JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

# On the Applicability of Elastic Network Normal Modes in Small-Molecule Docking

Matthias Dietzen,*[,†] Elena Zotenko,[†,‡] Andreas Hildebrandt,[§] and Thomas Lengauer[†]

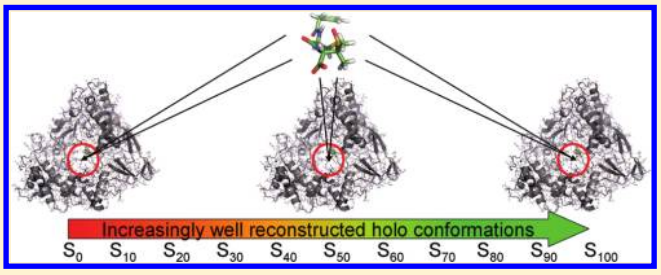
[†]Max-Planck-Institut für Informatik, 66123 Saarbrücken, Germany

[‡]Garvan Institute of Medical Research, Darlinghurst NSW 2010, Australia

[§]Johannes-Gutenberg Universität, 55099 Mainz, Germany

Supporting Information

**ABSTRACT:** Incorporating backbone flexibility into protein−ligand docking is still a challenging problem. In protein−protein docking, normal mode analysis (NMA) has become increasingly popular as it can be used to describe the collective motions of a biological system, but the question of whether NMA can also be useful in predicting the conformational changes observed upon small-molecule binding has only been addressed in a few case studies. Here, we describe a large-scale study on the applicability of NMA for protein−ligand docking using 433 apo/holo pairs of the Astex data sets. On the basis of sets of the first normal modes from the apo structure, we first generated for each paired holo structure a set of conformations that optimally reproduce its $C_\alpha$ trace with respect to the underlying normal mode subspace. Using AutoDock, GOLD, and FlexX we then docked the original ligands into these conformations to assess how the docking performance depends on the number of modes used to reproduce the holo structure. The results of our study indicate that, even for such a best-case scenario, the use of normal mode analysis in small-molecule docking is restricted and that a general rule on how many modes to use does not seem to exist or at least is not easy to find.




## INTRODUCTION

The molecular basis of diseases resides in processes pertaining to the function and interaction of proteins and other biological macromolecules. With the help of computer-aided drug design, potent drugs can be developed that influence the activity of these molecules and lessen or neutralize the effects of functional disorders. Such drugs can, for example, stimulate signal-transduction pathways, inhibit a protein's catalytic function, modulate protein−protein interactions, or change the rate at which a gene is transcribed.[1−4]

However, the accurate and fast prediction of promising candidate molecules and the correct protein−ligand complex conformation—preferably combined with an accurate estimation of the respective binding affinities—is still an unsolved problem. To be useful in a high-throughput virtual screening, docking must be computationally very efficient but still produce reliable results.[5,6] However, fast and accurate scoring functions that efficiently guide the search for the final protein−ligand complex are currently not available, and probably will not be in the near future. To make matters worse, the energy landscape of the protein changes through the binding of a ligand: on the one hand, the ligand itself changes the landscape by establishing interactions with the protein, and on the other hand, both protein and ligand are not rigid bodies but are able to undergo substantial conformational changes.[7−9] To tackle these problems in a reasonable computation time, docking algorithms are forced to apply simplifications that, however, typically reduce the accuracy of the predictions.

In this context, the treatment of ligand and protein flexibility must usually be different. While ligands are chemically much more diverse than proteins, they contain significantly fewer degrees of freedom. Thus, approaches to handling ligand flexibility use different global optimization schemes, such as Monte Carlo methods, e.g. ICM[10] or LigandFit,[11] genetic algorithms such as AutoDock[12] or GOLD,[13] incremental construction like FlexX,[14] or grid-based methods like Glide.[15,16] Protein flexibility, in contrast, requires a different strategy, since here the number of degrees of freedom becomes forbiddingly large, even for small proteins. In a first step, the protein's movements are decomposed into side-chain rearrangement and backbone movement. Side-chain conformers can be described suitably by a discrete set of rotamers,[17] but optimizing their arrangement via exhaustive sampling incurs the risk of combinatorial runtime explosion. Most of the earliest methods able to capture small induced fit effects[18] thus concentrated on locally sampling the side-chain conformations of the active site during the docking process.[19−27] Others used pregenerated protein ensembles and, in some cases, also considered different backbone conformations.[28−35] Such molecular structure ensembles are assumed to be representative

**Received:** October 12, 2011
**Published:** February 10, 2012

for the conformational space the protein can explore, but they usually fail when large-scale backbone movements are involved, as for example observed in HIV-1 protease[36] or aldose reductase.[37]

To describe such conformational alternatives of proteins, different techniques can be applied. One possibility is to carry out a principal component analysis of movements, determined by MD simulations.[38,39] These so-called essential modes have been shown to improve the docking performance significantly.[40,41] But because MD simulations are computationally expensive, approximating the global dynamics of proteins by normal mode analysis (NMA) on the basis of coarse-grained elastic network models (ENM)[42] has become increasingly popular over the last years. Due to its ability to reproduce the collective motions of proteins without significant loss of accuracy,[43−51] NMA has been applied to many different problems, for example protein domain decomposition,[44,52,53] guiding MD simulations along normal modes,[54,55] or fitting proteins into electron density maps obtained from cryo-EM or X-ray crystallography.[56−59]

Besides the convenient strategy to use the $C_\alpha$ trace of the backbone, sets of atoms or residues can be combined into blocks.[60−64] Other approaches explore subsets of protein components,[65−67] use additional grains that represent side-chain centroids,[68] or employ a mixed coarse-graining with a higher resolution in important protein regions.[69] So far, several studies have established the ability of elastic network models to also predict conformational changes during protein−protein docking.[70−73] Moreover, a recent paper has proposed a sophisticated method to sample protein conformations using an ENM.[74]

However, in protein−ligand docking, the binding interfaces are typically much smaller than in protein−protein docking and there are theoretical arguments for the assumption that normal modes may not be suitable to model backbone movements involved in ligand binding. On the one hand, the primary purpose of normal modes is to describe large-scale collective motions of a system and thus they may not be able to model the more local movements related to ligand binding. On the other hand, differences between several conformations of the same protein which are interpreted to be due to protein motion may, in fact, be the result of uncertainties in the coordinates of experimentally determined structures. Thus it may not be adequate to use normal modes to interpret such differences. However, despite these assumptions, normal modes have already been successfully applied in select cases of protein−ligand docking as, for example, in two studies using normal modes from heavy-atom and all-atom ENM.[75,76]

These studies indicate that conformational changes observed in protein-small molecule binding may, in some cases, only be modeled when using noncollective modes. The question arises how this observation translates to coarse-grained ENMs using $C_\alpha$ atoms, where the number of modes is drastically reduced, and how suitable they are for such problems where alternative protein conformations are needed to improve docking results. In this context, May et al. have shown in a cross-docking study with six different CDK2 inhibitors[77] that NMA can significantly improve docking results while Cavasotto et al. have reported similar results in their study with cAPK.[78]

In this study, we thus investigate on a larger scale how suitable binding-pocket restricted normal modes from a $C_\alpha$-ENM are for protein−ligand docking, with a focus on high-throughput applications. By establishing a best-case scenario for conformational sampling on a diverse data set derived from the Astex Diverse[79] and Non-Native[80] Set, we evaluate how the number of modes used to reproduce a ligand-bound (holo) conformation from its respective unbound (apo) state influences the docking accuracy. The corresponding ligands are docked into the reproduced holo structures using AutoDock,[12] GOLD,[13] and FlexX.[14]

## MATERIALS AND METHODS

**Normal Mode Analysis for Elastic Network Models.** The aim of performing a normal mode analysis (NMA) on biological macromolecules[81,82] is to determine the global and energetically most favorable motions of a system close to its energetic minimum. Here, the assumption is that a protein in an energetically stable conformation oscillates harmonically around this equilibrium. A normal mode represents a collective motion within such a system at a certain oscillation frequency. Each mode has its own unique frequency which is proportional to the energy required with respect to the underlying potential to perform a unit length motion along the mode.

To determine these modes, we use an elastic network model (ENM),[46] an extension of the Gaussian network model (GNM),[43−45,47] that accounts for the anisotropy of motions of a system's components in Cartesian space. Here, an artificial harmonic potential $V$ is constructed around an assumed minimum energy protein conformation $\mathbf{R}^0$ consisting of $N$ mass points ($C_\alpha$ atoms in our case), such that $\mathbf{R}^0$ becomes the minimum conformation of $V$. The potential energy of a conformation $\mathbf{R}$ in this elastic network model is given by

$$V(\mathbf{R}) = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} k(|\mathbf{R}_i{}^0 - \mathbf{R}_j{}^0|) \\ (|\mathbf{R}_i - \mathbf{R}_j| - |\mathbf{R}_i{}^0 - \mathbf{R}_j{}^0|)^2 \quad (1)$$

The quadratic term describes a spring between two mass points $i$ and $j$ that is relaxed in $\mathbf{R}^0$. It can be easily seen that $V(\mathbf{R}) = 0$ when $\mathbf{R} = \mathbf{R}^0$ and $V(\mathbf{R}) > 0$ anywhere else. Furthermore, to scale the influence of each spring according to the distance of the participating mass points, each spring is assigned a spring constant by a function $k(d)$ which decreases with the distance $d$ between the involved mass points. In this way, close spatial neighbors can be made to contribute more strongly to the potential than remote atoms.

Taylor expansion of $V(\mathbf{R})$ around the minimum $\mathbf{R}^0$ up to second order yields

$$V(\mathbf{R}) = \frac{1}{2}(\mathbf{R} - \mathbf{R}^0)^T\mathbf{H}(\mathbf{R} - \mathbf{R}^0) \quad (2)$$

where $\mathbf{H}$ is the Hessian matrix containing all second partial derivatives of $V$ evaluated at $\mathbf{R}^0$ (a detailed derivation of the Hessian in an ENM can, for example, be found in the work of Atilgan et al.[46]).

By construction, $\mathbf{H}$ is positive semidefinite and hence has real eigenvectors and all eigenvalues are either positive or zero. The normal modes are defined as the eigenvectors $\mathbf{U}$ of $\mathbf{H}$, which, together with the eigenvalues $\mathbf{\Lambda}$, are obtained by carrying out an eigenvalue decomposition on the Hessian:

$$\mathbf{H} = \mathbf{U}^T\mathbf{\Lambda}\mathbf{U} \quad (3)$$

Usually, the normal modes are sorted in ascending order with respect to their eigenvalues as these correspond to the energetic


dx.doi.org/10.1021/ci2004847 | J. Chem. Inf. Model. 2012, 52, 844−856

cost required to perform a unit length movement along a mode in the energetic model of the ENM. Modes that require little energy are considered the most collective ones, i.e. the associated motion involves a large part of the underlying system. The first six modes have zero eigenvalues and correspond to the translational and rotational degrees of freedom of the whole system—in the case of an elastic network model, these obviously require no energy.

**Extracting Binding Pocket Normal Modes.** In protein−ligand docking, we are especially interested in the conformational changes of the binding site. But by calculating the normal modes for the whole protein, we will obtain many normal modes that are associated with collective movements elsewhere in the protein. Hence, to restrict the normal mode set to those modes that are collective for the binding pocket and thus relevant for protein−ligand docking, we use an approach described in the work of Zheng et al.[65] and Ming et al.:[66] we divide the protein into two components, the binding pocket and the remaining protein, by rearranging $\mathbf{H}$ such that we obtain four submatrices:

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_{pp} & \mathbf{H}_{pe} \\ \mathbf{H}_{ep} & \mathbf{H}_{ee} \end{bmatrix} \tag{4}$$

$\mathbf{H}_{pp}$ and $\mathbf{H}_{ee}$ contain the couplings within binding pocket and environmental protein, respectively, and $\mathbf{H}_{ep}$ and $\mathbf{H}_{pe}$, those between pocket and the remaining protein. We can now assume that upon a conformational change $\mathbf{r}_p$ in the binding pocket, the environmental residues perform an adaptive movement $\mathbf{r}_e$ that minimizes the total energy in the elastic network. $\mathbf{r}_e$ is given by

$$\mathbf{r}_e = -\mathbf{H}_{ee}^{-1}\mathbf{H}_{ep}\mathbf{r}_p \tag{5}$$

Under this assumption, we define the effective Hessian for our binding pocket as

$$\mathbf{H}_{eff} = \mathbf{H}_{pp} - \mathbf{H}_{pe}\mathbf{H}_{ee}^{-1}\mathbf{H}_{ep} \tag{6}$$

The normal modes $\mathbf{U}_{eff}$ for the binding pocket can then be obtained from $\mathbf{H}_{eff}$ using eq 3. The resulting modes provide an orthonormal basis set which is again sorted in ascending order according to the modes' eigenvalues and thus to their degree of collectivity. Hence, the first modes describe the most collective motions within the subsystem represented by $\mathbf{H}_{eff}$. The adaptive modes for the remaining protein can be calculated from $\mathbf{U}_{eff}$ and eq 5.

**Data Set.** The data set we used is derived from the Astex Diverse[79] and the Astex Non-Native Set.[80] The Astex Diverse Set comprises 85 diverse protein crystal structures, bound to drug-like ligands, with a resolution of less than 2.5 Å. The structures have been automatically and manually checked for structural problems, i.e. clashes, interactions with symmetry units, and dubious ligand binding. The Non-Native Set was set up analogously and consists of 1112 non-native (alternative) protein conformations − apo structures as well as holo conformations with corresponding ligands − for 65 of the reference structures contained in the Diverse Set. The binding pockets of the non-native structures are unmutated with respect to those of the reference structure and have been superimposed onto the reference pocket for a better comparability of the results of cross-docking studies.

In our study, the aim is to address the question whether and if so, how well, reduced sets of normal modes can model conformational changes observed during binding of small molecules. We thus chose only reference structures with at least one apo structure in the Non-Native Set to avoid bias of the protein backbone conformation toward any ligand.

For each of the 29 remaining reference structures, we determined the residues that are involved in substrate binding in any of the corresponding holo structures and are thus associated with possible conformational changes in the protein binding site: we first independently defined the binding pocket for each holo structure as consisting of those residues that have at least one heavy atom within a distance of 6.0 Å to a heavy atom of the corresponding ligand. These pockets were then aligned and merged into one extended active site (EAS) that comprises for each conformation all the residues that are in contact with any of the ligands. The respective residues make up the residues contributing to $\mathbf{H}_{eff}$; the remaining residues form the environment which is assumed to perform an adaptive movement that minimizes the global energy required for the conformational changes in the binding site.

For each of the 29 reference structures, we only kept those apo/holo pairs that have no mismatches or indels in the EAS between apo and holo structure. The resulting data set consisted of 283 apo/holo pairs from 20 reference structures, with 260 having a $C_\alpha$-rmsd (root mean square deviation) below 0.5 Å in the EAS.

To gain more data on structural differences exceeding a $C_\alpha$-rmsd of 0.5 Å while keeping the effect of structural mutations on the protein dynamics small, we decided to also incorporate apo/holo pairs with a $C_\alpha$-rmsd of at least 0.5 Å and at most five mismatches which, however, must not occur within the EAS. In this way, we augmented our data set by 150 additional apo/holo pairs while ensuring that the docking results do not suffer from mutations in the binding pocket. The complete data set contains 433 apo/holo pairs from 21 different reference structures. A distribution of the $C_\alpha$-RMSDs can be found in Figure 1. Each structure in this data set was converted into pdb format; missing atoms and side-chains were added with BALL.[83] The corresponding ligands were converted into mol2 format using OpenBabel.[84]

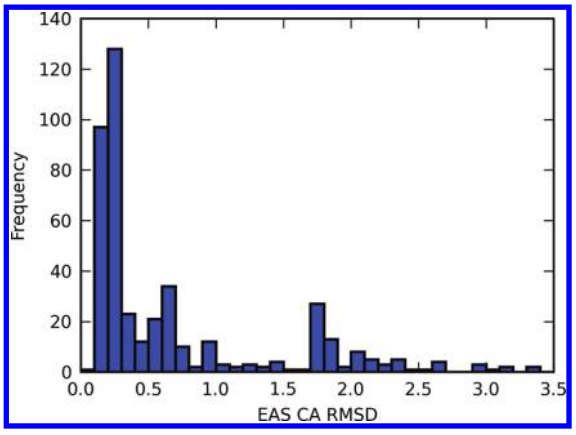



**Figure 1.** Distribution of $C_\alpha$-rmsds in the extended active site.

**Establishing a Best-Case Scenario.** From the apo/holo pairs of the data set derived in the previous section, we then generated intermediate structures that optimally reproduce the holo conformation. For each apo structure, we first calculated the effective modes $\mathbf{U}_{eff}$ for the residues contained in its EAS as


dx.doi.org/10.1021/ci2004847 | J. Chem. Inf. Model. 2012, 52, 844−856

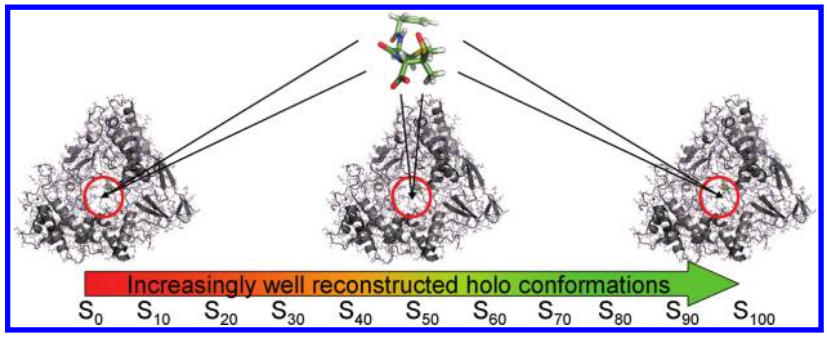



**Figure 2.** Schematic description of our best-case scenario (shown for acetylcholinesterase, pdb code 1gpk). The native ligands are sequentially docked into a set of increasingly well reconstructed holo conformations to assess how the docking performance depends on the number of modes used for the reconstruction.

described in the previous sections and established subsets of the first $m$ ($m$ = 10%, 20%, ..., 100%) of these modes.

Given the space **S** spanned by such a subset,

$$\mathbf{S} = \text{span}\{\mathbf{U}_1, ..., \mathbf{U}_m\} \quad (7)$$

we then orthogonally projected the conformational difference between apo and holo binding pockets, $\mathbf{R}_A$ and $\mathbf{R}_H$ onto **S**:

$$\mathbf{P} = \mathbf{S}(\mathbf{R}_H - \mathbf{R}_A) \quad (8)$$

From this projection, we can then obtain the amplitudes **A** for the modes that minimize the distance between this projection and the space **S**:

$$\mathbf{A} = (\mathbf{S}^T\mathbf{S})^{-1}\mathbf{P} \quad (9)$$

We can then generate an approximate conformation $\mathbf{R}_H^*$ of the holo binding pocket from that of the apo structure with respect to the underlying space **S** by

$$\mathbf{R}_H^* = \mathbf{R}_A + \mathbf{AS} \quad (10)$$

Due to the orthogonal projection, the distance between the $C_\alpha$ trace of the holo conformation, $\mathbf{R}_H$, and $\mathbf{R}_H^*$ is minimal with respect to **S**. Thus, a conformational sampling in the same subspace can never yield a conformation that is closer to the original holo structure than our intermediate structure. Our intermediate structures can thus be considered as an upper bound for the accuracy achievable with conformational sampling algorithms with respect to the underlying normal mode subspace.

Applying the above procedure to each of the generated mode subsets leads to increasingly well reproduced holo $C_\alpha$ conformations. This makes it possible to investigate how the docking performance relates to the number of modes used to reconstruct the holo conformation and to estimate how many modes are needed to sufficiently reproduce the conformational change upon ligand binding.

To prepare these conformations for docking, the all-atom structures were reconstructed by translating the side-chains and remaining backbone atoms according to the displacement of the corresponding $C_\alpha$ atom. We then relaxed the resulting structure for 0, 10, and 50 steps using the AMBER96[85] force field and an L-BFGS minimizer[86,87] to resolve possible steric clashes between side-chains and/or backbone atoms while keeping $C_\alpha$ atoms fixed. In this context, when using all available modes ($m$ = 100%), we obtain a so-called 100% reconstructed holo conformation, i.e. the $C_\alpha$ trace of the holo conformation binding pocket is exactly reproduced while the remaining part of the protein (side-chains, non-$C_\alpha$ backbone atoms) may show deviations from the original holo structure due to the reconstruction procedure. A validation of this reconstruction procedure can be found in the Supporting Information.

With this procedure, we obtained 33 intermediate structures per apo/holo pair, in total. The full data set to be docked contained 15 304 protein conformations derived from 433 apo/holo pairs and the original holo structures (see the Supporting Information for more details on the data set composition). A schematic representation of our best-case scenario is given in Figure 2.

**Docking Experiments.** To analyze the quality of the structures with respect to docking, we performed two different docking rounds. In the first round, we investigated the capability of normal modes without considering the side-chain conformations. We established six different docking protocols, consisting of a standard and a soft docking setup for each of the docking programs AutoDock,[12] GOLD,[13] and FlexX.[14] The standard protocols used the default parameters of the respective docking program; for AutoDock and GOLD, the number of runs was set to 25 in both the standard and soft docking protocol. Furthermore, for the soft docking protocols, we adjusted the parameters to reduce the impact of steric clashes. In AutoDock, FE_coeff_vdW was reduced by a factor of 0.5 while in GOLD, start_vdw_linear_cutoff was set to 4 and the binding pocket residues were assigned a 2−4 vdW potential. In FlexX, MAX_OVERLAP_VOL and DOT_OVERLAP_VOL were increased by a factor of 1.5.

In the second round, side-chain flexibility was explicitly taken into account as it can have a significant impact on the docking performance. We therefore selected those apo/holo pairs from the first round for which the ligand could be successfully redocked into the original holo structures (i.e., with a minimum pose rmsd below 2.0 Å) with at least one of the above-described docking protocols, but failed to do so for the 100% reconstructed holo conformation. We established three additional docking protocols that account for side-chain flexibility. AutoDock directly incorporates side-chain flexibility (at the cost of greatly increased runtimes), but the maximum number of torsions is restricted to 32. We thus iteratively chose binding pocket residues with growing distance to the ligand as long as the total number of torsions did not exceed this threshold. The second protocol uses FlexX with binding pocket side-chain conformations generated with SCWRL,[88] the third employs FlexE[29] with side-chain ensembles derived with IRECS[89] (rotamer density 3 and at most 3 additional side-chain conformations per binding pocket residue). All protocols used the default parameters of the respective docking


dx.doi.org/10.1021/ci2004847 | J. Chem. Inf. Model. 2012, 52, 844−856

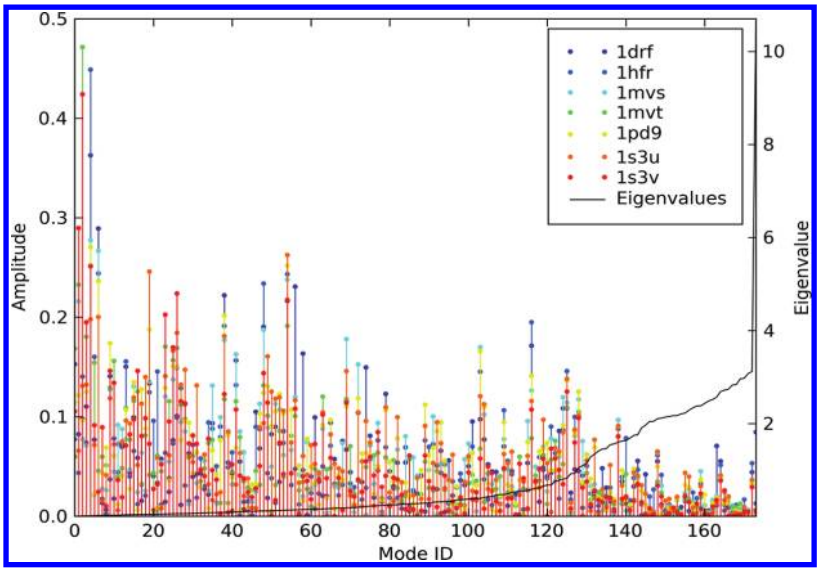



**Figure 3.** Mode amplitude spectra for a reconstruction of different dihydrofolate reductase holo structures from one single apo structure (1pdb). The modes are sorted by increasing eigenvalues. Modes and amplitudes (dimensionless scalar factors) differ for a reconstruction of the individual holo conformations.

algorithm; the number of runs for AutoDock was again set to 25.

In contrast to the EAS which is used to define the residues supposedly involved in the backbone movements relevant for the binding pocket, for both docking rounds, we used the smaller, ligand-specific binding pocket which we defined as consisting of all residues with heavy atoms within a distance of at most 6.0 Å to any of the ligand's heavy atoms.

The resulting docking poses were evaluated by calculating the symmetry-corrected rmsd to the crystallized ligand structure using the smartrms program included in GOLD. However, the docking poses also depend on the used protein conformation. The reconstructed structures are not identical with the original holo structure, and the rmsd between the crystallized ligand and the docked pose may thus be biased toward the input protein conformation. But the ligand may nevertheless be able to adapt itself to slightly different protein conformations and establish the same interactions as present in the crystal structure. We thus additionally calculated the symmetry-corrected fraction of native ligand contacts of the crystal structure realized in each docked pose. The ligand contacts were determined using HBPLUS and HBADD[90] as implemented in LigPlot;[91] the symmetry-corrected fraction of a pose is given as the maximum fraction of native ligand contacts over all its automorphisms as calculated by OpenBabel.[84]

## RESULTS AND DISCUSSION

**Analysis of Normal Mode Amplitude Spectra.** The general assumption in protein−protein docking as well as conformational studies of proteins is that only a few modes are required to reproduce most collective, global conformational changes that the protein is able to perform.

To investigate whether this assumption also holds in the protein−small molecule docking case, we thus first compared the mode amplitude spectra for a full reconstruction of the $C_\alpha$ trace of different holo structures using the normal modes obtained from the effective Hessian (for the selection of an optimal spring force function and a comparison to normal modes computed from force-field derived Hessians, please see the Supporting Information) of one common apo conformation. If the initial assumption also holds for protein−small molecule docking, the used modes and the corresponding amplitudes should be similar for a reconstruction of different holo conformations. Figure 3 shows such a spectrum for the protein dihydrofolate reductase (for a better insight into the differences, the absolute amplitude values are shown). The modes are sorted by increasing eigenvalues, such that only the first few modes should suffice to represent a conformational change if the fundamental assumption behind the normal modes procedure is valid.

The main difference between the displayed structures comprises a conformational change in two loops, while the rest of the system remains relatively rigid. While these movements do not involve the entire protein, they exhibit a certain amount of collectivity on the scale of the binding pocket, on which we focus with our effective Hessian approach. Hence, the spectrum should contain regions with a clearly similar behavior for all holo structures.

It can be clearly seen that this is not the case: not only do the amplitudes differ, but also modes that strongly contribute to reproducing one conformation have almost no influence for other conformations and vice versa, a fact which makes an a priori selection of relevant modes difficult. For example, a conformational change from 1pdb to 1s3v (red) requires a large amplitude for mode 23, while a reconstruction of 1drf (blue) can be performed very accurately without using that mode. On the other hand, generating the backbone conformation of 1drf cannot be achieved without using mode 58 (with an even higher amplitude than mode 23 for 1s3v) while this mode plays almost no role for 1s3v.

One reason for this difference in the relevance of modes is that the eigenvalues which correspond to the energy required to perform a movement along a mode are very similar for a large fraction of the modes: a protein that is excited by a certain amount of energy distributes this energy evenly among all its degrees of freedom. Movements that require less energy thus dominate those that need much energy. However, if there are many energetically similarly demanding modes, the space of

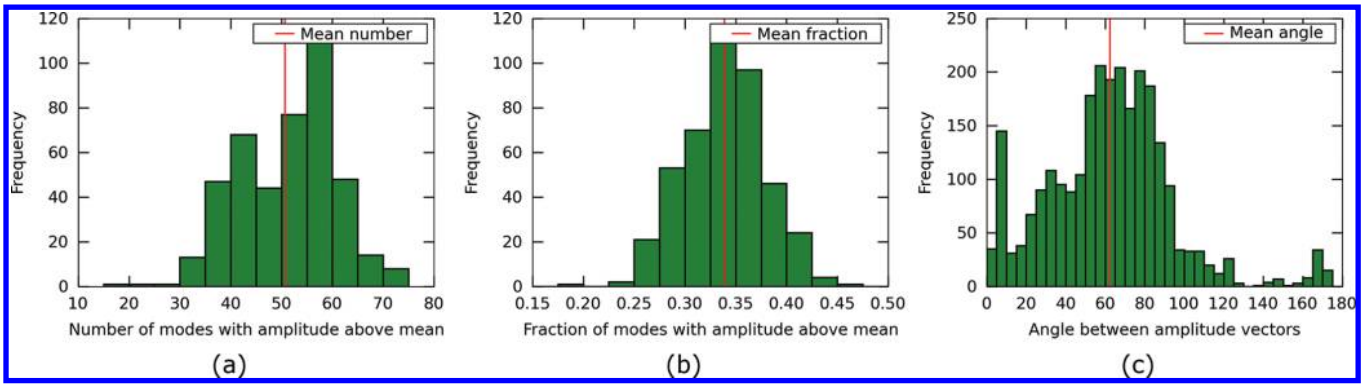



**Figure 4.** Distributions of the number (a) and fraction of modes (b), as well as the pairwise angles between amplitude vectors (c) for a reconstruction of different holo structures from the same apo structure.

possible motions grows exponentially and the conformations become more diverse, as a result. In addition, elastic network models do not account for anharmonic motions, which may become especially important for small-scale, local backbone movements; a fact that may also contribute to the different relevance of the modes. Furthermore, a bound ligand can shift the minima on the protein energy landscape and, thus, conformations that are less likely in the absence of a ligand may attain a lower overall energy in the bound complex due to favorable interactions with a ligand. This is in concordance with the observation that, as soon as the eigenvalues increase significantly (approximately mode 130 in the case of 1pdb), the mode amplitudes decrease for the whole set of reconstructed holo conformations.

The results for the whole set of apo/holo pairs are summarized in Figure 4 and confirm the findings detailed for dihydrofolate reductase: the distribution of the number of modes with an absolute amplitude greater than average indicates that a sampling in a small set of modes is not sufficient to reconstruct a holo structure with high accuracy. On average, 51 modes are responsible for the largest part of the conformational shift, an observation that is consistent with the results shown in Figure 3. The corresponding fractions of modes (Figure 4b) to be used range from 0.19 to 0.46 with a mean value of 0.34. Accordingly, at least one-third of all modes must be considered in a conformational sampling.

However, an a priori selection of modes seems hard at best, as the amplitude vectors greatly differ in their composition between complexes with different ligands. Figure 4c shows the distribution of pairwise angles between amplitude vectors for a reconstruction of different holo structures from the same apo conformation. The main fraction of angles lies between 50° and 90° (mean value 63.1°), a fact that not only confirms the assumption that the relevance of modes highly depends on the bound ligand, but also shows that many amplitude vectors are almost perpendicular to each other and that modes which are switched off in one complex are essential in another. Thus, in the normal case that the bound conformation is unknown and an a priori predictor for the modes relevant for individual ligands does not exist, the number of modes that has to be sampled can be expected to be well above the average number of 51.

**Docking into Reconstructed Holo Structures.** The analysis of the mode amplitudes has shown the highly diverse nature of transitions from apo to different holo structures and demonstrated that the important modes are distributed over almost the full range of modes. However, it has yet to be clarified whether all these modes or only a subset of the most collective ones (i.e., those with the lowest eigenvalues) are required to achieve a conformational change that results in a successful docking of the ligand.

We established 18 docking series, consisting of 6 different protocols using AutoDock, GOLD, and FlexX each in a standard and a soft setup, for each of the 3 different minimization lengths (0, 10, and 50 steps). While many (partially contradictory) studies that compare the performance of different docking tools exist, our primary aim of using different docking programs here is to ensure that the obtained results are not due to peculiarities of any of these tools. We thus do not compare the actual performances of the different programs but rather use them to frame a stable picture of the capability of normal modes to improve small-molecule docking. In some cases, the docking failed due to structural problems, e.g. when the reconstruction procedure produced irresolvable clashes or the atom types could not be assigned properly. The missing results were interpolated using natural splines; eight apo/holo pairs were excluded because they produced five or more missing values in at least one of the docking series.

Figure 5 illustrates the docking results for the remaining 425 pairs for each docking series. The results have been normalized to account for the unbalanced distribution of the number of apo conformations associated with each holo structure and the number of holo structures per protein. In addition to the data series for the three minimization protocols, a minimum envelope (ME) curve which considers only the optimum value obtained from the three protocols for each apo/holo pair and a linear least-squares fitted line for the ME curve are shown in each plot. For the data points on the ME curve in the pose rmsds, the standard errors in the mean are also shown.

In the ideal case, one would expect a steep decrease of the minimum rmsd for the most collective modes that diminishes as the normal mode subspace grows. This would indicate that only the first modes are required to produce a conformational change that is sufficient for a successful docking. However, in our best-case scenario, the overall drop in minimum rmsd is small and essentially linear (Figure 5a). This tendency is observable in each of the 18 docking series and even more imminent in the ME curves. Regardless of the actual docking performance of the different protocols, the reduction in rmsd compared to the docking results with the apo structure is at most 0.6 Å (mean value 0.4 Å) when including 100% of the modes in the holo reconstruction; for the first 20% of the modes the average reduction amounts to only 0.26 Å. This behavior is also reflected by the fitted lines: the steepest decay


dx.doi.org/10.1021/ci2004847 | J. Chem. Inf. Model. 2012, 52, 844−856

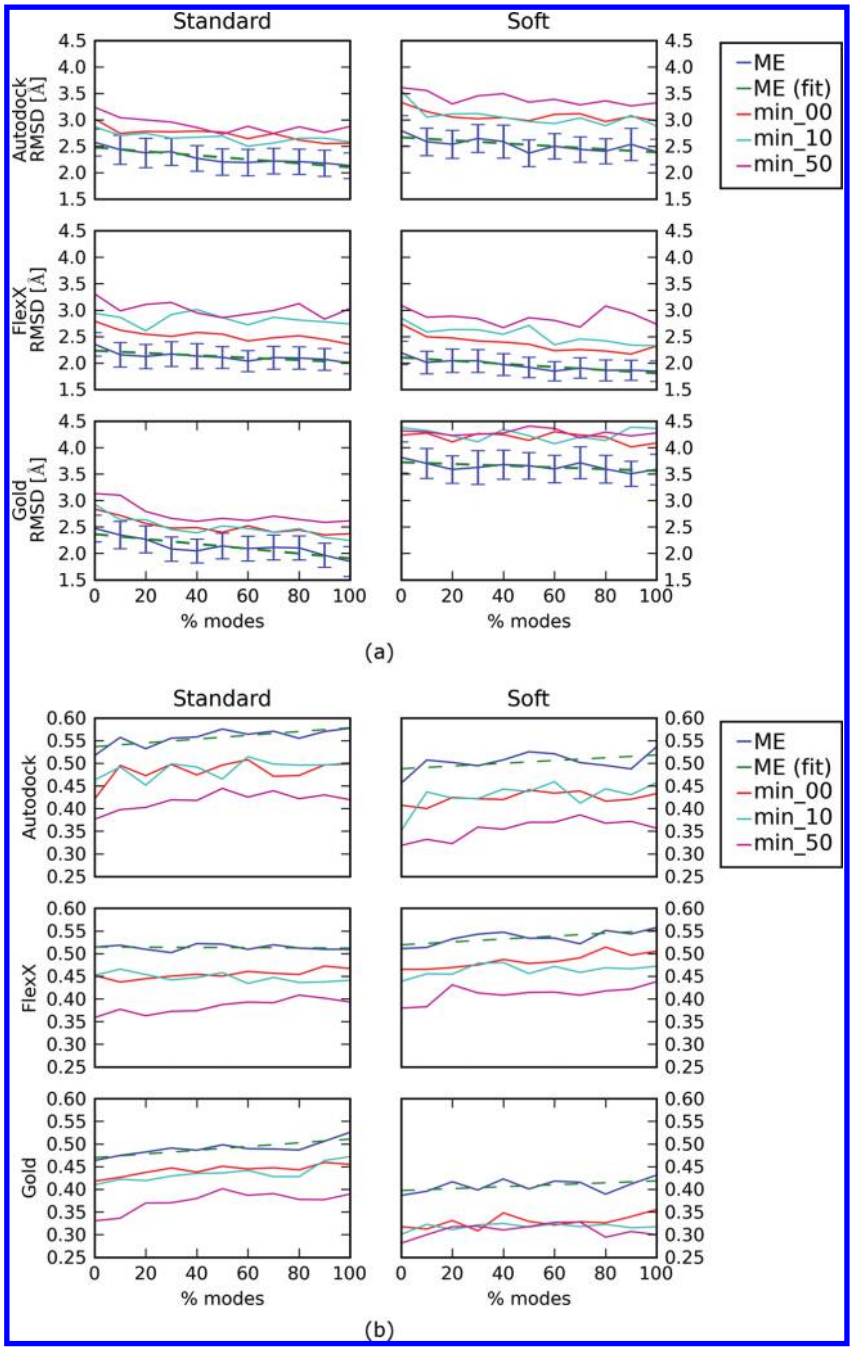



**Figure 5.** Best pose rmsd (a) and maximum fraction of contacts (b) for holo reconstructions with an increasing normal mode subset size averaged over the full data set.

was found to be −0.0046 with a residual sum of squares (RSS) of 0.07. The standard errors of the means for the 100% (and 0%) reconstruction in the standard protocols of AutoDock, FlexX, and GOLD are 0.24 (0.26), 0.20 (0.22), and 0.28 (0.25), respectively. The values for the soft protocols are comparable, indicating a significant improvement in all six protocols.

Similarly, the maximum fraction of native contacts (Figure 5b) grows linearly and increases by at most 0.081 when including all modes, and only by 0.049 for the first 20% of the modes. The fitted ME lines have a maximum slope of 0.00042 with an RSS of 0.0015.

We also investigated the top scores and the corresponding poses (data not shown), and the results reveal another factor that negatively affects the usability of normal modes for sampling binding pocket conformations: the top scores differ by ≈8%, on average, between apo and the reconstructed holo structures, which is far below the standard deviation of scores obtained from a typical docking run. A linear least-squares fit gave a maximum decay of −0.023 with an RSS of 1.68, showing that there is basically no decline in the top scores. Likewise, the top-pose RMSDs were reduced by at most 0.5 Å using a full reconstruction and only 0.17 Å for the first 20% of the modes. These results show that, even in the case that a conformational sampling in the most collective modes would reproduce the original holo structure, it will be difficult to find the correct protein−ligand complex in the set of generated protein conformations with today's scoring functions, unless additional


dx.doi.org/10.1021/ci2004847 | J. Chem. Inf. Model. 2012, 52, 844−856

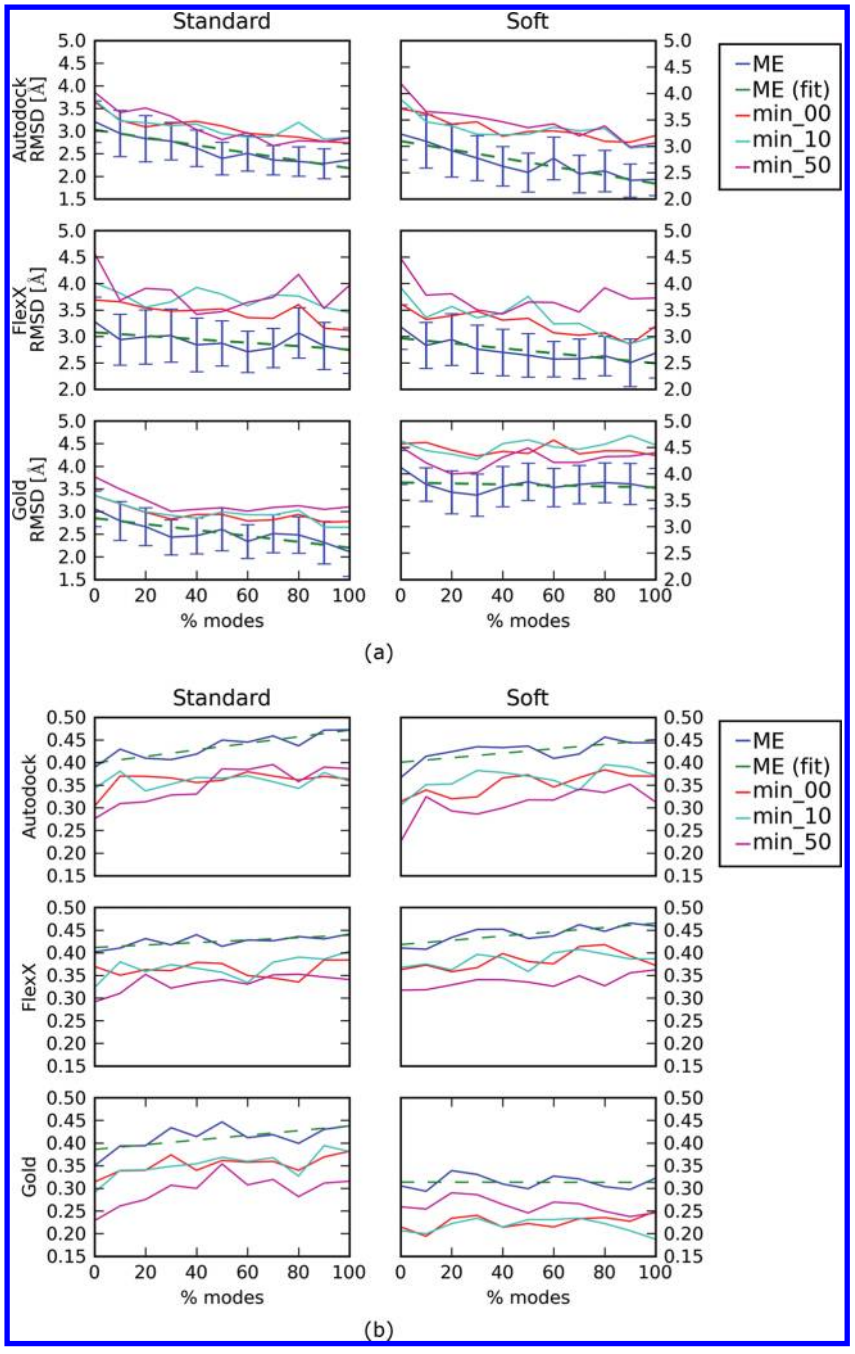



**Figure 6.** Best pose rmsd and maximum fraction of contacts for holo reconstructions with an increasing normal mode subset size averaged over the subset of apo/holo pairs with a $C_\alpha$ rmsd >0.5.

terms that estimate the plausibility of the different protein conformations are incorporated.

Because normal modes are expected to reproduce the large-scale motions of a protein especially well, we also investigated the subset of 165 apo/holo pairs with a $C_\alpha$ rmsd >0.5 Å (Figure 6). The results are mostly comparable to those on the full data set: Due to the larger conformational difference between apo and holo structure, the best-pose rmsds obtained from docking into the apo conformation are larger on average than for the full data set (compare Figure 5).

The standard errors of the means for the 100% (and 0%) reconstruction in the standard protocols of AutoDock, FlexX, and GOLD have values of 0.36 (0.46), 0.43 (0.47), and 0.54 (0.39), respectively. This implies that, at least in some cases, the significance of improvement in docking performance is questionable; however, both AutoDock protocols and the GOLD standard protocol can be considered to achieve a significant improvement.

In comparison to the full data set, the best-pose rmsd decreases faster as more modes are used to reconstruct the holo structures. But the decay in rmsd is, in essence, still linear (maximum slope of the linear least-squares fit −0.0085, RSS 0.12) and the best-pose rmsds for the 100% reconstructed holo conformations are 0.36 Å greater than those of the full data set, on average.

The best results for a single protein (data not shown) were obtained for aldose reductase (pdb code in the Astex Diverse Set: 1t40) where the best-pose rmsd resulting from docking


dx.doi.org/10.1021/ci2004847 | J. Chem. Inf. Model. 2012, 52, 844−856

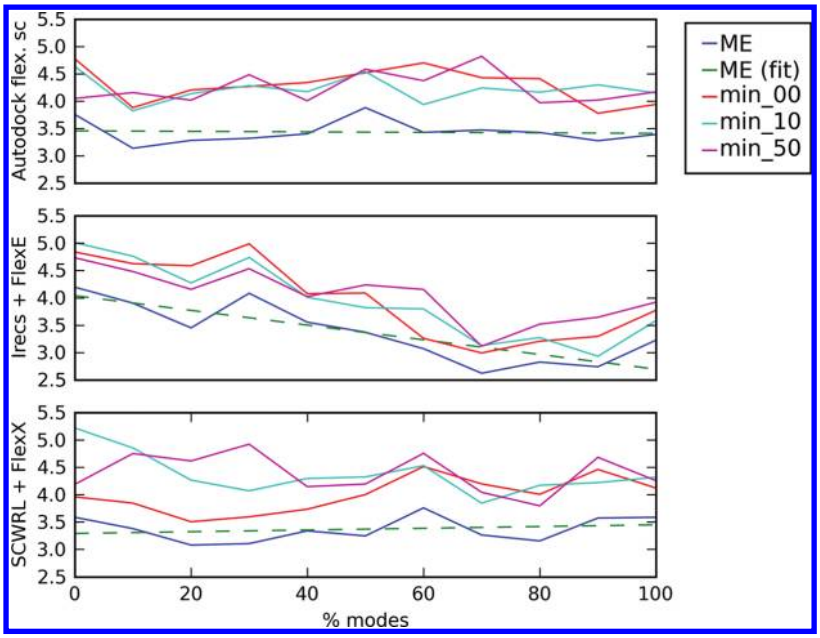



**Figure 7.** Performances of the three docking protocols explicitly accounting for side-chain flexibility.

into the apo structure consistently dropped from values of between 3.89 and 4.34 Å to below 2.0 Å after an inclusion of 50% of the modes in four of the six docking protocols. Although such an increase in performance seems encouraging, at first sight, using 50% of the modes is still an infeasible task for a conformational sampling, especially in a high-throughput setting.

These findings indicate that, even for larger $C_\alpha$ RMSDs, the movements in the binding pocket upon ligand binding are not collective enough to be represented by a small set of normal modes, in general. Thus it may be indispensable to use the ligand in some way to perform a preselection of the required modes, or to directly guide the conformational change upon ligand binding. To do so seems difficult, however, given the problems of today's scoring functions in discriminating between correct poses and decoys.

**Docking with Side-Chain Flexibility.** In this section, we study how strongly the previous results depend on the side-chain conformations in the binding pocket. To this end, we selected the 59 apo/holo pairs from our data set that were successfully redocked into the original holo structure with a best-pose rmsd of less than 2.0 Å in at least one of the docking series but failed to do so for the corresponding 100% reconstructed holo structure. Because the $C_\alpha$ trace of this structure is identical to that of the crystal holo structure, the problem reduces to non-$C_\alpha$ backbone atoms and, more importantly, the side-chain conformation.

Figure 7 shows the results for the three additional docking protocols: AutoDock with flexible side-chains, FlexE with IRECS-computed side-chain ensembles, and FlexX using side-chains generated with SCWRL. AutoDock with flexible side-chains as well as SCWRL+FlexX show no clear tendency toward an improvement, the fitted ME lines have slopes (RSS) of −0.0005 (0.46) and 0.0016 (0.49), respectively. In contrast, the line for IRECS+FlexE demonstrates that using a side-chain ensemble may help to improve the docking performance on a protein conformation generated using normal modes. The corresponding slope and RSS are −0.0134 and 0.97, respectively.

Partial improvements over the original best-pose rmsd and successful dockings could be obtained with all docking protocols, as can be seen in Table 1. In total, improvements

**Table 1. Results for the Docking Protocols with Flexible Side-Chains on the Fully Reconstructed Holo Structures**

| | | successes (best pose rmsd <2.0 Å) | |
|---|---|---|---|
| protocol | improvements | overall | best in comparison |
| AutoDock, flex. side-chains | 25 | 8 | 7 (27%) |
| Irecs + FlexE | 25 | 19 | 15 (58%) |
| SCWRL + FlexX | 14 | 10 | 4 (15%) |
| | 46 (78%) | | 26 (44%) |

in best-pose rmsd were achieved for 46 of the 59 apo/holo. AutoDock and FlexE both gave better results in 25 cases, FlexX in 14 cases. Altogether, at least one docking pose with an rmsd below 2.0 Å could be obtained for 26 of the 59 apo/holo pairs. FlexE was most successful with 19 poses in total and had the absolute minimum rmsd in comparison to the other docking results in 15 of these cases (last column). For the AutoDock protocol, these numbers were 8 and 7, respectively, for FlexX they were 10 and 4, respectively.

Nevertheless, not all docking results could be improved. The side-chain rotamers are backbone dependent, but even in the 100% reconstructed conformations, where the $C_\alpha$ trace is equal to that of the original holo structure, the conformation of the non-$C_\alpha$ backbone atoms can differ slightly from that of the original structure since the elastic network only acts on the $C_\alpha$ atoms and the remaining ones are approximately reconstructed from these. Using normal modes obtained from a backbone heavy-atom ENM instead yields the correct backbone conformation for a 100% reconstruction; however, doing so increases the set of resulting normal modes by a factor of 16, which leads to the conclusion that the chance to achieve better results in such a scenario is small, at best.

These results imply that the poor docking performance on normal-mode generated protein conformations may be


dx.doi.org/10.1021/ci2004847 | J. Chem. Inf. Model. 2012, 52, 844−856

improved when including side-chain flexibility. For docking algorithms that do not model side-chain flexibility explicitly, it may even be necessary to include not only one side-chain rotamer but an ensemble thereof to effectively increase the chances of a good docking result. However, this can greatly increase the computational effort required for conformational sampling, both for the generation of protein conformations and the dockings to be performed. But even when including side-chain flexibility, 70% of the modes were required, on average, to obtain a pose with an rmsd below 3.0 Å. This shows that the capability of normal modes to model binding pocket rearrangements is strongly limited even when accounting for side-chain conformations.

## CONCLUSIONS

The aim of this study was to empirically gain insight into the usability of binding-pocket normal modes obtained from $C_\alpha$-ENMs in protein−small molecule docking. We have established a scenario that provides an upper bound for conformational sampling algorithms: for known holo structures, we have generated optimal reconstructions with respect to differently sized normal mode subspaces retrieved from corresponding apo structures.

The analysis of mode amplitude spectra and the subsequent docking experiments have shown that the use of normal modes in protein−small molecule docking is limited: the amplitude vectors to be used differ greatly when reconstructing holo structures for different ligands from the same apo structure. This may not always be the case. If the conformation changes globally upon ligand binding, an improvement in docking accuracy can be achieved with a small set of modes, as shown in the work of May et al.[77] and Cavasotto et al.[78] In this study, Cavasotto et al. also introduced a measure of relevance to determine the modes that are involved in binding pocket conformational changes. This method makes is possible to narrow down the sampling space to a small set of modes and does not necessarily require the calculation of the effective Hessian as done in our study. Ensembles generated from such mode sets in this study have shown to improve docking results for several ligands of cAPK Kinase ligands. This approach is especially powerful if the mobility of a binding pocket is well-defined and mostly independent of the ligand, as for example in the conformational selection stage during protein movement. However, in case of local ligand-specific induced fit movements, if the binding pocket motions are unknown or cannot be well captured by a small set of relevant modes, this method is difficult to apply.

Even if the conformational change is not fully represented by the most collective modes, the amplitude vectors for the conformational change upon binding different ligands may show a high degree of similarity. For example in the case of calmodulin, a protein that changes its tertiary structure from an elongated form to a globular conformation when binding a ligand,[92] the amplitude vectors are very similar—even if the ligands are highly diverse—due to the dominating complexity and the highly distinct conformations in the bound and unbound state.[76] However, in our study, the conformational changes are less extensive and our results give rise to the assumption that, in such cases, the ligand information is of great importance when selecting the relevant modes, as most of the modes are energetically almost equivalent and nearly equally likely to be activated when binding a ligand.

But the problem how to include information on the ligand in the selection procedure is unsolved: state-of-the-art scoring functions are hardly accurate enough to even reliably select the original pose from the set of generated solutions in the redocking case. Hence, their general usability in helping to find the relevant modes for the given ligand during a conformational search is more than questionable. To our knowledge, approaches to directly use the ligand as a predictor for the relevant normal modes do not exist, and the question of whether this is possible at all has not even been tackled.

Due to the mutual dependence of finding the correct ligand conformation and determining the true protein conformation, researchers are currently forced to apply conformational sampling strategies. Our docking results show that sampling with a fraction of only the first and most collective modes is not sufficient to significantly improve the docking performance in general, i.e. when no large-scale motions are involved in ligand binding. Furthermore, although sampling with large fractions of the normal mode space can improve the results, the computational effort increases exponentially with the number of modes and thus the docking may become infeasible.

The additional docking experiments accounting for flexible side-chains show that it is often indispensable to adjust the side-chain conformations upon backbone movement and that doing so can enhance the docking performance when applied in combination with normal modes. But while the docking results could be improved using flexible side-chains, the number of modes required for obtaining reasonable results was still too large to be applicable in high-throughput settings.

Summarizing these observations, the general reduction in the complexity of modeling protein flexibility with normal modes in protein−ligand docking is small if relevant modes cannot be determined from some external criterion, because even in our best-case scenario, where the actual holo conformation and the path from the apo conformation are known, the gain in docking performance is small and will be hard to achieve in an actual sampling scenario. Moreover, structural uncertainties in the atom coordinates or the fact that normal modes are designed to mainly detect collective motions of a system and thus have problems to describe local changes involving single atoms can cause the normal modes to fail in protein−ligand docking. This leads to the strong assumption that the use of normal modes in protein−small molecule docking may be restricted to select cases where only few collective motions are responsible for binding a ligand.

## ASSOCIATED CONTENT

### Supporting Information

Supporting Information S1: detailed information on the data set composition. S2: selection of the spring force function in detail. A comparison of the selected spring force function with normal modes obtained from a molecular mechanics force field is provided in S3. The validation of the reconstruction performance is reported in S4. This material is available free of charge via the Internet at http://pubs.acs.org.

## AUTHOR INFORMATION

### Corresponding Author

*E-mail: mdietzen@mpi-inf.mpg.de. Phone: +49 681 9325 3020.

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

The authors thank the International Max Planck Research School for Computer Science (IMPRS-CS), Saarbrücken, Germany, for funding this work.

## ■ REFERENCES

(1) Goodey, N. M.; Benkovic, S. J. Allosteric regulation and catalysis emerge via a common route. *Nat. Chem. Biol.* **2008**, *4*, 474−482.
(2) Schenk, P. W.; Snaar-Jagalska, B. E. Signal perception and transduction: the role of protein kinases. *Biochim. Biophys. Acta Mol. Cell Res.* **1999**, *1449*, 1−24.
(3) Arkin, M. R.; Whitty, A. The road less traveled: modulating signal transduction enzymes by inhibiting their protein-protein interactions. *Curr. Opin. Chem. Biol.* **2009**, *13*, 284−290.
(4) Chang, E. C.; Charn, T. H.; Park, S.-H.; Helferich, W. G.; Komm, B.; Katzenellenbogen, J. A.; Katzenellenbogen, B. S. Estrogen Receptors $\alpha$ and $\beta$ as Determinants of Gene Expression: Influence of Ligand, Dose, and Chromatin Binding. *Mol. Endocrinol.* **2008**, *22*, 1032−1043.
(5) Sotriffer, C.; Mannhold, R.; Kubinyi, H.; Folkers, G. *Virtual Screening: Principles, Challenges, and Practical Guidelines*; John Wiley & Sons: New York, 2011.
(6) Zoete, V.; Grosdidier, A.; Michielin, O. Docking, virtual high throughput screening and in silico fragment-based drug design. *J. Cell. Mol. Med.* **2009**, *13*, 238−248.
(7) Tsai, C.-J.; Ma, B.; Nussinov, R. Folding and binding cascades: Shifts in energy landscapes. *Proc. Natl Acad. Sci.* **1999**, *96*, 9970−9972.
(8) Okazaki, K.-i.; Takada, S. Dynamic energy landscape view of coupled binding and protein conformational change: Induced-fit versus population-shift mechanisms. *Proc. Natl Acad. Sci.* **2008**, *105*, 11182−11187.
(9) Kondo, H. X.; Okimoto, N.; Morimoto, G.; Taiji, M. Free-Energy Landscapes of Protein Domain Movements upon Ligand Binding. *J. Phys. Chem. B* **2011**, *115*, 7629−7636.
(10) Abagyan, R.; Totrov, M.; Kuznetsov, D. ICM—A new method for protein modeling and design: Applications to docking and structure prediction from the distorted native conformation. *J. Comput. Chem.* **1994**, *15*, 488−506.
(11) Venkatachalam, C. M.; Jiang, X.; Oldfield, T.; Waldman, M. LigandFit: a novel method for the shape-directed rapid docking of ligands to protein active sites. *J. Mol. Graph. Model.* **2003**, *21*, 289−307.
(12) Goodsell, D. S.; Morris, G. M.; Olson, A. J. Automated docking of flexible ligands: applications of AutoDock. *J. Mol. Recognit.* **1996**, *9*, 1−5.
(13) Jones, G. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267*, 727−748.
(14) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.* **1996**, *261*, 470−489.
(15) 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.
(16) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004**, *47*, 1739−1749.
(17) Canutescu, A. A.; Shelenkov, A. A.; Dunbrack, R. L. A graph-theory algorithm for rapid protein side-chain prediction. *Protein Sci.* **2003**, *12*, 2001−2014.
(18) Koshland, D. E. Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proc. Natl Acad. Sci.* **1958**, *44*, 98−104.
(19) Leach, A. R. Ligand docking to proteins with discrete side-chain flexibility. *J. Mol. Biol.* **1994**, *235*, 345−356.
(20) Fernández-Recio, J.; Totrov, M.; Abagyan, R. ICM-DISCO docking by global energy optimization with fully flexible side-chains. *Proteins* **2003**, *52*, 113−117.
(21) Meiler, J.; Baker, D. ROSETTALIGAND: protein-small molecule docking with full side-chain flexibility. *Proteins* **2006**, *65*, 538−548.
(22) Källblad, P.; Dean, P. M. Efficient conformational sampling of local side-chain flexibility. *J. Mol. Biol.* **2003**, *326*, 1651−1665.
(23) Zavodszky, M. I.; Kuhn, L. A. Side-chain flexibility in protein-ligand binding: the minimal rotation hypothesis. *Protein Sci.* **2005**, *14*, 1104−1114.
(24) Frimurer, T. M.; Peters, G. H.; Iversen, L. F.; Andersen, H. S.; Møller, N. P.; Olsen, O. H. Ligand-induced conformational changes: improved predictions of ligand binding conformations and affinities. *Biophys. J.* **2003**, *84*, 2273−2281.
(25) Dominguez, C.; Boelens, R.; Bonvin, A. M. J. J. HADDOCK: A Protein−Protein Docking Approach Based on Biochemical or Biophysical Information. *J. Am. Chem. Soc.* **2003**, *125*, 1731−1737.
(26) Zacharias, M. ATTRACT: protein-protein docking in CAPRI using a reduced protein model. *Proteins* **2005**, *60*, 252−256.
(27) May, A.; Zacharias, M. Protein-protein docking in CAPRI using ATTRACT to account for global and local flexibility. *Proteins* **2007**, *69*, 774−780.
(28) Knegtel, R. M.; Kuntz, I. D.; Oshiro, C. M. Molecular docking to ensembles of protein structures. *J. Mol. Biol.* **1997**, *266*, 424−440.
(29) Claussen, H.; Buning, C.; Rarey, M.; Lengauer, T. FlexE: efficient molecular docking considering protein structure variations. *J. Mol. Biol.* **2001**, *308*, 377−395.
(30) 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.
(31) Osterberg, 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* **2002**, *46*, 34−40.
(32) Bottegoni, G.; Kufareva, I.; Totrov, M.; Abagyan, R. Four-dimensional docking: a fast and accurate account of discrete receptor flexibility in ligand docking. *J. Med. Chem.* **2009**, *52* (2), 397−406.
(33) Lorber, D. M.; Shoichet, B. K. Flexible ligand docking using conformational ensembles. *Protein Sci.* **1998**, *7*, 938−950.
(34) Zhao, Y.; Sanner, M. F. FLIPDock: docking flexible ligands into flexible receptors. *Proteins* **2007**, *68*, 726−737.
(35) Ferrari, A. M.; Wei, B. Q.; Costantino, L.; Shoichet, B. K. Soft docking and multiple receptor conformations in virtual screening. *J. Med. Chem.* **2004**, *47*, 5076−5084.
(36) Hornak, V.; Okur, A.; Rizzo, R. C.; Simmerling, C. HIV-1 protease flaps spontaneously open and reclose in molecular dynamics simulations. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 915−920.
(37) Sotriffer, C. A.; Krämer, O.; Klebe, G. Probing flexibility and "induced-fit" phenomena in aldose reductase by comparative crystal structure analysis and molecular dynamics simulations. *Proteins* **2004**, *56*, 52−66.
(38) Amadei, A.; Linssen, A. B. M.; Berendsen, H. J. C. Essential dynamics of proteins. *Proteins* **1993**, *17*, 412−425.
(39) Balsera, M. A.; Wriggers, W.; Oono, Y.; Schulten, K. Principal Component Analysis and Long Time Protein Dynamics. *J. Phys. Chem.* **1996**, *100*, 2567−2572.
(40) Zacharias, M. Rapid protein-ligand docking using soft modes from molecular dynamics simulations to account for protein deformability: binding of FK506 to FKBP. *Proteins* **2004**, *54*, 759−767.
(41) Smith, G.; Sternberg, M.; Bates, P. The Relationship between the Flexibility of Proteins and their Conformational States on Forming Protein-Protein Complexes with an Application to Protein-Protein Docking. *J. Mol. Biol.* **2005**, *347*, 1077−1101.
(42) Case, D. A. Normal mode analysis of protein dynamics. *Curr. Opin. Struct. Biol.* **1994**, *4*, 285−290.
(43) Tirion, M. M. Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis. *Phys. Rev. Lett.* **1996**, *77*, 1905−1908.

(44) Hinsen, K. Analysis of domain motions by approximate normal mode calculations. *Proteins* **1998**, *33*, 417−429.
(45) Bahar, I.; Atilgan, A. R.; Erman, B. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Fold. Des.* **1997**, *2*, 173−181.
(46) Atilgan, A. R.; Durell, S. R.; Jernigan, R. L.; Demirel, M. C.; Keskin, O.; Bahar, I. Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys. J.* **2001**, *80*, 505−515.
(47) Tama, F.; Sanejouand, Y. H. Conformational change of proteins arising from normal mode calculations. *Protein Eng.* **2001**, *14*, 1−6.
(48) Chennubhotla, C.; Rader, A. J.; Yang, L.-W.; Bahar, I. Elastic network models for understanding biomolecular machinery: from enzymes to supramolecular assemblies. *Phys. Biol.* **2005**, *2*, 173−181.
(49) Ahmed, A.; Villinger, S.; Gohlke, H. Large-scale comparison of protein essential dynamics from molecular dynamics simulations and coarse-grained normal mode analyses. *Proteins* **2010**, *78*, 3341−52.
(50) Yang, L.; Song, G.; Jernigan, R. L. How well can we understand large-scale protein motions using normal modes of elastic network models? *Biophys. J.* **2007**, *93*, 920−929.
(51) Ahmed, A.; Gohlke, H. Multiscale modeling of macromolecular conformational changes combining concepts from rigidity and elastic network theory. *Proteins* **2006**, *63*, 1038−1051.
(52) Kundu, S.; Jernigan, R. L. Molecular Mechanism of Domain Swapping in Proteins: An Analysis of Slower Motions. *Biophy. J.* **2004**, *86*, 3846−3854.
(53) Kundu, S.; Sorensen, D. C.; Phillips, G. N. Automatic domain decomposition of proteins by a Gaussian Network Model. *Proteins: Struct. Funct. Bioinf.* **2004**, *57*, 725−733.
(54) Tatsumi, R.; Fukunishi, Y.; Nakamura, H. A hybrid method of molecular dynamics and harmonic dynamics for docking of flexible ligand to flexible receptor. *J. Comput. Chem.* **2004**, *25*, 1995−2005.
(55) Zhang, Z.; Shi, Y.; Liu, H. Molecular dynamics simulations of peptides and proteins with amplified collective motions. *Biophys. J.* **2003**, *84*, 3583−3593.
(56) Tama, F.; Miyashita, O.; Brooks, C. L. Flexible multi-scale fitting of atomic structures into low-resolution electron density maps with elastic network normal mode analysis. *J. Mol. Biol.* **2004**, *337*, 985−999.
(57) Tama, F.; Miyashita, O.; Brooks, C. L. Normal mode based flexible fitting of high-resolution structure into low-resolution experimental data from cryo-EM. *J. Struct. Biol.* **2004**, *147*, 315−326.
(58) Delarue, M. Dealing with structural variability in molecular replacement and crystallographic refinement through normal-mode analysis. *Acta Crystallogr. D Biol. Crystallogr.* **2008**, *64*, 40−48.
(59) Hinsen, K.; Reuter, N.; Navaza, J.; Stokes, D. L.; Lacapère, J. J. Normal mode-based fitting of atomic structure into electron density maps: application to sarcoplasmic reticulum Ca-ATPase. *Biophys. J.* **2005**, *88*, 818−827.
(60) Durand, P.; Trinquier, G.; Sanejouand, Y.-H. A new approach for determining low-frequency normal modes in macromolecules. *Biopolymers* **1994**, *34*, 759−771.
(61) Tama, F.; Gadea, F. X.; Marques, O.; Sanejouand, Y. H. Building-block approach for determining low-frequency normal modes of macromolecules. *Proteins* **2000**, *41*, 1−7.
(62) Li, G.; Cui, Q. A coarse-grained normal mode approach for macromolecules: an efficient implementation and application to Ca(2+)-ATPase. *Biophys. J.* **2002**, *83*, 2457−2474.
(63) Schuyler, A. D.; Chirikjian, G. S. Normal mode analysis of proteins: a comparison of rigid cluster modes with C[alpha] coarse graining. *J. Mol. Graph. Model.* **2004**, *22*, 183−193.
(64) Doruker, P.; Jernigan, R. L.; Bahar, I. Dynamics of large proteins through hierarchical levels of coarse-grained structures. *J. Comput. Chem.* **2002**, *23*, 119−127.
(65) Zheng, W.; Brooks, B. R. Probing the local dynamics of nucleotide-binding pocket coupled to the global dynamics: myosin versus kinesin. *Biophys. J.* **2005**, *89*, 167−178.
(66) Ming, D.; Wall, M. E. Allostery in a coarse-grained model of protein dynamics. *Phys. Rev. Lett.* **2005**, *95*, 198103.
(67) Eom, K.; Baek, S. C.; Ahn, J. H.; Na, S. Coarse-graining of protein structures for the normal mode studies. *J. Comput. Chem.* **2007**, *28*, 1400−1410.
(68) Micheletti, C.; Carloni, P.; Maritan, A. Accurate and efficient description of protein vibrational dynamics: Comparing molecular dynamics and Gaussian models. *Proteins* **2004**, *55*, 635−645.
(69) Kurkcuoglu, O.; Jernigan, R. L.; Doruker, P. Mixed levels of coarse-graining of large proteins using elastic network model succeeds in extracting the slowest motions. *Polymer* **2004**, *45*, 649−657.
(70) Dobbins, S. E.; Lesk, V. I.; Sternberg, M. J. E. Insights into protein flexibility: The relationship between normal modes and conformational change upon protein−protein docking. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 10390−10395.
(71) May, A.; Zacharias, M. Accounting for global protein deformability during protein-protein and protein-ligand docking. *Biochim. Biophys. Acta* **2005**, *1754*, 225−231.
(72) Bastard, K.; Prévost, C.; Zacharias, M. Accounting for loop flexibility during protein-protein docking. *Proteins* **2006**, *62*, 956−969.
(73) Zacharias, M. Accounting for conformational changes during protein-protein docking. *Curr. Opin. Struct. Biol.* **2010**, *20*, 180−186.
(74) Ahmed, A.; Rippmann, F.; Barnickel, G.; Gohlke, H. A Normal Mode-Based Geometric Simulation Approach for Exploring Biologically Relevant Conformational Transitions in Proteins. *J. Chem. Inf. Model.* **2011**, *51*, 1604−1622.
(75) Rueda, M.; Bottegoni, G.; Abagyan, R. Consistent improvement of cross-docking results using binding site ensembles generated with elastic network normal modes. *J. Chem. Inf. Model.* **2009**, *49*, 716−725.
(76) Petrone, P.; Pande, V. S. Can conformational change be described by only a few normal modes? *Biophys. J.* **2006**, *90*, 1583−1593.
(77) May, A.; Zacharias, M. Protein-ligand docking accounting for receptor side chain and global flexibility in normal modes: evaluation on kinase inhibitor cross docking. *J. Med. Chem.* **2008**, *51*, 3499−3506.
(78) 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.
(79) Hartshorn, M. J.; Verdonk, M. L.; Chessari, G.; Brewerton, S. C.; Mooij, W. T. M.; Mortenson, P. N.; Murray, C. W. Diverse, High-Quality Test Set for the Validation of Protein−Ligand Docking Performance. *J. Med. Chem.* **2007**, *50*, 726−741.
(80) Verdonk, M. L.; Mortenson, P. N.; Hall, R. J.; Hartshorn, M. J.; Murray, C. W. Protein-ligand docking against non-native protein conformers. *J. Chem. Inf. Model.* **2008**, *48*, 2214−2225.
(81) Levitt, M.; Sander, C.; Stern, P. S. The normal modes of a protein: Native bovine pancreatic trypsin inhibitor. *Int. J. Quantum Chem.* **1983**, *24*, 181−199.
(82) Brooks, B.; Karplus, M. Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor. *Proc. Natl. Acad. Sci. USA* **1983**, *80*, 6571−6575.
(83) Hildebrandt, A.; Dehof, A.; Rurainski, A.; Bertsch, A.; Schumann, M.; Toussaint, N.; Moll, A.; Stockel, D.; Nickels, S.; Mueller, S.; Lenhof, H.-P.; Kohlbacher, O. BALL - biochemical algorithms library 1.3. *BMC Bioinf.* **2010**, *11*, 531.
(84) Open Babel: The Open Source Chemistry Toolbox v. 2.2.2. http://www.openbabel.org/
(85) Kollman, P.; Dixon, R.; Cornell, W.; Fox, T.; Chipot, C.; Pohorille, A. The development/application of a "minimalist" organic/biochemical molecular mechanic force field using a combination of ab initio calculations and experimental data. In *Computer simulations of biomolecular systems*; van Gunsteren, W. F., Weiner, P. K., Wilkinson, A. J., Eds.; Kluwer/Escom Academic Publishers: Dordrecht, The Netherlands, 1997; Vol. 3, pp 83−96.
(86) Nocedal, J. Updating Quasi-Newton Matrices with Limited Storage. *Math. Comput.* **1980**, *35*, 773−782.
(87) Liu, D. C.; Nocedal, J. On the limited memory BFGS method for large scale optimization. *Math. Program.* **1989**, *45*, 503−528.
(88) Krivov, G. G.; Shapovalov, M. V.; Dunbrack, R. L. Improved prediction of protein side-chain conformations with SCWRL4. *Proteins* **2009**, *77*, 778−795.

(89) Hartmann, C.; Antes, I.; Lengauer, T. IRECS: a new algorithm for the selection of most probable ensembles of side-chain conformations in protein models. *Protein Sci.* **2007**, *16*, 1294−1307.
(90) McDonald, I. K.; Thornton, J. M. Satisfying Hydrogen Bonding Potential in Proteins. *JMB* **1994**, *238*, 777−793.
(91) Wallace, A. C.; Laskowski, R. A.; Thornton, J. M. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng.* **1995**, *8*, 127−134.
(92) Vandonselaar, M.; Hickie, R. A.; Quail, W.; Delbaere, L. T. J. Trifluoperazine-induced conformational change in Ca2+-calmodulin. *Nat. Struct. Mol. Biol.* **1994**, *1*, 795−801.