JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

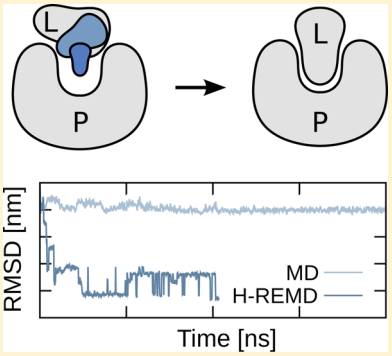
Protein−Ligand Docking Using Hamiltonian Replica Exchange Simulations with Soft Core Potentials

Manuel P. Luitz and Martin Zacharias*

Physik-Department T38, Technische Universität München, James Franck Str. 1, 85748 Garching, Germany

Supporting Information

ABSTRACT: Molecular dynamics (MD) simulations in explicit solvent allow studying receptor−ligand binding processes including full flexibility of the binding partners and an explicit inclusion of solvation effects. However, in MD simulations, the search for an optimal ligand−receptor complex geometry is frequently trapped in locally stable non-native binding geometries. A Hamiltonian replica-exchange (H-REMD)-based protocol has been designed to enhance the sampling of putative ligand−receptor complexes. It is based on softening nonbonded ligand−receptor interactions along the replicas and one reference replica under the control of the original force field. The efficiency of the method has been evaluated on two receptor−ligand systems and one protein−peptide complex. Starting from misplaced initial docking geometries, the H-REMD method reached in each case the known binding geometry significantly faster than a standard MD simulation. The approach could also be useful to identify and evaluate alternative binding geometries in a given binding region with small relative differences in binding free energy.



INTRODUCTION

The design and identification of putative drug molecules can greatly benefit from the prediction of the structure of ligand−receptor complexes.[1,2] It involves the computational docking of possible ligand molecules to cavities or ligand binding pockets on the surface of a target biomolecule. The docking search is typically performed without explicit inclusion of the surrounding solvent and no or only limited inclusion of receptor flexibility.[3,4] Predicted complexes are evaluated by simple scoring functions that include geometrical fit and different energetic contributions.[1,3,4] It is frequently possible to correctly predict the recognition site on the receptor protein, but the placement of the ligand can significantly deviate from the experimentally observed geometry.[5,6] The approximations inherent to the scoring function may not allow precise and realistic distinction between different sterically possible complex structures.[7] Furthermore, explicit solvent molecules are mostly neglected in docking approaches to accelerate the computations. However, water molecules can play a significant role in ligand−receptor recognition.

Molecular dynamics (MD) simulations in explicit solvent offer a possible route for the refinement of incorrectly placed ligands in the binding site. In an MD simulation, full atomistic flexibility of both partners (ligand and receptor) can be considered and solvent effects can be included more realistically by explicit waters compared to the approximate scoring functions used in typical docking searches.[8,9] Indeed, in ultralong MD simulations, it has been demonstrated that the complete process of ligand binding to a receptor cavity can be explored.[10,11] However, such simulations require many microseconds of simulation time, and the ligand can still be trapped for long times in metastable states separated by high-energy barriers from other realistic binding geometries. For practical applications, it would be desirable to accelerate the MD search without loss in the accuracy of representing the system and the aqueous environment. In Hamiltonian replica exchange MD (H-REMD) simulations, the force field of the system is modified in parallel replica runs that exchange with a reference simulation under the control of the original force field.

An advantage of H-REMD compared to the more common variation of the temperature among replicas (T-REMD) is the possibility to vary only part of the Hamiltonian of the system among replicas. This allows acceleration of transitions along selected degrees of freedom of the system typically requiring fewer replicas compared to T-REMD.[12] A variety of H-REMD approaches have been described scaling different parts of the force field depending on the specific sampling task.[13−21] Due to the variety of choices, it is not always clear which part of the Hamiltonian to modify among replicas in order to best tackle a given sampling problem. However, in case of ligand−receptor interactions, the trapping of the ligand on the surface of a receptor molecule at a non-native locally stable placement is due to nonbonded interactions of the ligand and receptor (or with the solvent). In order to improve the sampling of relevant ligand−receptor structures in MD-refinement simulations, we have applied H-REMD based on softening the ligand−receptor interactions along the replicas. Such softening of nonbonded interactions has been used already to improve the conformational sampling of molecules in solution[13,14] but not for

Received: February 26, 2014

   dx.doi.org/10.1021/ci500296f | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

ligand−receptor docking in explicit solvent. It allows for a broader sampling and accelerated barrier crossing in the replicas, which in turn improves the sampling of relevant states in the reference replica. The technique is able to generate a pool of possible ligand orientations in a known binding pocket and to discriminate among those which are energetically favorable. Significantly better sampling of relevant states compared to standard MD simulations was observed. It is further demonstrated that the method is applicable to the refinement of peptide−protein complexes, where side chain placement is uncertain and only backbone conformations are known approximately.

METHODS

H-REMD Docking. Energetic and entropic driving forces of ligand−receptor complex formation are closely linked to the representation of solvent molecules and ions and often demand a more accurate treatment than given by simplified docking algorithms. MD simulations have the potential to become a valuable tool for elucidating the association of ligand−receptor complexes. Seminal studies of ultralong MD simulations on special purpose machines reproduced in a proof of principle manner the binding event of small ligands to receptor molecules.[10,11] Recent speedup of MD algorithms with graphics processing units (GPUs) allowed the production of μs binding trajectories on readily affordable hardware.[22] The computational effort, however, is still tremendous. The major drawback of MD simulations is the incapability of scanning the relevant phase space region in reasonable computing time due to energy barriers. In order to accelerate the generation of conformations and to rapidly scan relevant areas in the ligand−receptor-free energy landscape, a Hamiltonian replica exchange (H-REMD) method was employed.

The method accelerates the searching of possible ligand binding modes in a known binding site by scaling the nonbonded interactions of the ligand with the rest of the system. A coupling coordinate λ is introduced to connect the unmodified Hamiltonian H_A of the system with a Hamiltonian H_B, where all nonbonded interactions of the ligand with the rest of the system are turned off (dummy ligand). The transition from the H_A to the H_B regime is mapped to parallel replica simulations at different λ_i values. The reference simulation at $\lambda = 0$ is connected to replicas at higher λ values by the H-REMD technique[23] as depicted in Figure 1. The increment of λ between the replica has to be chosen such that sufficient phase space overlap ensures reasonable exchange rates. To avoid numerical instabilities at λ values close to 1 and 0, the Hamiltonians are connected with soft core scaling potentials.[24,25]

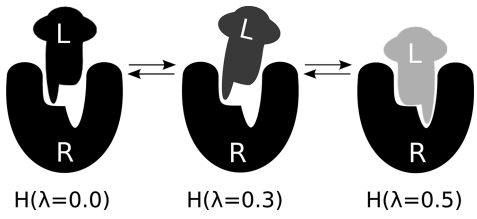


Figure 1. Schematics of protein−ligand docking with H-REMD-based soft core scaling of nonbonded protein−ligand interactions. Ligand (L) and Receptor (R) are depicted in a key−lock representation. Nonbonded interactions of the ligand are scaled along the replicas to overcome energetic barriers of misplaced conformations in the higher replicas (represented by large λ). Energetically favorable conformations are swapped back to the reference replica at $\lambda = 0$.

Simulation Setup. Simulations were performed with an in-house implementation of H-REMD in GROMACS release 4.5 and with release 4.6 of GROMACS, which includes the implementation of H-REMD and is significantly faster than release 4.5. Structural data was taken from PDB files and processed with the GROMACS toolchain[26,27] to solvate the proteins in a truncated octahedron water box with periodic boundary conditions. Crystal water and ions were removed from the PDB file in order to not overlap with manually generated initial ligand placements for FKBP-52, but crystal water was kept for FKBP-12 and MHC class I. FKBP-52 was cleaved after residue Gly139 to obtain the ligand binding subdomain solely. Hydrogen atoms provided by the PDB files were discarded and automatically assigned by the GROMACS tool PDB 2gmx, which derived protonation states for all residues. No possible differences in protonation states between bound and unbound receptor proteins were considered. The Amber parm99SB force field[28] was used in conjunction with the TIP3P water model.[29] Force field parameters for the ligand molecules were parametrized in the general Amber force field GAFF[30] with the Antechamber software package.[31] Counterions were added to render the system electrostatically neutral. The box dimensions were adjusted with an initial minimum distance of 1 nm of the solute to the box boundaries. Energy minimization via steepest decent was performed with a convergence criterion of 20 k steps or 100 kJ mol^{-1} nm^{-1} followed by an equilibration simulation in the NVT ensemble for 250 ps and a time step size of 1 fs. A second equilibration phase of the same duration and step size adapted the system to 1.01 bar in a constant pressure NPT ensemble. The equations of motion during equilibration were integrated with the leapfrog integrator (MD). Temperature was coupled to the system with the velocity rescale thermostat[32] at 298 K, and the Berendsen barostat[33] was used for NPT. During energy minimization and equilibration simulations, non-hydrogen atoms of the protein were restrained in space with a harmonic potential at force constant of 1000 kJ mol^{-1} nm^{-2}. The bond lengths of H atoms were constrained with the linear constraint solver[34] with a coupling matrix extension order of 12 (4 in production run) as proposed in ref 32. Electrostatic interactions were calculated with the particle−mesh Ewald algorithm[35] with a grid spacing of 0.12 nm and cubic B-spline interpolation. The Lennard−Jones interactions were switched to zero after 0.8 nm and a cutoff at 0.9 nm. The long-range dispersion correction resulting from the truncated Lennard−Jones interactions was applied to pressure and energy.

Hamiltonian Replica Exchanges. The replica exchange docking simulations were performed with the stochastic velocity Verlet[36] integrator, which handles the temperature coupling implicitly (298 K). The Parrinello−Rahman barostat[37] (1.01 bar) was used to generate an NPT ensemble. A 2 fs time step was used. The nonbonded ligand interactions (with receptor and solvent) were scaled in each replica using the soft core scaling method[24,25] with $\alpha = 0.3$, soft core power $p = 1.0$, and interaction radius of $\sigma = (C_{12}/C_6)^{1/6}$ or $\sigma = 0.25$ when van der Waals parameters C_{12} or C_6 were zero.

dx.doi.org/10.1021/ci500296f | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

$$V_{\text{soft core}}(r) = (1-\lambda)V_A(r_A) + \lambda V_B(r_B)$$
$$r_A = (\alpha\sigma_A^6\lambda^p + r^6)^{1/6}$$
$$r_B = (\alpha\sigma_B^6(1-\lambda)^p + r^6)^{1/6} \quad (1)$$

The parameter λ allows smooth scaling of the nonbonded interactions with $\lambda = 0$ representing the reference states with all interactions switched on and $\lambda = 1$ the state with all ligand atoms represented as (noninteracting) dummy atoms. To avoid sampling of completely unrealistic ligand placements (with large possible overlap with receptor atoms), a maximum $\lambda = 0.54$ was used in the highest of 10 replicas, and λ increased in each replica by $\Delta\lambda = 0.06$. Test simulations indicated that this step size resulted in a reasonable exchange acceptance rate of about 20%. Replica exchanges were attempted every 1000 steps among neighboring replicas (λ_i and λ_{i+1}, where i was alternating between odd or even replica numbers). A further increase in the maximum $\lambda > 0.54$ resulted in a rapid drop of the exchange acceptance rates because many conformations with close ligand−receptor atom−atom distance (and large nonbonded overlap energies in neighboring replicas with smaller λ values) were sampled. Note that the soft core scaling of the electrostatic interactions was only performed for the real space interaction. Due to the shift factor α, the nonbonded potential is shifted in real space, which can lead to slight discontinuities of the electrostatic interactions at the cutoff radius (in the replicas but not in the reference replica with the original force field that was used for all the analysis of the simulations).

Starting positions for protein−ligand compounds were generated either by manually rotating and translating the ligand in its binding pocket or by running a continuous MD simulation at $\lambda = 0.54$ and subsequent energy minimization using the full force field. The downscaling of ligand−receptor interactions can cause dissociation of the ligand from the receptor surface. In order to prevent the dissociation of the ligand from the receptor especially in the high λ replica windows, a position restraint based on the center of mass (COM) pull code was implemented in GROMACS. A harmonic potential $f(r)$ with force constant $k_{\text{com}} = 1000$ kJ mol^{-1}nm^{-2} along the connection vector of the COMs of ligand and protein was applied when the actual COM distance r moved farther than a penalty value d away from the equilibrium distance r_{eq}.

$$f(r) = \begin{cases} k_{\text{com}}(r - r_{\text{eq}} + d)^2, & r < r_{\text{eq}} - d \\ k_{\text{com}}(r - r_{\text{eq}} - d)^2, & r > r_{\text{eq}} + d \\ 0, & r_{\text{eq}} - d \le r \le r_{\text{eq}} \end{cases} \quad (2)$$

We adjusted the penalty value to $d = 0.7$ nm.

Test Systems. The human FKBP protein (FKBP-52) in complex with the high-affinity ligand FK506 (PDB code 4LAX) and a complex with the lower affinity ligand SB3[38] (in complex with FKBP-12, PDB code 1FKG) served as test systems for the H-REMD docking simulations. The dissociation constant of SB3 (1,3-diphenyl-1-propyl-1-(3,3-dimethyl-1,2-dioxyphenyl)-2-piperidine carboxylate) and the FKBP domain is $K_d = 10$ nM compared to $K_d = 0.4$ nM of the native FKBP inhibitor FK506. Refinement of docked peptides was tested on the peptide binding domain of a murine MHC class I molecule in complex with a viral antigen (PDB code 2VAB). The antigen sequence is FAPGNYPAL and is derived from the Sendai virus nucleoprotein (324−332), SEV-9.[39]

RESULTS AND DISCUSSION

FKBP Ligand−Receptor Complexes. MD simulations have been used in the past to refine docked ligand−receptor complexes.[40−42] In order to evaluate the capability of continuous MD (cMD) simulations for refining docked ligand−receptor complexes, simulations were started from two initial ligand placements denoted as W1 and W2 that deviated by 0.5 and 0.4 nm of $\text{RMSD}_{\text{ligand}}$ from the native FKBP-FK506 complex (Figure 2). The root mean-square deviation ($\text{RMSD}_{\text{ligand}}$) corresponds to the deviation of the ligand atoms from the native structure after best superposition of the receptor structure onto the receptor of the native complex. A $\text{RMSD}_{\text{ligand}}$ of less than 0.3 nm was considered as a near-native binding mode. In order to provide a fair comparison with the H-REMD approach, 10 independent simulations of 20 ns duration for each starting structure W1 and W2 were performed. Starting velocities for each run were assigned randomly according to the Maxwell distribution at temperature 298 K. The cMD simulation setup was chosen such that approximately the same computational resources were used for the comparative H-REMD run with 10 replicas.

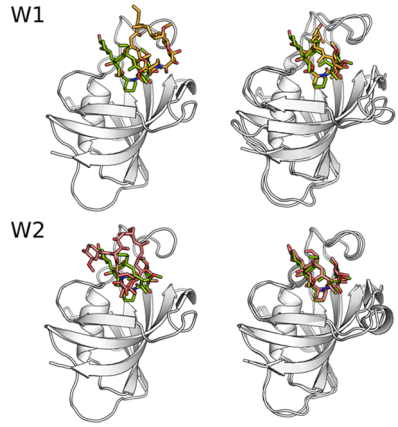


Figure 2. Starting (left panels) and final (right panels) W1 and W2 structures for H-REMD simulations of FKBP-52 protein in complex with ligand FK506 Tacrolimus (sticks). Protein backbone structures of FKBP-52 are aligned with the crystal structure (PDB code 4LAX), and the native ligand binding mode is colored green.

For the first starting structure W1, all cMD simulations were unsuccessful to move the ligand closer to the known binding configuration within 20 ns simulation time (Figure 3A). A cluster analysis revealed that simulations were kinetically trapped in the initial ligand configuration or moved away from the native binding mode (Figure 3B).

For the second structure W2, two cMD simulations explored the native binding mode after 5 and 10 ns; however, the majority of trajectories moved away from the native ligand conformation and were trapped in metastable non-native binding modes (Figure 3A). The cluster representing the native binding mode covers approximately 10% of the joined 200 ns W2 trajectories but is not the largest cluster.

In contrast to the cMD simulations, the H-REMD simulations resulted in both cases in a rapid drop of the $\text{RMSD}_{\text{ligand}}$ with respect to the native complex structure within less than 1 ns in case of initial structure W2 and 1.3 ns

dx.doi.org/10.1021/ci500296f | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

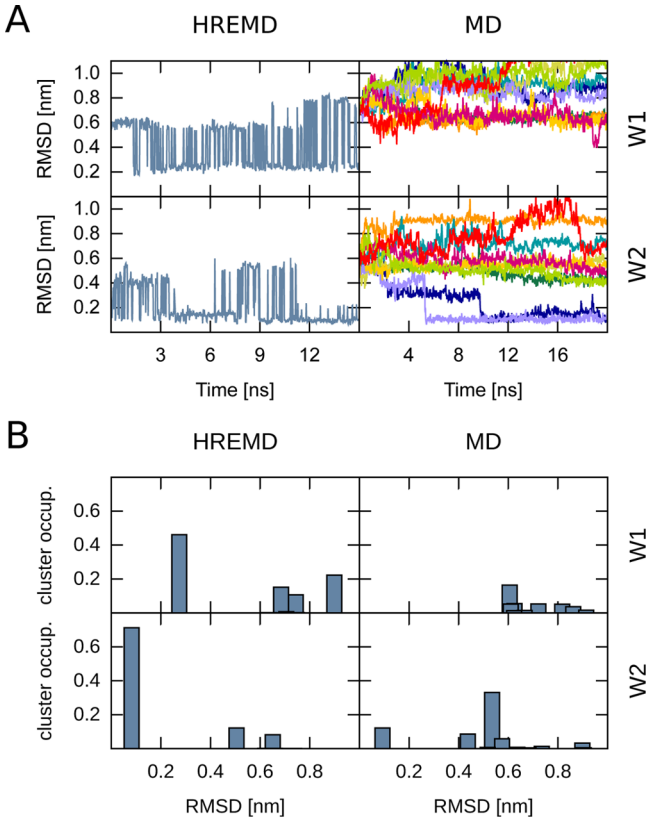


Figure 3. H-REMD simulations of FK506 binding to the FKBP-52 protein. Two starting conformations W1 (upper row) and W2 (lower row) of FK506 ligand were used to initially feed H-REMD and control MD simulations (start structures are shown in Figure 2). (A) $RMSD_{ligand}$ of FK506 versus crystal structure for the H-REMD reference replica (left panel) and 10 continuous MD runs (right panel). The FKBP-52 backbone was aligned with respect to the crystal structure prior to $RMSD_{ligand}$ calculations. (B) Relative population of the seven larges clusters for H-REMD and MD simulations, respectively, vs $RMSD_{ligand}$. The cluster occupancy is the fraction of cluster members over all frames and gives a measure of the relative cluster size. Binding modes are clustered with the single-linkage method, and the cluster $RMSD_{ligand}$ to reference crystal structure was calculated for the cluster member with smallest average $RMSD_{ligand}$ to all other structures of the cluster. MD runs have been concatenated to one 200 ns trajectory for W1 and W2, respectively, before clustering.

simulation time in case of start structure W1 (Figure 3). A cluster analysis of the sampled states in the replica under the control of the original force field identified in both cases ligand placements with a $RMSD_{ligand}$ < 0.3 nm as most populated states within 15 ns H-REMD simulations (Figure 3B).

The ligand binding modes sampled at the final stage of the two H-REMD simulations starting from different initial placements differed in $RMSD_{ligand}$. The placement that was reached from the second start structure W2 reached a considerably lower $RMSD_{ligand}$ than the simulation starting from W1 (Figure 3). Inspection of the binding pocket indicated that differences in the side chain conformations of the receptor in the binding pocket are a likely reason for the alternative ligand placements. The Tyr92 side chain hindered the (1*S*,2*S*)-2-methoxycyclohexanol motive of FK506 from moving in the correct configuration (see W1, right panel, Figure 2). It should be noted that during H-REMD only the receptor−ligand interactions were scaled along the replicas but not the side chain−side chain interactions in the binding region. This indicates that the H-REMD approach rapidly identifies near-native binding modes of the FKBP-52−FK506 complex. However, a full equilibration could not be achieved because alternative binding modes still competed during the 15 ns in the reference replica, and for both start structure cases, slightly different receptor conformations evolved during the search.

For the second test system FKBP-12 with ligand SB3, an alternative strategy was evaluated. Instead of starting from only one ligand−receptor complex, different start structures (one in each replica) were used. Each start structure had a $RMSD_{ligand}$ > 0.4 nm from the native binding mode. Application of 10 independent cMD simulations (each 20 ns) resulted in kinetically trapped complex structures, and none of the simulations reached any near-native complex geometry (Figure S1, Supporting Information). In contrast, cluster analysis of the H-REMD application resulted in several notable near-native binding modes in the reference replica (Figure 4). Cluster 3 was close to the crystal structure binding mode with $RMSD_{ligand}$ < 0.2 nm. Noteworthy, cluster 5, which showed a quasi-symmetric flip of the two phenyl rings in the SB3 1,3-diphenylprobyl subunit compared to crystal structure, was otherwise very similar to the experimental geometry (Figure 4). A 20 ns H-REMD simulation on the crystal structure swapped after 15 ns an alternative binding mode with a rearranged loop region into the reference replica. This binding mode survived for the last 5 ns in the lower λ replica and was present in the reference replica after its first occurrence with a 5% probability. This mode was not found in a standard MD simulation starting from the FKBP-12-SB3 crystal structure (data not shown). However, the observation of alternative binding modes for the SB3 ligand with lower affinity compared to the first ligand−receptor system in line with previous ultralong MD simulations on the same system.[43] It indicates that the present approach could be especially useful to identify putative alternative binding modes for a ligand in shorter simulation time than standard MD simulations.

For a fully equilibrated simulation, the relative populations of conformational clusters should directly reflect the relative free energies of the clusters. Hence, the accumulation of ligand placements in the most populated cluster can be used to identify the most likely binding mode in the typical case when the native complex is not known. Indeed, in the case of the high-affinity FKBP−FK506 complex, the most populated clusters corresponded to the near-native complex (although not with a large gap relative to alternative placements), and its population was larger for the second half compared to the first half of the H-REMD simulation. However, for the second system, the near native geometry was sampled but not as most populated cluster (Figure 4). In practical applications, one is interested to identify putative binding modes as quickly as possible and will not be able to run simulations to full equilibration. It is therefore useful to look into alternative strategies to identify the most likely near-native geometry, e.g., by rescoring of sampled ligand−receptor geometries. Interestingly, for the present systems, the average ligand−receptor interaction energy showed a correlation with the deviation of the ligand from the near-native placement (Figure S2, Supporting Information). Near-native sampled complexes of FKBP−FK506 showed on average a more favorable interaction energy compared to complexes with larger deviation from the native placement. Presumably, this is due to an optimal complementarity of interactions in the native binding placement. For the FKBP−SB3 system, the near native placement

dx.doi.org/10.1021/ci500296f | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

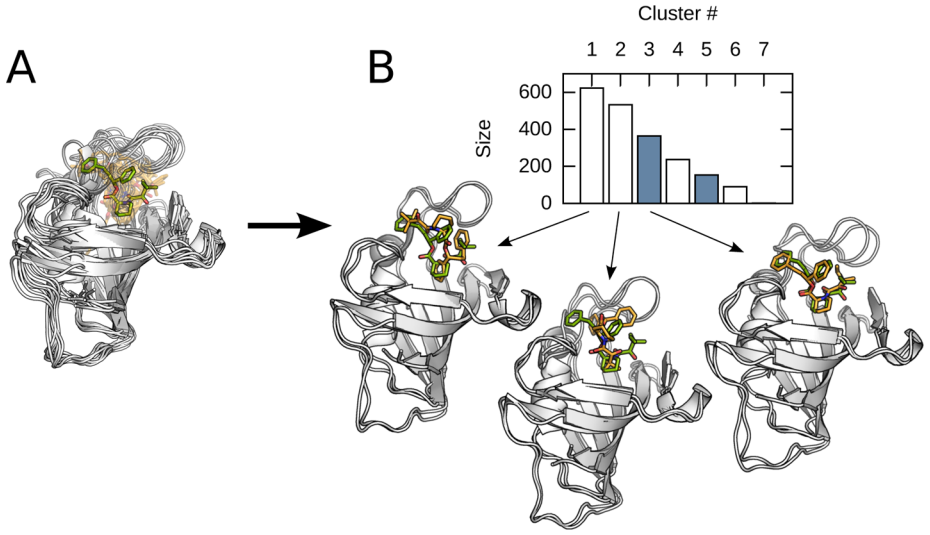


Figure 4. H-REMD simulations of FKBP-12 in complex with ligand SB3. (A) Pool of initial SB3 conformations for H-REMD (shaded orange) superimposed onto the crystal structure (green). (B) Cluster analysis of SB3 binding modes observed in the reference H-REMD window (blue boxes correspond to near native solutions). Average structures of the three most populated clusters are illustrated (predicted ligand as orange vs native ligand in green stick representation).

and an alternative configuration at $RMSD_{ligand} \approx 0.5$ nm showed similar interaction energies (Figure S2, Supporting Information). Hence, for the present cases, the interaction energy could be used (besides of cluster population) as a criterion to select possible near-native binding modes. Note, however, that the binding affinity is in general not only determined by direct ligand−receptor interactions but also by other energetic and entropic contributions.

Refinement of MHC Class I Peptide−Protein Complexes. MHC class I proteins play a central role in the recognition of antigenic peptides and their representation to the immune system. Class I proteins contain a narrow cleft between two α-helices to bind antigenic peptides of 8−10 residues with an extended backbone structure.[44] The receptor conformation and bound peptide backbone conformation is similar for most antigenic peptides of the same length, which is confirmed by over 170 crystallized structures in the Protein Data Bank. It offers the possibility to test the H-REMD method on predicting the side chain structure for an approximately known peptide backbone conformation. A peptide start structure with extended backbone and randomized side chain conformations was generated and energy minimized in the binding groove to remove possible sterical clashes (Figure 5). The ligand backbone conformation was weakly restrained to the reference coordinates of the crystal structure with a force constant of 1000 kJ mol^{-1} nm^{-2}, which allowed backbone conformational fluctuations of ±0.05 nm. In a standard MD simulation of 20 ns, the peptide was trapped over the whole simulation time close to the initial conformation due to sterical barriers in the narrow binding cleft of MHC class I protein (Figure 6). The cluster analysis for ligand conformations found only one cluster with more than one member. The average structure was similar to the starting structure (Figure 5). In contrast, the H-REMD simulations sampled the correct binding mode with most side chains in near-native rotameric states in less than 3 ns. Already during the equilibration phase, a conformation with lower ligand RMSD than the starting structure was swapped in the reference replica, and only two relevant clusters survived for the rest of the simulation. The two

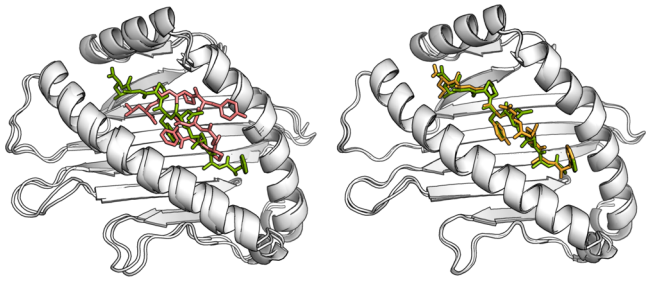


Figure 5. MHC class I protein in complex with viral antigen SEV-9. The MHC class I protein backbone structure is aligned with the crystal structure (PDB code 2VAB). The native antigen peptide configuration is colored green. On the left side, the starting structure for MD and H-REMD simulations is indicated in red. The right panel shows the average structure representing the second largest cluster (orange) obtained after H-REMD simulation (green: native ligand conformation). The $RMSD_{ligand}$ of the predicted structure is below 0.2 ns with respect to the native ligand placement.

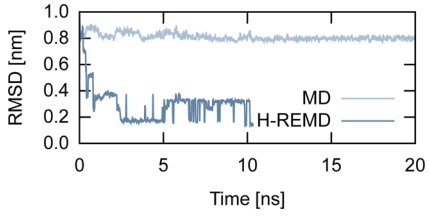


Figure 6. $RMSD_{ligand}$ vs simulation time for MD simulation and reference replica of the H-REMD simulation of MHC class I protein in complex with viral protein fragment SEV-9. Simulations were started from a starting structure shown in Figure 5. The ligand conformation is trapped near its starting conformation by sterical barriers during the whole MD simulation. In the H-REMD reference window, a lower RMSD configuration is already found within <0.5 equilibration. Only two main binding modes survive in the reference replica after 0.5 ns, which differ mainly in the side chain rotamer of Phe1.

clusters differed mainly in the side chain rotamer of the Phe1 peptide residue. Similar to the FKBP systems (previous paragraphs), a noteworthy correlation of the peptide−receptor

dx.doi.org/10.1021/ci500296f | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

interaction energy with respect to deviation from the native complex structure was found (Figure S2, Supporting Information), indicating that also in this case the interaction energy could serve as an additional criterion to select realistic docking structures.

CONCLUSIONS

The realistic prediction of ligand−receptor binding geometries is an important goal of in silico drug discovery. Molecular docking methods are widely used to generate and evaluate putative binding geometries mostly employing simple scoring functions that include flexibility of the binding partners and solvation of the ligand and receptor binding site only approximately. MD simulations on the other hand include both full flexibility of binding partners and explicit solvation of partner molecules and are increasingly being used for studying ligand−receptor association events. The major drawback of standard MD simulations, however, is the large computational demand coupled with sampling mostly irrelevant states before reaching a native binding mode. High energy barriers separating favorable putative binding geometries prevent the identification of near-native binding modes in reasonable sampling times. In order to accelerate the search, we tested an H-REMD approach based on soft-core scaling nonbonded interactions between partners along the replicas. Similar approaches have already been used to enhance conformational sampling of isolated molecules[13,14] in explicit solvent but not to refine putative ligand−receptor complex structures. The approach requires a smaller number of replicas compared to standard temperature (T-)REMD because the scaling affects only a small fraction of the system. For the present test cases, 10 replicas were sufficient to significantly reduce barriers, enhance sampling, and still allow reasonable exchange rates between neighboring replicas. The method was tested on three systems including a peptide−protein complex. In all cases, significant improvement of initially misplaced ligand−receptor complexes and side chain placements in case of the peptide−protein complex was obtained. In contrast, standard MD simulations largely failed on the test cases to refine the initial conformations. The approach is especially powerful to sample a broader range of possible ligand−receptor complex structures compared to standard MD simulation. A full equilibration of sampled states was still not achieved for the ligand−receptor complexes because different distributions of sampled states were obtained depending on the start structure (e.g., in the FKBP−FK506 system). This can be attributed in part to limited sampling but may also be due to alternative receptor conformations evolving during the simulations starting from different initial ligand placements.

A possible extension of the approach is to not only scale ligand−receptor interactions among replicas but also the interactions between side chains of the receptor in the vicinity of the ligand binding site, which could result in a simultaneous enhanced sampling of alternative receptor conformations. It should be noted that the computational demand of the approach is much higher than of standard docking tools. However, it should also be emphasized that the approach is not intended for systematic evaluation of thousands of putative ligands or binding modes but to refine a small fraction of preselected putative docking geometries (e.g., obtained form a systematic docking run). With increasing efficiency of MD simulations, such an approach can become a valuable alternative to simple scoring methods of docking geometries that largely neglect the dynamics of partners and do not include explicit solvent molecules.

Another interesting possible application is the generation of alternative binding geometries for a given ligand in a binding site that are of similar binding free energy. We identified an alternative binding mode of the FKBP-12-SB3 system in a H-REMD simulation that started from crystal structure configuration. The finding agrees with previous MD studies that identified alternative binding modes for the same complex.[43] The population of different binding modes in a given binding site reflects the relative stability of these modes and may give hints of which interactions (chemical groups on the ligand or possible chemical ligand modifications) may stabilize or destabilize a given binding geometry.

ASSOCIATED CONTENT

Supporting Information

$RMSD_{ligand}$ plots of 10 independent MD simulations on the FKBP-12-SB3 complex. The $RMSD_{ligand}$ was calculated with respect to the native complex. In addition, the correlation of $RMSD_{ligand}$ with interaction energy between ligand and receptor for all three test systems is shown. The correlation was calculated from the reference replica of the H-REMD simulations. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: zacharias@tum.de. Phone: +49 (0)89-289-12335. Fax: +49 (0)89-289-12444.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the financial support by grant SFB1035/project B02 of the DFG (Deutsche Forschungsgemeinschaft). Computer resources for this project have been provided by the Gauss Centre for Supercomputing/Leibniz Supercomputing Centre under grant pr86ri.

REFERENCES

(1) Park, H.; Lee, J.; Lee, S. Critical assessment of the automated AutoDock as a new docking tool for virtual screening. *Proteins: Struct., Funct., Bioinf.* **2006**, *65*, 549−554.

(2) Ewing, T. J. A.; Makino, S.; Skillman, A. G.; Kuntz, I. D. DOCK 4.0: Search strategies for automated molecular docking of flexible molecule databases. *J. Comput.-Aided Mol. Des.* **2001**, *15*, 411−428.

(3) Totrov, M.; Abagyan, R. Flexible ligand docking to multiple receptor conformations: A practical alternative. *Curr. Opin. Struct. Biol.* **2008**, *18*, 178−184.

(4) May, A.; Sieker, F.; Zacharias, M. How to efficiently include receptor flexibility during computational docking. *Curr. Comput.-Aided Drug Des.* **2008**, *4*, 143−153.

(5) Zacharias, M. ATTRACT: Protein−protein docking in CAPRI using a reduced protein model. *Proteins: Struct., Funct., Bioinf.* **2005**, *60*, 252−256.

(6) Kuzu, G.; Keskin, O.; Gursoy, A.; Nussinov, R. In *Computational Drug Discovery and Design*; Methods in Molecular Biology; Baron, R., Ed.; Springer: New York, 2012; Vol. *819*; pp 59−74.

(7) Keeble, A. H.; Joachimiak, L. a.; Maté, M. J.; Meenan, N.; Kirkpatrick, N.; Baker, D.; Kleanthous, C. Experimental and computational analyses of the energetic basis for dual recognition of immunity proteins by colicin endonucleases. *J. Mol. Biol.* **2008**, *379*, 745−759.

(8) Samsonov, S.; Teyra, J.; Anders, G.; Pisabarro, M. T. Analysis of the impact of solvent on contacts prediction in proteins. *BMC Struct. Biol.* **2009**, *9*, 1−11.
(9) Ahmed, M. H.; Spyrakis, F.; Cozzini, P.; Tripathi, P. K.; Mozzarelli, A.; Scarsdale, J. N.; Safo, M. A.; Kellogg, G. E. Bound water at protein−protein interfaces: Partners, roles and hydrophobic bubbles as a conserved motif. *PLoS One* **2011**, *6*, e24712.
(10) Dror, R. O.; Pan, A. C.; Arlow, D. H.; Borhani, D. W.; Maragakis, P.; Shan, Y.; Xu, H.; Shaw, D. E. Pathway and mechanism of drug binding to G-protein-coupled receptors. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 13118−13123.
(11) Shan, Y.; Kim, E. T.; Eastwood, M. P.; Dror, R. O.; Seeliger, M. A.; Shaw, D. E. How does a drug molecule find its target binding site? *J. Am. Chem. Soc.* **2011**, *133*, 9181−9183.
(12) Ostermeir, K.; Zacharias, M. Advanced replica-exchange sampling to study the flexibility and plasticity of peptides and proteins. *Biochim. Biophys. Acta* **2013**, *1834*, 847−853.
(13) Affentranger, R.; Tavernelli, I.; Di Iorio, E. E. A novel Hamiltonian replica exchange MD protocol to enhance protein conformational space sampling. *J. Chem. Theory Comput.* **2006**, *2*, 217−228.
(14) Hritz, J.; Oostenbrink, C. Hamiltonian replica exchange molecular dynamics using softcore interactions. *J. Chem. Phys.* **2008**, *128*, 144121.
(15) Wang, L.; Friesner, R. A.; Berne, B. J. Replica exchange with solute scaling: A more efficient version of replica exchange with solute tempering (REST2). *J. Phys. Chem. B* **2011**, *115*, 9431−9438.
(16) Itoh, S. G.; Okumura, H.; Okamoto, Y. Replica-exchange method in van der Waals radius space: Overcoming steric restrictions for biomolecules. *J. Chem. Phys.* **2010**, *132*, 134105.
(17) Kannan, S.; Zacharias, M. Enhanced sampling of peptide and protein conformations using replica exchange simulations with a peptide backbone biasing-potential. *Proteins: Struct., Funct., Bioinf.* **2007**, *66*, 697−706.
(18) Kannan, S.; Zacharias, M. Application of biasing-potential replica-exchange simulations for loop modeling and refinement of proteins in explicit solvent. *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 2809−2819.
(19) Curuksu, J.; Zacharias, M. Enhanced conformational sampling of nucleic acids by a new Hamiltonian replica exchange molecular dynamics approach. *J. Chem. Phys.* **2009**, *130*, 104110.
(20) Zacharias, M. Combining elastic network analysis and molecular dynamics simulations by Hamiltonian replica exchange. *J. Chem. Theory Comput.* **2008**, *4*, 477−487.
(21) Piana, S.; Laio, A. A bias-exchange approach to protein folding. *J. Phys. Chem. B* **2007**, *111*, 4553−4559.
(22) Buch, I.; Giorgino, T.; De Fabritiis, G. Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 10184−10189.
(23) Woods, C. J.; Essex, J. W.; King, M. a. The development of replica-exchange-based free-energy methods. *J. Phys. Chem. B* **2003**, *107*, 13703−13710.
(24) Zacharias, M.; Straatsma, T. P.; McCammon, J. A. Separation−shifted scaling, A new scaling method for Lennard−Jones interactions in thermodynamic integration. *J. Chem. Phys.* **1994**, *100*, 9025−9031.
(25) Beutler, T. C.; Mark, A. E.; van Schaik, R. C.; Gerber, P. R.; van Gunsteren, W. F. Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations. *Chem. Phys. Lett.* **1994**, *222*, 529−539.
(26) Hess, B.; Kutzner, C.; Spoel, D. V. D.; Lindahl, E. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.
(27) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, flexible, and free. *J. Comput. Chem.* **2005**, *26*, 1701−1718.
(28) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Struct., Funct., Bioinf.* **2006**, *65*, 712−725.
(29) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(30) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general Amber force field. *J. Comput. Chem.* **2004**, *25*, 1157−1174.
(31) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bondtype perception in molecular mechanical calculations. *J. Mol. Graph. Model.* **2006**, *25*, 247−260.
(32) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, 014101−7.
(33) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(34) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997**, *18*, 1463−1472.
(35) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577−8593.
(36) van Gunsteren, W. F.; Berendsen, H. J. C. A leap-frog algorithm for stochastic dynamics. *Mol. Simul.* **1988**, *1*, 173−185.
(37) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182−7190.
(38) Holt, D. A.; Luengo, J. I.; Yamashita, D. S.; Oh, H. J.; Konialian, A. L.; Yen, H. K.; Rozamus, L. W.; Brandt, M.; Bossard, M. J. a. Design, synthesis, and kinetic evaluation of high-affinity FKBP ligands and the X-ray crystal structures of their complexes with FKBP12. *J. Am. Chem. Soc.* **1993**, *115*, 9925−9938.
(39) Fremont, D.; Matsumura, M.; Stura, E.; Peterson, P.; Wilson, I. Crystal structures of two viral peptides in complex with murine MHC class I H-2Kb. *Science* **1992**, *257*, 919−927.
(40) 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.
(41) van Dijk, A. D. J.; Bonvin, A. M. J. J. Solvated docking: Introducing water into the modelling of biomolecular complexes. *Bioinformatics* **2006**, *22*, 2340−2347.
(42) Smith, G. R.; Sternberg, M. J.; Bates, P. A. 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.
(43) Fujitani, H.; Tanida, Y.; Ito, M.; Jayachandran, G.; Snow, C. D.; Shirts, M. R.; Sorin, E. J.; Pande, V. S. Direct calculation of the binding free energies of FKBP ligands. *J. Chem. Phys.* **2005**, *123*, 084108.
(44) Zacharias, M.; Springer, S. Conformational flexibility of the MHC class I α1-α2 domain in peptide bound and free states: A molecular dynamics simulation study. *Biophys. J.* **2004**, *87*, 2203−2214.