atomistic ANM modes given in parenthesis. [g] Mode 6 is excluded from docking calculations due to its correspondence with atomistic modes 7 and 8 at the same time as shown in Figure 2.

high DF is used to generate conformations with an average RMSD of 1.7 ± 0.3 Å from x-ray structure. DF values are adjusted so that the extent of deformation before energy minimization is almost identical for the coarse-grained and the highly-deformed atomistic conformers.

All reverse-mapped conformers are aligned in Figure 3 with their flexible loop regions (residues 65-76, 99-108, and 143-155) shown in different colors. The x-ray structure is also shown in red. The mean RMSD values with respect to crystal structure are 0.4 ± 0.1 Å, 0.9 ± 0.2 Å, and 0.8 ± 0.2 for the low, highly deformed atomistic and coarse-grained ANM conformers after energy-minimization, respectively (see Tables I and II for final RMSDs).

The correspondence between the atomistic and coarse-grained ANM modes can be observed from the overlap matrix plotted in Figure 2. This matrix gives the absolute values of the dot products between the two eigenvector sets based on only the α-carbons, where values close to one indicate high correspondence. First column of Table II also indicates the coarse-grained ANM modes with their corresponding atomistic ANM modes given in parenthesis.

Among the atomistic reverse-mapped conformers, those generated along modes 4 and 6 have the lowest RMSD values for the loop regions with respect to the crystal structure (see Table I). As a result, we initially excluded modes 4 and 6 from the ensemble. However, mode 4 for the high deformation cases was later included because of the high correspondence between atomistic mode 4 and coarse-grained mode 1 (similarly, between coarse-grained mode 4 and atomistic mode 3).

The initial energies, *i.e.*, those after deformation of the x-ray structure along harmonic modes, are quite high due to the unrealistic bonded and nonbonded interactions, ranging between 10^2 to 10^{12} kcal/mol. However, the conformers after energy minimization have reasonable energy values, when compared with the x-ray structure, indicating low energy conformations around the native state. The implicit solvation model employed during minimization allows a direct comparison of the energies obtained for different conformers. The final energies fall between -4765 and -4623 kcal/mol for conformers from atomistic ANM and -4747 and -4587 kcal/mol for conformers from coarse-grained ANM, most of which are only

slightly higher than the x-ray's energy which is -4750 kcal/mol. Final energies of the conformers indicate that the nodes of the network can be successfully reverse-mapped to realistic atomistic structures.

A strict validation of the energy-minimized conformers is performed using the MolProbity web-server (49, 50) and given in Table III. MolProbity provides both all-atom contact analysis and dihedral angle diagnostics on main and side-chains. No steric overlaps have been detected in the reverse-mapped conformers due to the energy minimization procedure. Our conformers give a few Ramachandran outliers (< 1.3%) and rotamer outliers (< 3.2% for atomistic with high DF and < 4.7% for coarse-grained ANM conformers). There are slightly more outliers in conformers generated by the coarse-grained ANM, even though their deformation factors are lower than the atomistic case. Considering the existence of rotamer outliers (~1.6%) in the native structure, some of the outliers stem from the original x-ray model.

Docking to CypA Conformers

Next, docking of the cyclic ligand cyclosporin A to certain conformers is performed, which would serve as a further test of the validity of the reverse-mapping procedure employed, in addition to incorporating full flexibility of the protein. Table I and II list the docking results for twenty-two protein conformations from atomistic ANM and twelve conformations from coarse-grained ANM, respectively. There is a strong correlation between the average RMSD of a set of conformers with respect to x-ray and the success rate of docking, which is measured as the number of conformers in the set that docked to its ligand at least once. The average RMSDs for conformers obtained from atomistic and coarse-grained ANM are 0.9 (atomistic, high DF), 0.8 (coarse-grained), and 0.4 (atomistic, low DF), with their corresponding success rates of 60%, 75%, and 100%, respectively. We assume that docking to conformers

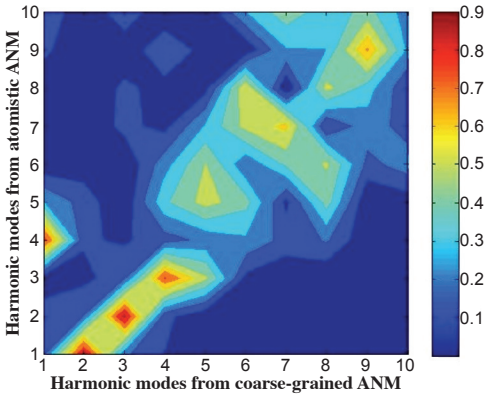

Figure 2: Overlap matrix for harmonic mode directions between coarse-grained and atomistic ANM. Values close to one indicate perfect overlap of corresponding mode pairs.

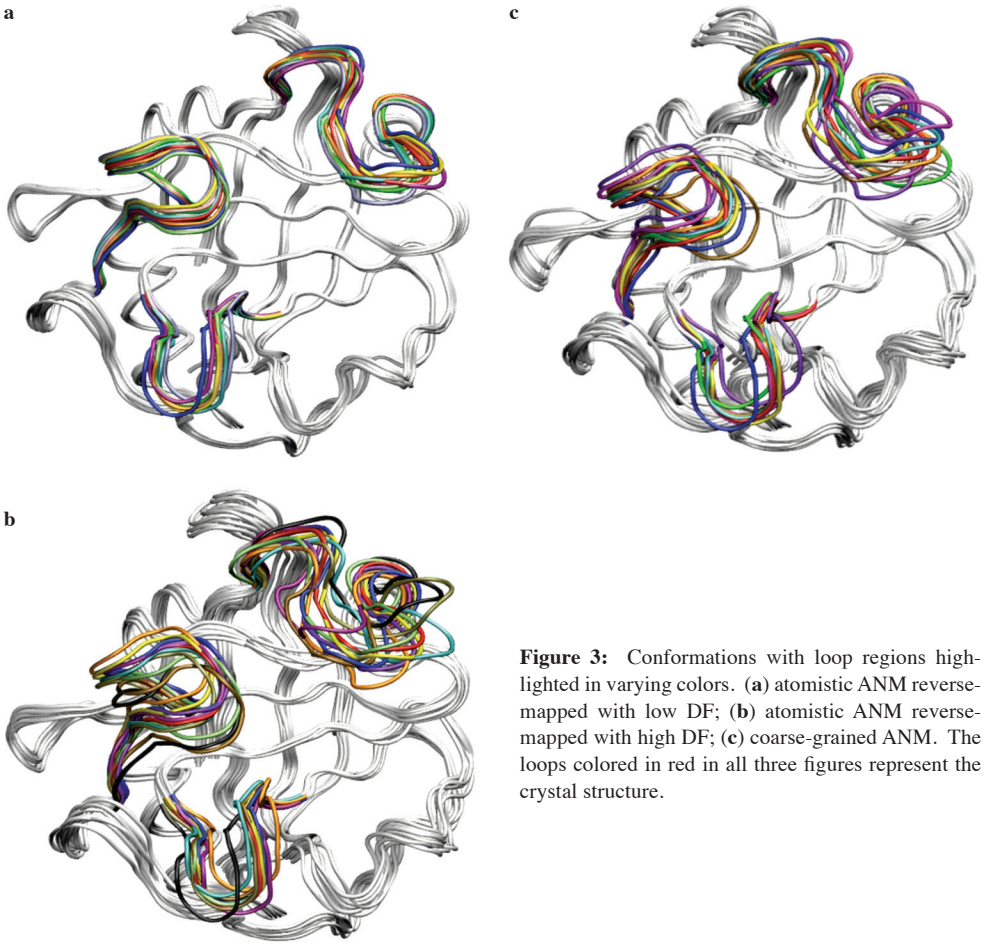

Figure 3: Conformations with loop regions highlighted in varying colors. (**a**) atomistic ANM reverse-mapped with low DF; (**b**) atomistic ANM reverse-mapped with high DF; (**c**) coarse-grained ANM. The loops colored in red in all three figures represent the crystal structure.

Table III
MolProbity[42] results for geometric validation of conformers generated by atomistic and coarse-grained ANM.

Atomistic ANM Conformers	Rotamer outliers (%)	Ramachandran outliers (%)	Coarse-Grained ANM Conformers	Rotamer outliers (%)	Ramachandran outliers (%)
x-ray	1.56	0.00	x-ray	1.56	0.00
H-1a	2.34	0.63	1a	3.13	1.27
H-1b	1.56	0.00	1b	2.34	0.63
H-2a	2.34	1.27	2a	3.13	0.63
H-2b	2.34	0.63	2b	3.91	0.63
H-3a	3.13	0.00	3a	3.13	1.27
H-3b	1.56	0.63	3b	1.56	0.63
H-4a	2.34	0.63	4a	1.56	0.00
H-4b	2.34	0.00	4b	2.34	1.27
H-5a	2.34	0.00	5a	4.69	0.00
H-5b	2.34	0.00	5b	1.56	1.27
H-7a	2.34	0.00	7a	1.56	1.27
H-7b	3.13	1.27	7b	2.34	0.00

from mode 4 with low DF value would be successful due the low RMSD values in the loop regions (see Table I), thus leading to 100% success for low DF. However, it is difficult to detect any strong correlation between the energies of the ligand-free conformers and their number of correct docks, simply because the difference in their energy values is not significant.

Also, conformers of coarse-grained ANM being close to highly deformed conformers of atomistic ANM in terms of their RMSDs, produced a considerable amount of successful docks, namely 37, with an average scoring energy of -13.4 ± 1.1 kcal/mol, compared to highly deformed conformations with 40 success docks and a scoring energy of -14.1 ± 0.9 kcal/mol. This result strongly suggests that ANM, even with a relatively more simplified approach, can successfully produce reliable conformations in the same manner for being used in molecular recognition studies.

Among 34 different cyclophilin conformers, only eight of them have failed to recognize the ligand (Table I and II). The failed conformations were with largest RMSD values (1.4 ± 0.5 Å in average) in their critical loop regions (residues 65 to 76, 99 to 108, and 143 to 155) with respect to crystal structure. Loop regions surround the binding pocket and have a direct influence on conformational search during docking that consequently determines the docking results (Fig. 3). Also, the average distance between two sets of residues (Arg54, Gln62) and (Asn101, Ala102, His125) situated at two opposite sites of the binding pocket was measured and found to be smaller than that of crystal structure for all failed conformations (see last column of Table I and II). Arg54, Gln62, and Asn102 are key interaction sites for hydrogen bond formation with the ligand (see Table SI). For a favorable binding, the ligand needs to be well placed inside the cavity in order to make all five hydrogen bonds simultaneously. Failed conformations clearly present a more closed up binding pocket (see Figure 4), which hinders a favorable ligand fit and consequently an efficient conformational search.

For some conformations, even with large overall and loop RMSDs (H-1a and H-2b in the first group, Table I), successful dockings were obtained mainly due to their more open binding pockets, allowing enough space for efficient conformational search. This can be observed in Figures 4a-d, which represents four distinct modes with their binding pocket highlighted. For each mode, two conformations obtained from deformation in two opposite directions from x-ray are aligned for comparison. For all seventeen pairs of conformers listed in Tables I and II, except for H-7a and H-7b, higher number of correctly docked ligands is obtained for the conformer which has the wider binding pocket than the other conformer in the pair (Tables I and II). Overall, it is clear that the shape of the cavity that contains the key interaction sites is a crucial determining factor for the success of a dock.

Considering the fact that the protein was held rigid during docking, the overall success rate for 34 conformations is ~76%, which is quite satisfactory and well confirms the accuracy of the generated conformations. In addition, 11 out of 34 conformations have binding energies equal to or greater than that of the crystal structure. One important case is shown in Figure 5a, where the conformation H-7a has the strongest binding with a scoring energy of -16.6 kcal/mol (which is the sum of intermolecular and internal energies), well above the average scoring energy of the crystal complex, -14.5 ± 1.1 kcal/mol. In H-7a, the side chain of Arg147 makes multiple hydrogen bonds via its positively charged guanidium group with the bound ligand's carbonyl oxygens of residues 6 and 8, due to orientation of its loop region covering residues 143 to 155. However, this type of loop orientation is absent in conformation H-7b, where the side chain of Arg147 is facing away from the binding pocket and the guanidium group is displaced by about 18 Å, therefore missing favorable contacts that would secure the ligand in its binding mode. Consequently, conformation H-7b failed to dock any ligand. Figure 5a also shows the crystal structure with side chain of Arg147 situated between that of H-7a and H-7b. Compared to H-7a, crystal structure docked more number of ligands, yet with weaker binding energies; this can be explained by the fact that in crystal conformation, the position of the side chain of Arg147 is further away from the ligand, therefore weakening the strength of the contact energy. However, allowing more space for conformational search, the chance of a successful dock becomes higher for crystal structure than for H-7a. Clearly, a small orientation of a single loop can create a large deviation of up to 18 Å on the position of its side chains and consequently determine the fate of a dock.

Figure 5b shows the aligned conformations with their corresponding ligands with strongest binding energy. All successful docks from the atomistic ANM case are shown here. It clearly illustrates that in all conformations, the portion of cyclosporin A (residues 1, 2, 3, 9, 10, 11) which interacts with CypA preserves its

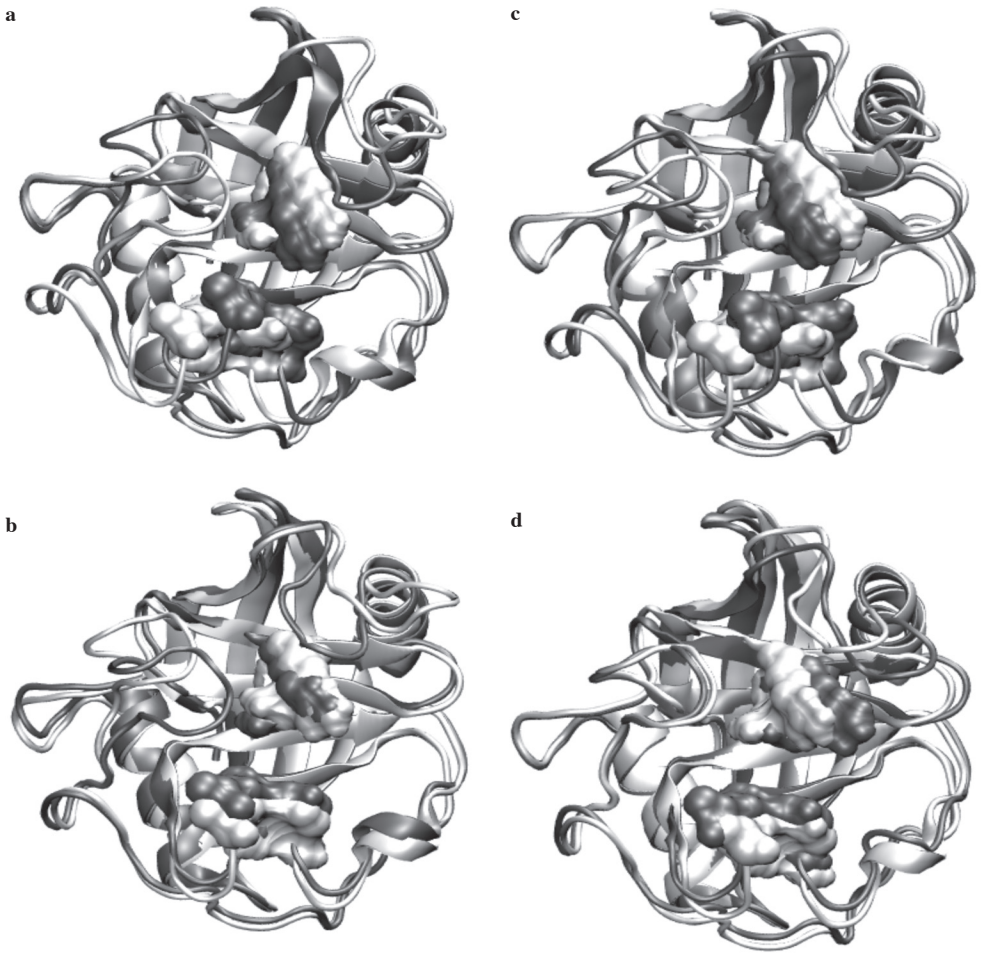

Figure 4: Binding pocket for four modes highlighted with critical residues (Arg54, Gln62, Asn101, Ala102, and His125) involved in the interaction between protein and the ligand, represented in surface representation and colored in white and gray for open and closed form, respectively. (**a**) H-1a (open)/H-1b (closed), (**b**) H-5a (open)/H-5b (closed), (**c**) 2a (open)/2b (closed), and (**d**) 3a (open)/3b (closed) (The last two are generated from the coarse-grained ANM).

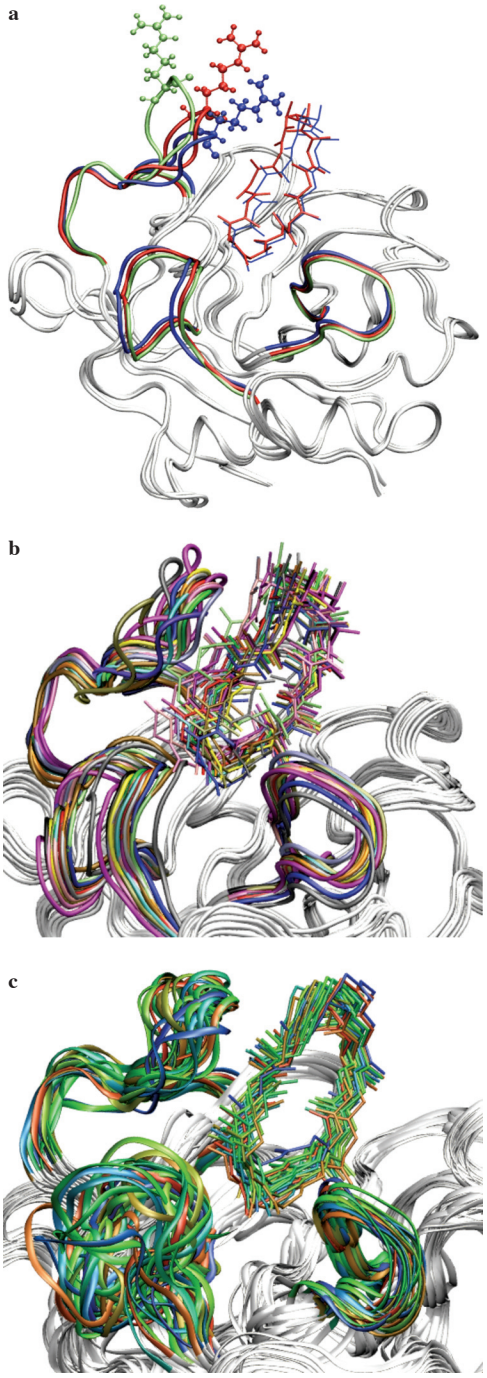

Figure 5: (**a**) Conformer H-7a, H-7b, and the crystal structure with flexible loop regions and Arg147 colored in blue, green, and red, respectively. (**b**) Alignment of generated CypA conformers with their corresponding docked ligand CsA. Both low and high DF conformers are shown. (**c**) Alignment of CsA/CypA conformers obtained from NMR studies (data taken from PDB code: 3cys.pdb).

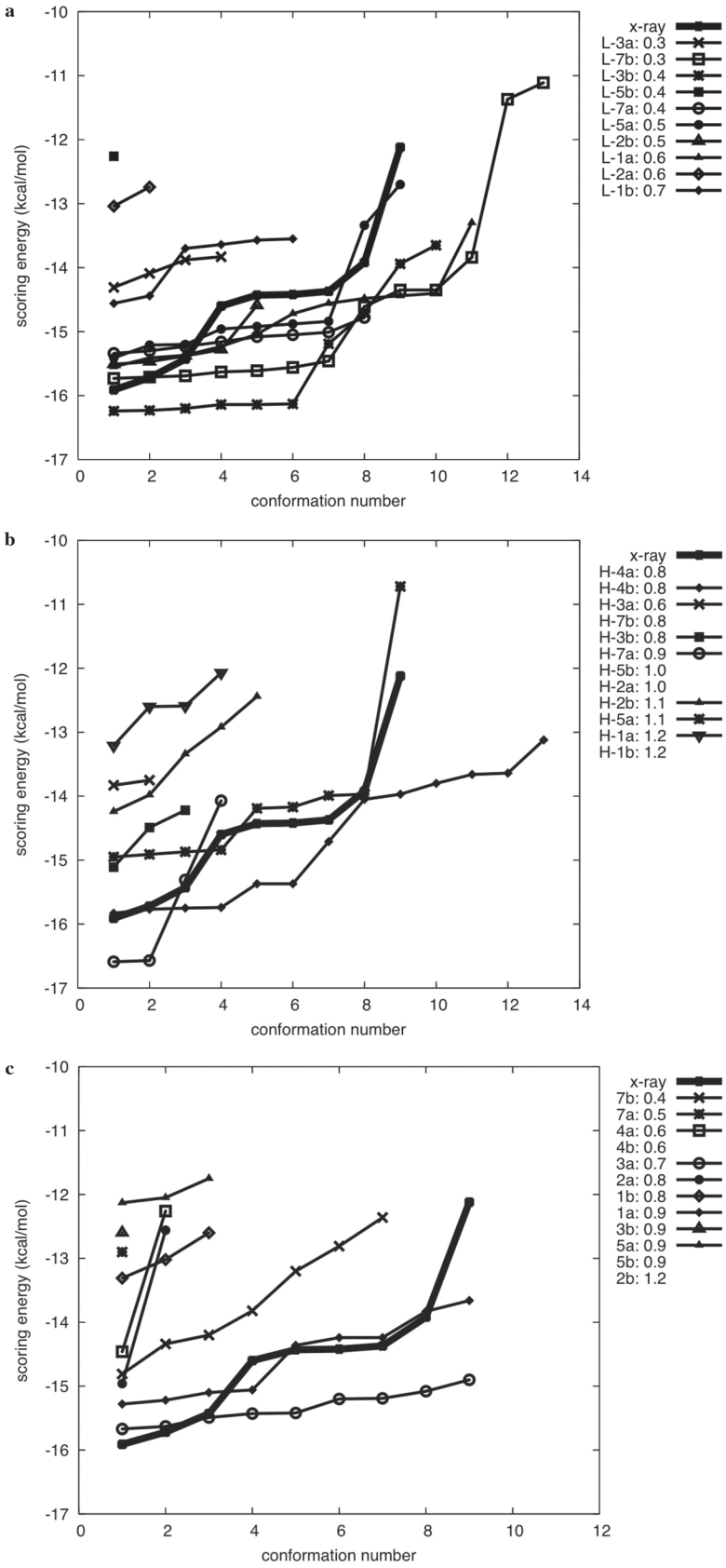

Figure 6: Scoring energy of correctly docked ligands for protein conformers generated from (**a**) atomistic ANM with low deformation, (**b**) atomistic ANM with high deformation, (**c**) coarse-grained ANM. The crystal structure is shown in bold lines in all three figures. Conformers shown without symbols in the right legend could not dock any ligand.

position in the binding cavity, while the orientation of the portion protruding out from the CypA surface, also known as the effector loop, varies in different protein conformations. These results are in good agreement with the ensemble determined from NMR studies shown in Figure 5c (51). Similar results are obtained for the coarse-grained conformers (not shown).

For every conformer tested, correctly docked ligands are sorted by their corresponding scoring energy, which is composed of intermolecular and internal components only and plotted in Figure 6. For the group of atomistic ANM conformers with low deformation (Fig. 6a), where all ten conformers were successfully docked, an average of 6.9 ± 3.9 ligands are correctly docked for each conformer. Six out of ten conformers achieved average binding interactions equal to or stronger than that of the crystal structure. The set of atomistic ANM conformers with high deformation (Fig. 6b) was less successful than the first group due to their higher RMSD values, yet highly favorable binding conformations were achieved. Equally successful docking results were obtained for coarse-grained ANM conformations as shown in Figure 6c, where both the number of successful docks and their corresponding scoring energy were within acceptable ranges compared to atomistic ANM conformers. However, for all three sets, it is difficult to observe any correlation between the success rate of docking and the energies of the conformers in their ligand-free state, because the difference between their energy values does not exceed 3%.

Concluding Remarks

A reverse-mapping technique is employed to generate realistic conformers by deforming the native structure along low-frequency ANM modes (first seven modes) and performing subsequent energy minimization using implicit solvation. Both atomistic and coarse-grained representations of the network model are used to generate the harmonic modes. With this procedure, both low- and high-resolution ANM can generate distinct low-energy conformations for cyclophilin A. The energies of most conformers are found to be comparable or lower than that of the x-ray structure. These conformers are also validated in terms of torsional angle preferences and atom-atom contacts using the MolProbity server (49, 50).

For another crucial structural validation and exploration of better binding modes, all conformers are subjected to an ensemble docking experiment. Compared to the crystal structure, the number of cyclosporin conformations correctly bound to the protein well exceeds the expectations, several of them having stronger binding modes, considering the fact that the protein was held rigid and the number of allowed rotatable bonds in the ligand was substantial for a typical docking experiment. In addition, conformations generated from coarse-grained ANM modes were equally successful in producing successfully docked ligands. Also, docking with alternative conformations allowed us to explore the binding site in more detail, such as the binding pocket's shape and special interactions that become determining factors for the success of a dock.

Our results encourage the potential use of reverse-mapped ANM conformers in docking experiments. The approach of using multiple conformations as an equivalent to flexible protein docking, was first introduced by Kuntz and coworkers who have used ensembles of NMR and X-ray structures as target structures (52). However, the ensemble of conformers either determined by experiments or generated by MD simulations may be limited and may not represent the whole range of a protein's global conformational variations for large systems. The new methodology presented here was successful in generating an ensemble of distinct low-energy conformations with significant fluctuations in the backbone of the protein's loop regions near the binding site, providing plenty of flexibility. As a result, a considerable amount of protein conformations have produced acceptable binding modes in a reasonable time frame due to the computational efficiency of elastic network

model. In the future, we plan to apply this technique to sample conformations for proteins that show large conformational changes upon binding with the aim of further validating its use in docking experiments.

Cavasotto *et al*. (17) have recently used elastic network model to incorporate loop flexibility in ligand docking to cAMP-dependent protein kinase. Several relevant modes (falling in the range from 144th to 660th modes) that serve for the relatively localized loop rearrangements at the binding site of protein kinase have produced reliable conformers for ensemble docking. Our work also based on elastic network models shows that realistic flexible conformers can be generated using the most collective seven modes of motion for CypA, derived from either atomistic or coarse-grained ANM.

Even though most flexible dockings allow only side chains to be flexible, it has become crucial to take into account the flexibility of the protein's main-chain for accuracy in docking experiments. The method's success presented in this work leads the way to study proteins with major conformational variations, up to 10 Å, and potentially determine better binding modes that would lead to novel stronger inhibitors. For this aim, successive deformation and energy minimization cycles along a single or multiple modes may need to be performed. Specifically, we used this approach to generate realistic conformers for triosephosphate isomerase (pdb code: 8TIM), which is a homodimeric enzyme of almost 500 residues (results not shown here). Here, it was possible to obtain conformations by successive deformations along the first mode, which corresponds to the closure direction of the important catalytic loop 6, in conformity with our previous MD simulations (52).

Based on the satisfactory results of low-resolution ANM conformers, our ultimate goal is to combine this reverse-mapping technique with the mixed resolution elastic network models (54,55), where the parts of the structure with functional importance, such as catalytic site or interacting interfaces, are modeled at a higher resolution than the remainder of the structure, so that flexible conformations of large proteins and supra-molecular assemblies can be generated for ensemble docking within a shorter time frame and without loosing too much accuracy. High-throughput virtual drug screening can also benefit from an accurate representation of the conformational ensemble in order to increase the number of hits with fewer false positives, which would save money and time in drug discovery and development.

Figure Preparation

Figures 1, 3-5 are created by the graphical visualization program VMD (53).

Acknowledgments

This work has been partially supported by TUBITAK Project 104M247, EU-FP6-ACC-2004-SSA-2 contract no. 517991, and Bogazici University BAP Project (08A507).

Supplementary Materials

Supplemental materials can be acquired, free of charge, from the contact author or from Adenine Press for $50.

References and Footnotes

1. Koshland, D. E. *Proc Natl Acad Sci USA 44*, 98-104 (1958).
2. Daniel, R. M., Dunn, R. V., Finney, J. L., Smith, J. C. *Ann Rev of Biophys Biomol Struct 32*, 69-92 (2003)
3. Ma, B., Kumar, S., Tsai, C. J., Nussinov, R. *Protein Eng 12*, 713-720 (1999).
4. Tsai, C. J., Ma, B., Nussinov, R. *Proc Natl Acad Sci USA 96*, 9970-9972 (1999).
5. Ma, B., Shatsky, M., Wolfson, H. J., Nussinov, R. *Protein Sci 11*, 184-197 (2002).

6. Gsponer, J., Christodoulou, J., Cavalli, A., Bui, J. M., Richter, B., Dobson, C. M., Vendruscolo, M. *Structure 16*, 736-746 (2008).
7. Lange, O. F., Lakomek, N. A., Fares, C., Schroder, G. F., Walter, K. F., Becker, S., Meiler, J., Grubmuller, H., Griesinger, C., de Groot, B. L. *Science 320*,1471-1475 (2008).
8. Carlson, H. A. *Curr Opin Chem Biol 6*, 447-452 (2002).
9. Halperin, I., Ma, B., Wolfson, H., Nussinov, R. *Proteins 47*, 409-443 (2002).
10. Claussen, H., Buning, C., Rarey, M., Lengauer, T. *J Mol Biol 308*, 377-395 (2001).
11. Vajda, S., Camacho, C. J. *Trends in Biotechnology 22*, 110-116 (2004).
12. Kuntz, I. D., Blaney, J. M., Oatley, S. J., Langridge, R., Ferrin, T. E. *J Mol Biol 161*, 269-288 (1982).
13. Rarey, M., Kramer, B., Lengauer, T., Klebe, G. *J Mol Biol 261*, 470-489 (1996).
14. Jones, G., Willett, P., Glen, R. C., Leach, A. R., Taylor, R. *J Mol Biol 267*, 727-748 (1997).
15. Leach, A. R. *J Mol Biol 235*, 345-356 (1994).
16. Bastard, K., Prevost, C., Zacharias, M. *Proteins 62*, 956-969 (2006).
17. Cavasotto, C. N., Kovacs, J. A., Abagyan, R. A. *JACS 127*, 9632-9640 (2005).
18. Zacharias, M. *Proteins 54*, 759-767 (2004).
19. Gray, J. J., Moughon, S., Wang, C., Schueler-Furman, O., Kuhlman, B., Baker, R., Baker, D. *J Mol Biol 33*, 281-299 (2003).
20. Meiler, J., Baker, D. *Proteins 65*, 538-548 (2006).
21. Cavasotto, C. N., Abagyan, R. A. *J Mol Biol 337*, 209-225 (2004).
22. Kamiya, N., Yonezawa, Y., Nakamura, H., Higo, J. *Proteins 70*, 41-53 (2008).
23. Nakajima, N., Nakamura, H., Kidera, A. *J Phys Chem B 101*, 817-824 (1997).
24. Bahar, I., Rader, A. J. *Curr Opin Struct Biol 15*, 586-592 (2005).
25. Atilgan, A. R., Durell, S. R., Jernigan, R. L., Demirel, M. C., Keskin, O., Bahar, I. *Biophys J 80*, 505-515 (2001).
26. Tirion, M. M. *Phys Rev Lett 77*, 1905-1908 (1996).
27. Ma, J. P. *Structure 13*, 373-380 (2005).
28. Hinsen, K. *Proteins 33*, 417-429 (1998).
29. Schneidman-Duhovny, D., Nussinov, R., Wolfson, H. J. *Proteins 69*, 764-773 (2007).
30. May, A., Zacharias, M. *Proteins 70*, 794-809 (2008).
31. Tama, F., Sanejouand, Y. H. *Protein Eng 14*, 1-6 (2001).
32. Yang, L., Guang, S., Jernigan, R. L. *Biophys J 93*, 920-929 (2007).
33. Tirion, M. M. *Phys Rev Lett 77*, 1905-1908 (1996).
34. Doruker, P., Jernigan, R. L., Bahar, I. *J Comp Chem 23*, 119-127 (2002).
35. Schreiber, S. L., Crabtree, G. R. *Immunol Today 13*, 136-142 (1992).
36. Zavodsky, M. I., Lei, M., Thorpe, M. F., Ray, A. R., Kuhn, L. A. *Proteins 57*, 243-261 (2004).
37. Jacobs, D. J., Rader, A. J., Kuhn, L. A., Thorpe, M. F. *Proteins 44*, 150-165 (2001).
38. Tama, F., Gadea, F. X., Marques, O., Sanejouand, Y. V. *Proteins 41*, 1-7 (2000).
39. Case, D. A., Darden, T. A., Cheatham, T. E. I., Simmerling, C. L., Wang, J., Duke, R. E., Luo, R., Merz, K. M., Wang, B., Pearlman, D. A. AMBER, Version 8. University of California, San Francisco (2004).
40. 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 Comput Chem 26*, 1668-1688 (2005).
41. Duan, Y., Wu, C., Chowdhury, S., Lee, M. C., Xiong, G., Zhang, W., Yang, R., Cieplak, P., Luo, R., Lee, T. *J Comput Chem 24*, 1999-2012 (2003).
42. Hawkins, G. D., Cramer, C. J., Truhlar, D. G. *Chem Phys Lett 246*, 122-129 (1995).
43. Hawkins, G. D., Cramer, C. J., Truhlar, D. G. *J Phys Chem 100*, 19824-19839 (1996).
44. Srinivasan, J., Trevathan, M. W., Beroza, P., Case, D. A. *Theor Chem Acc 101*, 426-434 (1999).
45. Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K., Olson, A. J. *J Comp Chem 19*, 1639-1662 (1998).
46. Goodford, P. J. *J Med Chem 28*, 849-857 (1985).
47. Kallen, J., Mikol, V., Taylor, P., Walkinshaw, M. D. *J Mol Biol 283*, 435-449 (1998).
48. Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Jr, Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., Kollman, P. A. *J Am Chem Soc 117*, 5179-5197 (1995).
49. Lovell, S. C., Davis, I. W., Arendall, III, W. B., de Bakker, P. I. W., Word, J. M., Prisant, M. G., Richardson, J. S., Richardson, D. C. *Proteins 50*, 437-450 (2003).
50. Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, III, W. B., Snoeyink, J., Richardson, J. S., Richardson, D. C. *Nucleic Acids Research 35*, W375-W383 (2007).
51. Spitzfaden, C., Braun, W., Wider, G., Widmer, H., Wuthrich, K. *J Biomol NMR 4*, 463-482 (1994).
52. Knegtel, R. M., Kuntz, I. D., Oshiro, C. M. *J Mol Biol 266*, 424-440 (1997).
53. Cansu, S., Doruker, P. *Biochemistry 47*, 1358-1368 (2008).
54. Kurkcuoglu, O., Jernigan, R. L., Doruker, P. *QSAR & Combinat Sci 24*, 443-448 (2005).
55. Kurkcuoglu, O., Jernigan, R. L., Doruker, P. *Polymer 45*, 649-657 (2004).
56. Humphrey, W., Dalke, A., Schulten, K. *J Molec Graphics 14*, 33-38 (1996).

Date Received: December 20, 2008

Communicated by the Editor
Ramaswamy H. Sarma