Representing Receptor Flexibility in Ligand Docking through Relevant Normal Modes
Claudio N. Cavasotto,*† Julio A. Kovacs,‡ and Ruben A. Abagyan‡

Abstract: Inspired by the current representation of the ligand–receptor binding process, a normal-mode-based methodology is presented to incorporate receptor flexibility in ligand docking and virtual screening. However, the systematic representation of the deformation space grows geometrically with the number of modes, and furthermore, midscale loop rearrangements like those found in protein kinase binding pockets cannot be accounted for with the first lowest-frequency modes. We thus introduced a measure of relevance of normal modes on a given region of interest and showed that only very few modes in the low-frequency range are necessary and sufficient to describe loop flexibility in cAMP-dependent protein kinase. We used this approach to generate an ensemble of representative receptor backbone conformations by perturbing the structure along a combination of relevant modes. Each ensemble conformation is complexed with known non-native binders to optimize the position of the binding-pocket side chains through a full flexible docking procedure. The multiple receptor conformations thus obtained are used in a small-scale virtual screening using receptor ensemble docking. We evaluated this algorithm on holo and apo structures of cAMP-dependent protein kinase that exhibit backbone rearrangements on two independent loop regions close to the binding pocket. Docking accuracy is improved, since the ligands considered in the virtual screening docked within 1.5 Å to at least one of the structures. The discrimination between binders and nonbinders is also enhanced, as shown by the improvement of the enrichment factor. This constitutes a new step toward the systematic integration of flexible ligand–flexible receptor docking tools in structure-based drug discovery.

1. Introduction
Computer-aided drug discovery through ligand docking-based virtual screening is already a key component in the lengthy and costly process of developing new drugs.1,2 The capability to correctly predict ligand–protein interactions is fundamental to any accurate docking algorithm and the necessary starting point for any reliable virtual screening protocol. Molecular flexibility is critical for a thorough understanding of the principles that govern ligand binding to proteins. Structural changes in the receptor upon ligand binding is a very common phenomenon;3 hence, ignoring this effect might have a strong impact on ligand docking4 and virtual screening. The implications of protein flexibility in drug discovery have been recently reviewed.5,6 Dealing with receptor flexibility is in many cases crucial to accurately predict the orientation and interactions of ligands within the binding pocket.7 So the big challenge ahead is to routinely incorporate flexibility considerations into structure-based drug discovery in an affordable computing time.

Early attempts to include protein flexibility in ligand docking, such as soft docking8 and partial side-chain flexibility9,10 among others, have been reviewed.11–13 However, most of these methods do not include backbone rearrangements, and explicit sampling of side chains is an unsurmountable drawback in virtual screening of large chemical libraries.

The use of multiple receptor conformations (MRCs) (either experimental or in silico generated) in ligand docking seems probably the best choice to date.11,13 However, questions about how this should be efficiently accomplished are open. An important advantage of this approach is that the structural space of the binding pocket can be represented in a virtual screening process, even in the case of loop displacements.14 In the early days, experimental structures were used to derive average interaction grids that became targets for rigid-ligand docking.15 The FlexE method incorporates flexibility by combining rigid-

‡ The Scripps Research Institute.
† Molsoft LLC.
Receptor Flexibility in Ligand Docking

A method that recombines multiple flexible regions into a discrete set of receptor conformations and that scales linearly with receptor flexibility has been recently presented and evaluated. The ensemble of structures collected from molecular dynamics (MD) were also used to determine the correct geometry of a ligand–enzyme complex by docking and energy evaluation of selected complex structures. In another interesting development, snapshots taken from MD simulations were employed to construct a receptor-based pharmacophore model of the HIV-1 integrase that was validated experimentally. This method has been recently extended to the case of the apo HIV-1 protease. Normal-mode analysis (NMA) has also been used to study the induced fit in HIV integrase, and very recently a method of molecular dynamics and harmonic dynamics has been proposed and tested to study the docking of HIV-1 protease and its ligand.

There is growing evidence about the relationship between pre-existing conformations of the receptor unbound state in equilibrium and structural changes upon ligand binding. In those cases that cannot adjust to the rigid “lock-and-key” model, ligand binding is seen as a combination of a conformational selection stage of partially fitting structures followed by minor structural rearrangements within the complex (induced-fit stage). This led very recently to the formulation of the ligand binding process in terms of linear response theory, whereby the response of structures to ligand binding is predicted using the conformational ensemble of the unbound (unperturbed) state. It has also been demonstrated that equilibrium conformations of a protein can be represented using low-frequency normal modes.

On the basis of this evidence, we propose a low-frequency normal-mode-based algorithm to generate MRCs and thus incorporate receptor flexibility in ligand docking and virtual screening. In an attempt to drastically reduce the dimension of the conformational space that grows geometrically with the number of modes considered, we introduce a measure of relevance of normal modes on selected regions of interest known to be important for ligand binding. We found that very few modes are critical and sufficient to represent binding pocket plasticity in protein kinases, and we describe one algorithm to find those modes. Perturbation along the relevant modes followed by full flexible docking of known ligands to optimize positioning of side chains is then used to generate an ensemble of de novo alternative conformations. The structurally different ligand-binding pockets thus generated were used as starting points for a receptor ensemble docking (RED). We tested this methodology in both holo and uncomplexed structures of the cAMP-dependent protein kinase (cAPK, PKA). A small-scale virtual screening showed a significantly better discrimination of binders from nonbinders, which was evident by the improvement in the enrichment factors. Importantly, our procedure did not make custom-fit pockets that only accommodated the ligand used in the optimization. This constitutes a new advance in the challenging task of taking into account protein flexibility in structure-based drug discovery.

2. Results and Discussion

2.1. Foundations of the Normal-Mode-Based Approach in Receptor Ensemble Docking. It has been shown that ignoring protein flexibility in cAPK and other protein kinases is the reason ligands belonging to certain chemical spaces fail to dock correctly using the rigid receptor approach. Our goal is to show that even when only one crystal structure (apo or holo) and a few binders are known, alternative structural conformations can be obtained by distortion along normal modes followed by full flexible docking of non-native ligands. These two steps are inspired in the ligand-binding process outlined in the Introduction: by perturbing along normal modes, we intend to represent equilibrium conformations of the receptor, while docking of known binders with flexible side chains is a way to generate structural rearrangements within the binding pocket (induced-fit stage). The lowest-energy complex should correspond to the structural conformer to be found experimentally. The receptor structures thus obtained could be used in a small-scale virtual screening using RED. We acknowledge that side-chain flexibility and other local motions might be not completely uncoupled from the docking process. In the future we plan to couple continuous changes along relevant modes with the docking step. In this paper, we are making the first step toward that direction by taking into account partial coupling through the side-chain optimization stage with known non-native binders.

We should note, however, that in certain cases like protein kinases, where binding pocket plasticity upon ligand binding is mainly concentrated in the gly-rich and C-term loops, any attempt to reproduce this sort of movements with normal modes should take into account the following:

(a) the first very low-frequency modes are associated with very low-energy large-scale dynamics, and thus do not represent the more localized and intermediate-amplitude loop rearrangement;

(b) the representation of the conformational space grows geometrically with the number of normal modes considered; and

(c) inclusion of low-energy modes not relevant to the region of interest will add unnecessary noise to the energy calculation of the system.

For this reason, we introduced the notion of measure of relevance for each mode on a region of interest, thus limiting the number of modes to be used in generating MRCs.

The main steps of our methodology can be summarized as follows:

Figure 1. Notation used in the definition of the measure of relevance $\rho(n)$.

1. Determination of the relevant normal modes necessary to represent binding pocket flexibility (section 2.2).

2. Generation of a de novo ensemble of MRCs by perturbing the structure along the relevant modes (section 2.3).

3. Side-chain optimization by complexing the receptor conformational ensemble with known binders, followed by global-energy minimization using a flexible ligand–flexible side chains approach (section 2.4).

4. Receptor ensemble docking (RED) against the generated MRCs using a flexible ligand–grid receptor docking procedure, combination of the screening results, and keeping the best rank per compound (this is the merging–shrinking procedure that has been already described and validated (25) (section 2.5).

We tested our methodology on the apo (PDB entry 1JLU) and holo (PDB entry 1FM0) structures of cAPK. After an ensemble of diverse backbone conformations was generated through distortion along relevant normal modes, the alternative structures generated from 1JLU and 1FM0 were complexed with non-native binders staurosporine and balanol, respectively.

It should be noted that these ligands failed to dock using the rigid receptor approach due to structural changes upon ligand binding. Side-chain conformations were then optimized through global-energy minimization. Ligands in their lowest energy state were obtained (see Table 1 for details). Remarkably, these conformations were selected solely on the basis of energy, with an energy-based discrimination $\Delta G > 10$ kcal/mol between the best energy conformations and the first structurally diverse one. The alternative receptor conformations were then used in a flexible ligand–rigid receptor virtual screening, and the results thus obtained were combined with those from the original PDB structures. RMSDs of seeded known binders together with enrichment factors were significantly improved by using this methodology.

2.2. Relevant Modes vs First Low-Frequency Modes. The NMA was performed on crystal structures 1FM0 and 1JLU, based on a simplified spring model using $C_\alpha$ atoms only (see section 3 for details). This approximation, originally developed by Tirion (31) and extended by Hinsen (32) and Bahar (33), was found to reproduce very well slow protein dynamics, while being very fast and less noisy than the all-atom calculation, and has been used in a number of different problems (see ref 34 and references therein). These are the reasons we chose a $C_\alpha$ model rather than an all-atom model. In fact, the computation time was only 12 min on a standard workstation, including the time needed for the computation of the deformability and mobility functions (used in the measure of relevance).

To select the smallest number of normal modes that are most "concentrated on", or "relevant" to the region of interest (not necessarily the first ones with lowest frequency), we introduced a measure of relevance $\rho(n)$ that expresses in relative terms how much each mode $n$ is active on a specific region. We defined two adjacent regions along the chain: region $A$ surrounds the ends of the loop, and region $B$ includes the central part of the loop (Figure 1). The relevance of mode $n$ on the loop is then defined as

$$\rho(n) = \frac{|d_n|^2_{2,A}}{|d|^2_{2,A}} + \frac{|d - d_n|^2_{2,B}}{|d|^2_{2,B}} + \frac{|m_n|^2_{2,B}}{|m|^2_{2,B}} + \frac{|m - m_n|^2_{2,A}}{|m|^2_{2,A}}$$

(1)

where $m_n$ is defined in eq 12, $d_n$ and $d$ are defined in eq 14 of section 3, and

$$|d|^2_{2,A} = (\sum_{j=1}^A d(j)^2)^{1/2}, \text{ etc.}$$

(2)

The first term in eq 1 represents modes that bend the chain near the ends of the loop. The second term tends to exclude those modes that distort the central part of the loop. The third term favors those modes that move the central part of the loop the most, while the last term avoids modes that would move the loop and the surroundings at the same time.

A few words might be in order here regarding the definition of relevance. While all four terms are reasonable, it is, in principle, not clear that all of them are needed in order to detect relevant modes. Hence, we performed a test consisting of dropping one term at a time and comparing the best-ranking modes—according to the various relevance measures obtained—with the modes furnished by a control. (The control is described in section 3.4.1.) The result of this test was that, except when the trial measure was the sum of the second and third terms, the overlap with the control modes worsened with respect to the overlap of the modes given through eq 1. On the other hand, one can easily think of examples of motions where the presence of the fourth term is essential. (In the test cases considered, this did not happen.) Therefore, one needs to include the first term as well.

In Figure 2, the distributions of mode frequencies are displayed. It should be noted that, due to the simplified model used, high-frequency modes from side-chain fast atomic motion are not present. To understand the two-bell shape of the distribution, we performed an analysis with augmented spring constants (data not shown). When $k \to \infty$ (no residue bond stretching or bending), the frequency distribution overlapped very well with the left-bell portion of Figure 2. Thus, the left-bell part corresponds to low-frequency distortions of the dihedral angles of the backbone.

Once the normal modes had been ranked by relevance, the first $s$ modes (after which a significant decrease in $\rho$ was observed) were selected (see Table 1 for details). Remarkably, in each of the receptors considered, less than 10 modes have significant contribution in the loops of interest, thus reducing dramatically the number of modes necessary to represent their conformational space. Furthermore, since the first relevant modes are numbers 144 (1FM0) and 166 (1JLU), it is evident that the first lowest-frequency modes cannot represent this type of midrange loop rearrangement, making it necessary to include modes with higher frequency.

2.3. De Novo Generation of an Ensemble of Structurally Diverse Backbone Conformations. With the $s$ relevant modes, we formed a linear combination of them:

$$u = \sum_{k=1}^s \alpha_k u_k$$

(3)
Conformations per mode: \( k \) (the same for all \( \text{Phys. Re} \))

Possible linear combinations of the \( R \) where each of the coefficients \( \alpha \) used for each mode in the linear combination, the number \( s \) of linear combinations, the number \( \kappa \) of conformations (resulting after clustering of the \( s \) conformations) used in the docking, and the per-mode displacement limit. Each structure is deformed along the \( s \) directions, so that the resulting displacement of any \( C_i \) atom does not exceed \( 1.5 \text{Å} \). When near-duplicate conformations from these \( s \) are eliminated, \( \kappa \) conformations are obtained, which are then minimized and used as templates in receptor ensemble docking. Actual residue numbers in the PDB structure.

| receptor | \( R_1, R_2, R_3, R_4 \) | relevant normal modes | \( s \) | \( \kappa \) | \( d \) | \( d \) |
|----------|-----------------|----------------|-----|-----|-----|
| 1FMO     | 46, 51, 56, 60  | 144, 147, 559, 609, 610, 624, 627, 655, 660 | 9   | 512 | 2   |
| J1LU     | 322, 325, 327, 330 | 166, 168, 176, 333, 336 | 5   | 243 | 3   |

* Shown are the receptor, loop parameters (see Figure 1), the relevant normal modes in the region of interest, the number \( s \) of normal modes used to make linear combinations, the number \( \kappa \) of conformations per mode.

Prior to making these linear combinations, each mode \( \mathbf{u}^k \) is normalized so that \( \max_{1 \leq j \leq k} | \mathbf{u}^k_j | = d_k \), where \( d_k \) is a prescribed displacement limit. Now, the resulting linear combinations may exceed the limit \( d_k \), so each of them (call it \( \mathbf{u} \)) is renormalized in the following way: let \( M = \max_{1 \leq j \leq k} | \mathbf{u}_j | \), if \( M > d_k \), compute

\[
\frac{d_k}{M} (d_k - d_k \left( 1 - e^{-(M-d_k)/(d_k-d_1)} \right) + d_1) \]

and replace \( \mathbf{u} \) by \( f(M)/\mathbf{u} \). The new \( \mathbf{u} \) satisfies \( \max_{1 \leq j \leq k} | \mathbf{u}_j | < d_k \). The second displacement limit \( d_2 \) is set to be 50% larger than \( d_1 \).

Next, each linear combination \( \mathbf{u} \) is applied to the original structure by displacing all atoms of residue \( j \) by \( \mathbf{u}_j \). Since deformation is performed in the Cartesian coordinate space, covalent geometry was restored by tethering and minimizing a 3D structure of ideal-geometry residues to each of the \( \kappa \) generated structures. This procedure is consistent with the ECEPP/3 force field, which considers bond lengths and planar angles fixed. Minimization in the Cartesian coordinates space (or in the complete internal coordinates space) using for this purpose another force field would not be consistent. Once covalent geometry was restored, the structures were tethered to an identical copy and further energy-minimized in vacuo to avoid residue inter-locking and to relieve energy strain, while keeping to a minimum the deviations from the original \( \kappa \) conformations. This was done in successive steps where the tether weight was reduced from 50 kcal/mol Å² to 0, thus avoiding excessive deformation from highly strained regions.

Structures within each ensemble were then superimposed and clustered according to their backbone RMSD in the region of interest using a threshold of 0.8 Å, giving an acceptable balance between number of conformations and their structural diversity. After clustering, the ensemble was reduced from 512 to 5 conformations for 1FMO and from 243 to 3 conformations for J1LU. The final numbers \( \kappa \) of conformations to be used in the docking, as well as other parameters, are listed in Table 1. Figure 3 displays the representative structures thus generated.

It should be clear now that few relevant modes (±10) are the natural way to represent intermediate-scale loop movements (for which selecting the first lowest-frequency modes will fail), thus avoiding the geometric explosion in the number of possible combinations of normal eigenvectors. It is also evident that the noise in the energy calculation of the system will be significantly reduced, since distortion was mostly localized in the region of interest.

2.4. Side-Chain Optimization within the Ligand-Binding Pocket. The optimized structures from the previous step (five
for 1FMO and three for 1JLU) were complexed with non-co-crystallized ligands [1FMO with balanol (complex A) and 1JLU with staurosporine (Complex B)], and the global energy was minimized according to the double-energy scheme biased probability Monte Carlo (BPMC) method (see section 3.2). A solvation energy contribution was included by means of atomic probability Monte Carlo (BPMC) method (see section 3.2). A

for 1FMO and three for 1JLU) were complexed with non-co-crystallized ligands [1FMO with balanol (complex A) and 1JLU with staurosporine (Complex B)], and the global energy was minimized according to the double-energy scheme biased probability Monte Carlo (BPMC) method (see section 3.2). A solvation energy contribution was included by means of atomic probability Monte Carlo (BPMC) method (see section 3.2). A

was kept rigid during the global-energy optimization. Torsional variables associated with side chains in zone II were included in the local energy minimization step of the global-energy optimization, while only those associated with zone I were sampled in the Monte Carlo procedure. Zone II serves as a “buffer” to propagate disturbances in zone I. Since no significant rearrangement is expected within zone III, it is kept rigid during the Monte Carlo step for computing time reasons and relaxed through minimization at the end (see below). Initially, the positional and torsional variables of the ligand were randomly perturbed, and 10 independent parallel simulations were performed for each of the complexes. An upper limit of 3000 steps was set for the local energy minimization stage. The global-energy optimization stage was stopped after 3.5 million energy evaluations as used in similar studies.\(^\text{14}\) During the simulations, conformational stacks of low-energy states\(^\text{37}\) were collected and then merged. To further optimize ligand—receptor contacts, a relaxation step was then performed by a full local energy minimization of the complexes represented in the conformational stack.

\[ G_{\text{proj}} = E_{\text{int}}^{\text{er}} + E_{\text{int}}^{\text{elec}} + G_{\text{II}}^{\text{SASA}} + G_{\text{II}}^{\text{elec}} - TS_{\text{II}} \]  

The first term on the right-hand side represents the interaction energy (van der Waals and hydrogen bond) of zone II with the rest of the system (including its self-energy). The second term represents the torsional energy within zone II, while the last term represents the configurational entropy of zone II. The third

\[ G_{\text{II}} \]  

\[ E_{\text{int}}^{\text{er}} \]  

\[ E_{\text{int}}^{\text{elec}} \]  

\[ G_{\text{II}}^{\text{SASA}} \]  

\[ G_{\text{II}}^{\text{elec}} \]  

\[ TS_{\text{II}} \]  


(38) Totrov, M.; Abagyan, R. *Biopolymers* 2001, 60, 124–133.
symmetry of the Green function. We define the projection of II and III. These two terms have the same value, due to the last two terms represent the interaction energy between zones the electrostatic self-energy of zone II and zone III, while the term is calculated as

\[ \Delta G_{\text{proj}} = G_{\text{II,II}} + G_{\text{II,rest}} \]

where \( q_i \) and \( q_j \) are the charges associated with atomic centers at \( \vec{r}_i \) and \( \vec{r}_j \), and \( G_{\text{RF}}(\vec{r}_i, \vec{r}_j) \) is the reaction field Green function matrix at \( \vec{r}_i, \vec{r}_j \). The \( G_{\text{RF}} \) is independent of the charges and depends only on the geometry of the system and the dielectric constant at each point of space. It represents the electric potential due to the induced charges on the solvent boundary. Moreover, \( G_{\text{RF}}(\vec{r}_i, \vec{r}_j) \) corresponds to the value of the electric potential due to the reaction field at site \( \vec{r}_i \), provided a unit charge is located at \( \vec{r}_j \). By grouping indices in eq 6, it is clearly seen that \( G_{\text{II}}^{\text{elec}} \) can be split into four terms:

\[ G_{\text{II}}^{\text{elec}} = G_{\text{II,II}}^{\text{elec}} + G_{\text{II,rest}}^{\text{elec}} + G_{\text{II,rest}}^{\text{elec}} + G_{\text{rest,II}}^{\text{elec}} \]

where “rest” refers to zone III. The first two terms represent the electrostatic self-energy of zone II and zone III, while the last two terms represent the interaction energy between zones II and III. These two terms have the same value, due to the symmetry of the Green function. We define the projection of \( G_{\text{II}}^{\text{elec}} \) onto zone II as

\[ G_{\text{II}}^{\text{elec}} = G_{\text{II,II}}^{\text{elec}} + G_{\text{II,rest}}^{\text{elec}} \]

\[ = \frac{1}{2} \left( G_{\text{II}}^{\text{elec}} + G_{\text{II,II}}^{\text{elec}} - G_{\text{II,rest}}^{\text{elec}} \right) \]

where eq 7 has been used.

The three \( G_{\text{II}}^{\text{elec}} \) terms on the right-hand side of the second eq 8 were calculated considering the complete set of charges, setting those of zone III to zero, and setting those of zone II to zero, respectively.

In Table 2 we show the RMSD of the top-ranking solutions for the systems considered, along with the relative free energy with respect to the best-ranking complex. It is clearly seen that in the cases considered, \( G_{\text{proj}} \) constitutes a satisfactory measure to discriminate the correct complex geometry solely on the basis of energy, with an energy gap > 10 kcal/mol between the correct pose and the first missed docked one.

2.5. Small-Scale Virtual Screening Using RED: Improvement of RMSD Values and Enrichment Factors. Receptor
not identical, the differences in enrichment factors could be due to the scoring function.

Staurosporine and H89 cannot be docked to the apo structure 1JLU using the rigid receptor approach. However, in the alternative receptor conformation generated with the normal-mode-based procedure, and by docking staurosporine (receptor B), both compounds dock with an RMSD < 1.2 Å, while staurosporine has the best score (see Figure 5). Considering the merged set, it is interesting that 100% of the compounds are docked within 1.5 Å. The enrichment factors are comparable to those obtained with IFREDA, and in the three cases considered, there is an improvement with respect to those of the individual receptors. However, the fact that all of the compounds are docked correctly but only half of them appear in the top 10% of the ranking list could be due to the scoring function. It should also be pointed out that, in the cases displayed in Table 4, there is no “dilution” of the top hits after the merging and shrinking procedure, since the enrichment factors of the merged set are larger than or equal to those corresponding to the individual structures.

3. Computational Methods

3.1. Complex Preparation. Receptor structures of 1FMO and 1JLU were taken from the Protein Data Bank, and hydrogens and missing heavy atoms were added. Peptide PKI-(5–24) was removed from 1FMO. The system was then subjected to a local minimization step. Asn and Gln residues were flipped whenever necessary to optimize hydrogen bonding. Polar hydrogens in the vicinity of the ligand-binding pocket were also optimized.

The 3D structures of cAPK ligands balanol (1BX6), staurosporine (1STC), adenosine (1FMO), and isoquinolinesulfonamides analogues H7 (1YDR), H8 (1YDS), and H89 (1YDT) were taken from their co-crystalized native structures (specified in parentheses), and their correct stereochemistry and formal charges were assigned. The staurosporine amino group and terminal amino substituents in H7, H8, and H89 were regarded as protonated, in agreement with an environment of pH 7.4. Ligands were assigned MMFF atom types and then subjected to global-energy optimization using the MMFF energy terms.

3.2. Energy Evaluation and Minimization. The molecular system is described in the internal coordinate space using a modified version of the ECEPP/3 force field. Entropy and solvation energy terms were added to the in vacuo energy. The stochastic global-energy optimization method consists of (i) random conformational change of free variables according to a predefined probability distribution as described in the biased probability Monte Carlo (BPMC) method; (ii) local energy minimization of analytically differentiable terms, followed by total energy re-evaluation including nondifferentiable terms, like entropy and solvation energy; and (iii) acceptance or rejection on the basis of the Metropolis criterion applied to the total energy. The temperature of the simulations was set to $T = 600$ K. The water

---

Figure 4. Flexible ligand—rigid receptor docking of compounds balanol and H89 to receptor A generated from 1FMO by (i) perturbation of the structure along relevant normal modes and (ii) fully flexible docking of balanol into the ligand-binding pocket. Receptor A is displayed in magenta superimposed onto the native receptors, displayed in green-yellow, of balanol (1BX6, a) and of H89 (1YDT, b). Carbon atoms of docked compounds are displayed in white. Color code: blue, nitrogen; red, oxygen; yellow, sulfur.

Figure 5. Flexible ligand—rigid receptor docking of compounds staurosporine and H89 to receptor B generated from 1JLU by (i) perturbation of the structure along relevant normal modes and (ii) fully flexible docking of staurosporine into the ligand-binding pocket. Receptor B is displayed in magenta superimposed onto the native receptors, displayed in green yellow, of staurosporine (1STC, a) and of H89 (1YDT, b). Carbon atoms of docked compounds are displayed in white. Color code: blue, nitrogen; red, oxygen; yellow, sulfur.

---

dielectric constant was set to $\varepsilon_{\text{water}} = 78.5$. The internal dielectric constant was $\varepsilon_{\text{int}} = 2\varepsilon$ during the global-energy optimization process and $\varepsilon_{\text{int}} = 2$ for electrostatic calculations when solving the Poisson equation.

3.3. Library Preparation and Flexible Ligand–Rigid Receptor Docking. A random chemical library was extracted from the Diverse Set of ChemBridge (ChemBridge, Inc., San Diego, CA). These compounds were prepared in a similar way as the native ligands. Key chemical descriptors of the random library (molecular weight, number of rotatable bonds, number of hydrogen bond donors, and number of hydrogen bond acceptors) overlap with those of cAPK ligands. 14 The random library thus extracted was seeded with cAPK ligands to get a docking library of 1000 compounds.

The flexible ligand–rigid receptor docking algorithm, as implemented in ICM, 39–41 consists of (i) representation of the receptor with six potential energy maps (three for van der Waals, electrostatic, hydrogen bond, and hydrophobic); (ii) global-energy minimization of the flexible ligand in the field of the receptor, so that both the self-energy of the ligand and its interaction with the receptor are optimized in ICM, 39,42 each virtual screening experiment was repeated four times, and the best score for each compound among the four was kept. The scoring function was not optimized for protein kinases. Computing time for ligand docking and scoring was $\sim 1$–2 min on a 700 MHz processor (1 Mb RAM dual-processor node).

The enrichment factor (EF) for a library built with the $n$ top compounds of the ranked library is defined as

\[ \text{EF}(n) = \frac{\text{Hits}_n}{N_n} \times \frac{\text{Hits}_{\text{total}}}{N_{\text{total}}} \]  

and expresses the relative change in the probability of finding a ligand in the focused library when compared to a random pick from the complete library.

3.4. Normal-Mode Analysis. 3.4.1. Protein Vibrational Analysis. We used a reduced C$_N$ harmonic network protein model, 11 whereby the C$_N$ atoms of the molecule are interconnected with springs defined as follows:

\[ C_{ij} = 225 \delta_{ij} + \exp \left( \frac{E_{ij} - E_i^T}{kT} \right) \quad (1 \leq i < j \leq N) \]  

where $\delta_{ij}$ is the Kronecker delta, $i$ and $j$ denote residue numbers, $N$ is the total number of residues, $k$ is the Boltzmann constant, $T$ is the temperature (fixed $T = 600$ K), $E_i$ is the interaction energy between residues $i$ and $j$ (which was computed using the ICM algorithm), 31 I and J denote the residue types of residues $i$ and $j$, respectively ($1 \leq I, J \leq 20$), and

\[ E_i^T = \Bar{E}_{ij} - 1.65 \sigma_i \]  

where $\Bar{E}_{ij}$ and $\sigma_i$ denote the mean and standard deviation of the distribution of nonzero energy values corresponding to the pair of types $I, J$.

The masses $m_i$ of the pseudo-atoms, located at the C$_N$ atom positions, were set to the total mass of the corresponding residues.

The units in eq 10 are kcal/(mol Å$^2$). The value

\[ \text{EF}(n) = \frac{\text{Hits}_n}{N_n} \times \frac{\text{Hits}_{\text{total}}}{N_{\text{total}}} \]  

and expresses the relative change in the probability of finding a ligand in the focused library when compared to a random pick from the complete library.

3.4.2. Mobility and Deformability Functions. The mobility function is given by the classical formula for atomic fluctuations\(^{47}\) (excluding the factor $kT$):

\[ \text{EF}(n) = \frac{\text{Hits}_n}{N_n} \times \frac{\text{Hits}_{\text{total}}}{N_{\text{total}}} \]  

and expresses the relative change in the probability of finding a ligand in the focused library when compared to a random pick from the complete library.

As in a previous work, 38 we view the normal modes as vector fields over the molecule. For any such vector field $u$, we can define the corresponding "conformal tensor" $S_{ij}$ with components

\[ S_{ij} = \frac{1}{2} \left( \frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right) - \frac{1}{3} \delta_{ij} \text{div} u \quad (1 \leq i, j \leq 3) \]  

where $\text{div} u$ denotes the divergence of $u$: $\text{div} u = \sum_1^3 \frac{\partial u_i}{\partial x_i}$.

Note: $u_1$, $u_2$, and $u_3$ stand for the three components of the vector field $u$ as functions of the spatial coordinates $x_1$, $x_2$, and $x_3$. This should not be confused with an expression such as $u_i$, which means "$u$ at the $i$th residue" (a three-dimensional vector).

The tensor $S_{ij}$ describes how the vector field $u$ affects (locally) the shape of the molecule. 46 And it does so in a "relative" way, since it involves only derivatives of $u$. Thus, we call it the relative conformal tensor. Accordingly, the deformability measure given previously\(^{46}\) is called here relative deformability.

In this work we want to define and use an "absolute" deformability measure, which takes into account not only the derivatives of the normal modes but also their amplitudes. The rationale for this is that high-frequency/low-amplitude modes should have a small contribution to the total deformability measure. Therefore, the (relative) scalar deformation function\(^{48}\) corresponding to mode $n$ (after scaling it by its thermal amplitude $\omega_n^2$), $|S_n|$, (the "derivative" of the normal mode), should be multiplied by the wavelength $\lambda_n$ of that mode in order to get the corresponding "absolute" scalar deformation function. The wavelength is not known, but assuming, as a first-order approximation, that the speed of the vibrations is independent of their frequency, we can take (disregarding constant factors) $\lambda_n = \omega_n^{-1}$. Thus, our absolute deformability function $d^2 M \Rightarrow \lambda$ is defined as

\[ d^2 = \sum_{n=1}^{3N} \frac{|S_n|^2}{\lambda_n^2} \approx \sum_{n=1}^{3N} \frac{d^2}{\lambda_n^2} \]  

The method for the numerical computation of the partial derivatives has been previously described.\(^{48}\)

4. Conclusions

Ligand binding to a receptor is currently best described as a selection process of partially fitting structures (conformer

selection stage) followed by minor structural changes upon ligand binding (induced-fit stage). This concept inspired our methodology to incorporate receptor flexibility in ligand docking and virtual screening through a normal-mode-based algorithm to generate multiple receptor backbone conformations, followed by a flexible ligand—flexible side chain docking of non-native ligands. The alternative receptor conformations thus generated could be used for virtual screening using the receptor ensemble docking (RED) approach.

However, intermediate-scale loop motions like those found in the binding pocket of protein kinases cannot be represented by picking the first lowest-energy modes. Furthermore, adequate representation of the conformational space grows geometrically with the number of modes considered. To overcome these limitations, we introduced a measure of relevance that expresses how active a given mode is on a region of interest. In this way, we showed that very few normal modes ($\lesssim10$) are necessary to describe loop flexibility in protein kinases. Remarkably, the relevant modes are not those with first lowest frequency, but are among the low-frequency modes (see Figure 2). We validated our methodology on holo and apo structures of cAPK protein kinase, where loop rearrangement of $\sim2$ Å takes place. Alternative receptor conformations were generated by perturbing the structures along a combination of relevant modes, followed by a flexible ligand—flexible side chain docking of known non-native ligands in order to optimize side-chain conformations. In a second stage, the receptor conformations thus obtained were used as starting points for a virtual screening using RED.

The lowest-energy de novo ligand—receptor complexes generated with our procedure from 1FMO and balanol (complex A), and from 1JLU and staurosporine (complex B), exhibited a ligand RMSD within 1.2 Å compared to the experimental structures 1BX6 and 1STC, respectively, while their $\Delta G$ was lower than 10 kcal/mol with respect to the first misdocked structure (see Table 2). It should be emphasized that balanol and staurosporine cannot dock to 1FMO and 1JLU, respectively, in the rigid receptor approach. The small-scale virtual screening performed against the multiple receptor conformations 1FMO and receptor A (from complex A), 1JLU, and receptor B (from complex B) showed improved enrichment factors when compared to those obtained using a single receptor conformation. Altogether, this indicates that the structural diversity of the pocket was correctly represented. Moreover, most of the known ligands were able to dock within 1.5 Å to each of the multiple receptor ensembles, showing that our procedure does not necessarily make custom-fit pockets that accommodate only those ligands used in the generation of the ensemble structures. Thus, the alternative backbone conformations generated represent actual states of the system.

This methodology is applicable to both holo and apo structures. The use of known binders to optimize the position of side chains has the advantage of reproducing likely binding pockets, introducing however a limitation in the applicability of the method, since it cannot be used as it stands for an ab initio protein de-orphanization. This opens a large space for improvement of our method. Although side-chain flexibility and local motions might be not always uncoupled from the docking process, in this paper we do take into account partial coupling through the side-chain optimization stage with known non-native binders. We plan to couple continuous changes along relevant modes with the docking step in the future.

Virtual screening using RED is a new avenue toward the consideration of protein flexibility in computer-aided drug discovery. In many cases, receptor structural diversity can be successfully represented with a few receptor conformations, either experimentally or in silico generated. Screening against a few structures is computationally affordable, using actual receptor conformations, with the possibility of incorporating diversity from both side-chain and loop rearrangements.

Acknowledgment. We thank Dr. Andrew Orry for helpful discussions.

Supporting Information Available: Complete list of authors for refs 7 and 46. This material is available free of charge via the Internet at http://pubs.acs.org.

JA042260C