Conformational Transitions upon Ligand Binding: Holo-Structure Prediction from Apo Conformations

Daniel Seeliger, Bert L. de Groot*

Computational Biomolecular Dynamics Group, Max-Planck-Institute for Biophysical Chemistry, Göttingen, Germany

Abstract

Biological function of proteins is frequently associated with the formation of complexes with small-molecule ligands. Experimental structure determination of such complexes at atomic resolution, however, can be time-consuming and costly. Computational methods for structure prediction of protein/ligand complexes, particularly docking, are as yet restricted by their limited consideration of receptor flexibility, rendering them not applicable for predicting protein/ligand complexes if large conformational changes of the receptor upon ligand binding are involved. Accurate receptor models in the ligand-bound state (holo structures), however, are a prerequisite for successful structure-based drug design. Hence, if only an unbound (apo) structure is available distinct from the ligand-bound conformation, structure-based drug design is severely limited. We present a method to predict the structure of protein/ligand complexes based solely on the apo structure, the ligand and the radius of gyration of the holo structure. The method is applied to ten cases in which proteins undergo structural rearrangements of up to 7.1 Å backbone RMSD upon ligand binding. In all cases, receptor models within 1.6 Å backbone RMSD to the target were predicted and close-to-native ligand binding poses were obtained for 8 of 10 cases in the top-ranked complex models. A protocol is presented that is expected to enable structure modeling of protein/ligand complexes and structure-based drug design for cases where crystal structures of ligand-bound conformations are not available.

Introduction

Interactions between proteins and small molecules are involved in many biochemical phenomena. Insight into these processes relies on detailed knowledge about the structure of protein/ligand complexes, e.g. how enzymes stabilize substrates and cofactors in close proximity. Moreover, almost all drugs are small-molecule ligands that interact with enzymes, receptors or channels. Accordingly, ligand-bound receptor complex structures are a critical prerequisite for understanding biological function and for structure based drug design. However, structure determination of protein/ligand-complexes can be difficult, time-consuming and expensive. Crystal structures of protein/ligand complexes are usually obtained either by co-crystallization or soaking and it is a common problem that even when conditions for crystallizing the apo-protein are well established these might not be transferable to the protein/ligand complex [1–4]. Particularly, conformational transitions of the receptor associated with ligand binding pose a severe challenge to the structure elucidation of holo complexes [5–8].

When structures of ligand-bound protein conformations are not available, structure-based drug design becomes highly challenging. Several studies showed that virtual screening to an apo-structure usually results in a poor enrichment factor (the ability to discriminate between binders and non-binders) compared to the holo-structure even when the structural difference between both is comparably small [9–11]. Therefore, the development of docking programs aims at allowing a certain degree of receptor flexibility

either by using an ensemble of structures instead of a single receptor conformation [12–15] or by explicitly modeling flexibility such as sidechain variations (Autodock4 [16,17], Gold [18,19], FlexX [20], RosettaLigand [21]), predefined flexibility of certain parts of the structure (FlipDock [22]) and also small variations of the backbone (Glide/Prime [23], RosettaLigand [24], ICM [25,26]). Incorporating receptor flexibility in molecular docking is a substantial progress and has been shown to enhance both enrichment factors and the ability to predict correct binding poses, particularly in cases when docking a compound to a receptor structure that has been crystallized with a different ligand (cross-docking) which is usually the case when searching for novel drugs. However, the degree of flexibility thus far is limited to either sidechain motions or small variations of the backbone and thus, the availability of a holo-structure or an apo-structure that is highly similar to the holo conformation is currently a prerequisite for a successful docking, severely limiting structure-based drug design.

Particularly, receptors that undergo a substantial conformational transition upon ligand binding are currently precluded from structure based drug design.

Although protein-ligand crystals suitable for diffraction might not be accessible, several experimental techniques exist to detect conformational changes. In many cases where proteins undergo domain reorientations upon ligand binding they adopt a different shape in the ligand bound state, corresponding to a change in the radius of gyration that can be studied either by NMR, where a

Author Summary

Structure-based drug design has become a powerful tool in modern drug discovery pipelines. A critical prerequisite is a structure of the target protein close to its ligand bound conformation which is often difficult to determine experimentally. In many cases, a structure of the unbound receptor is available, but conformational changes with respect to the ligand-bound form preclude it from being used as a basis for structure-based drug design. We have developed a computational approach to predict the structure of protein/ligand complexes based solely on the unbound conformation, the ligand, and easy-to-assess experimental data. We tested our protocol on proteins that undergo substantial structural rearrangements upon binding a ligand and were able to predict structures of protein/ligand complexes which are in good agreement with experimentally determined structures. The ability to predict ligand bound receptor conformations based on structures in the unbound state enables structure-based drug design for cases where crystallization of the complex has not been successful so far.

more compact shape causes a descrease in the rotational correlation time [27–29] or by small-angle scattering of x-rays (SAXS) or neutrons (SANS) [30–33]. These shape descriptions provide invaluable information for modeling of structures [34–36] and macromolecular assemblies [37,38] as well as insight into protein dynamics.

Here, we present a method to predict the structure of protein/ligand complexes for proteins that undergo a large conformational change upon ligand binding. The protocol solely requires the apo-structure, a known ligand and experimental data on the shape of the holo-structure. Here, we apply the radius of gyration as shape information, a quantity that can frequently be readily assessed more easily than an x-ray structure. We developed a simulation protocol that combines biased conformational sampling, docking and molecular dynamics simulations and applied it to ten ligand-binding proteins (see Table 1). We chose cases where both, the unbound conformation and the bound conformation are known from x-ray crystallography in order to be able to a-posteriori validate the predicted receptor conformations and docking poses.

The conformational changes involved range from 2.1 to 7.1 Å backbone RMSD (see table 1 and fig. 1) and the binding site geometries differ substantially between the apo and the holo conformations. In nine of the ten cases we predict holo receptor conformations close to the native ligand-bound conformation and in eight cases we predict ligand binding poses close to the native state, rendering our method suitable for blind predictions of protein/ligand complexes involving large conformational transitions.

Results

The proteins used in this study undergo large conformational changes upon ligand binding. The conformational changes involve large domain reorientations, accompanied by loop motions and sidechain movements. Since we use the radius of gyration as an example of an experimentally determined constraint we chose cases where a distinct change of this observable takes place upon ligand binding. This is usually the case for protein motions which are classified as hinge-bending motions according to the database of macromolecular movements [6]. However, such a classification is not necessary for the described protocol and is not used for the predictions made in this work. Here, we focus at cases that undergo domain closure upon ligand binding. With that aim, constraints which enforce domain closure were incorporated into the tCONCOORD program, thereby enhancing the sampling around the closed state significantly (see Text S1).

The ten proteins used for this study belong to five different SCOP [39–41] superfamilies.

- Periplasmic-binding protein type I: ALLO, LEU, RIBO
- Periplasmic-binding protein type II: OSMO, ALGI, GLUR2
- EPT/RTPC-like: EPSP
- P-loop containing nucleoside triphosphate hydrolase: GUA
- Actin-like ATPase domain: HEXO, GLUCO

Periplasmic-binding proteins (PBPs) mediate a wide range of fundamental processes and are ubiquitous in bacteria. Therefore, they are discussed as potential targets for antmicrobial agents. Moreover, the general "Venus-flytrap"- architecture of PBPs is also found in the extracellular ligand-binding domain of class C G-protein coupled receptors (GPCRs) many of which are of great

Table 1. Comparison of apo/holo pairs of proteins.

Receptor	Abbreviation	PDB_{apo}	PDB_{holo}	# Res.	Ligand	$RMSD_{BB}$	$RMSD_{BS}$
GluR2 ligand binding core	GLUR2	1fto	1ftm	257	AMPA	2.2	2.0
DNA Beta-Glucosyl-transferase	GLUCO	1jej	1jg6	351	Uridine-5′-diphosphate	2.1	2.6
D-Allose binding protein	ALLO	1gud	1rpj	288	D-Allose	4.4	4.0
D-Ribose binding protein	RIB	1urp	2dri	271	β-D-Ribose	4.3	3.5
L-Leucine binding protein	LEUB	1usg	1usi	345	Phenylalanine	7.1	6.8
5-Enolpyruvylshikimate-3-phosphate synthase	EPSP	1rf5	1rf4	427	SPQ	3.7	4.6
Osmo-protection protein	OSMO	1sw5	1sw2	270	Glycine-betaine	5.0	4.4
Guanylate kinase	GUA	1ex6	1ex7	186	GMP	3.6	3.9
Hexokinase	HEXO	2e2n	2e2o	298	Glucose	3.0	1.9
Alginate binding protein	ALGI	1y3q	1y3n	490	Alginate Disaccharide	4.8	3.6

$RMSD_{BB}$: Root mean square deviation of all backbone atoms after least square fit. $RMSD_{BS}$: Root mean square deviation of the binding site. The binding site was defined as all residues which have at least one atom within 6 Å of any ligand atom.
doi:10.1371/journal.pcbi.1000634.t001
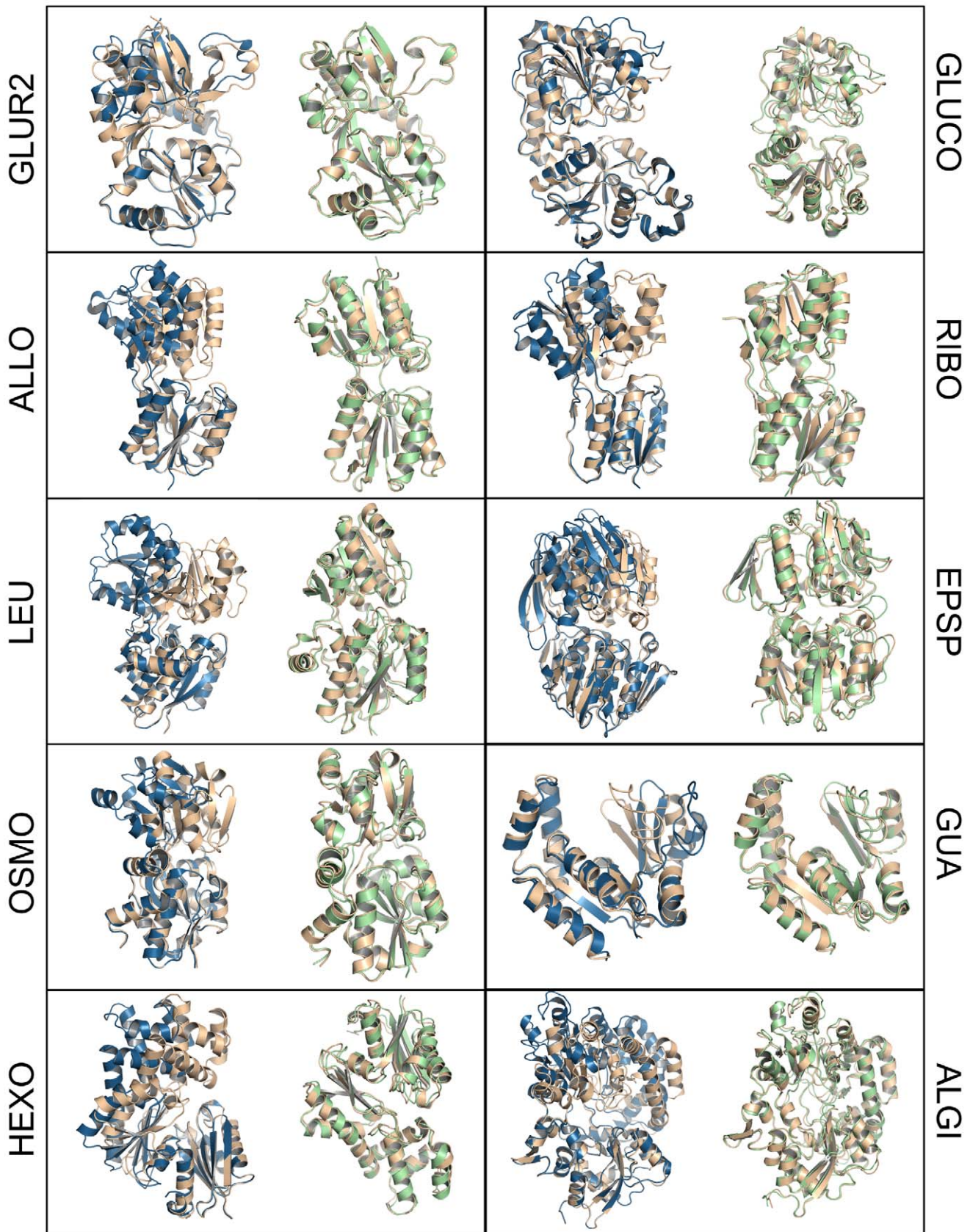
Figure 1. Overlay of apo/holo pairs and holo/model pairs. Receptor models were predicted based on the apo-structure, the ligand and the radius of gyration of the holo conformation. Apo x-ray structures are shown in blue, holo x-ray structures are colored wheat. The best modelled receptor structure is shown in green.
doi:10.1371/journal.pcbi.1000634.g001

interest as drug targets [42]. The other cases belong to different classes of kinases which represents another large group of pharmacologically interesting receptors.

Table 1 lists the protein data bank [43] (PDB) codes of the apo/holo pairs, their respective ligands and the global backbone root mean square deviations (RMSD) as well as the RMSDs of the binding site.

In all cases, the ligand binding pocket differs substantially between the bound and the unbound state or is virtually not present in the apo conformation. A structure prediction protocol therefore needs to precisely predict the domain motions as well as local structural rearrangements. We use the tCONCOORD program [44] to generate an initial ensemble of conformations which share a predefined radius of gyration. Since the constraints tCONCOORD uses for the structure prediction are derived from a given protein structure (here the apo conformation) the resulting ensemble represents an approximation of the conformational space that is accessible within the predefined constraints. Hence, if interactions like hydrogen bonds are defined as constraints in the input structure they will be conserved in the resulting ensemble. If conformational changes are associated with partial unfolding or refolding, resulting in major changes of, e.g. the hydrogen bond network, the generated ensemble will be likely to miss those conformations. Domain motions however, are ususaly well predicted [44]. The first part of the structure prediction protocol therefore aims at predicting the correct orientation of domains. The molecular dynamics simulations carried out subsequently basically serve as a filter to distinguish between correct models and those with correct geometry but unfavourable energetics which tend to re-open quickly.

Prediction of holo receptor structures. The results are summarized in Table 2. After the refinement cycles we find a good correlation between the radius of gyration and the RMSD to the holo structures (see fig 2). In all of the 10 cases the global backbone RMSD of the best model to the known holo structure is below 1.6 Å (fig. 1). For the prediction of ligand binding poses, however, it is more important to focus on the geometry of the binding site. Here, we find that the prediction of the arrangement of the backbone atoms belonging to the binding site is equal or below 1 Å RMSD accuracy in six of the ten cases and below 1.4 Å

Table 2. Model accuracy.

Receptor	Overall BB-RMSD	BS BB-RMSD	BS all-atom RMSD	Rank <1.5 Å BB
GLUR2	0.91	0.46	0.84	1
GLUCO	1.38	1.35	1.78	10
ALLO	1.07	0.96	1.55	2
RIBO	0.98	0.76	1.2	3
LEU	1.15	1.2	1.9	5
EPSP	1.58	2.8	3.58	-
OSMO	1.27	1.3	1.96	2
GUA	1.45	1.0	1.76	2
HEXO	1.42	0.87	1.38	3
ALGI	1.27	0.77	1.5	3

The table shows RMSD values for the best models. In all cases the global backbone RMSD is below 1.6 Å. The backbone RMSD of the binding site is below 1 Å in 5 cases and the the RMSD calculated over all heavy atoms in the binding site is below 2 Å in 9 of 10 cases. Rank <1.5 Å BB gives the rank of the first model to get within 1.5 Å RMSD to the holo x-ray structure.
doi:10.1371/journal.pcbi.1000634.t002

in nine cases. An RMSD calculated over all atoms (including the sidechains) corresponding to the binding site reveals RMSD values below 2 Å for nine of the ten cases. EPSP, where the backbone deviation is 2.8 Å represents a case were the current protocol reaches its limitations since here, the structural rearrangment between apo and holo structure involves refolding of a helix into a loop, a transition which is not predicted by the current sampling protocol. This causes some residues in the predicted structures to be more than 10 Å away from their target destination, hence resulting in a large RMSD.

Prediction of ligand binding poses. Obtaining close-to-native ligand binding poses with receptor structures obtained from structure modeling or low-resolution data is a non-trivial task even when close-to-native receptor conformations are available. Scoring functions in docking programs are usually very sensitive to even small conformational variations or structural inaccuracies and a single misplaced sidechain can preclude successful docking. Therefore, low binding site RMSD's do not automatically translate into low ligand RMSD's. Fig. 3 shows a plot of the ligand RMSD versus the RosettaLigand energy. As can be seen, ligand binding poses close to the native state were sampled in all cases but, due to structural variations in the binding site, these were not always among the top ranked configurations. In 8 of the 10 cases, however, ligand binding poses below 3 Å RMSD to the X-ray structure are found within the top ranked solutions (see fig. 4). In docking studies to rigid receptors ligand RMSDs of <2Å to the experimentally determined binding pose are usually regarded as successful. In our case the ligand RMSD was calculated after least square fitting of the modeled complex to the target structure. Hence, deviations of both, the modeled receptor structure and the ligand pose contribute to the calculated ligand RMSDs.
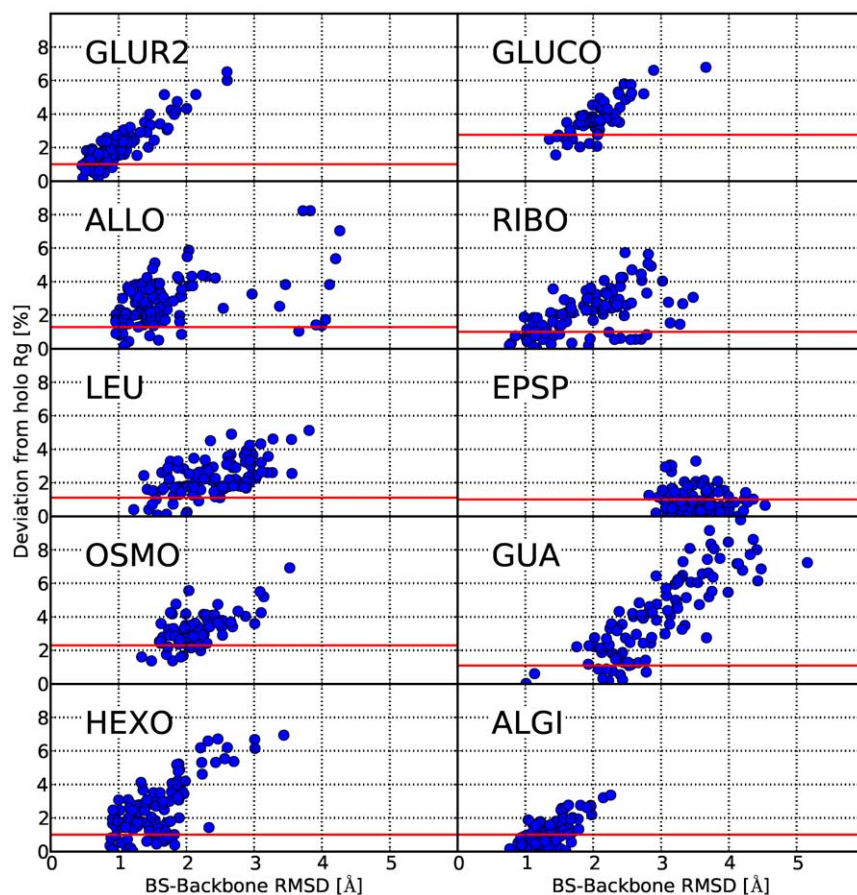
Discussion

In this work, we present a proof of concept study that allows successful prediction of holo-receptor conformations and ligand binding poses. With an approach based solely on the receptor in its unbound conformation, a known binder, such as the natural substrate, and relatively easy to assess shape information on the ligand-bound state, we predicted receptor models within 1.6 Å RMSD to the x-ray holo structure starting from apo conformations that deviate up to 7.1 Å RMSD from their respective ligand-bound state. For 8 of the 10 cases ligand binding poses within <3Å RMSD to the x-ray structure were predicted, underpinning the applicability of our method to predict receptor models suitable for virtual screening.

With the described protocol we bridge a gap arising from the difficulties associated with structure determination of protein/ligand complexes, the limited conformational flexibility that can be incorporated in docking programs and the limited sampling that can be achieved with molecular dynamics simulations. If apo and holo conformations of a protein differ substantially, current computational structure prediction approaches are usually unable to provide accurate receptor models. As docking protocols need to be fast to be suitable for virtual high-troughput screening they always represent a trade-off between speed and accuracy. Incorporation of sidechain conformational flexibility and backbone plasticity is a substantial progress but each additional degree of freedom is accompanied with growing computational demand and, hence, allowing for large conformational changes has thus for not been feasible. Molecular dynamics simulations are in principle capable of sampling holo conformations when starting from an apo protein. But, in addition to the general sampling problem MD simulations suffer from, ligands may cause alterations in the free

Figure 2. Binding-site RMSD vs. deviation from holo Rg. The deviation from the target radius of gyration correlates well with low RMSD's to the experimentally determined holo structures. The best receptor models (indicated by the red line) were used for binding site refinement with Rosetta.
doi:10.1371/journal.pcbi.1000634.g002

energy landscape. Depending on the differences in the free energy landscapes of the apo protein and the protein bound to a ligand, the sampling of close-to-holo receptor conformations without a ligand present may be particularly poor.

The protocol described in this work combines the strengths of docking, conformational sampling and molecular dynamics while overcoming their particular limitations. In contrast to previous studies [45–48] it is neither restricted to a particular type of protein motion nor does it require the definition of rigid parts or hinges in a structure or any other approximation. Protein flexibility is accounted for on each level from domain motions to sidechain fluctuations and receptors and ligands are treated as all atom or heavy atom models in each state of the protocol. As the sampling of large conformational transitions and crossing of potential energy barriers is handled by the efficient concoord [49] algorithm and MD simulations are used in refinement we gain both efficient sampling and high accuracy within the limitations of the force field.

The protocol furthermore allows incorporation of various experimental data. In this work we used the radius of gyration of the holo conformation but in principle any kind of data can be used that is transferable into geometrical constraints, e.g. data from FRET experiments or mutagenesis studies. Each additional piece of information confines the sampled conformational space and, hence enhances sampling of conformations around a conformational state of interest (see also Text S1). Additional

improvement may be achieved by applying filters to the docked configurations. For example, in the initial docking screens all poses were taken into account without further screening. Analysis of the generated poses and discarding e.g. those that are attached to the surface rather than docked into the binding site would further contribute to the enrichment of promising models. As in all structure prediction protocols, sampling is the limiting factor. In most cases, docking of the ligand to the holo x-ray structure yields a lower score than in the modelled receptor structures. Hence, the docking score is usually sufficient to discriminate between the native structure and a receptor model and if the *true* conformation is sampled it is most likely detectable via the docking score. However, despite the relative simplicity to distinguish between the true structure and a model it is very difficult to distinguish between a model with 1 Å RMSD and a model with 5 Å RMSD to the native state. When applying the current protocol to a blind prediction it would therefore be beneficial to incorporate as much experimental data as possible to reduce the conformational space that needs to be explored and to enhance sampling around the *true* ligand-bound conformation. The problem of conformational sampling furthermore increases with the conformational flexibility of the ligand. Although the ligands discussed in this work are less complex than typical drugs they depict the necessity of very accurate binding sites for a successful docking. Sugar ligands and their derivatives, which some of our ligands belong to, contain many hydroxyl groups and are almost symmetric. Thus, rotating
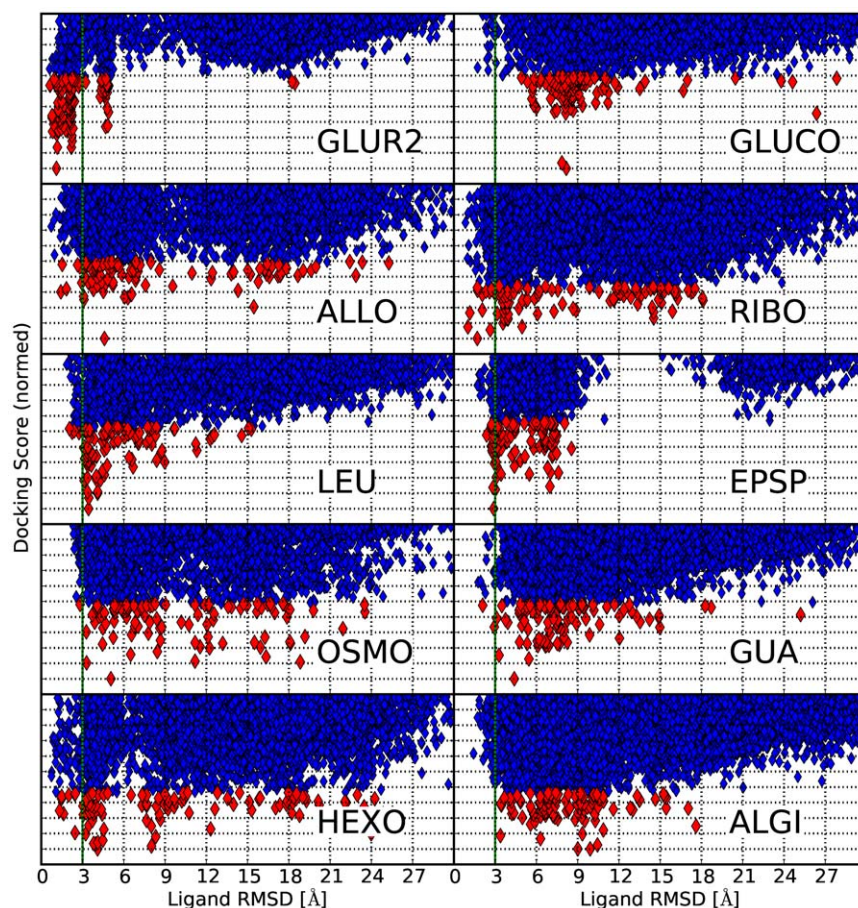
Figure 3. Ligand RMSD's vs. Rosetta Score. The plot shows the RMSD's of ligand poses generated with RosettaLigand. In all cases ligand poses below 3 Å to the target structure have been sampled and in 8 cases a pose below 3 Å RMSD is found within the top 100 (red) of the ranked poses.
doi:10.1371/journal.pcbi.1000634.g003

the molecule by 180 degrees offers almost the same scope of interaction to the receptor but the resulting RMSD would not characterize such a pose as a successful docking. Water molecules can also play an important role in ligand binding. In the present study we removed all crystallographic water molecules from the input structures since it is quite challenging to judge which water molecules might also be present in the ligand bound state, which are replaced by parts of the ligand or which might even be essential for ligand binding.

Concluding remarks. Overall, we find that our protocol is successful in predicting close-to-native receptor models and, with limitations, close-to-native ligand poses. Starting from apo structures displaying global backbone RMSDs up to 7.1 Å we were able to predict holo structures within 1.6 Å for all of the ten cases and ligand poses within 3 Å RMSD for 8 of the 10 proteins. It offers a robust protocol for accurate structure prediction of protein/ligand complexes and allows structure based drug design for proteins that undergo large conformational changes upon ligand binding, for which only an apo-structure and a shape description of the holo-structure is available.

Methods

The workflow of the presented protocol combines biased conformational sampling, docking and molecular dynamics simulations (see fig. 5).

Biased conformational sampling. In a first step, a conformational ensemble is generated based on the apo-structure and the radius of gyration of the target. In this study we did not use experimentally determined radii of gyration but calculated them from the ligand-bound conformations according to

$$\mathrm{Rg} = \left(\frac{\sum_i ||\mathbf{r_i}||^2 m_i}{\sum_i m_i} \right)^{\frac{1}{2}}$$

where $\mathbf{r_i}$ denotes the position of atom i with respect to the center of mass of the protein and m_i the respective mass. When using experimentally determined values of the radius of gyration one should be aware of the fact that these usually closely match calculated values but are systematically larger due to the solvation shell of the protein [50]. Larger deviations may be a hint towards oligomerization in solution.

The conformational ensembles were generated using tCON-COORD [44,51]. In previous studies [44,51] we showed that tCONCOORD is able to generate structure ensembles of proteins from a single input structure, representing a reasonable approximation to the functionally relevant conformational space. The method rests on a thorough analysis of the input structure and the translation of conserved interactions like topological restrictions
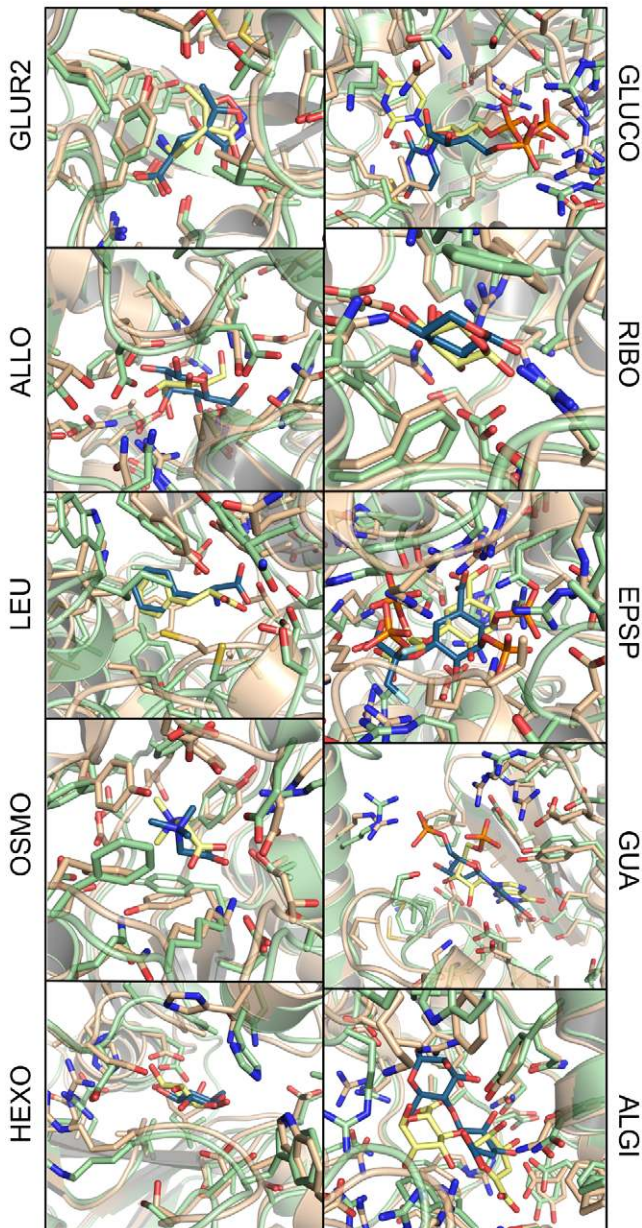
Figure 4. Modeled protein/ligand complexes. X-ray structures of receptor and ligand are shown in wheat and yellow, modeled complexes are colored green/blue.
doi:10.1371/journal.pcbi.1000634.g004

(bonds, angles), hydrogen bonds and hydrophobic clusters into geometrical constraints with upper and lower bounds. Crystallographic water molecules were removed from the input structures since these are often replaced by the ligand in the bound state. In a second step, these constraints are used to rebuild the structure starting from random atomic coordinates yielding a new structure that fulfills all predefined constraints but differs from the input structure, hence representing an alternative conformation. Typically, an ensemble of 100–10000 independent structures is generated. When generating such an ensemble starting from an apo-protein structure conformations close to the holo receptor conformation are usually included in the ensemble.

If structural information of the ligand-bound state is available from experimental data, it can be used as constraint to bias the

ensemble to a conformational subspace of interest [51]. A novel implementation in tCONCOORD allows to additionally use the radius of gyration as geometrical constraint. Thus, with a tCONCOORD simulation using the apo-structure as input, the radius of gyration of the ligand-bound state can be used to enhance the probability of sampling of a conformational subspace of interest, here the ligand-bound state. A radius of gyration bias, however, does not represent such a strong bias as, e.g a distance constraint which enforces a domain closure. Therefore, the ensemble properties are relatively robust against variations of the target radius of gyration and the tolerance used for sampling (see Text S1 for details about biased tCONCOORD samplings).

Initial ensembles of 1000 conformations were generated with tCONCOORD using heavy-atom representations of the receptor structure. Constraints were defined for bonds, angles, 1–4 pairs, hydrogen bonds that meet the desolvation stability criteria [44] and hydrophobic clusters. The radius of gyration of the target structure was used as additional constraint with a tolerance of 0.1 Å to enhance sampling around the ligand-bound conformation. The resulting ensemble was subjected to an RMSD-clustering to reduce the dataset to ≈ 100 conformations and each conformation was subsequently refined using a 250 steps steepest descent energy minimization in vacuum and 500 steps conjugate gradient minimization in explicit solvent. Afterwards the structures were subjected to a Rosetta refinement and scored with the ddfire [52] scoring function. The best scoring 50 structures were selected for resampling. The tCONCOORD resampling protocol employs random perturbations on sidechain rotamers which are predicted to be flexible based on interaction and packing considerations while keeping the rest of the structure fixed. From each model, 100 conformations where generated, yielding a total number of 5000 protein conformations.

Docking and refinement. Afterwards, the known ligand was docked into each of the 5000 models using Autodock VINA [53]. Receptor and ligand preparations were carried out using the AutoDockTools [16]. The binding site was defined as a cubic box of 26.25 Å length placed at the center of the reference ligand position. (Hence, for blind predictions a rough knowledge about the location of the binding site would be required). In all cases the docking box was large enough to guarantee independence of the docking results from small variations of the binding site definitions. From the docking solutions, the best 50 scoring protein/ligand complexes were selected and subjected to 50 ps molecular dynamics simulation with position restraints, followed by 2 ns free simulation and 200 ps simulated annealing to 100 K. The MD simulations were carried out with the Gromacs-4.0 [54] molecular dynamics package in explicit solvent using the Amber03 force field [55,56] and the tip3p water model [57]. The systems were solvated and NaCl was added to achieve a 150 mM concentration. Ligand parameters were obtained with the generalized Amber force field (GAFF) approach [58]. Electrostatic interactions were calculated at every step with the particle-mesh Ewald method [59], short-range repulsive and attractive dispersion interactions were simultaneously described by a Lennard-Jones potential, which was cut off at 1.0 nm. The SETTLE [60] algorithm was used to constrain bonds and angles of water molecules, and LINCS [61] was used for all other bonds, allowing a time step of 2 fs. The temperature was kept constant at 300 K by weakly coupling the system to an external heat bath [62,63] (time constant $\tau = 0.1\,\mathrm{ps}$) and the pressure was kept constant at 1 atm by weak isotropic coupling to a pressure bath ($\tau = 1\,\mathrm{ps}$).

The resulting structures were ranked by their compliance with the target radius of gyration and the best 10 models were subjected to a second refinement cycle consisting of resampling, docking
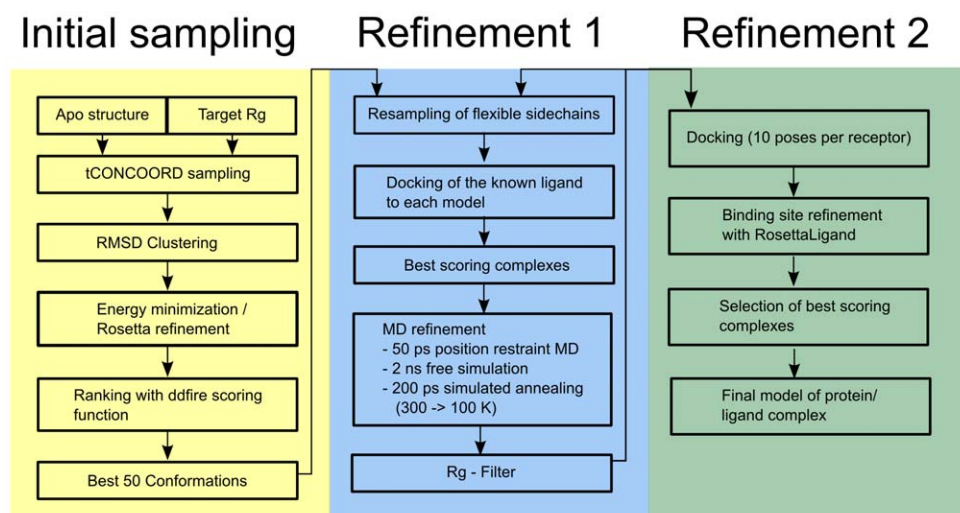
Figure 5. Flowchart of the modeling process. The procedure starts with a biased tCONCOORD sampling that takes the apo conformation and the radius of gyration of the target structure as input. An ensemble of 1000 conformations is generated and, after energy minimization, RMSD clustering and Rosetta refinement, ranked according to the ddfire scoring function. The best 50 conformations were subjected to a tCONCOORD-based sidechain resampling protocol that generates 100 conformations from each input structure with fixed backbone. Afterwards the ligand is docked into each of the 5000 conformations and the 50 best scoring solutions were subjected to molecular dynamics refinement protocol, consisting of 50 ps position restrained MD, followed by 2 ns free simulation and 200 ps simulated annealing. From those models that remained in a closed conformation (close to the holo radius of gyration) another ensemble was generated and subjected to a second refinement cycle. The best models were taken and initial docking poses generated with VINA and refined using the RosettaLigand protocol. The final ranking was carried according to the RosettaLigand score.
doi:10.1371/journal.pcbi.1000634.g005

and, after filtering the best poses, to molecular dynamics refinement.

Model assessment and binding site refinement.

From the resulting receptor structures the radius of gyration was calculated and those models that had less than 1% deviation from the holo radius of gyrations were chosen for a binding site refinement (In two cases we chose a larger Rg-cutoff since no models were obtained within 1%). For the refinment we carried out docking to the rigid receptor structures and used RosettaLigand to refine the docking pose while simultaneously repacking the sidechains around the ligand. 100 conformations were generated based on each ligand pose and the resulting complexes were ranked according to the RosettaLigand score.

The current protocol requires about four days per case on a 50 node cluster with the most time spent in the molecular dynamics part and the QM based ligand parametrization with gaussian [58,64].

References

1. Danley DE (2006) Crystallization to obtain protein-ligand complexes for structure-aided drug design. Act Cryst D 62: 569–575.
2. McNea IW, Kan D, Kontopidis G, Patterson A, Tayler P, et al. (2005) Studying protein-ligand interactions using protein crystallography. Crystallography Reviews 11: 61–71.
3. Zhu X, Kim JL, Newcomb JR, Rose PE, Stover DR, et al. (1999) Structural analysis of the lymphocyte kinase Lck in complex with non-selective and Src family selective kinase inhibitors. Structure 7: 651–661.
4. Hiller N, Fritz-Wolf K, Deponte M, Wende W, Zimmermann H, et al. (2006) Plasmodium falciparum glutathion S-transferase- structural and mechanistic studies on ligand binding and enzyme inhibition. Protein Sci 15: 281–289.
5. Brylinski M, Skolnick J (2008) What is the relationship between the global structures of apo and holo proteins? PROTEINS: Struct Funct Bioinf 70: 363–377.
6. Gerstein M, Krebs W (1998) A database of macromolecular motions. Nucleic Acids Res 15: 4280–4290.
7. Teague SJ (2003) Implications of protein flexibility for drug discovery. Nature Rev Drug Discovery 2: 527–541.
8. Ma B, Shatsky M, Wolfson HJ, Nussinov R (2002) Multiple diverse ligands binding at a single protein site: a matter of pre-existing populations. Protein Sci 11: 184–197.
9. McGovern SL, Shoichet BK (2003) Information decay in molecular docking screens against holo, apo and modeled conformations of enzymes. J Med Chem 46: 2895–2907.
10. Warren GL, Andrews CW, Capelli AM, Clarke B, LaLonde J, et al. (2006) A critical assessment of docking programs and scoring functions. J Med Chem 49: 5912–5931.
11. Murray C, Baxter C, Frenkel A (1999) The sensitivity of the results of molecular docking to induced fit effects: application to thrombin, thermolysin and neuraminidase. J Comput Aided Mol Des 13: 547–562.
12. Claušen H, Buning C, Rarey M, Lengauer T (2001) FlexE: efficient molecular docking considering protein structure variations. J Mol Biol 308: 377–395.
13. Totrov M, Abagyan R (2008) Flexible ligand docking to multiple receptor conformations: a practical alternative. Curr Opin Struct Biol 18: 178–184.
14. Bolstad ESD, Anderson AC (2009) In pursuit of virtual lead optimization: pruning ensembles of receptor strutcures for increased efficiency and accuracy during docking. PROTEINS: Struct, Funct and Bioinf 75: 62–74.
15. Knegtel RMA, Kuntz ID, Oshiro CM (1997) Molecular docking to ensembles of protein structures. J Mol Biol 266: 424–440.

16. Morris GM, Goodsell DS, Halliday DS, Huey R, Hart WE, et al. (1998) Automated docking using a lamarckian genetic algorithm and and empirical binding free energy function. J Comp Chem 19: 1639–1662.

17. Huey R, Morris GM, Olson AJ, Goodsell DS (2007) A semiempirical free energy force field with charge-based desolvation. J Comp Chem 28: 1145–1152.

18. Jones G, Willett P, Glen R (1995) Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. J Mol Biol 245: 43–53.

19. Verdonk M, Cole J, Hartshorn M, Murray C, Taylor R (2003) Improved protein-ligand docking using GOLD. Proteins: Struct Funct and Genet 52: 609–623.

20. Hartmann C, Antes I, Lengauer T (2009) Docking and scoring with alternative side-chain conformations. PROTEINS: Struct, Funct and Bioinf 74: 712–726.

21. Meiler J, Baker D (2006) ROSETTALIGAND: protein-small molecule docking with full side-chain flexibility. PROTEINS: Struct, Funct and Bioinf 65: 538–548.

22. Zhao Y, Sanner MF (2007) FLIPDock: Docking flexible ligands into flexible receptors. PROTEINS: Struct, Funct and Bioinf 68: 726–737.

23. Sherman W, Day T, Jacobson M, Friesner R, Farid R (2006) Novel procedure for modeling ligand/receptor induced fit effects. J Med Chem 49: 534–553.

24. Davis IW, Baker D (2009) ROSETTALIGAND docking with full ligand and receptor flexibility. J Mol Biol 385: 381–392.

25. Cavasotto CN, Abagyan RA (2004) Protein flexibility in ligand docking and virtual screening to protein kinases. J Mol Biol 337: 209–225.

26. Cavasotto CN, Kovacs JA, Abagyan RA (2005) Representing receptor flexibility in ligand docking through relevant normal modes. J Am Chem Soc 127: 9632–9640.

27. Olson AL, Yao H, Herdendorf TJ, Miziorko HM, Hannongbua S, et al. (2009) Substrate induced structural and dynamics changes in human phosphomevalonate kinase and implications for mechanism. PROTEINS: Struct, Funct and Bioinf 75: 127–138.

28. Yang S, Rogers KM, Johnson DA (2005) MgATP-induced conformational change in the catalytic subunit of cAMP-dependent protein kinase. Biophys Chem 113: 193–199.

29. Lee D, Hilty C, Wider G, Wüthrich K (2006) Effective rotational correlation times of proteins from NMR relaxation interference. J Magn Reson 178: 72–76.

30. Shiau AK, Harris SF, Southworth DR, Agard DA (2006) Structural analysis of E.coli hsp90 reveals dramatic nulceotide-deoendent conformational rearragements. Cell 127: 329–340.

31. Forstner M, Kriechbaum M, Laggner P, Wallimann T (1996) Changes of creatine kinase structure upon ligand binding as seen by small-angle scattering. J Mol Struct 383: 217–222.

32. Perkins S (1988) Structural studies of proteins by high-flux X-ray and neutron solution scattering. Biochem J 254: 313–327.

33. Svergun D, Koch M (2002) Advances in structure analysis using small-angle scattering in solution. Curr Opin Struct Biol 12: 654–660.

34. Zhou G, Ellington WR, Chapman MS (2000) Induced fit in arginine kinase. Biophys J 78: 1541–1550.

35. Glatter O, Kratky O (1982) Small-Angle X-ray Scattering. London: Academic Press.

36. Bernado P, Mylonas E, Petoukhov M, Blackledge M, Svergun D (2007) Structural characterization of flexible proteins using small-angle X-ray scattering. J Am Chem Soc 129: 5656–5664.

37. Förster F, Webb B, Krukenberg KA, Tsuruta H, Agard DA, et al. (2008) Integration of small-angle x-ray scattering data into structural modeling of proteins and their assemblies. J Mol Biol 382: 1089–1106.

38. Krukenberg KA, Förster F, Rice LM, Sali A, Agard DA (2008) Multiple conformations of E.coli hsp90 in solution: Insights into the conformational dynamics of hsp90. Structure 16: 755–765.

39. Murzin A, Brenner S, Hubbard T, Chothia C (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. J Mol Biol 247: 536–540.

40. Lo Conte L, Brenner SE, Hubbard TJP, Chothia C, Murzin A (2002) SCOP database in 2002: refinements accomodate structural genomics. Nucl Acid Res 30: 264–267.

41. Andreeva A, Howorth D, Brenner SE, Hubbard T, Chothia C, et al. (2004) SCOP database in 2004: refinements integrate structure and sequence familiy data. Nucl Acid Res 32: 226–229.

42. Felder C, Graul R, Lee A, Merkle H, Sadee W (1999) The Venus flytrap of periplasmic binding proteins: an ancient protein module present in multiple drug receptors. The AAPS Journal 1: 7–26.

43. Berman H, Westbrook J, Feng Z, Gilliland G, Bhat T, et al. (2000) The protein data bank. citeseer.ist.psu.edu/berman02protein.html.

44. Seeliger D, Haas J, de Groot BL (2007) Geometry-based sampling of conformational transitions in proteins. Structure 15: 1482–1492.

45. Schneidman-Duhovny D, Nussinov R, Wolfson H (2007) Automatic prediction of protein interactions with large scale motion. Proteins: Struct Funct and Bioinf 69: 764–773.

46. Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson H (2005) Geometry-based flexible and symmetric protein docking. Proteins: Struct Funct Bioinf 60: 224–231.

47. Sandak B, Wolfson HJ, Nussinov R (1998) Flexible docking allowing induced-fit in proteins: Insights from an open to closed isomers. Proteins: Struct Funct Genet 32: 159–174.

48. Sandak B, Nussinov R, Wolfson H (1998) A method for biomolecular structural recognition and docking allowing conformational flexibility. J Comp Biol 5: 631–654.

49. de Groot BL, van Aalten DMF, Scheek RM, Amadei A, Vriend G, et al. (1997) Prediction of protein conformational freedom from distance constraints. Proteins: Struct Funct Genet 29: 240–251.

50. Huang X, Powers R (2001) Validity of using the radius of gyration as a restraint in NMR protein structure determination. J Am Chem Soc 123: 3834–3835.

51. Seeliger D, de Groot B (2009) tCONCOORD-GUI: Visually supported conformational sampling of bioactive molecules. J Comp Chem 30: 1160–1166.

52. Yang Y, Zhou Y (2008) Specific interactions for ab initio folding of protein terminal regions with secondary structures. Proteins: Struct Funct and Bioinf 72: 793–803.

53. Trott O, Olson A (2009) AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comp Chem.

54. Hess B, Kutzner C, van der Spoel D, Lindahl E (2008) Gromacs 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. J Chem Theory Comput 4: 435–447.

55. Ponder JW, Case DA (2003) Force fields for protein simulations. Adv Prot Chem 66: 27–85.

56. Sorin E, Pande V (2005) Exploring the helix-coil transition via all-atom equilibrium ensemble simulations. Biophys J 88: 2472–2493.

57. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. J Chem Phys 79: 926–935.

58. Wang J, Wolf RM, Caldwell JW, Kollmann PA, Case DA (2004) Development and testing of a generalized amber force field. J Comp Chem 25: 1157–1174.

59. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, et al. (1995) A smooth particle mesh ewald potential. J Chem Phys 103: 8577–8592.

60. Miyamoto S, Kollman PA (1992) SETTLE: An analytical version of the SHAKE and RATTLE algorithms for rigid water models. J Comp Chem 13: 952–962.

61. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM (1997) LINCS: A linear constraint solver for molecular simulations. J Comp Chem 18: 1463–1472.

62. Berendsen HJC, Postma JPM, DiNola A, Haak JR (1984) Molecular dynamics with coupling to an external bath. J Chem Phys 81: 3684–3690.

63. Bussi G, Donadio D, Parrinello M (2007) Canonical sampling through velocity rescaling. J Chem Phys 126: 014101-1–014101–7.

64. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, et al. Gaussian 03, Revision C.02. Wallingford, CT: Gaussian, Inc., 2004.