

# The intrinsic dynamics of enzymes plays a dominant role in determining the structural changes induced upon inhibitor binding

Ahmet Bakan and Ivett Bahar<sup>1</sup>

Department of Computational Biology, School of Medicine, University of Pittsburgh, 3064 BST3, 3501 Fifth Avenue, Pittsburgh, PA 15213

Edited by Ken A. Dill, University of California, San Francisco, CA, and approved July 8, 2009 (received for review April 17, 2009)

**The conformational flexibility of target proteins continues to be a major challenge in accurate modeling of protein–inhibitor interactions. A fundamental issue, yet to be clarified, is whether the observed conformational changes are controlled by the protein or induced by the inhibitor. Although the concept of induced fit has been widely adopted for describing the structural changes that accompany ligand binding, there is growing evidence in support of the dominance of proteins' intrinsic dynamics which has been evolutionarily optimized to accommodate its functional interactions. The wealth of structural data for target proteins in the presence of different ligands now permits us to make a critical assessment of the balance between these two effects in selecting the bound forms. We focused on three widely studied drug targets, HIV-1 reverse transcriptase, p38 MAP kinase, and cyclin-dependent kinase 2. A total of 292 structures determined for these enzymes in the presence of different inhibitors and unbound form permitted us to perform an extensive comparative analysis of the conformational space accessed upon ligand binding, and its relation to the intrinsic dynamics before ligand binding as predicted by elastic network model analysis. Our results show that the ligand selects the conformer that best matches its structural and dynamic properties among the conformers intrinsically accessible to the protein in the unliganded form. The results suggest that simple but robust rules encoded in the protein structure play a dominant role in predefining the mechanisms of ligand binding, which may be advantageously exploited in designing inhibitors.**

anisotropic network model | conformational flexibility | ensemble of structures | pre-existing equilibrium | principal component analysis

**T**he dynamic nature of proteins plays a critical role in molecular recognition. Understanding the determinants of ligand-recognition and -binding dynamics is a major challenge with impact on drug discovery. Yet, progress in this field has been impeded by the complexity and specificity of interactions, the multiplicity of conformations accessible under equilibrium conditions, and insufficient data on the structure and energetics of protein–ligand interactions.

Two different models have been proposed for explaining the conformational changes observed between the bound and unbound forms of proteins. The classical view, which dates back to the original work of Koshland in 1958, proposes an induced fit mechanism whereby ligand binding induces conformational changes on the target protein. Such an onset of conformational changes could be plausible on a local scale, i.e., slight rearrangements in side chain reorientations or even transitions between isomeric states could be triggered by the ligand. However, the more cooperative changes observed in other complexes, including concerted rearrangements of entire domains, have challenged this classical concept. The second, alternate view, pioneered by Monod, Wyman, and Changeux (MWC model), has gained broad acceptance in the last decade, supported by experimental and computational studies (1–10), and consistent with the accessibility of a host of conformational substates under

native state conditions. Accordingly, the protein samples an ensemble of conformations (preexisting equilibrium), a fraction of which is predisposed to recognize and bind a particular ligand (conformational selection). Therefore, observed structural rearrangements would not occur if it were not for the predisposition or intrinsic dynamics of the protein to fluctuate between multiple conformers including those prone to readily bind the ligand (7).

A number of more recent studies suggest a more complex interplay between intrinsic dynamics and ligand-induced motions. For example, Okazaki and Takada reported that stronger and long-range interactions favor induced fit, whereas shorter-range interactions favor conformational selection (11). Even if binding occurs via conformational selection, additional rearrangements may be induced to stabilize the complex (6, 12). And although protein–protein interactions may be strongly affected by their intrinsic dynamics, it is unclear which effect, intrinsic dynamics vs. induced dynamics, plays a dominant role, in protein–small molecule interactions, which may entail in many cases highly specific, localized interactions. Sullivan and Holyoak argued for example that the presence of a lid at the binding site implies an induced fit mechanism (13), and folding upon binding is commonly observed in intrinsically disordered protein segments (14).

With the rapid accumulation of multiple liganded structures for a given protein in the Protein Data Bank (PDB) and with the development of analytical models for rapid estimation of intrinsic dynamics, we are now in a position to (*i*) critically examine sets of conformations assumed by the same protein in the presence of different ligands and (*ii*) compare these conformational changes to those predicted for the unliganded protein using simplified, physics-based models. Although such comparisons between experimental and computational data may be obscured by the heterogeneities of the accessible conformations and uncertainties in atomic coordinates, there exist powerful methods to extract dominant patterns from complex data. With regard to experimental data, principal component analysis (PCA) is an old but powerful method to unveil the principal variations in structure. An excellent application is the recent examination of the ensemble of ubiquitin X-ray structures complexed with different substrates, compared with the ensemble of NMR models determined by residual dipolar coupling measurements (9). This study showed that the conformational changes assumed in different complexes, and those observed for the isolated protein in solution show close overlap, and essentially represent displacements along a well-defined (combined)

Author contributions: I.B. designed research; A.B. performed research; A.B. analyzed data; and A.B. and I.B. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

<sup>1</sup>To whom correspondence should be addressed. E-mail: bahar@pitt.edu.

This article contains supporting information online at [www.pnas.org/cgi/content/full/0904214106/DCSupplemental](http://www.pnas.org/cgi/content/full/0904214106/DCSupplemental).

principal mode of deformation intrinsically favored by the unbound protein.

As to structural dynamics, again a classical approach to retrieve dominant modes of motion is normal mode analysis (NMA) (15, 16). NMA has seen a revival in recent years with the realization that highly simplified models, such as the anisotropic network model (ANM) (17, 18), can be used to efficiently predict global modes of motions. These motions are characterized by a high degree of collectivity, and usually lie at the lowest frequency end of the mode spectrum. They are insensitive to structural details or underlying force field, but defined by the overall architecture, or topology of interresidue contacts in the native structure (16, 19–20). Applications of NMA are becoming increasingly popular in modeling protein-drug interactions (21–23).

In the present study, we focus on three proteins widely studied as drug targets, HIV-1 reverse transcriptase (RT), p38 mitogen-activated protein (MAP) kinase (p38), and cyclin-dependent kinase 2 (Cdk2). We have  $\approx 300$  structures deposited in the PDB for these three enzymes. This large amount of structural data permits us to perform an extensive comparative analysis of the conformational space being accessed while binding different ligands, and its relation to the intrinsic dynamics of the protein, predicted by the ANM analysis of the unliganded structure. The questions we ask are: how heterogeneous are the structures of the same protein in different complexes? Can we describe them in terms of a few dominant modes accessible under equilibrium conditions, which can be extracted by PCA of the set of structures? To what extent the protein selects from this preexisting equilibrium when binding the ligand; or to what extent does the particular ligand induce a substrate-specific rearrangement? Our results provide insights into the determinants of structural changes that accompany ligand binding. Essentially, the protein topology offers a few well defined, energetically favorable, mechanisms/modes of structural rearrangements along preferential mode axes; and the ligand selects the one that best matches its structural and dynamic properties. As a further test, we analyzed ensembles of NMR models determined for ubiquitin (9) and calmodulin (CaM) (8), to verify that the conformational space sampled by these proteins in solution, in the absence of ligand-binding, is consistent with ANM-predicted global modes. The results suggest that simple but robust rules control ligand binding, which may be systematically exploited in designing drugs.

## Results and Discussion

Results are presented for RT, p38, and Cdk2, using the datasets, I, II, and III, comprised of 112 (RT), 74 (p38) and 106 (Cdk2) X-ray crystallographic structures, respectively. The complete list of PDB identifiers may be found in the [supporting information \(SI\) Table S1](#), and the corresponding distributions of RMSDs may be seen in the [Fig. S1](#). The RMSD distributions for the examined NMR ensembles are presented in the [Fig. S2](#). Below is an outline of the approach adopted for each dataset. More details can be found in the *Methods* and the *SI Text* and [Fig. S3](#).

The goal is to compare two sets of data: experimental structural data for the same protein in different functional forms, including various ligand-bound forms; and computational data predicted using a representative unliganded structure in the dataset.

The experimental structural data are analyzed as follows: (i) The ensemble of structures are superimposed using the Kabsch algorithm in an iterative procedure (see *SI Text*), and mean positions  $\langle \mathbf{R}_i \rangle = [\langle x_i \rangle \langle y_i \rangle \langle z_i \rangle]^T$  are determined for  $\alpha$ -carbons  $1 \leq i \leq N$  (or those with known coordinates), (ii) Departures from their mean positions,  $\Delta \mathbf{R}_i^s = [\Delta x_i^s \Delta y_i^s \Delta z_i^s]^T$  (where  $\Delta x_i^s = x_i^s - \langle x_i \rangle$ ) are organized in a  $3N$ -dimensional deformation vector  $\Delta \mathbf{R}^s$  (where  $(\Delta \mathbf{R}^s)^T = [(\Delta \mathbf{R}_1^s)^T (\Delta \mathbf{R}_2^s)^T \dots (\Delta \mathbf{R}_N^s)^T]$ ), for all structures,  $s$ , in the dataset; and their cross-correlations, averaged over the entire set are combined in a  $3N \times 3N$  covariance matrix

$\mathbf{C}$ , and (iii)  $\mathbf{C}$  is diagonalized to determine the principal modes of structural variations,  $\mathbf{p}^{(i)}$ , observed in experiments. The principal modes ( $m$  of them, for an ensemble of  $m < 3N-6$  structures) are rank-ordered: PCA mode 1 (PC1),  $\mathbf{p}^{(1)}$ , refers to the direction of maximal variance, succeeded by PC2, etc. Of interest is to view the distribution of dataset structures in the subspace spanned by PC1 and PC2, which permit us to discriminate, or cluster, the conformations based on their most distinctive structural similarities and/or dissimilarities.

The computational data are generated using the ANM (see *Methods*). In this case,  $\mathbf{C}$  is simply the inverse of the  $3N \times 3N$  Hessian matrix  $\mathbf{H}$  (17, 18, 24) (see the *SI Text*). Eigenvalue decomposition of  $\mathbf{H}$  gives the ANM modes  $\mathbf{u}^{(1)}, \mathbf{u}^{(2)}, \dots, \mathbf{u}^{(3N-6)}$ , intrinsically favored by the native contact topology. Each eigenvector represents a normalized direction of motion away from the original energy minimum in the  $3N$ -dimensional space. The easiest (or energetically least expensive) direction is, by definition,  $\mathbf{u}^{(1)}$ . This is because the associated uphill curvature of the energy surface or effective force constant (eigenvalue  $\lambda_l$ ) is the smallest.

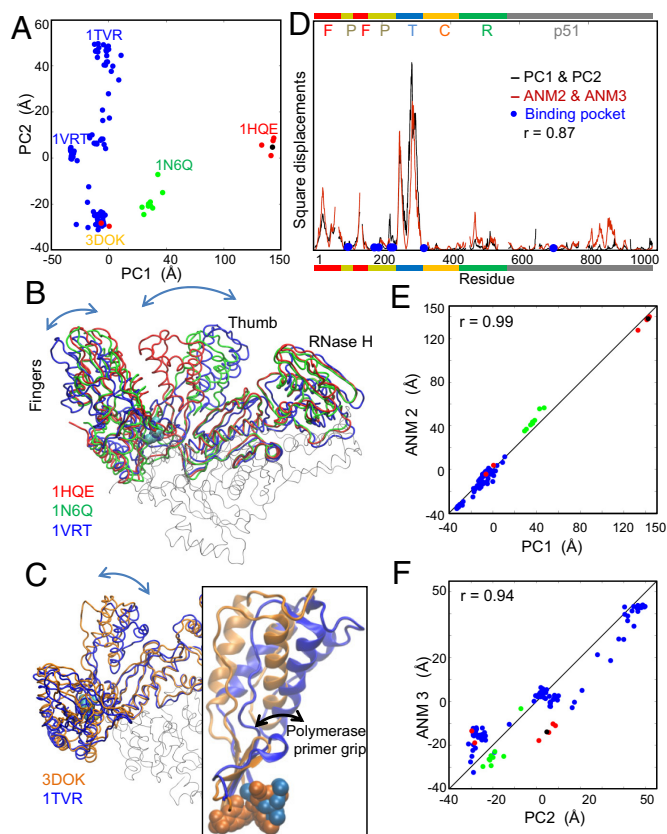
Our objective is to compare the functional variations in structures observed experimentally, and those expected from a physical theory and method (ANM) based on native contact topology. We focus on the top-ranking PCA modes and top-ranking ANM modes, which reflect the dominant features in each group of results. In all three proteins, we will show how the ensembles of conformations observed in experiments (in the presence of different ligands) may be explained by the intrinsic dynamics of the protein (in the absence of ligands). These findings lend support to the view that the “functional changes in structure” predominantly obey ‘structure-encoded preferences’. Alternatively, one might argue that structures have evolved to preferentially sample functional changes in conformation. We now proceed to detailed analyses of the three cases.

**HIV-1 RT.** HIV-1 RT is a multifunctional enzyme composed of two subunits, p66 and p51 (25). p66 contains the polymerase and RNase H domains. The polymerase domain is described as a right hand containing fingers, palm, thumb, and connection subdomains. The connection subdomain links the polymerase and RNase H domains. The p51 subunit bears only the polymerase domain, comprised of the same subdomains, arranged in a different tertiary structure. Non-nucleoside RT inhibitors (NNRTIs) are known to act allosterically by interfering with the functional motions of p66 fingers and thumb upon binding a pocket at a global hinge site (25–28).

**PCA Results.** Fig. 1A shows the projection of RT structures onto the subspace spanned by the first two principal axes, PC1 and PC2, determined for the examined dataset ([Table S1](#)). The points therein represent 112 RT structures. These two PCA modes were found to account for 71% of the total variance in structure. Notably PC1 provides a clear separation of the structures into three clusters according to the types of ligands.

The structural variation represented by PC1 is shown in Fig. 1B. The most distinctive feature is the large movement of the thumb. More careful examination reveals the anti-correlated displacements of the fingers and thumb of  $\approx 10\text{\AA}$  and  $20\text{\AA}$  at their most exposed regions, respectively. The fluctuations along this mode contribute by 47% to the total variance, and span a cumulative displacement of  $176\text{\AA}$  along this axis. Note that this value refers to the cumulative displacement summed over all residues (see *Methods* and [Fig. S3](#)). The tendency of RT to sample conformations along this mode in the absence of ligands is evidenced by site-directed spin labeling experiments (29) and supported by ENMs (26, 27) and MD simulations (10).

PC2 describes the out-of-plane fluctuations of the thumb (Fig. 1C), which would not be obvious from comparisons of arbitrarily chosen structures. These fluctuations are approximately orthog-



**Fig. 1.** Results for HIV-RT. (A) Projection of 6 unliganded (red), 97 NNRTI-bound (blue), 8 dsDNA/RNA-bound (green), and 1 ATP-bound (black) RT structures onto PC1 and PC2. (B) Structural variation along PC1, illustrated using selected structures labeled in A. (C) Structural variation along PC2. *Inset* shows a closer view of the thumb subdomain. Inhibitors are shown by the same color as the corresponding RT conformation. (D) Comparison of the weighted sum of square displacements along PC1 and PC2, with those predicted along ANM modes 2 and 3. (E) Projections of the 112 structures onto PC1 and ANM2 directions. (F) Projections onto PC2 and ANM3.

onal to those described by PC1. Resulting structural differences are illustrated in Fig. 1C using two structures separated by 81 Å. Fig. 1C *Inset* shows a close up view of the thumb. The thumb and the polymerase primer grip (part of the palm), move together as a rigid body.

As a further test, we analyzed the NNRTI-bound subset. The PC1 in this case was found to be almost identical (correlation coefficient of 0.99) to the PC2 of the complete set, and contributed by 50% to the total variance (Table S2). It was also interesting to note that the PC1 from the complete ensemble did not have a counterpart in the NNRTI-bound subset. This supports the view that NNRTI binding depresses the anticorrelated fluctuations of the fingers and thumb, and stimulates thumb fluctuations in an orthogonal direction. This observation is in parallel with the previously proposed view that NNRTI inhibition is achieved by imparting a change in the direction of the thumb movements (26, 27).

How does a small molecule perturb the global dynamics of such a large structure? The answer lies in the location of the NNRTI binding pocket. As shown in the displacement profile in Fig. 1D, the binding pocket residues show minimal, if any, variations in their positions in these two dominant PCA modes. ANM calculations presented below also confirm that the NNRTI binding residues are severely constrained in the global modes of RT. Perturbation of such a severely constrained region in the

**Table 1.** Overlap between PCA and ANM modes

PCA modes (fractional contribution)	ANM modes		
	ANM1	ANM2	ANM3
HIV-RT			
PC1 (0.47)	0.25	0.89	0.11
PC2 (0.24)	0.47	0.08	0.63
PC3 (0.09)	0.42	0.06	0.15
p38			
PC1 (0.29)	0.39	0.05	0.71
PC2 (0.16)	0.79	0.21	0.22
PC3 (0.11)	0.06	0.57	0.05
Cdk2			
PC1 (0.23)	0.36	0.73	0.09
PC2 (0.16)	0.06	0.05	0.26
PC3 (0.12)	0.48	0.09	0.34

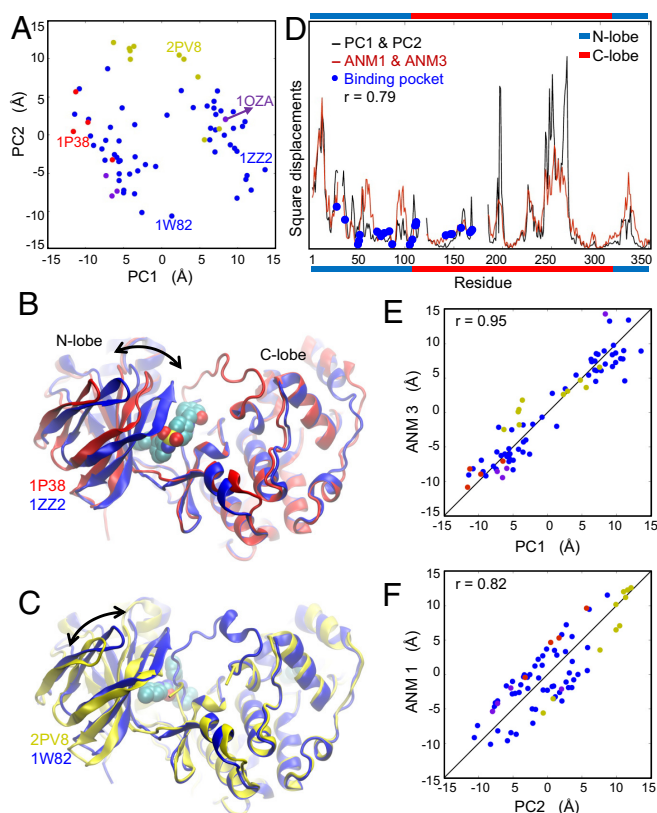
global modes has a dramatic effect, hence the inhibition of functional movements by NNRTI binding at this critical site.

**Comparison with ANM Results.** Fig. 1D displays the joint effect of ANM modes 2 and 3. These two ANM modes were found to yield the highest correlation (among all ANM modes) with PC1 and PC2, respectively (see Table 1). ANM mode 1, however, refers to the anticorrelated fluctuations of fingers subdomain and RNase H domain. Comparison with PCA modes showed that this mode shows a weak correlation (0.52) with the PC5. The directions of first three ANM modes are shown in Fig. S4A. As a direct comparison of the structural changes along PC1 and motions induced in ANM mode 2 (ANM2), we examined the level of correlation between the projections of the structures onto these two collective displacement directions. Fig. 1E displays the results. Strikingly, the structures perfectly align along these two axes (correlation coefficient of 0.99), demonstrating the equivalence of the predicted (ANM2) and experimentally observed (PC1) global modes. Similarly, by projecting the structures onto ANM3, and its PCA counterpart, PC2, we find a correlation coefficient of 0.94, again supporting the view that the most distinctive structural changes assumed by the NNRTI-bound RTs simply originate from intrinsically favorable ANM modes. When put together, these results suggest that RT samples conformations predisposed to NNRTI binding, and NNRTI binding shifts RT dynamics from one mode to another, both being intrinsically favored.

**p38 MAP Kinase.** The p38 MAP kinases are serine/threonine kinases activated in response to external stress, regulate the production of proinflammatory cytokines, and hence serve as targets in the treatment of inflammatory diseases (30). The structure and interactions of p38s with different inhibitor classes are well characterized. Their dynamics, however, is not as well understood and poses challenges in inhibitor docking (31).

**PCA Results.** The results from the PCA are presented in Fig. 2. Fig. 2A displays the distribution of 74 structures of p38 isoform  $\alpha$ . The p38 structure has a canonical kinase fold composed of a  $\beta$ -sheets rich N-terminal lobe (N-lobe) and an  $\alpha$ -helix rich C-terminal lobe (C-lobe) (32). The catalytic site, the binding site for competitive inhibitors, is the cleft between these two lobes. Example structures from different subgroups are displayed in Fig. 2B and C to illustrate the major structural changes along PC1 and PC2, respectively. PC1 refers to anticorrelated movements of the two lobes, as shown by the superimposition of an unliganded (red) and an inhibitor-bound (blue) structures in panel B. These movements map to a separation of >25 Å between the two conformers along PC1 axis (Fig. 2A). PC2, however, involves twisting motion of N-lobe with respect to the C-lobe illustrated by a glucoside- and an inhibitor-bound structure (Fig. 2C). The size of the motions along PC2 is comparable





**Fig. 2.** Results for p38 MAP kinase. (A) Projection of 4 unliganded (red dots), 56 inhibitor-bound (blue), 10 glucoside-bound (yellow), and 4 peptide-bound (violet) p38 structures onto PC1 and PC2. Distant structures along PC1 and PC2 are selected to illustrate in the respective B and C the structural variations represented by these PCs. (D) Square displacements of residues along PC1 and PC2, compared with those driven by ANM modes 1 and 3. (E) Projections of the 74 structures onto PC1 and ANM3. (F) Projections onto PC2 and ANM1.

with that along PC1 (Fig. 2A). The global backbone changes observed in the ensemble are well described by the first two PCs. Notably, although these modes represent only 2 degrees of freedom of a total of 963, they account for 45% of the variance in the backbone structure observed in the dataset (Table 1), and reconfigurations along these two directions allow for reducing the average RMSD with respect to the mean structure from  $1.0 \pm 0.3 \text{ \AA}$  to  $0.53 \pm 0.17 \text{ \AA}$ .

Fluctuations along PC1 determine the exposure of the binding cleft. In the unliganded state, the cleft is wide open, presumably to facilitate ATP/inhibitor recognition and optimal interaction. Upon ligand binding, the cleft closes down. The closure of the cleft increases the packing interactions with the bound molecule. Movements along PC1 are therefore functionally relevant. In one of the peptide-bound structures, p38 is complexed with its primary substrate MAPK-activated protein kinase 2 (MK2), which is located  $20 \text{ \AA}$  away from the unliganded p38 along PC1 (labeled as 1OZA in Fig. 2A). Other peptide-bound structures contain short stretches, corresponding to docking sites of p38 regulators/substrates, with comparably much less impact on p38. The conformational change observed when p38 is crystallized with MK2 potentially affects its function (33). Fluctuations along PC2, however, change the relative positioning of the atoms lining the cleft between the lobes, affecting in particular the  $\beta$ -strands 1–3 and the connecting hairpins.

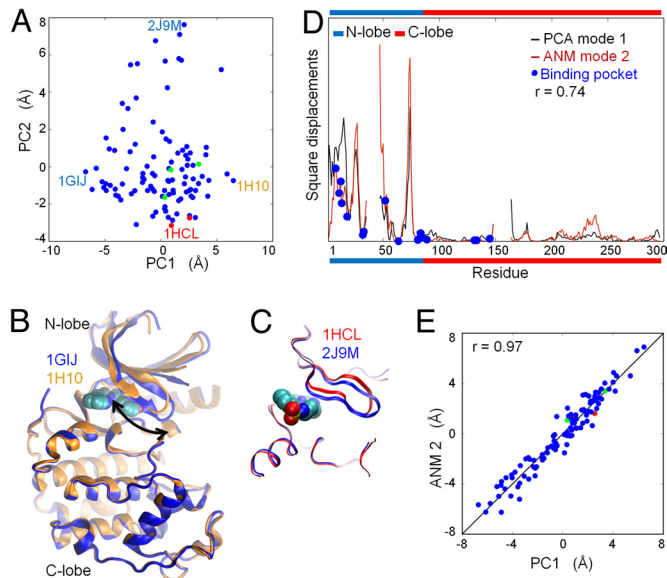
Fig. 2D displays the square displacements of residues resulting from the weighted contributions of movements along PC1 and PC2. The binding pocket residues are marked by the blue dots.

Some of them, corresponding to hinge sites in these global modes, are extremely constrained, and others show moderate displacement. The structural diversity of inhibitors may be a primary reason for the conformational heterogeneity along these modes. The average pairwise similarity among the 55 compounds in the dataset is  $0.40 \pm 0.15$  based on standard cheminformatics metrics. Apparently, the movements of the lobes allow for optimizing their interactions with the small-molecules, and hence the poor results in *in silico* cross-docking experiments, when the target backbone is assumed to be rigid (31).

**Comparison with ANM Results.** The correlations between the lowest three ANM modes and the top-ranking three PC directions are listed in Table 1. The counterparts of PC1 and PC2 are found to be ANM3 and ANM1, with respective correlation coefficients of 0.71 and 0.79. ANM2, a twisting motion of the two lobes, was found to correspond to PC3 with a coefficient 0.57 (Fig. S4B illustrates these three ANM modes). The displacements of residues induced by ANM modes 1 and 3 are compared in Fig. 2D with those driven by PC1 and PC2. The two profiles agree with a correlation coefficient of 0.79. Also important is the directionality of ANM predicted fluctuations. Hence, we projected the ensemble onto these ANM modes and distributions along corresponding PCs. Distributions of the structures along the PC1-ANM3 (Fig. 2E) and PC2-ANM1 (Fig. 2F) yielded correlation coefficients of 0.95 and 0.82, respectively. This excellent correspondence underscores the robustness of the low-frequency modes, and shows their functional importance in accommodating the binding of structurally diverse inhibitors and the substrate MK2 which presents a much larger interface, at the P38 active site cleft.

**Cdk2.** Cdks are serine/threonine kinases involved in the regulation and progression of the cell cycle. Cdk activity is regulated by activator (cyclin family) or inhibitor (INK and Cip families) protein binding and phosphorylation (34). In cyclin- or Ink-bound structures, the N-lobe adopts a host of distinct conformations intrinsically favored by Cdk (6). This conformational flexibility impacts the ATP/small-molecule pocket, and hence needs to be considered in designing drugs (31).

**Results from PCA.** We have assembled an ensemble of 106 Cdk2 structures. This set was more narrowly distributed compared with the previous two cases (average RMSD of  $0.50 \pm 0.14 \text{ \AA}$  with respect to the reference structure; Fig. S1) and therefore the structural variations may not be large and precise enough to classify them into distinctive PC modes. Yet, the first two PCA modes were able to account for 39% of the variance in the dataset (Table 1). The projection of Cdk2 structures onto PC1 and PC2 showed a diffuse distribution, not clearly distinguishing the different bound forms, although the two unliganded structures clustered together at a distinctive end of the subspace (Fig. 3A). We selected the Cdk2 structures that fall in the extremes of this distribution to visualize the conformational differences along these two PCA directions. PC1 describes the twisting motion of N and C lobes, which is comparable with p38 movements along PC2. This is illustrated by two structures that are  $13.2 \text{ \AA}$  apart along this axis (Fig. 3B). In Fig. 3A, the unliganded structures fall close to the center on PC1 coordinate, which indicates that twisting motion occurs in either direction. The structural variation described by the PC2, however, was localized at chain termini and flexible loops. Fig. 3C illustrates such a local movement for the so-called glycine loop that lines the binding cleft. The two structures in Fig. 3C are  $10.8 \text{ \AA}$  apart when projected onto PC2. The heterogeneity of the ensemble along these modes presumably originates from the physicochemical diversity of bound inhibitors. The ensemble contained 100 different compounds with an average pairwise similarity metric of  $0.43 \pm 0.14$ .

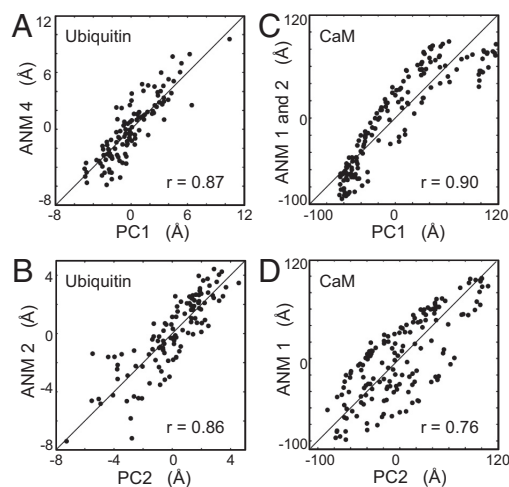


**Fig. 3.** Results for Cdk2. (A) Projection of 2 unliganded (red), 3 ATP-bound (green), and 101 inhibitor-bound (blue) Cdk2 structures onto PC1 and PC2. (B) Structural variation along PC1. (C) Structural variation along PC2. (D) Comparison of the square displacements of residues along PC1 and ANM2. (E) Projection of 106 Cdk2 structures onto PC1 and ANM2.

**Comparison with ANM Predictions.** The top-ranking ANM modes predicted for the unliganded structure (1HCL) are displayed in Fig. S4C. Among them, ANM2 yields a correlation of 0.73 with PC1 (Table 1). The square displacements of residues along these modes are compared in Fig. 3D. Some of the binding pocket residues are located at the minima of these profiles, while the rest are located in highly mobile regions of N-lobe. Those at the minima include Val-64, Phe-80, Asn-132, and Ala-144-Asp-145, which are buried deep into the cleft between the two lobes, while those exhibiting moderate-to-high flexibility are at the periphery of the cleft. This indicates that fluctuations along this mode allow for the positioning of recognition site residues, while other functional residues, such as Asn-132 and Asp-145 that coordinate the metal atom for the catalytic reaction, and the catalytic residues Asp-127 and Lys-129 are almost fixed. The projections of the ensemble of structures onto ANM2 and PC1 yield a correlation of 0.97 (Fig. 3E). These suggest that despite the minimal changes in structure, the observed structural heterogeneity is not random, but geared toward the intrinsically favored mode of motion. The changes along the experimentally observed PC2, however, seem to be rather local and would rather be viewed as induced by inhibitor binding.

**NMR Ensembles.** The ensembles of NMR models deposited for ubiquitin (9) and for CaM (8) have been pointed out to populate conformations comparable to those observed in their substrate-bound forms. The question is: how do principal modes of structural change extracted from the PCA of these ensembles correlate with those predicted by the ANM?

Results are presented in Fig. 4. The projection of the NMR models onto PC1 and PC2 are shown in both cases to exhibit a close agreement with one or two ANM global modes. We note that CaM samples a very broad conformational space in its unliganded form (Fig. S2) as the N- and C-terminal domains (NTD and CTD) are connected by a helix that is readily flexed. We examined the principal modes accessible to the NTD and CTD in relation to their ANM counterparts, as a further comparison. Correlations in the range 87% to 94% were observed in both cases (Fig. S5A–D), demonstrating that the ANM provides a good



**Fig. 4.** Comparison of the principal modes from NMR ensembles with ANM-predicted global modes. The projections of NMR models for ubiquitin (panels A and B), and CaM (C and D) onto PC1 and PC2 are compared with the projections onto ANM global modes. ANM calculations are performed for the model that has the lowest RMSD with respect to all others in each ensemble. See Fig. S2 and Fig. S6 for the respective RMSD distributions and ANM modes.

representation of the variability of the structures even within domains, provided that they are sufficiently decoupled.

Finally, we also analyzed the complex CaM-MLCK with a myosin light chain kinase (MLCK) peptide, which further confirmed that principal modes derived from NMR models are in accord with ANM predictions (correlations of 0.88 and 0.77 in Fig. S5 E and F). The results for all examined NMR ensembles are compiled in Table S3 and Table S4, and illustrated in Fig. S6. The cumulative overlaps between subsets of 3, 6, and 20 ANM modes with the top three PCs (Table S3) further support the consistent correlation between the subspace of conformations seen in the experimental structures and those predicted by computations.

## Conclusion

We presented here a detailed analysis of conformational changes experimentally observed for three enzymes upon binding a broad range of ligands, and those predicted by simple physics-based models based on their native fold contact topology. In all three cases, the first principal mode of structural change, PC1, observed in experiments exhibits a correlation of  $0.78 \pm 0.1$  with a top-ranking mode (ANM1-ANM3) intrinsically preferred by the unliganded protein. If, we further consider the correlation between subsets of PC and ANM modes (Table S3), we see that PC1 is accounted for with a cumulative overlap of  $0.85 \pm 0.06$  by the top three ANM modes in all three cases. Similarly, the principal modes of structural variability observed in NMR ensembles exhibited remarkable correlations with top-ranking ANM modes.

We note that three PCs describe between 50% (Cdk2) and 80% (RT) of the structural variance observed in the datasets of enzymes. The structures display further heterogeneities beyond those along the first three PCs, specific to particular inhibitors, which would rather fall in the category of induced changes, succeeding the initial recognition driven by target proteins' intrinsic dynamics.

The top-ranking ANM modes are by definition collective modes of motions and they are also known to be highly robust against sequence and structure variations. The strong correlation of experimentally observed structural changes with these ANM modes demonstrates the collectivity and robustness of the structural changes undergone by these enzymes upon binding their ligands, even if the sizes of these concerted motions are small in many cases.



What is the physical basis of ANM modes? These modes are purely based on native contact topology, or geometry. No specific interactions, other than the absence/presence of interresidue contacts (equivalent to an excluded volume effect) are taken into consideration. The basic driving potential is entropic in origin, i.e., the directions of motions predicted by the ANM are those entropically favored, where the uphill curvature away from the original energy minimum is minimal. The close correspondence with experimentally observed deformations suggests that the conformational changes undergone for ligand binding are dominated by entropic effects.

On a practical side, that the structures assumed by the target proteins for recognizing their substrates comply with the global modes of motions ‘predictable’ by simple models such as the ANM opens the way to possible generation of representative ensembles of conformers with the ANM. Adequate consideration of target proteins’ backbone flexibility has been a major bottleneck in computer-aided drug design. A proposed solution has been to dock ligands onto multiple receptor conformations (ensemble docking) (35). Sets of crystal structures, each bound to a distinct ligand (36), or, ensembles of NMR structures (37) have been considered to this aim. However, both sets may provide incomplete, if not inaccurate, information. The absence of a PCA counterpart for the ANM1s predicted for RT or for Cdk2 may signal such deficiencies. Such shortcomings are likely to be alleviated by resolving multiple X-ray structures for a given protein (38). As to NMR structures, the physical meaning of the ensemble of models deposited in the PDB, whether they represent a “mathematical solution” that satisfies experimental (and semiempirical) constraints, or physically accessible conformations, has yet to be established (39). The correlations of the global ANM modes with the PCA modes extracted from NMR

ensembles (40), consistent with present observations for ubiquitin and CaM, and with the structural variations observed in X-ray structures (6, 41) are in support of the use of ANM for consolidating existing structural data, or gaining insights into potential inhibitor binding mechanisms.

## Methods

**PCA of Experimental Structural Data.** Principal modes were obtained by decomposing the covariance matrix  $C$  (see *SI Text*) for each dataset as  $C = \sum_{i=1}^{3N} \sigma_i \mathbf{p}^{(i)} \mathbf{p}^{(i)T}$  where  $\mathbf{p}^{(i)}$  and  $\sigma_i$ , are the respective  $i$ th eigenvalue and eigenvector of  $C$ ,  $\sigma_1$  corresponding to the largest variance component. The fractional contribution of  $\mathbf{p}^{(i)}$  to structural variance in the dataset is given by  $f_i = \sigma_i / \sum_j \sigma_j$  where the summation is performed over all  $m$  components. The square displacement of the  $k$ th residue along  $\mathbf{p}^{(1)}$  and  $\mathbf{p}^{(2)}$  (or PC1 and PC2) is  $(\Delta R_k)^2|_{1 \leq i \leq 2} = \text{tr}\{\{\sum_{i=1}^{3N} \sigma_i \mathbf{p}^{(i)} \mathbf{p}^{(i)T}\}_{kk}\}$  where the subscript  $kk$  denotes the  $k$ th diagonal element (a  $3 \times 3$  matrix) of the  $3N \times 3N$  matrix enclosed in square brackets.

**Projection of Conformations onto the Subspace Spanned by the PCs.** The projection of a given conformational change  $\Delta R^s$  onto  $\mathbf{p}^{(i)}$  is found from  $c_i^s = (\Delta R^s)^T \mathbf{p}^{(i)}$ . The points in the Figs. 1A, 2A, and 3A, and those along the abscissa of Fig. 1E and 2E and F, and 3E, represent the projections  $c_1^s$  and  $c_2^s$  of each structure  $s$  onto PC1 and PC2. In the extreme case of  $(\Delta R^s)^T$  perfectly aligned along  $\mathbf{p}^{(i)}$ ,  $c_i^s = \|\Delta R^s\|$ , where the double bars designate the magnitude. See Fig. S3 for an illustration.

**ANM Analysis and Overlap with PCA Modes.** In the ANM, the Hessian  $H$  is decomposed to yield  $3N-6$  nonzero eigenvalues  $\lambda_i$  and corresponding eigenvectors  $\mathbf{u}^{(i)}$ , i.e.,  $H = \sum_{i=1}^{3N-6} \lambda_i \mathbf{u}^{(i)} \mathbf{u}^{(i)T}$ . ANM covariance is  $C_{\text{ANM}} = H^{-1}$  such that  $1/\lambda_i$  is the counterpart of the PCA  $\sigma_i$ , and  $\mathbf{u}^{(i)}$  is the counterpart of  $\mathbf{p}^{(i)}$ . The overlap between ANM and PCA modes is given by the correlation cosine  $O_{ij} = \mathbf{p}^{(i)} \cdot \mathbf{u}^{(j)}$  (24).

**ACKNOWLEDGMENTS.** This work was supported in part by National Institutes of Health Grant 5R01GM086238-02.

- Ma B, Kumar S, Tsai CJ, Nussinov R (1999) *Protein Eng Des Sel* 12:713–720.
- Kern D, Zuiderweg ER (2003) The role of dynamics in allosteric regulation. *Curr Opin Struct Biol* 13:748–757.
- Tang C, Schwieters CD, Clore GM (2007) Open-to-closed transition in apo maltose-binding protein observed by paramagnetic NMR. *Nature* 449:1078–1082.
- Mittermaier A, Kay LE (2006) New tools provide new insights in NMR studies of protein dynamics. *Science* 312:224–228.
- Showalter SA, Bruschweiler R (2007) Quantitative molecular ensemble interpretation of NMR dipolar couplings without restraints. *J Am Chem Soc* 129:4158–4159.
- Tobi D, Bahar I (2005) Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state. *Proc Natl Acad Sci USA* 102:18908–18913.
- Bahar I, Chennubhotla C, Tobi D (2007) Intrinsic dynamics of enzymes in the unbound state and relation to allosteric regulation. *Curr Opin Struct Biol* 17:633–640.
- Gsponer J, et al. (2008) A coupled equilibrium shift mechanism in calmodulin-mediated signal transduction. *Structure* 16:736–746.
- Lange OF, et al. (2008) Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. *Science* 320:1471–1475.
- Ivetac A, McCammon JA (2009) Elucidating the inhibition mechanism of HIV-1 non-nucleoside reverse transcriptase inhibitors through multicopy molecular dynamics simulations. *J Mol Biol* 388:644–658.
- Okazaki K, Takada S (2008) Dynamic energy landscape view of coupled binding and protein conformational change: Induced-fit versus population-shift mechanisms. *Proc Natl Acad Sci USA* 105:11182–11187.
- James LC, Tawfik DS (2005) Structure and kinetics of a transient antibody binding intermediate reveal a kinetic discrimination mechanism in antigen recognition. *Proc Natl Acad Sci USA* 102:12730–12735.
- Sullivan SM, Holyoak T (2008) Enzymes with lid-gated active sites must operate by an induced fit mechanism instead of conformational selection. *Proc Natl Acad Sci USA* 105:13829–13834.
- Turjanski AG, Gutkind JS, Best RB, Hummer G (2008) Binding-induced folding of a natively unstructured transcription factor. *PLoS Comput Biol* 4:e1000060.
- Kitao A, Go N (1999) Investigating protein dynamics in collective coordinate space. *Curr Opin Struct Biol* 9:164–169.
- Cui Q, Bahar I (2006) *Normal Mode Analysis: Theory and applications to biological and chemical systems* (CRC, Boca Raton, FL).
- Atilgan AR, et al. (2001) Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys J* 80:505–515.
- Eyal E, Yang LW, Bahar I (2006) Anisotropic network model: Systematic evaluation and a new web interface. *Bioinformatics* 22:2619–2627.
- Tama F, Brooks CL (2006) Symmetry, form, and shape: Guiding principles for robustness in macromolecular machines. *Annu Rev Biophys Biomol Struct* 35:115–133.
- Nicolay S, Sanejouand YH (2006) Functional modes of proteins are among the most robust. *Phys Rev Lett* 96:078104.
- Cavasotto CN, Kovacs JA, Abagyan R (2005) Representing receptor flexibility in ligand docking through relevant normal modes. *J Am Chem Soc* 127:9632–9640.
- Floquet N, et al. (2006) Normal mode analysis as a prerequisite for drug design: Application to matrix metalloproteinases inhibitors. *FEBS Lett* 580:5130–5136.
- May A, Zacharias M (2008) Protein-ligand docking accounting for receptor side chain and global flexibility in normal modes: Evaluation on kinase inhibitor cross docking. *J Med Chem* 51:3499–3506.
- Tama F, Sanejouand YH (2001) Conformational change of proteins arising from normal mode calculations. *Protein Eng* 14:1–6.
- Kohlstaedt LA, Wang J, Friedman JM, Rice PA, Steitz TA (1992) Crystal structure at 3.5 Å resolution of HIV-1 reverse transcriptase complexed with an inhibitor. *Science* 256:1783–1790.
- Bahar I, Erman B, Jernigan RL, Atilgan AR, Covell DG (1999) Collective motions in HIV-1 reverse transcriptase: Examination of flexibility and enzyme function. *J Mol Biol* 285:1023–1037.
- Temiz NA, Bahar I (2002) Inhibitor binding alters the directions of domain motions in HIV-1 reverse transcriptase. *Proteins* 49:61–70.
- Zhou Z, Madrid M, Evansek JD, Madura JD (2005) Effect of a bound non-nucleoside RT inhibitor on the dynamics of wild-type and mutant HIV-1 reverse transcriptase. *J Am Chem Soc* 127:17253–17260.
- Kensch O, Restle T, Wohrl BM, Goody RS, Steinhoff HJ (2000) Temperature-dependent equilibrium between the open and closed conformation of the p66 subunit of HIV-1 reverse transcriptase revealed by site-directed spin labelling. *J Mol Biol* 301:1029–1039.
- Kumar S, Boehm J, Lee JC (2003) p38 MAP kinases: Key signalling molecules as therapeutic targets for inflammatory diseases. *Nat Rev Drug Discov* 2:717–726.
- Cavasotto CN, Abagyan RA (2004) Protein flexibility in ligand docking and virtual screening to protein kinases. *J Mol Biol* 337:209–225.
- Wang Z, et al. (1997) The structure of mitogen-activated protein kinase p38 at 2.1-Å resolution. *Proc Natl Acad Sci USA* 94:2327–2332.
- White A, Pargellis CA, Studts JM, Werneburg BG, Farmer BT (2007) Molecular basis of MAPK-activated protein kinase 2:p38 assembly. *Proc Natl Acad Sci USA* 104:6353–6358.
- Pavletich NP (1999) Mechanisms of cyclin-dependent kinase regulation: Structures of Cdk, their cyclin activators, and Cip and Ink4 inhibitors. *J Mol Biol* 287:821–828.
- Totrov M, Abagyan R (2008) Flexible ligand docking to multiple receptor conformations: A practical alternative. *Curr Opin Struct Biol* 18:178–184.
- Barril X, Morley SD (2005) Unveiling the full potential of flexible receptor docking using multiple crystallographic structures. *J Med Chem* 48:4432–4443.
- Huang SY, Zou X (2007) Efficient molecular docking of NMR structures: Application to HIV-1 protease. *Protein Sci* 16:43–51.
- Levin EJ, Kondrashov DA, Wesenberg GE, Phillips GN, Jr. (2007) Ensemble refinement of protein crystal structures: Validation and application. *Structure* 15:1040–1052.
- Best RB, Lindorff-Larsen K, DePristo MA, Vendruscolo M (2006) Relation between native ensembles and experimental structures of proteins. *Proc Natl Acad Sci USA* 103:10901–10906.
- Yang LW, Eyal E, Bahar I, Kitao A (2009) Principal component analysis of native ensembles of biomolecular structures (PCA-NEST): Insights into functional dynamics. *Bioinformatics* 25:606–614.
- Yang L, Song G, Carriquiry A, Jernigan RL (2008) Close correspondence between the motions from principal component analysis of multiple HIV-1 protease structures and elastic network modes. *Structure* 16:321–330.