

Application of Drug-Perturbed Essential Dynamics/Molecular Dynamics (ED/MD) to Virtual Screening and Rational Drug Design

Rima Chaudhuri,[†] Oliver Carrillo,[†] Charles Anthony Laughton,[‡] and Modesto Orozco^{*,†,§,||}

[†]Joint IRB-BSC Program on Computational Biology, Institute for Research in Biomedicine, Barcelona, Spain

[‡]School of Pharmacy and Centre for Biomolecular Sciences, University of Nottingham, Nottingham, England

[§]National Institute of Bioinformatics and ^{||}Departament de Bioquímica, Facultat de Biologia, University of Barcelona, Barcelona, Spain

Supporting Information

ABSTRACT: We present here the first application of a new algorithm, essential dynamics/molecular dynamics (ED/MD), to the field of small molecule docking. The method uses a previously existing molecular dynamics (MD) ensemble of a protein or protein–drug complex to generate, with a very small computational cost, perturbed ensembles which represent ligand-induced binding site flexibility in a more accurate way than the original trajectory. The use of these perturbed ensembles in a standard docking program leads to superior performance than the same docking procedure using the crystal structure or ensembles obtained from conventional MD simulations as templates. The simplicity and accuracy of the method opens up the possibility of introducing protein flexibility in high-throughput docking experiments.

■ INTRODUCTION

Accommodating protein flexibility in virtual screening is a major challenge^{1–10} in present day drug design. Varying conformations of flexible active site loops, defining the correct conformation from the *apo*-collapsed binding sites, or simply tracing large backbone and side-chains conformational changes induced by ligand binding are some of the dynamic challenges that we are facing in high throughput in silico screenings. The shortcomings of current experimental techniques like X-ray crystallography and nuclear magnetic resonance (NMR) to account explicitly for flexibility have prompted the application of theoretical methods to the study of structure and flexibility of the target receptors in search of complementary insight.^{11–26}

Among the different approaches to introduce flexibility, molecular dynamics simulation (MD) is probably the most rigorous and in principle universal,²⁷ and examples of its success in the description of drug–protein binding are very common.^{14,15,28–33} However, despite its success, we cannot ignore the fact that MD also suffers from intrinsic limitations. It entails a heavy computational burden, and the proper set up of the simulated systems is complex and requires high levels of expertise, as does the analysis of trajectories once they have been obtained. Lastly, the accessible trajectories (typically in the nanosecond to microsecond time scale) suffer from not being able (within standard simulation time scales) to overcome high conformational barriers in the potential energy surface of the receptors and complexes. This limits the technique to exploring only conformations of the ensemble that are around the starting structures. In summary, MD has become extremely powerful in the study of specific protein–drug complexes with reasonable accuracy, but it is far from being a high-throughput technique in theoretical drug design. In the same light, we can discuss induced fit docking (IFD) procedures where conformational changes induced by the bound ligand are accounted for in the final modeled active site

geometry of the protein–ligand complex. This procedure can be effectively used for predicting protein–ligand interactions, but one observes a trade-off between the accuracy achieved and the computing time required. Although receptor flexibility and induced fit are allowed by some commercial modeling packages, this method is computationally expensive and currently not suitable for virtual high-throughput screening purposes.

Methods less rigorous than MD and IFD, but more efficient from a computational point of view, such as normal mode analysis (NMA), have recently opened up new ways of studying receptor flexibility,^{21–24,34–38} especially when ligand binding induces low frequency collective movements in the protein. Conceptually similar to NMA is essential dynamics (ED),^{25,39–43} a technique based on the process of atomistic MD trajectory by principal component analysis (PCA) to derive the essential movements of the protein. The creators of ED suggested the use of the essential modes to bias atomistic MD simulations a decade ago,^{39–42} as a way to enrich the sampling of collective movements obtained from short trajectories. More recently, our groups⁴⁴ have demonstrated how the essential movements computed from a system can be manipulated to obtain, with good accuracy and high computational efficiency, ensembles of related systems which can be considered as perturbations of the original system.

A clear example where this technique (named essential dynamics/molecular dynamics; ED/MD⁴⁴) is expected to be useful is in the study of drug binding, since the dynamics of a given drug–protein complex is expected to be described as a perturbation of the dynamics of the protein alone or bound to a related drug. In this paper, we present for the first time a systematic study on the performance of the ED/MD approach

Received: March 17, 2012

Published: June 15, 2012

to improve docking results in the high-throughput regime. We demonstrate that by coupling ED/MD ensembles obtained from the target protein under the influence of a generic perturbing drug with a standard docking algorithm (Schrodinger's Glide program),^{45–47} systematic improvements can be obtained with respect to random search and docking experiments carried out using experimental structures or standard MD ensembles.

METHODOLOGICAL APPROACH

ED/MD Basic Method. Following the ED paradigm,³⁹ the essential movements describing the dynamics of a protein are defined by principal component analysis of atomistic MD simulations. Accordingly, covariance matrices derived from MD trajectories are diagonalized to obtain a set of eigenvectors ($\{\hat{e}_i\}$) and eigenvalues ($\{\lambda_i\}$) describing the original dynamics in terms of orthogonal movements. Note that a (Cartesian coordinate) trajectory^{39,48} can be projected into the entire set of eigenvectors with no loss of accuracy and that a new approximated trajectory can be obtained by a back transformation that only considers projections along the most important eigenvectors (for example, those explaining the largest amount of variance). Furthermore, within the harmonic limit, the protein Hamiltonian can be defined as

$$H^{(M)} = \sum_{i=1}^{3N} \frac{p_i^2}{2m_i} + \frac{1}{2} \sum_{i=1}^M \frac{k_B T}{\lambda_i} (\Delta \vec{r} \cdot \hat{e}_i)^2 \quad (1)$$

where i stands for a particle, p_i is the i th component of the momentum vector, m_i is the particle mass, $\Delta \vec{r} = \vec{r} - \vec{r}_0$ is the displacement from the equilibrium position of atoms (\vec{r}_0), k_B is Boltzmann's constant, and T is the absolute temperature. It is worth noting that this Hamiltonian can be evaluated to a given level of accuracy by simply restricting the sum to the M most significant deformation modes (instead of considering all the $3N - 6$ modes). The restriction of the eigenvector space ($M < 3N - 6$) allows the user to concentrate the computational effort on relevant modes (for example, those explaining most of the protein variance or most of the movements at binding site). Note also that it is trivial to restrict the Hamiltonian to a given set of atoms or to reproduce a part of the molecule at the coarse-grained level (for example, at the C_α level), keeping all of the atomistic detail in another region preserved. In both cases, the region of interest is reproduced with full atomistic detail, while a large reduction of computational effort is obtained.

In the presence of a perturbation, for example a drug, a hybrid Hamiltonian can be defined:

$$H = H^{(M)} + V \quad (2)$$

where V stands for the perturbation (typically defined in the Cartesian space). Derivation of these Hamiltonian leads to the forces defined by

$$\vec{F}_i^{(M)} = - \sum_{k=1}^M \frac{k_B T e_k^i}{\lambda_k} (\Delta \vec{r} \cdot \hat{e}_i) + \vec{F}_i^* \quad (3)$$

where the first term stands for forces in the ED space and the following terms account for the reduced perturbational forces defined for each particle as

$$\vec{F}^* = -\vec{\nabla} V - \sum_{k=M+1}^{3N} \varphi_k \hat{e}_k \quad (4)$$

where $\{\varphi_1, \dots, \varphi_M, \varphi_{M+1}, \dots, \varphi_{3N}\}$ are the components of \vec{F} in the base \hat{e}_k

The effective forces computed using eq 3 define a Langevin equation:^{49–51}

$$m_i \ddot{\vec{r}}_i = \vec{F}_i^{(M)} - \gamma \dot{\vec{r}}_i + \vec{\xi}_i^* \quad (5)$$

where T^* is the effective temperature, γ is a friction coefficient, and stochastic terms should satisfy

$$\begin{aligned} \langle \vec{\xi}_i^*(t) \rangle &= 0 \\ \langle \vec{\xi}_i^*(t) \vec{\xi}_j^*(t') \rangle &= 2k_B T^* \gamma \delta_{ij} \delta(t - t') \end{aligned} \quad (6)$$

Numerical integration of eq 5 is done using a slightly modified version of the Verlet algorithm,^{44,51} which implements an efficient multiple-time step integrator,⁴⁴ inspired by the RESPA algorithm.^{52,53} Overall, the procedure outlined in this study is based on our basic ED/MD protocol,⁴⁴ which assumes that the displacement of the reference protein along its essential dynamics modes is Gaussian. The ED/MD algorithm is designed to generate reasonable alternative protein configurations away from the equilibrium structure; hence an accurate form of the potential function is perhaps not critical as we are essentially using it as a conformation-generating tool. It should also be noted here that the presence of the ligand (as a probe in the perturbed ED/MD implementation) breaks the fully harmonic character of the dynamics.

ED/MD Implementation for Docking. For the selected set of proteins (see Table 1), atomistic MD simulations were

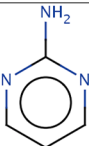
Table 1. The Eight Test Protein Systems Chosen to Validate the Drug-Induced Perturbed-EDMD Method^a

protein name	PDB ID	ligand bound
p38 MAP kinase	1A9U ⁶²	yes
cathepsin G	1KYN ⁶⁷	yes
trypsin	1DPO ⁶⁴	yes
serum albumin	1UOR ⁶⁰	no
carbonic anhydrase II	12CA ⁶⁸	no
p38 MAP kinase	1A9U	no
ribonuclease A	1ASP ⁶⁹	no
t-RNA transglycosylase	1PUD ⁷⁰	no

^aIn all cases, original trajectory was collected for 10–15 ns with or without ligand bound (depending on PDB state).

extracted from our MoDEL database and processed by means of PCA to derive the eigenvectors and eigenvalues that explain most of the variance of binding site residues (defined as all residues with at least 1 atom within 5 Å of the original bound ligand). The first 30 modes were used in all cases for ED/MD simulations. Sampling in a reduced Cartesian PCA space often leads to a distortion of chemical bonds. To avoid this problem, the chemical bond lengths are restrained within the binding site by adding a harmonic potential between the bonded atoms. A small pseudodrug (pyrimidin-2-amine; see Table 2) was used as a general probe to perturb the dynamics of the original protein (MD trajectories of proteins with or without ligand bound). For this purpose, the pseudodrug was docked in the protein-binding site using Glidev4.5^{45–47} to obtain the starting coordinates for perturbed ED/MD runs. A Lennard-Jones term was used to account for the perturbational terms in eq 2 (the probe was treated as a rigid body). ED/MD simulations were carried out for 10⁷ steps, generating 10 000 snapshots for

Table 2. The Perturbation Probe for the Drug-Induced Perturbed-EDMD Method

Structure	PDB	Ligand name
	2JJC ⁷¹	PYRIMIDIN-2-AMINE

each protein, which were then clustered using the K-means clustering module (kclust) of the MMTSB toolset.⁵⁴ The cluster representative snapshots were chosen from each of the cluster bins (closest structure to the centroid structure of the bin) to form the perturbed-EDMD generated cluster-representative-based protein ensemble for the eight protein systems studied here. These ensembles of five structures were

later used for Glide v4.5 docking experiments (no significant differences were obtained if a larger number of snapshots were considered).

Atomistic MD Ensembles for Docking. These were built using the same clustering process outlined above, but using as starting material the MD trajectories for the same proteins. As before, five snapshots representative of the entire ensemble were selected.

Protein and Ligands. Ensemble docking experiments were carried out using a set of eight proteins for which (i) at least 12 inhibitor-protein complexes are found in PDB and (ii) at least one trajectory is available in our MoDEL database⁵⁵ (see Table 1). The target protein structures were prepared using the Protein Preparation Wizard of MAESTRO8.5. All waters, ions, and metals were removed prior to docking. We defined a set of active compounds by selecting all available ligands from PDB complexes for each of the protein systems. This group of compounds was mixed with 99 diverse drug-like decoys (inactive molecules to test the methods) extracted as described

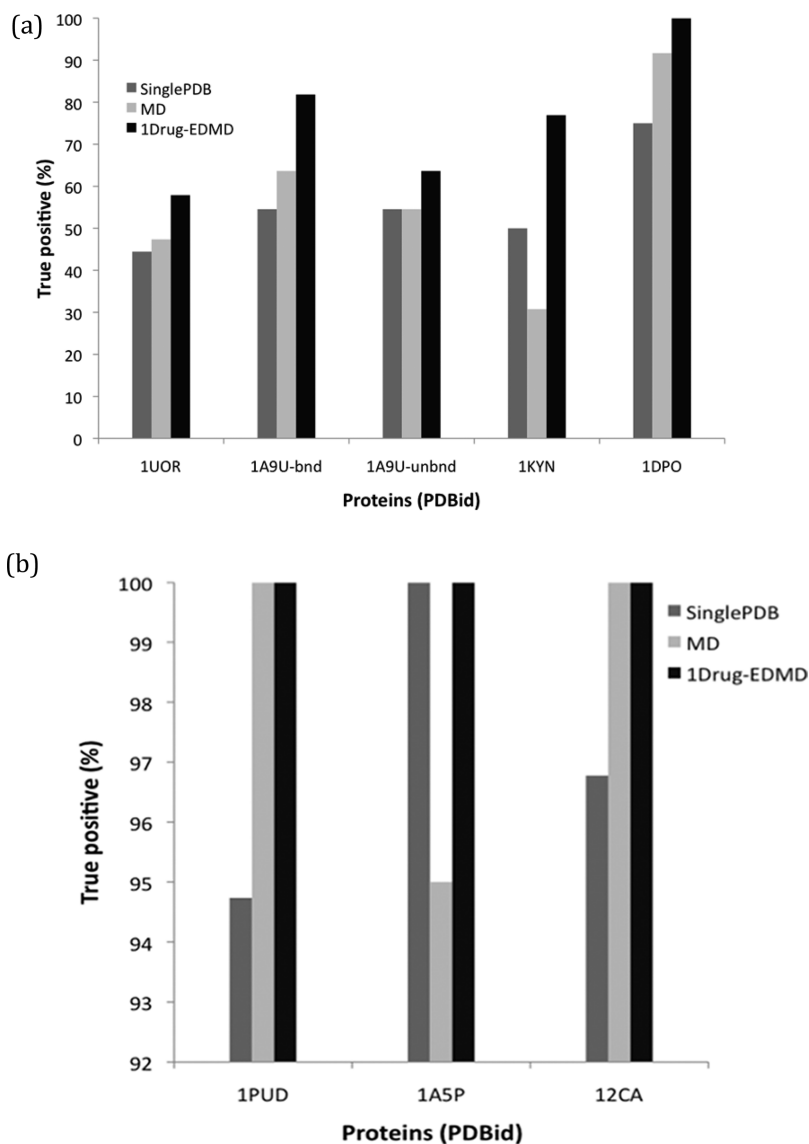


Figure 1. True positive (TP) enrichment plot for the eight protein systems studied here for a 15% cutoff (i.e., when 15% of database is analyzed). The top plot corresponds to cases where standard docking using X-ray structure shows moderate performance (between 40 and 80% TP), while the bottom plot corresponds to cases where standard docking using X-ray structure already performs well (more than 90% TP).

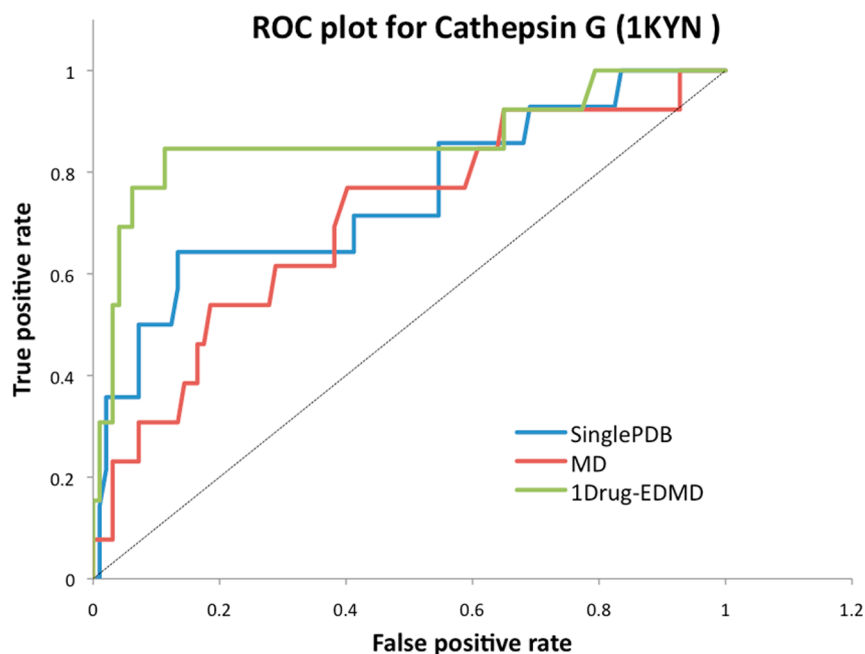


Figure 2. ROC curve obtained for cathepsin G system (1KYN). The straight line corresponds to the background model. The desired docking method is that where most TPs are recovered, while only a small number of false positives (FPs) are selected, i.e., those where the area under the curve of the ROC plot is larger.

elsewhere⁵⁶ from a local database of 1.7 million compounds. The final ligand data set for each protein was prepared using Ligprep version 2.2⁵⁷ by generating low-energy ionization and tautomeric states within the pH range of 7.0 ± 2.0 . All ligands were energy-minimized using the OPLS-2005 force field implemented in MAESTROv8.5.⁵⁸

Docking Experiments. Three sets of docking experiments were performed. One involved a standard docking procedure using crystal structures as targets, and two involved ensemble docking using either normal MD ensembles or ED/MD ensembles. In the first case, each crystal structure was manipulated with the Protein Preparation Wizard of MAESTROv8.5⁵⁸ to define a suitable ionic state, remove bad contacts, and locate hydrogen in suitable places. Docking was performed in all cases using Glidev4.5, with the Glide SP (Standard Precision)^{45,46} scoring function and all default settings of Glidev4.5. The receptor grid was constructed centered about the calculated centroid of the active site residues defined during grid generation. In the case of ensemble docking, each drug was associated with five final scores from the five cluster representative structures (from the ensemble) they were docked into; the highest score and pose was retained in all cases for further analysis and overall compound rank determination.

RESULTS AND DISCUSSION

The main objective of *in silico* virtual screening of large compound libraries is to derive a small set of compounds enriched with active ligands selected on the basis of their predicted high binding affinity to the target protein. The success of a docking protocol can be measured by the number of active compounds present in the final enriched set compared to the expected set from a random model. Thus, if we create a library containing, for example, 15% of the original compounds, we can expect that by chance it will also contain 15% of all active ligands (true positives; TP) present in the original

library; any improvement from this figure implies predictive power for the docking algorithm. Figure 1 summarizes the predicted power of docking algorithms for the eight proteins considered, when used to define a reduced set containing 15% of the database (similar results for a smaller set (5% of the database) are displayed in Supporting Information Figure S1).

Clearly, all docking procedures considered here have reasonable or even good predictive power (Figure 1), as noted in improvements from two to seven times over the background models, but we can also differentiate between two types of situations: (i) cases where protein flexibility is not crucial and simple docking procedures using standard X-ray structures perform well (Figure 1 bottom) and (ii) cases where flexibility should be accounted for in the docking process (Figure 1 top). In the first family of cases, the introduction of flexibility leads only to marginal improvements, since figures from standard docking are already very good. However, even in these cases the introduction of flexibility by our ED/MD-docking protocol leads to a systematic improvement in the results, and 100% of the TPs are recovered. It is worth noting that such a systematic improvement is not obtained when standard MD ensembles are used, since in at least one of the cases (1ASP) results obtained from MD docking are worse than those obtained by standard X-ray docking. For the five cases where the use of a rigid (even accurate) structure of the protein as a template for docking does not result in a high TP retrieval rate, clear improvements are obtained in all cases when docking is performed using the ED/MD ensembles as a template. Thus, ED/MD docking recovers 60%, 85%, 68%, 80%, and almost 98% of all TPs for 1UOR, 1A9U (bound and unbound), 1KYN, and 1DPO, while the X-ray docking recovers under identical conditions 45%, 55%, 55%, 50%, and 75% for the same proteins. It is worth noting that such a systematic improvement in the docking performance is not achieved when standard MD ensembles are used as templates, since in two cases the

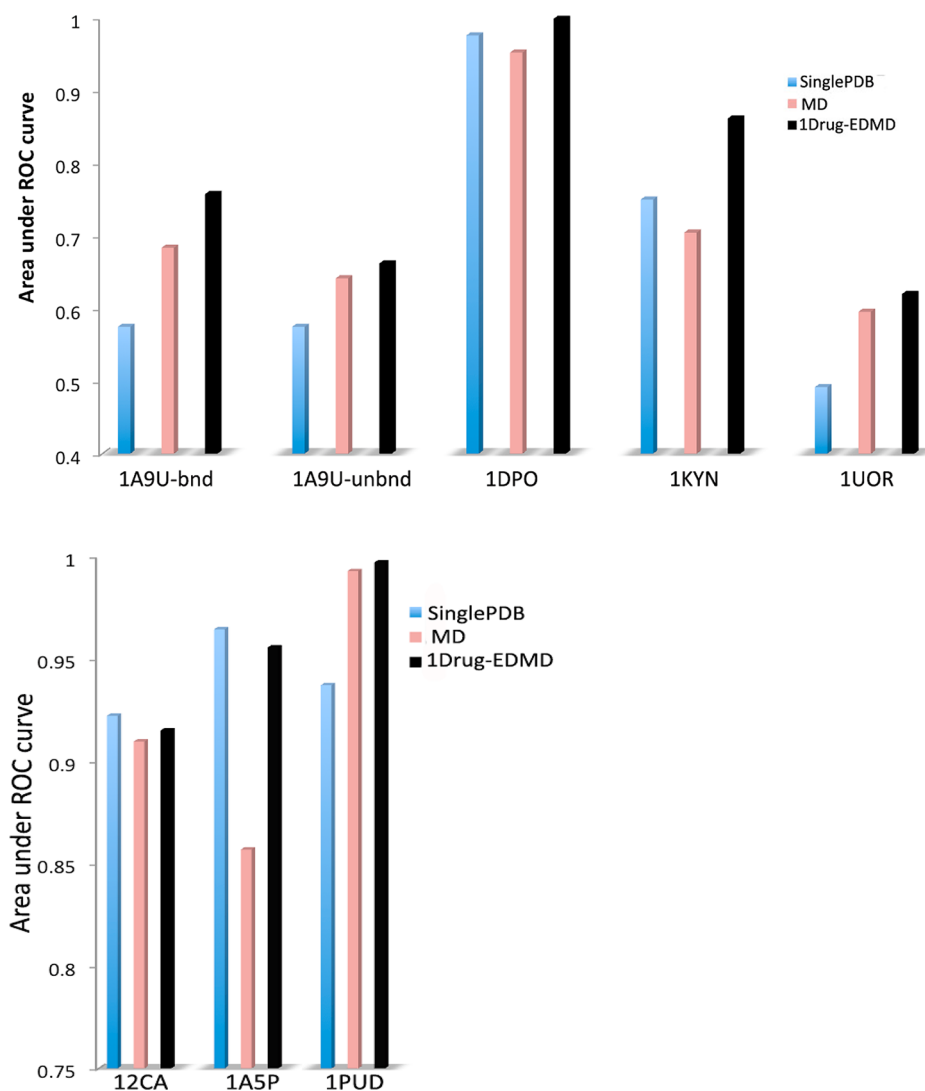


Figure 3. Area under the ROC curve (AUC) for the eight systems considered here and the three docking procedures considered. The top plot corresponds to cases where standard docking using X-ray structure shows moderate performance, while the bottom plot corresponds to cases where standard docking using X-ray structure already performs well (see Figure 1).

improvement is negligible and in another case MD docking (1KYN) performs significantly worse than X-ray docking.

An alternative measure of docking performance that is common in the field comes from the analysis of the receiver operating characteristic (ROC) curve.⁵⁹ This metric gives a global view of the performance of a method that selects real ligands (true positive; TP) without selecting inactive molecules (false positive; FP). A ROC curve representative of a good docking procedure is that for which with a small percentage of FPs a large number of TPs are retrieved, and this can be quantified by a sharper slope in the ROC curve in the low FP region and more precisely by a larger value of the area under the ROC curve (AUC). As seen in an example for 1KYN (Figure 2), ED/MD-docking leads to very clear improvements in the ROC curve. Thus, when we use X-ray or MD-docking we have to suffer a 0.6 FP rate to obtain a 0.8 TP rate, which means that we do not have impressive discriminatory power between ligands and decoys. Note, however, that when ED/MD docking is used, we can reach the 0.8 TP rate with a FP rate equal to only 0.1. That is, the ED/MD-docking procedure shows much better discriminative power than the other two

docking protocols and is more useful in generating a reduced database enriched in real ligands.

The superior performance of the ED/MD-docking procedure with respect to random selection or other docking procedures is clearly visible in Figure 3, where we display the AUCs obtained for the eight systems under study. As expected, the use of ED/MD does not produce major improvements on ROC curves for rigid proteins, but the introduction of flexibility in docking by means of the ED/MD procedure does not introduce a significant increase in the number of false positives, which would yield to a parallel reduction of the AUC. In fact, cases where this reduction occurs can be noted when MD samplings are used for docking (see Figure 3, bottom). For proteins where flexibility seems to be a requirement for ligand binding (Figure 3 top), the improvements obtained by using the ED/MD are strong and systematic, outperforming docking procedures performed considering either a single X-ray structure or MD ensembles as templates. It should also be noted that, out of the eight cases studied here, three of the original MD simulations were actually performed in the presence of a bound ligand (Table 1). In all three cases (1A9U-bnd, 1KYN, and 1DPO), the enrichment obtained

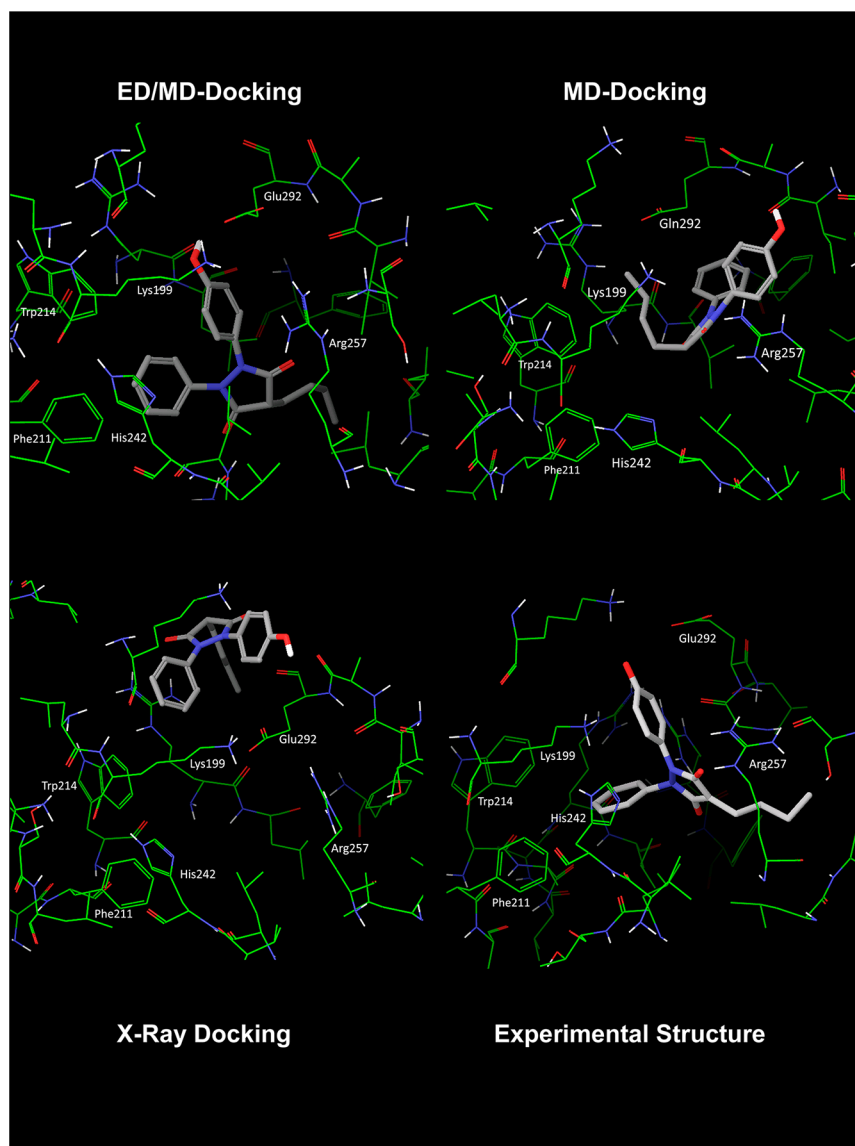


Figure 4. Best poses of the OPB–albumin complex obtained by different docking protocols (ED/MD, MD, and X-ray docking). All three screenings were done using the 1UOR protein structure. The experimental solution corresponding to the cocrystal of ligand OPB with albumin (PDB ID 2BXB) is displayed as a reference structure. See the text for discussion.

using the ED/MD docking protocol was higher than the MD based protocol. This might be attributed to the fact that the tightly bound and optimally fit native ligand leaves little room for protein side chain flexibility (since these side chains are already engaged in making strong interactions with the bound ligand), and hence sampling of alternate conformations becomes challenging during the MD simulations in the ligand-bound form. Hence, in spite of the fact that a holo binding site geometry is preserved during these three MD runs (and consequently the structural ensemble for docking), the lack of side chain diversity renders these snapshots less useful in picking ligands with diverse scaffolds. Overall, this means that ED/MD docking largely improves our ability to distinguish between ligands and decoys, compared to a random model and to alternative state-of-the-art docking procedures. In accordance with the above results, ED/MD can be safely used to select a reduced set of compounds enriched in active ligands for experimental validation.

The main output of a massive docking experiment is always a list of compounds ranked in such a way that the first few compounds have the highest probability of being active ligands. However, once one of the well-scored compounds is validated experimentally as an active ligand, the geometry of the drug–protein complex becomes interesting, since knowledge of the drug-binding mode can largely accelerate the lead optimization process. Therefore, we were interested in checking the ability of ED/MD docking to derive reliable binding modes and to compare them with those obtained from standard docking procedures. We found that in general the ED/MD-docking procedure is able to detect more reliable docking modes, while often other docking methods fail to detect the correct orientation mostly due to steric collapses. One typical example is the binding of OPB (an anticoagulant drug) to serum albumin. This is a challenging case since the 1UOR⁶⁰ entry corresponds to an apo structure, where the binding site appears buried, and slightly collapsed, which raises difficulties for standard docking protocols. Not surprisingly, the standard X-

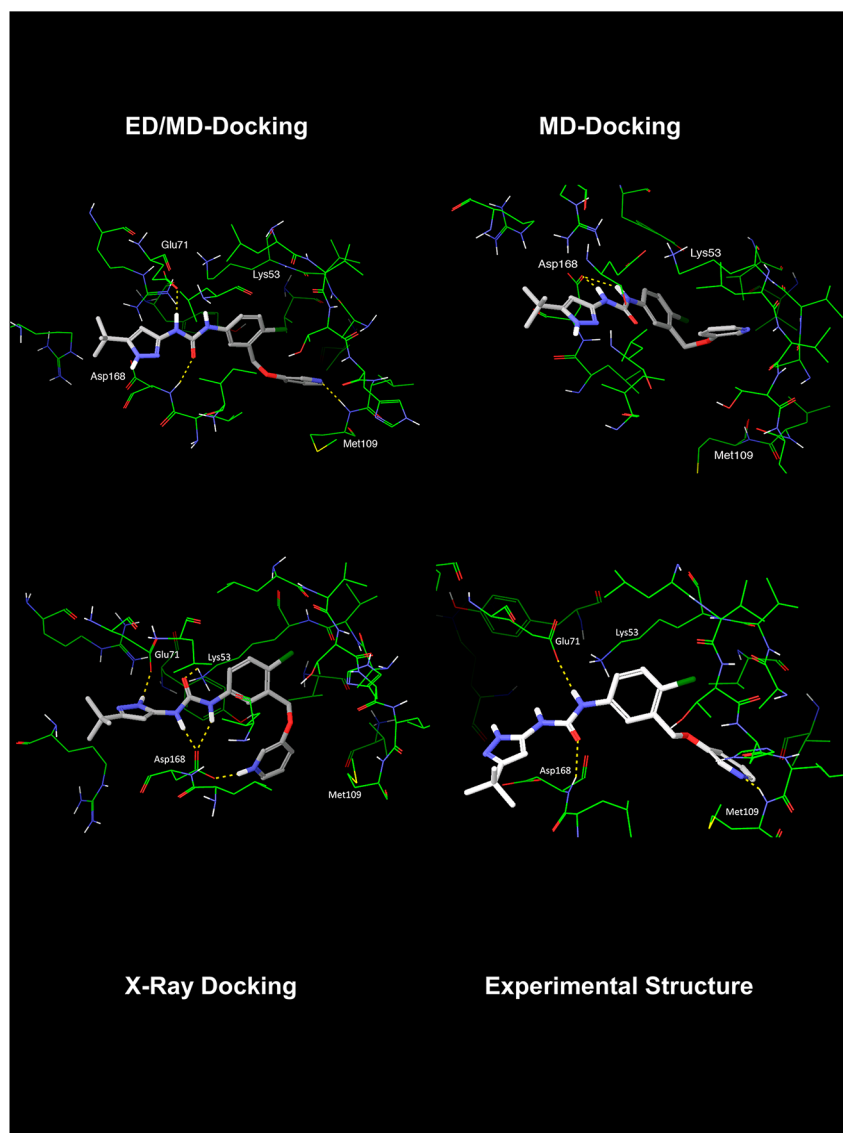


Figure 5. Best poses of the L09–p38 MAP kinase complex obtained by different docking protocols (ED/MD, MD, and X-ray dockings). The 1A9U structure was used for docking in all cases. The experimental solution corresponding to the cocrystal of L09–p38 MAP kinase (PDB ID 1WBN) is displayed as a reference. See the text for discussion.

ray docking is not very productive (see Figure 4), and the drug is placed far away from the real binding site (from 2BXB crystal structure),⁶¹ which is in fact associated with a poor Glide Score. The situation does not improve much when MD ensembles are used. Although the drug is placed in the binding site, the binding geometry is incorrect, and none of the true binding mode interactions are retained (Figure 4). Docking to the ED/MD ensemble reveals a binding mode very close to the real experimental structure and where the drug makes a variety of favorable interactions with the protein, for example, the nice fitting of the phenyl ring in the hydrophobic pocket formed by His242, Phe211, and Trp214, the hydrogen bond interaction between the pyrazolidine group and Arg257, or the cation– π interaction with Lys199. The RMSD of the ligand, IMA, bound to the ED/MD derived snapshot is closest to the ligand conformation in its native X-ray structure in 2BXB. The RMSD values for each of the ligand poses in the ED/MD, MD, and X-ray crystallography derived structures when compared to the native conformation of the ligand in the experimental structure are as follows: 4.0, 4.3, and 4.2 Å. It is imperative to mention

that besides RMSD values, which are average numbers, it is of more importance (and perhaps a better metric of measurement) to note whether the top ranking ligand conformation is able to capture the key interactions as observed in the native protein–ligand complex. The RMSD of binding site residues (defined by 30 residues within 5 Å of the ligand IMA in 2BXB) from the ED/MD snapshot deviates the most from the protein binding site conformation in the experimental structure in 2BXB (2.7 Å), in comparison to the RMSD of the MD derived structure (2.08 Å) and the X-ray crystal structure in 1UOR (2.1 Å) with the real binding site in 2BXB. This leads us to believe that the higher conformational diversity sampled by the ED/MD snapshot while preserving a holo conformation allows the binding site to accommodate the ligand in its correct binding mode. Clearly, in this case, models derived from X-ray docking or MD-docking will be of less use to direct lead optimization protocols, whereas a higher rate of success is perhaps possible using the models derived from the ED/MD docking.

A second example is shown for the binding of ligand L09 to p38-MAP kinase (1A9U-bound template).⁶² This is theoret-

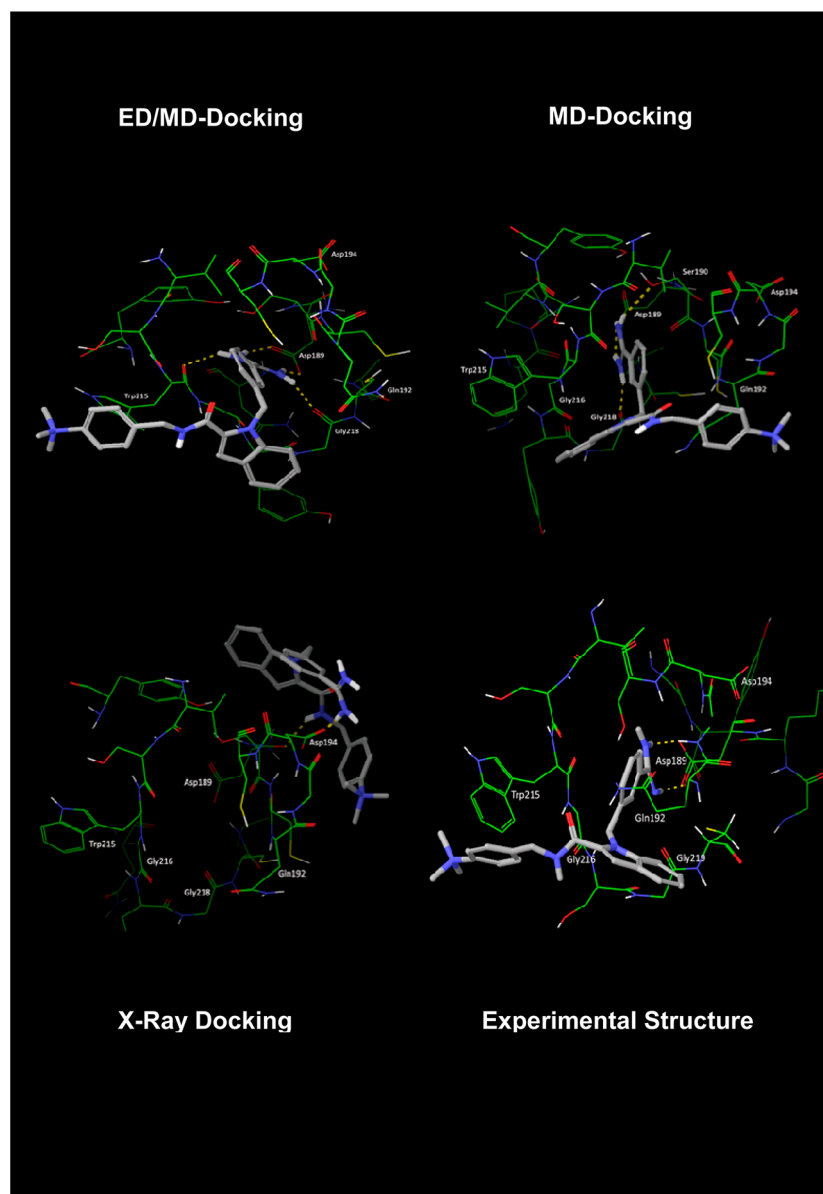


Figure 6. Best poses of the IMA–trypsin complex obtained by different docking protocols (ED/MD, MD, and X-ray dockings). All docking studies were done using the 1DPO structure. The experimental solution corresponding to the cocrystal of IMA–trypsin (PDB ID 1LQE) is displayed as a reference structure. See the text for discussion.

ically a less challenging example, since the template structure was bound to a ligand in the original MD trajectory and, hence, the binding site is not collapsed. We can see that the real binding mode (PDB ID 1WBN)⁶³ is dominated (see Figure 5) by hydrogen bonds of the drug with residues Met109, Asp168, and Glu71. The best pose obtained from X-ray docking (poorly scored by Glide) shows a wrong orientation of the ligand in the binding site of the protein, which leads to the loss of the hydrogen bond with Met109 and the generation of other artifacts, like a bifurcated hydrogen bond with the side chain of Asp168. The use of MD ensembles for docking makes the situation even worse: two of the three native hydrogen bonding interactions are not present, the drug moves away from Met109 adopting a wrong bioactive conformation, and several artifactual hydrogen bonding contacts and unfavorable steric clashes (around carboxamide and indole moieties) are detected. The compounds also rank and score poorer in the MD ensemble than the ED/MD docking generated results.

Fortunately, the ED/MD-docking procedure leads to a much more reasonable binding mode, where the native contacts are preserved. However, the receptor binding site in the ED/MD snapshot that ranks L09 the highest deviates the most (3.8 Å) from the 1WBN binding site (defined by 19 residues around ligand L09), sampling a slightly greater conformational space, when compared to 3.5 Å deviation in the case of the MD derived snapshot and 3.2 Å for the X-ray structure of the protein binding site in 1A9U-bound. The RMSD of the top ranking ligand conformation in the ED/MD snapshot is closest to the native conformation of ligand L09 in 1WBN. The RMSD values for each of the ligand poses in the ED/MD, MD, and X-ray derived structures when compared to the native conformation of the ligand in the experimental structure are as follows: 3.0, 3.3, and 3.8 Å. We can conclude that the ED/MD approach seems to produce ligand poses that are more likely to be a good starting point for a lead-optimization process

than those obtained by X-ray-docking or MD-docking procedures.

A third and compelling example is found in the docking of small molecule IMA to trypsin (1DPO).⁶⁴ This is one case where protein flexibility is important and X-ray-based docking simply fails to locate the binding site, yielding to a completely unrealistic binding mode, poorly scored by Glide (see Figure 6). The MD-docking procedure corrects part of the problem, and the first pose captures some of the native contacts (like the guanidinium–Asp189), even though the orientation of the drug is incorrect (the second best scored pose of IMA in the MD ensemble corresponds to the drug being docked out of the binding site as in the X-ray docking). From the true experimental structure (PDB ID 1LQE),⁶⁵ we note that the native π -cation interaction of the drug with Trp215 and the H-bonding with Asp189 can be reproduced through the ED/MD docking. The top ranking ligand pose in the ED/MD snapshot has the closest RMSD with the ligand conformation in its native X-ray structure in 1LQE. The RMSD values for each of the ligand poses in the ED/MD, MD, and X-ray derived structures when compared to the native conformation in the experimental structure are as follows: 3.1, 4.2, and 4.1 Å. In this case, the protein binding site (defined by 18 residues) conformation from the ED/MD snapshot is also the closest (3.8 Å) to the real binding site conformation in 1LQE. The MD snapshot (4.0 Å) and the X-ray structure of 1DPO (4.1 Å) show more divergence from the experimental binding site. In fact, here, the ED/MD snapshot is closer to the experimental holo structure (3.8 Å) than to the X-ray structure (4.7 Å). A realistic and well scored pose of the ligand IMA is generated; although not perfect, the model can help direct the lead-optimization process in the right course.

At this stage, we wished to determine if the success of this method is just due to the more diverse conformational sampling of the protein through ED/MD or to the actual presence of the “probe” or prototypical ligand, which generates candidate protein conformations for docking that are more biased toward the ligand-bound state (preferred for docking) than a simple MD derived structural ensemble. In addition, we also wanted to explore the probe space and characterize its effects on the performance of the drug-induced ED/MD method. In order to isolate the key reason for the success of this method, we conducted some additional experiments to confirm that the higher enrichment and method performance was indeed from the drug-induced perturbation and not from diverse sampling inherent in the ED/MD method. In summary, we find that (a) the presence of a prototypical perturbing ligand is vital for the success of the method—using the ED/MD approach—but without such a ligand is much less effective and (b) the exact structure of the prototypical ligand used to perturb the ED/M is not critical, though if it is too large the conformational ensemble becomes overbiased in a ligand-specific way and performance degrades. The details of these experiments and results can be found in the Supporting Information

CONCLUSIONS

We have explored in detail the ability of a new approach for high-throughput docking based on the combination of the standard docking procedure and a new sampling protocol based on the process of prestored molecular dynamics trajectories by means of a perturbed essential dynamics algorithm.⁴⁴ The method is able to generate, with extreme computational

efficiency, ensembles of configurations of the binding site which, then used as templates for the docking algorithm, increases the chances of finding active ligands in chemical databases and defines reasonable poses that can consequently guide lead optimization processes.

The ED/MD procedure outlined here performs better than standard docking procedures using X-ray structures as templates, and more surprisingly it clearly outperforms standard MD, which in the studied simulation time generates sampling with little diversity which is too narrowly centered around the starting structure. This effect is especially visible in the case of apo-proteins, since standard MD simulations of the apo form of the protein often start and end with a collapsed active site in the absence of a bound ligand. This collapsed state occludes part of the active site (usually with specific side chain orientations) dramatically decreasing the chances of successful docking. The ED/MD method, using a small drug-like molecule to perturb the equilibrium trajectory of the protein (obtained in the essential space by processing a previously stored MD trajectory), largely increases the diversity of the binding site, being able to generate configurations that can recognize diverse ligands. The ED/MD method is especially powerful in cases where ligand binding induces large conformational changes, but quite encouragingly, it improves, and in no case worsens, the results where the binding of the ligand does not produce sizable changes in the binding site configuration.

The simplicity of the method and its ability to take advantage of large structural databases of MD trajectories (such as MoDEL,⁵⁵ dynameomics⁶⁶) combined with its computational efficiency and easy integration into popular docking programs suggest that it can become a default feature in high-throughput docking routines.

ASSOCIATED CONTENT

Supporting Information

In order to isolate the key reason for the success of this method, we conducted some additional experiments to confirm that the higher enrichment and method performance was indeed from the drug-induced perturbation and not from diverse sampling inherent in the ED/MD method. The details of these experiments can be found in the Supporting Information. This information is available free of charge via the Internet at <http://pubs.acs.org>

AUTHOR INFORMATION

Corresponding Author

*E-mail: modesto.orozco@irbbarcelona.org.

Notes

The authors declare no competing financial interest.

REFERENCES

- (1) Claussen, H.; Buning, C.; Rarey, M.; Lengauer, T. FlexE: efficient molecular docking considering protein structure variations. *J. Mol. Biol.* **2001**, *308*, 377–395.
- (2) Cavasotto, C. N.; Abagyan, R. A. Protein Flexibility in Ligand Docking and Virtual Screening to Protein Kinases. *J. Mol. Biol.* **2004**, *337*, 209–225.
- (3) Huey, R.; Morris, G. M.; Olson, A. J.; Goodsell, D. S. A semiempirical free energy force field with charge-based desolvation. *J. Comput. Chem.* **2007**, *28*, 1145–1152.
- (4) Österberg, F.; Morris, G. M.; Sanner, M. F.; Olson, A. J.; Goodsell, D. S. Automated docking to multiple target structures:

Incorporation of protein mobility and structural water heterogeneity in AutoDock. *Proteins: Struct., Funct., Bioinf.* **2002**, *46*, 34–40.

(5) Jain, A. N. Surflex-Dock 2.1: robust performance from ligand energetic modeling, ring flexibility, and knowledge-based search. *J. Comput.-Aided Mol. Des.* **2007**, *21*, 281–306.

(6) Wei, B. Q.; Weaver, L. H.; Ferrari, A. M.; Matthews, B. W.; Shoichet, B. K. Testing a flexible-receptor docking algorithm in a model binding site. *J. Mol. Biol.* **2004**, *337*, 1161–1182.

(7) Najmanovich, R.; Kuttner, J.; Sobolev, V.; Edelman, M. Side-chain flexibility in proteins upon ligand binding. *Proteins* **2000**, *39*, 261–268.

(8) Zavodszky, M. I.; Kuhn, L. A. Side-chain flexibility in protein-ligand binding: the minimal rotation hypothesis. *Protein Sci.* **2005**, *14*, 1104–1114.

(9) Gunasekaran, K.; Nussinov, R. How different are structurally flexible and rigid binding sites? Sequence and structural features discriminating proteins that do and do not undergo conformational change upon ligand binding. *J. Mol. Biol.* **2007**, *365*, 257–273.

(10) Gutteridge, A.; Thornton, J. Conformational changes observed in enzyme crystal structures upon substrate binding. *J. Mol. Biol.* **2005**, *346*, 21–28.

(11) Cozzini, P.; Kellogg, G. E.; Spyraakis, F.; Abraham, D. J.; Costantino, G.; Emerson, A.; Fanelli, F.; Gohlke, H.; Kuhn, L. A.; Morris, G. M.; Orozco, M.; Pertinhez, T. A.; Rizzi, M.; Sotriffer, C. Target Flexibility: An Emerging Consideration in Drug Discovery and Design. *J. Med. Chem.* **2008**, *51*, 6237–6255.

(12) McCammon, J. A. Target flexibility in molecular recognition. *Biochim. Biophys. Acta, Proteins Proteomics* **2005**, *1754*, 221–224.

(13) Fanelli, F.; Benedetti, P. G. D. Computational Modeling Approaches to Structure–Function Analysis of G Protein-Coupled Receptors. *Chem. Rev.* **2005**, *105*, 3297–3351.

(14) Meagher, K. L.; Carlson, H. A. Incorporating Protein Flexibility in Structure-Based Drug Discovery: Using HIV-1 Protease as a Test Case. *J. Am. Chem. Soc.* **2004**, *126*, 13276–13281.

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

(16) Lin, J.-H.; Perryman, A. L.; Schames, J. R.; McCammon, J. A. The relaxed complex method: Accommodating receptor flexibility for drug design with an improved scoring scheme. *Biopolymers* **2003**, *68*, 47–62.

(17) Ahmed, A.; Gohlke, H. Multiscale modeling of macromolecular conformational changes combining concepts from rigidity and elastic network theory. *Proteins: Struct., Funct., Bioinf.* **2006**, *63*, 1038–1051.

(18) Jacobs, D. J.; Rader, A. J.; Kuhn, L. A.; Thorpe, M. F. Protein flexibility predictions using graph theory. *Proteins: Struct., Funct., Bioinf.* **2001**, *44*, 150–165.

(19) Tama, F.; Gadea, F. X.; Marques, O.; Sanejouand, Y.-H. Building-block approach for determining low-frequency normal modes of macromolecules. *Proteins: Struct., Funct., Bioinf.* **2000**, *41*, 1–7.

(20) Atilgan, A. R.; Durell, S. R.; Jernigan, R. L.; Demirel, M. C.; Bahar, I. Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys. J.* **2001**, *80*, 505–515.

(21) Case, D. A. Normal mode analysis of protein dynamics. *Curr. Opin. Struct. Biol.* **1994**, *4*, 285–290.

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

(23) Cavasotto, C. N.; Kovacs, J. A.; Abagyan, R. A. Representing receptor flexibility in ligand docking through relevant normal modes. *J. Am. Chem. Soc.* **2005**, *127*, 9632–9640.

(24) Ma, J. Usefulness and Limitations of Normal Mode Analysis in Modeling Dynamics of Biomolecular Complexes. *Structure* **2005**, *13*, 373–380.

(25) Rueda, M.; Chacón, P.; Orozco, M. Thorough validation of protein normal mode analysis: a comparative study with essential dynamics. *Structure* **2007**, *15*, 565–575.

(26) Camps, J.; Carrillo, O.; Emperador, A.; Orellana, L.; Hospital, A.; Rueda, M.; Cicin-Sain, D.; D'Abramo, M.; GelpiAnd, J. L.; Orozco,

M. FlexServ: an integrated tool for the analysis of protein flexibility. *Bioinformatics* **2009**, *25*, 1709–1710.

(27) Karplus, M.; Kuriyan, J. Molecular dynamics and protein function. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6679–6685.

(28) Dror, R. O.; Pana, A. C.; Arlowa, D. H.; Borhania, D. W.; Maragakisa, P.; Shana, Y.; Xua, H.; Shaw, D. E. Pathway and mechanism of drug binding to G-protein-coupled receptors. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 13118–13123.

(29) Dror, R. O.; Arlowa, D. H.; Maragakisa, P.; Mildorfa, T. J.; Pana, A. C.; Xua, H.; Borhani, D. W.; Shaw, D. E. Activation mechanism of the β_2 -adrenergic receptor. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 18684–18689.

(30) Buch, I.; Giorgino, T.; Fabritiis, G. D. Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 10184–10189.

(31) Soliva, R.; Almansa, C.; Kalko, S. G.; Luque, F. J.; Orozco, M. Theoretical studies on the inhibition mechanism of cyclooxygenase-2. Is there a unique recognition site? *J. Med. Chem.* **2003**, *46*, 1372–1382.

(32) Soliva, R.; Gelpi, J. L.; Almansa, C.; Virgili, M.; Orozco, M. Dissection of the recognition properties of p38 MAP kinase. Determination of the binding mode of a new pyridinyl-heterocycle inhibitor family. *J. Med. Chem.* **2007**, *50*, 283–293.

(33) Bowman, A. L.; Nikolovska-Coleska, Z.; Zhong, H.; Wang, S.; Carlson, H. A. Small molecule inhibitors of the MDM2-p53 interaction discovered by ensemble-based receptor models. *J. Am. Chem. Soc.* **2007**, *129*, 12809–12814.

(34) Cui, Q.; Bahar, I. *Normal Mode Analysis: Theory and Applications to Biological and Chemical Systems*; Chapman & Hall/CRC Press: Boca Raton, FL, 2006; p 406.

(35) Brooks, B.; Karplus, M. Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 6571–6575.

(36) Ben-Avraham, D.; Tirion, M. M. Normal modes analyses of macromolecules. *Phys. A (Amsterdam, Neth.)* **1998**, *249*, 415–423.

(37) Suhre, K.; Sanejouand, Y. H. ElNemo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement. *Nucleic Acids Res.* **2004**, *32*, 610–614.

(38) Hollup, S. M.; Salensminde, G.; Reuter, N. WEBnm@: a web application for normal mode analyses of proteins. *BMC Bioinf.* **2005**, *6*, 52.

(39) Amadei, A.; Linssen, A. B. M.; Berendsen, H. J. C. Essential dynamics of proteins. *Proteins* **1993**, *17*, 412–425.

(40) Amadei, A.; Linssen, A. B.; Groot, B. L.; De; Aalten, D. M.; Van; Berendsen, H. J. C. An efficient method for sampling the essential subspace of proteins. *J. Biomol. Struct. Dyn.* **1996**, *13*, 615–625.

(41) Aalten, D. M.; Van; Findlay, J. B. C.; Amadei, A.; Berendsen, H. J. C. Essential dynamics of the cellular retinol-binding protein - evidence for ligand-induced conformational changes. *Prot. Eng.* **1995**, *8*, 1129–1135.

(42) Aalten, D. M. F. V.; Groot, B. L. D.; Findlay, J. B. C.; Berendsen, H. J. C.; Amadei, A. A comparison of techniques for calculating protein essential dynamics. *J. Comput. Chem.* **1997**, *18*, 169–181.

(43) Hayward, S.; Groot, B. L. D. Normal modes and essential dynamics. *Methods Mol. Biol.* **2008**, *443*, 89–106.

(44) Carrillo, O.; Laughton, C. A.; Orozco, M. Fast Atomistic Molecular Dynamics Simulations from Essential Dynamics Samplings. *J. Chem. Theory Comput.* **2012**, *8*, 792–799.

(45) Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *J. Med. Chem.* **2004**, *47*, 1750–1759.

(46) *Glide*, version 4.5; Schrodinger LLC: New York, 2008.

(47) Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein-Ligand Complexes. *J. Med. Chem.* **2006**, *49*, 6177–6196.

(48) Meyer, T.; Ferrer-Costa, C.; Pérez, A.; Rueda, M.; Bidon-Chanal, A.; Luque, F. J.; Laughton, C. A.; Orozco, M. Essential

Dynamics: A tool for efficient trajectory compression and management. *J. Chem. Theory Comput.* **2006**, *2*, 251–258.

(49) Gardiner, C. *Handbook of Stochastic Methods: for Physics, Chemistry and the Natural Sciences*; Springer Eds: Berlin, 1989.

(50) van Kampen, N. G. *Stochastic Processes in Physics and Chemistry*; North-Holland: Amsterdam, 1981.

(51) Allen, M. P.; Tildesley, D. J. *Computer Simulations of Liquids*; Clarendon Press: Oxford, U. K., 1989.

(52) Tuckerman, M.; Berne, B. J.; Martyna, G. J. Reversible multiple time scale molecular dynamics. *J. Chem. Phys.* **1992**, *97*, 1990–2001.

(53) Martyna, G. J.; Tuckerman, M. E.; Tobias, D. J.; Klein, M. L. Explicit reversible integration algorithms for extended systems. *Mol. Phys.* **1996**, *87*, 1117–1157.

(54) Michael, F.; Karanickolas, J.; Brooks, C. L. I. MMTSB Tool Set 2001. <http://mmtsb.org/> (accessed Jun. 2012).

(55) Meyer, T.; D'Abramo, M.; Hospital, A.; Manuel, R.; Ferrer-Costa, C.; Pérez, A.; Carrillo, O.; Camps, J.; Fenollós, C.; Repchevsky, D.; Gelpi, J. L.; Orozco, M. MoDEL (Molecular Dynamics Extended Library): a database of atomistic molecular dynamics trajectories. *Structure* **2010**, *18*, 1339–1409.

(56) Novoa, E. M.; Ribas De Pouplana, L.; Barril, X.; Orozco, M. Ensemble Docking from Homology Models. *J. Chem. Theory Comput.* **2010**, *6*, 2547–2557.

(57) *LigPrep*, version 2.2; Schrodinger LLC: New York, 2008.

(58) *Maestro*, version 8.5.207; Schrodinger LLC: New York, 2008.

(59) Swets, J. A. *Signal Detection Theory and Roc Analysis in Psychology and Diagnostics. Collected Papers*; Lawrence Erlbaum Associates: Mahwah, NJ, 1996.

(60) He, X.; Carter, D. Atomic structure and chemistry of human serum albumin. *Nature* **1992**, *358*, 209–215.

(61) Ghuman, J.; Zunszain, P.; Petitpas, I.; Bhattacharya, A.; Otagiri, M.; Curry, S. Structural basis of the drug-binding specificity of human serum albumin. *J. Mol. Biol.* **2005**, *353*, 38–52.

(62) Wang, Z.; Canagarajah, B. J.; Boehm, J. C.; Kassisa, S.; Cobb, M. H.; Young, P. R.; Abdel-Meguid, S.; Adams, J. L.; Goldsmith, E. J. Structural basis of inhibitor selectivity in MAP kinases. *Structure* **1998**, *6*, 1117–1128.

(63) Gill, A. L.; Frederickson, M.; Cleasby, A.; Woodhead, S. J.; Carr, M. G.; Woodhead, A. J.; Walker, M. T.; Congreve, M. S.; Devine, L. A.; Tisi, D.; O'Reilly, M.; Seavers, L. C. A.; Davis, D. J.; Curry, J.; Anthony, R.; Padova, A.; Murray, C. W.; Carr, R. A. E.; Jhoti, H. Identification of Novel p38 α MAP Kinase Inhibitors Using Fragment-Based Lead Generation. *J. Med. Chem.* **2005**, *48*, 414–426.

(64) Earnest, T.; Fauman, E.; Craik, C. S.; Stroud, R. 1.59 Å structure of trypsin at 120 K: comparison of low temperature and room temperature structures. *Proteins* **1991**, *10*, 171–187.

(65) Matter, H.; Defossa, E.; Heinelt, U.; Blohm, P.; Schneider, D.; Müller, A.; Herok, S.; Schreuder, H.; Liesum, A.; Brachvogel, V.; Lönze, P.; Walser, A.; Al-Obeidi, F.; Wildgoose, P. Design and quantitative structure-activity relationship of 3-amidinobenzyl-1H-indole-2-carboxamides as potent, nonchiral, and selective inhibitors of blood coagulation factor Xa. *J. Med. Chem.* **2002**, *45*, 2749–2769.

(66) Kamp, M. W.; van der Schaeffer, R. D.; Jonsson, A. L.; Scouras, A. D.; Simms, A. M.; Toofanny, R. D.; Benson, N. C.; Anderson, P. C.; Merkley, E. D.; Rysavy, S.; Bromley, D.; Beck, D. A. C. Valerie Daggett Dynaomics: A comprehensive database of protein dynamics. *Structure* **2010**, *18*, 423–435.

(67) Greco, M.; Hawkins, M.; Powell, E.; Almond, H. J.; Corcoran, T.; Garavilla, L. de; Kauffman, J.; Recacha, R.; Chattopadhyay, D.; Andrade-Gordon, P.; Maryanoff, B. Nonpeptide inhibitors of cathepsin G: optimization of a novel beta-ketophosphonic acid lead by structure-based drug design. *J. Am. Chem. Soc.* **2002**, *124*, 3810–3811.

(68) Nair, S.; Calderone, T.; Christianson, D.; Fierke, C. Altering the mouth of a hydrophobic pocket. Structure and kinetics of human carbonic anhydrase II mutants at residue Val-121. *J. Biol. Chem.* **1991**, *266*, 17320–17325.

(69) Pearson, M.; Karplus, P.; Dodge, R.; Laity, J.; Scheraga, H. Crystal structures of two mutants that have implications for the folding of bovine pancreatic ribonuclease A. *Protein Sci.* **1998**, *7*, 1255–1258.

(70) Romier, C.; Reuter, K.; Suck, D.; Ficner, R. Crystal structure of tRNA-guanine transglycosylase: RNA modification by base exchange. *EMBO J.* **1996**, *15*, 2850–2857.

(71) Congreve, M.; Chessari, G.; Tisi, D.; Woodhead, A. Recent developments in fragment-based drug discovery. *J. Med. Chem.* **2008**, *51*, 3661–3680.