Addressing Suboptimal Poses in Nonequilibrium Alchemical Calculations

25CF•
2212-

Addressing Sub-optimal Poses in Non-equilibrium Alchemical Calculations

Maurice Karrenbrock,[†] Valerio Rizzi,[†] Piero Procacci,[‡] and Francesco Luigi Gervasio[*,†,¶,§,‖]

†*School of Pharmaceutical Sciences, University of Geneva, Rue Michel-Servet 1, CH-1206 Geneva, CH*

‡*Chemistry Department, University of Florence, Via della Lastruccia 3-13, 50019 Sesto Fiorentino, IT*

¶*Institute of Pharmaceutical Sciences of Western Switzerland, University of Geneva, CH-1206, Geneva, CH*

§*Chemistry Department, University College London (UCL), WC1E 6BT, London, UK*

‖*Swiss Bioinformatics Institute, University of Geneva, CH-1206, Geneva, CH*

E-mail: francesco.gervasio@unige.ch

Abstract

Alchemical transformations can be used to quantitatively estimate absolute binding free energies at reasonable computational cost. However, most of the approaches currently in use require knowledge of the correct (crystallographic) pose. In this paper we present a combined Hamiltonian replica exchange non-equilibrium alchemical method that allows to reliably calculate absolute binding free energies, even when starting from sub-optimal initial binding poses. Performing a preliminary Hamiltonian replica exchange enhances the sampling of slow degrees of freedom of the ligand and the target allowing the system to populate the correct binding pose when starting from an approximate docking pose. We apply the method on 6 ligands of the first bromodomain of the BRD4 bromodomain containing protein. For each ligand, we start non-equilibrium alchemical transformations from both the crystallographic pose and the top-scoring docked pose that are often significantly different. We show that the method produces statistically equivalent binding free energies making it a useful tool for computational drug discovery pipelines.

Introduction

The calculation of accurate absolute binding free energies of protein-ligand complexes is an essential and challenging task in computer-aided drug design[1] (CADD). Many algorithms have been developed for this purpose, ranging from end-point alchemical methods[2–15] to Collective Variable (CV) based enhanced sampling algorithms.[16–19] Ease of use, speed, automation and reliability of the algorithm are particularly important aspects in the context of drug discovery pipelines. In this respect, non-equilibrium alchemical techniques have been shown to strike a good balance[20–23] and be a valid alternative to traditional equilibrium alchemical techniques (FEP),[24,25] while they are generally easier to automate than CV-based enhanced sampling algorithms such as Metadynamics.[26,27]

While the absolute binding free energies predicted by alchemical methods are often in

good agreement with experimental results, they still suffer from a number of challenges, most notably a strong dependency on the initial binding pose and the difficulty in sampling slow degrees of freedom.[14] Slow side chain rearrangements and the presence of long-lived water molecules are known to play a significant role in ligand binding.[14,18] Thus, if the starting structure used in the alchemical transformation does not include the correct number of long-lived water molecules or its lateral chains are incorrectly positioned, the accuracy of the predicted free energies can be adversely affected.

One way to overcome this issue is to perform very long simulations[28] to fully sample the slow degrees of freedom associated with long-lived waters and the repositioning of lateral chains, but this approach is extremely computationally demanding. On the shorter timescales that are routinely employed, it has been recently shown[10] that carefully selected restraints between the protein and the ligand greatly help the convergence. However, the resulting binding affinities depend critically on the initial choice of binding pose, which determines the accuracy of the predictions.[29–31]

In a real-world drug discovery scenario, crystal binding poses are often not available, and alchemical calculations would have to start from predictions of the real poses obtained, for example, by docking.[30,32–36] In this context, a method capable of providing consistent binding free energy estimates when starting from a range of initial binding poses would be highly desirable.

Feeding an ensemble of configurations generated by enhanced sampling to alchemical transformations is one way to mitigate such challenges. We have recently introduced a non-equilibrium technique called Virtual Double System Single Box (vDSSB),[37,38] where a Hamiltonian Replica Exchange (HREM)[39–41] is performed on selected torsion angles prior to the alchemical transformations. The method aims to pre-sample the Boltzmann distribution of the ligand in the pocket and of the surrounding target residues and can be seen as an evolution of the non-equilibrium work fast switching double annihilation method NEW-FSDAM.[42–46]

In this paper we further develop the vDSSB algorithm by extending the sampling in the HREM stage, where both torsional angles and van der Waals interactions are scaled and the conformational landscape of the binding site is explored even more thoroughly. The proposed approach mitigates the effect of the initial binding pose on the free energy estimates by sampling slow degrees of freedom related to the ligand orientation and the the binding cavity conformation. Using a permissive restraint in the HREM simulations keeps the ligand in the cavity while allowing a wide sampling of accessible binding poses. The implementation of the algorithm in the HPC_Drug python middleware (V. 2.0) makes it easier to automate and include in ligand screening pipelines.

We test the method on a realistic and non-trivial set of protein-ligand systems consisting of 6 ligands of the first bromodomain of the BRD4 bromodomain containing protein (hereon BRD4 bromodomain) (Figure 1). The binding free energies for this system have been previously estimated by means of equilibrium alchemical transformations by Aldeghi et al.[25] and with non-equilibrium alchemical transformations by Gapsys et al.[22] For the sake of clarity, as we are using a subset of the ligands used in these papers, we keep the original numbering (the ligands we use are 1, 6, 7, 8, 9, 11).

To test the performance of the method, we run it both on the available crystal poses and on poses obtained from docking and compare the resulting conformational distributions and binding free energies. Notably, ligand 11 provides a real-world challenge as its crystallographic binding pose is unknown, therefore one pose was modelled upon ligand 9 and the other one was obtained by docking.[25] Two other challenging cases are ligands 6 and 9 where the poses obtained from docking are significantly different from the crystallographic ones, as they are flipped by 180 degrees. We show how the sampling provided by the HREM simulations produces comparable conformational and orientational distributions when starting from different poses. The calculated free energies are in good agreement with the experimental results and, more importantly, the results from different poses are strongly correlated, highlighting the ability of our algorithm to produce consistent and accurate free

energy predictions even when starting from sub-optimal binding poses.

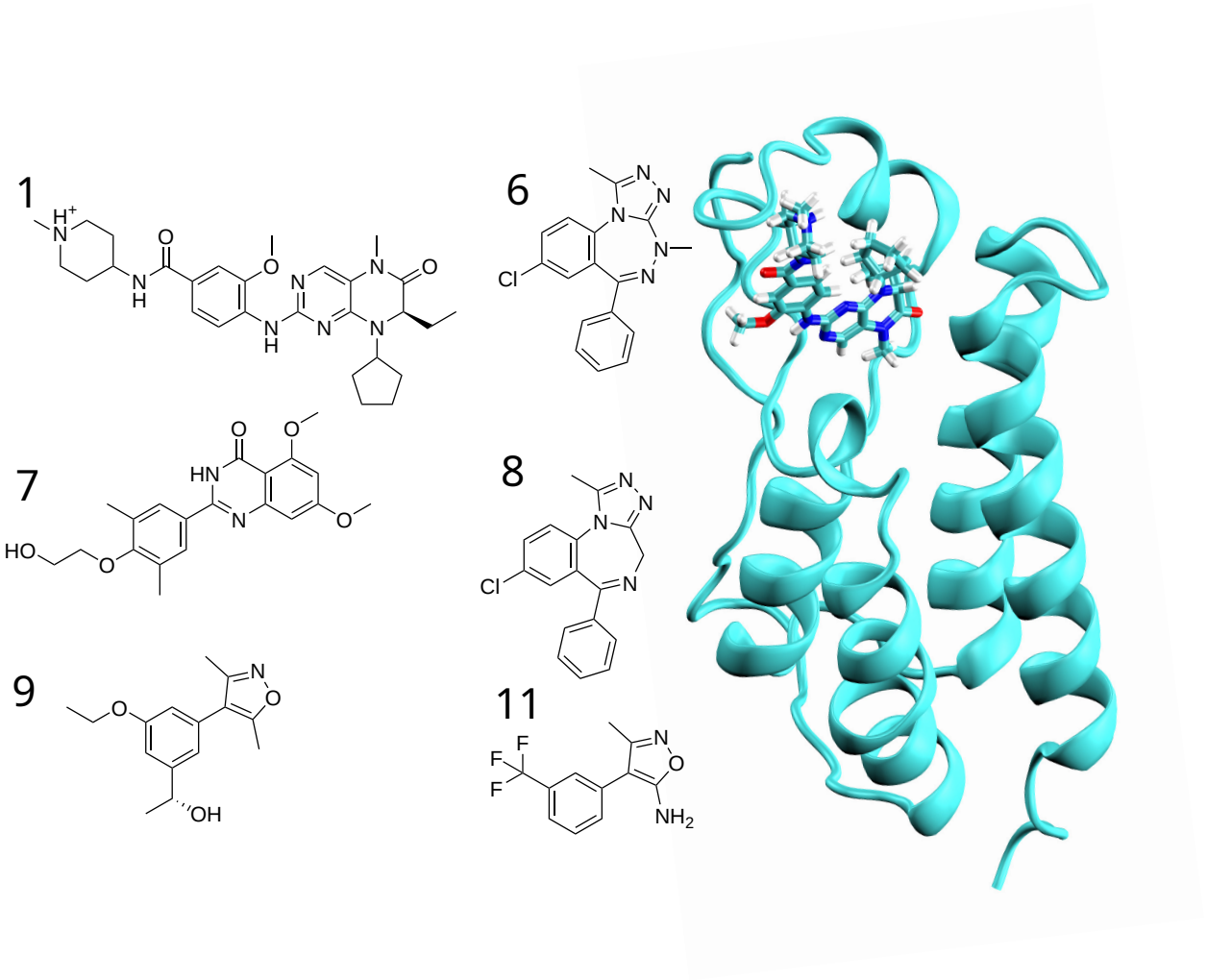

Figure 1: Structure of the BRD4 bromodomain HOLO protein and the ligands that we simulate. For clarity we kept the same ligand numbering as in the original paper by Aldeghi et al.[25] where we took the structures from.

Methods

The vDSSB[37,38] is a multi-step algorithm where, initially, a Hamiltonian Replica Exchange (HREM)[39–41] is used to explore the conformational phase space of the system. A weak position restraint is applied to the ligand so that it cannot escape from the binding pocket while it is still able to explore different bound conformations.

Subsequently, a number of snapshots is randomly extracted and used as the starting point for non-equilibrium alchemical transformations. From them a swarm of fully independent non-equilibrium alchemical transformations is run. One then extracts from these calculations a set of alchemical work values on which one employs Crooks' equation[47] through Bennett's acceptance ratio (BAR).[48] Here is a step-by-step description of the process.

Hamiltonian replica exchange HREM

This step requires the simultaneous simulation of a number of replicas where interactions are progressively tweaked. It is performed in the following systems:

- HREM of the protein-ligand complex where the torsions and the VdW interactions of the ligand and the first neighboring residues are scaled along 40 replicas according to a geometric progression where i is the zero-based replica index $0.2^{i/39}$.
- HREM of the APO protein where the torsions and the VdW interactions of the residues of the binding pocket are scaled along 40 replicas according to a geometric progression where i is the zero-based replica index $0.2^{i/39}$.
- HREM of the ligand in water where torsions and the VdW interactions of the ligand are scaled along 8 replicas according to a geometric progression where i is the zero-based replica index $0.2^{i/7}$.
- HREM of the ligand in vacuum where torsions of the ligand are scaled along 8 replicas according to a geometric progression where i is the zero-based replica index $0.2^{i/7}$, and the VdW and coulombic interactions are set to zero.

The HREMs of the ligand in vacuum and in water are run for 40ns. In the case of the complexes and the APO protein, each replica is initially equilibrated for 100ns applying restrains on the heavy atoms of the protein and ligand. This allows the water molecules to equilibrate. The HREM is then switched on with unrestricted atomic positions for a further

100ns of equilibration, during which high index replicas can diffuse down the replica ladder, allowing slow degrees of freedom sampling. The length of the equilibration is determined by checking the distribution of configurations visited by replica 0. The equilibration is stopped when the distribution becomes quasi-static (see Supporting Information (SI) Figure 1). As expected, ligands 6 and 9 when starting from a flipped docked pose require longer HREM equilibration times, respectively 400ns for ligand 6 and 300ns for ligand 9.

Finally, the production HREM simulations consist of three independent HREM runs of 100ns each, restarted from the equilibrated systems. The number of replicas was chosen to strike a balance between computational cost and a reasonably high exchange probability without creating diffusion bottlenecks between replicas.

Position Restraints

We perform the HREM runs with a weak harmonic restraint between the center of mass of the binding pocket and the center of mass of the ligand. This restraint allows the ligand to explore alternative binding poses while keeping it in the cavity.

The center of mass of the pocket is defined by a set of Cα atoms of some residues of the binding pocket that were selected by visual inspection (see Figure 2), favouring rather rigid protein portions. As the binding pocket is very flexible, the atoms that we chose are clustered in the few rigid regions available. The harmonic constant of the restraint is K = 28.7 kcal mol^{-1} nm^{-2}.

Figure 2: BRD4 bromodomain HOLO structure. The Cα atoms used to define the center of the binding pocket are shown as red spheres.

During the alchemical transformations, in order to prevent the ligand from diffusing out of the binding pocket or rotating into nonphysical conformations, we freeze three non collinear heavy atoms. To this end we use the isokinetic freezing algorithm[49] in which the mass of the three atoms is set to a high value (in our case 10^{20} atomic units of mass),

while their velocity is set to zero. It has been shown that isokinetic freezing does not affect the final free energy values[49] since the work done on the system to scale the mass and velocity is equal and opposite to the work required to return the system to its initial state $W(A \rightarrow B) = -W(B \rightarrow A)$. This approach can, therefore, be seen as an automatic on-the-fly Boresch[50] restraint where an ad-hoc restraint is automatically generated for each of the starting configurations, and that does not influence the final free energy value.

In a recent paper, Michel et al.[10] have shown that applying an orientational restraint on the ligand during bidirectional FEP[24,25] calculations delivers more accurate binding free energy estimates. This is due to the fact that the restraint keeps the ligand close to the binding pose when the VdW interactions are almost fully decoupled, preventing it from exploring less relevant poses during the alchemical transformation. To factor out the role of the introduced harmonic restraint between the protein and the ligand, we correct our free energy values with the equation suggested by Gilson et al.[12] (Equation 1) that for our setup returns $\Delta G_{\text{vol}} = -3.5$ kcal/mol.

$$
\begin{aligned}
\Delta G_{\text{vol}} &= RT \ln \left(\frac{V_{\text{r}}}{V_0} \right) \\
V_{\text{r}} &= \left(\frac{2\pi RT}{K} \right)^{3/2} \\
V_0 &= 1.661 \text{ nm}^3 \\
RT &= 0.593 \text{ kcal/mol} \\
K &= \text{harmonic constant}
\end{aligned}
\tag{1}
$$

Alchemical transformations

In all alchemical transformations we switch on/off both the intermolecular and intramolecular VdW and coulombic interactions,[51] at the same time in a linear fashion. As we are using a bidirectional approach, in one direction we mimic an unbinding process and in the other one a binding process. In the former case, we annihilate the ligand from the protein pocket (bound

annihilation) and we create the ligand in a box of equilibrated water (unbound creation).

$$PL_{\text{solvated}} \xrightarrow{\Delta G_{\text{bound annihilation}}} P_{\text{solvated}} + L_{\text{gas}}$$

$$L_{gas} \xrightarrow{\Delta G_{\text{unbound creation}}} L_{\text{solvated}}$$

P = protein, L = ligand, PL = protein-ligand complex.

In the latter case, we create the ligand in the APO protein pocket (bound creation) and we annihilate the ligand from a box of water (unbound annihilation). The strategy that we use to fit the vacuum phase ligand in the pocket of the APO protein is akin to the one adopted in Ref.[23] A randomly chosen frame of the ligand in vacuum HREM, and a randomly chosen frame of the APO protein HREM are combined. To orient the ligand properly in the APO structure, we superpose it to a randomly chosen frame from the protein-ligand HREM. We generate this way 200 starting conformations in total.

$$P_{\text{solvated}} + L_{\text{gas}} \xrightarrow{\Delta G_{\text{bound creation}}} PL_{\text{solvated}}$$

$$L_{\text{solvated}} \xrightarrow{\Delta G_{\text{unbound annihilation}}} L_{\text{gas}}$$

For the unbound case we execute 400 independent alchemical transformations of 0.5ns each for each direction. For the bound case instead, we perform 200 transformations of 8ns in each direction.

The rationale behind the length of the bound transformations lies on the necessity of bidirectional calculations to have a sufficient superposition between forward and backward work distributions. Transformations 4ns long did not display a sufficient superposition for ligands 1 and 11 (especially in the case of the docked poses), and in the case of ligand 8 the creation work distribution is a bimodal process and therefore tends to need longer transformations to converge even in the presence of work superposition[14] (see SI Figures 2, 3, 7, and 8).

Convolution of the probability distributions

The unbinding free energy is calculated from the convolution of the bound and unbound works obtained from our separate and independent alchemical transformations, improving the statistics with respect to the direct sum of the separate creation/annihilation free energy contributions, as discussed in Ref.[37] For comparison, the results without the application of the convolution are shown in the SI Table 1.

Free Energy Estimates

To obtain the free energy from the alchemical work values, we estimate the dissociation free energy by using Crooks' theorem[47] (Equation 2) through Bennett's acceptance ratio,[48] as implemented in Pymbar.[52]

$$\frac{P_{A\to B}(W)}{P_{B\to A}(-W)} = e^{\beta(W-\Delta G)} \tag{2}$$

The 95% confidence interval (CI) (CI = 1.96 * STD) is calculated via bootstrapping (10000 iterations with replacement). In the bootstrapping procedure, at every iteration we sample with replacement the alchemical work values from the bound and unbound distributions and perform the convolution. Afterwards, we calculate the binding free energy at each iteration in order to estimate the CI from the distribution of free energy values.

System setup and docking

The crystal structures of the protein ligand complexes, the two modelled structures of ligand 11, and the structure of the APO protein (PDB id 2OSS) are taken from Ref.[25] The docked poses were generated with Autodock v4.2.6,[53,54] using the protein structures extracted from the crystallographic complex. The parameters used for the docking were the default ones: the grid is of 40x40x40 points with a spacing of 0.0375nm, the search algorithm used was the "Genetic Algorithm" with randomised ligand configurations. For all ligands except for

Ligand 1 the top scoring pose was selected. For Ligand 1, the largest and one of the most flexible ligands, we chose a pose among the top scoring ones that was rather dissimilar from the crystallographic one (RMSD = 0.7 nm). (see Figure 3).

The parametrization is done with the Amber99SB-ILDN force field[55] for the protein and TIP3P for water.[56] For ligands, we use GAFF v2.0,[57] AM1-BCC for charges[58] and we re-parametrize the dihedrals of the ligands with the ANI-2.X machine learning forcefield[59] through Playmolecule's web interface[60] (`https://www.playmolecule.com/parameterize/`).

Simulation setup and parameters

All the molecular dynamics simulations were run with Gromacs-2022[61–68] patched with Plumed-2.9.[69–71] The input files can be found in the supporting material. For the equilibration of the complexes, an energy minimization with steepest descent was followed by 10ns of NVT and 10ns of NPT, for the ligand in water after the energy minimization 10ns of NVT and 10ns of NPT were done, and for the ligand in vacuum after the energy minimization 10ns of NVT were done. All the calculations were performed at standard temperature and pressure.

Pre-processing, post-processing, and data analysis

Most of the pre-processing, post-processing, and data analysis was done with HPC_Drug (V. 2.0). The topology files for the HREM runs were created with a modified Plumed[69–71] partial_tempering script[39] (see the supporting material section).

Results and Discussion

As can be seen in Figure 4 and SI Figure 1, the HREM simulations starting from the crystal structures and from the docked poses converged to mostly equivalent conformational distributions, highlighting that our approach is able to recover the correct binding poses

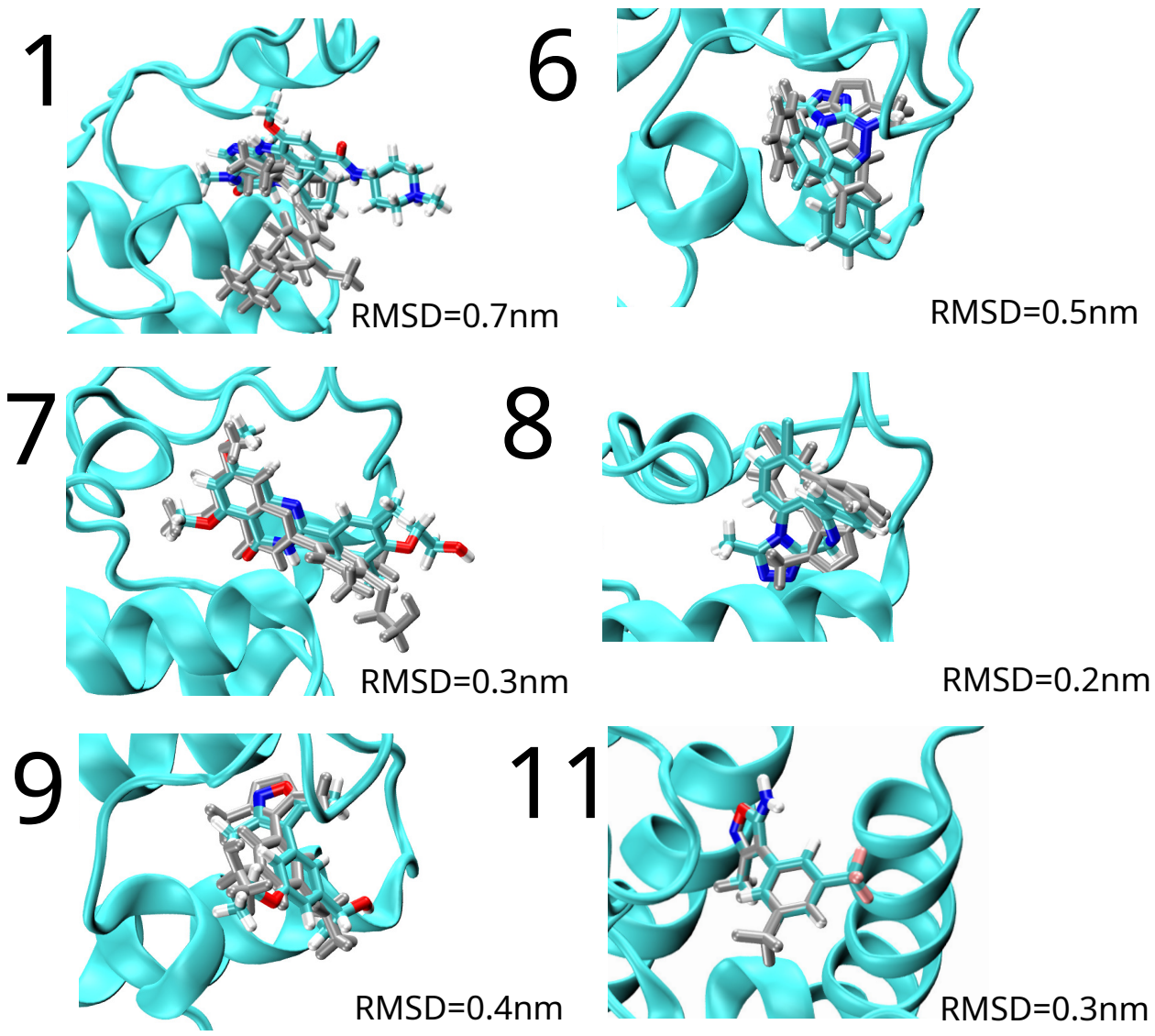

Figure 3: The crystallographic (coloured) and the docked (silver) poses of the 6 ligands. The only exception is ligand 11, for which no crystallographic poses are available and we have used the two modelled poses from Ref.[25] The reported RMSDs refer to the heavy atoms of the ligand in the docked position with respect to the crystallographic position.

even when starting from a sub-optimal docking pose. The case of ligand 11 is one of the most remarkable. It reflects a typical real application where there is no crystallographic information on the binding pose. The two independent HREM simulations starting from different poses[25] converge to comparable distributions in which a third pose is discovered and dominates the landscape.

Table 1: The free energy results in kcal/mol for the crystallographic pose, the docked pose, and the experimental value.

Ligand	ΔGcryst	ΔGdock	ΔGexp	PDB	Reference
1	8.0 +- 0.8	7.3 +- 0.7	9.8 +- 0.1	4OGI	[72]
6	9.3 +- 1.1	9.8 +- 1.1	8.2 +- 0.1	3U5L	[73]
7	3.9 +- 0.7	4.5 +- 0.5	7.8 +- 0.1	4MR4	[74]
8	7.8 +- 0.7	9.1 +- 0.7	7.4 +- 0.1	3U5J	[73]
9	4.0 +- 1.0	4.1 +- 1.3	7.3 +- 0.0	3SVG	[75]
11	3.7 +- 1.2	3.0 +- 1.2	5.6	Model	[76]

As shown in Fig. 5 and Table 1, the free energies obtained starting from the crystallographic or the docked poses are very close to the experimental values with mean unsigned errors (MUE) of 2.1 +- 0.4 kcal/mol for the calculations starting from the crystallographic poses and 2.5 +- 0.4 kcal/mol for those starting from the docked poses. They also agree well with previous alchemical calculations (SI Figure 9).

Most importantly, the correlation between the results obtained starting from the crystallographic and the docked poses shown in Figure 6 is outstanding (Pearson r = 0.96 +- 0.07). This highlights how our approach is able to address the issue due to starting from sub-optimal binding poses. In particular, we also obtain a good correlation for ligand 11 where both starting poses are modeled and there is no experimental information on the crystallographic binding pose.

In four out of six cases (ligands 1, 7, 9 and 11), the computed free energies underestimate the experimental values. This was previously reported in the literature in Ref.[77] where the authors suggested that the discrepancy with experiment was due to the sampling of the opening and closing of the ZA-loop of the protein coupled with the use of TIP3P waters (see

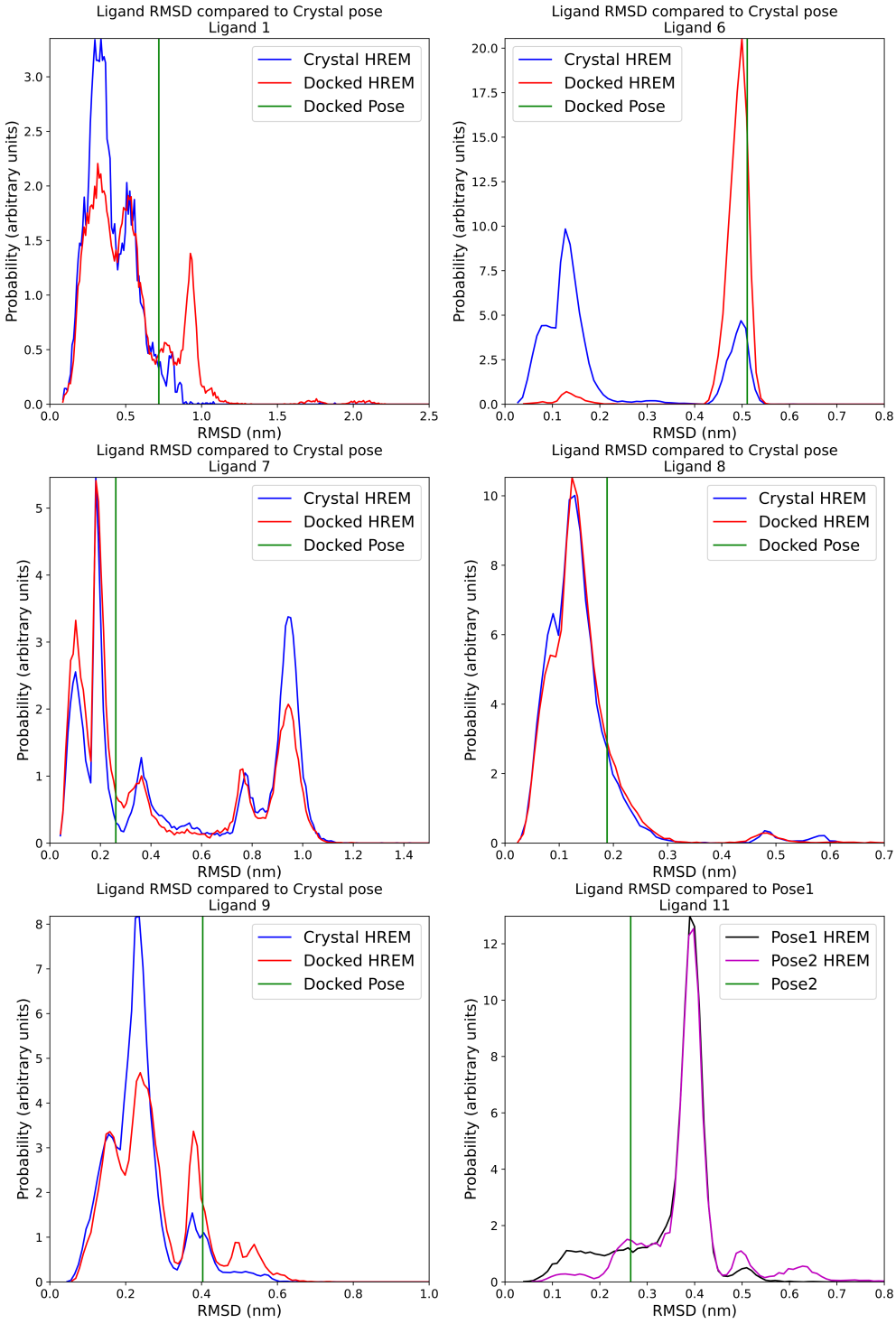

Figure 4: The probability distributions of the RMSD compared to the crystallographic pose during the HREMs starting both from the crystallographic and the docked poses. The RMSD of the docked pose is shown as a vertical green line for reference. As specified in the methods section, for ligands 6 and 9 longer equilibration times were used. For ligand 11, Pose 1 is modeled upon ligand 9 and Pose 2 is docked, both structures are taken from Ref.[25]

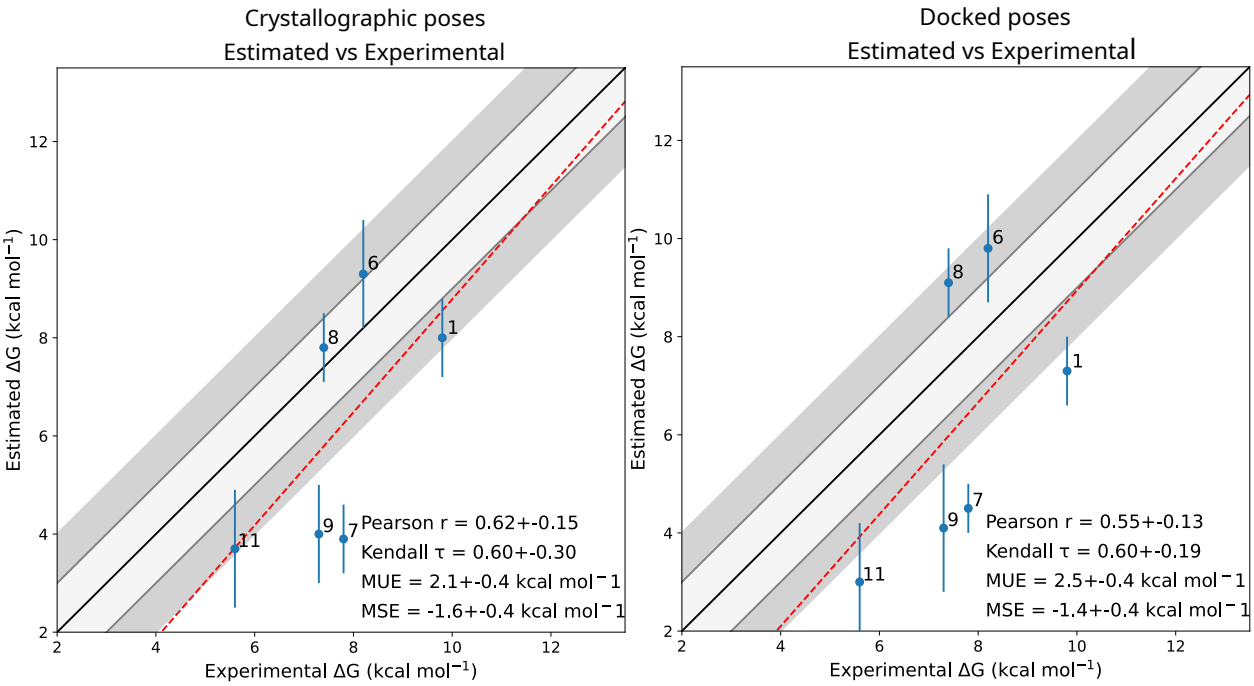

Figure 5: The binding free energies obtained starting from the crystallographic and docked poses compared to the experimental values.

the "Sampling slow degrees of freedom" section). It is clear that the quality of the force field, not only the one used for the water, but also the one used for the ligand and the protein, has a crucial effect on the results.

Challenging docking poses

Here we focus on ligands 1, 6 and 9, whose docked poses are the most distant from the crystallographic ones. The distribution of conformations of ligand 1, which is particularly large and flexible and whose pose is the most distant one, converges to the one obtained starting from the crystallographic pose in a relatively short time (Fig. 4) and the computed free energies are also in good agreement. This is likely because, both when starting from the crystallographic and docked pose, the ligand movements necessary to sample all the relevant binding poses do not require significant changes of slow degrees of freedom that are not accelerated by our HREM setup; additionally, there is no need for the ligand to exit and re-enter the binding pose. This is not the case for ligands 6 and 9. The preferred binding pose

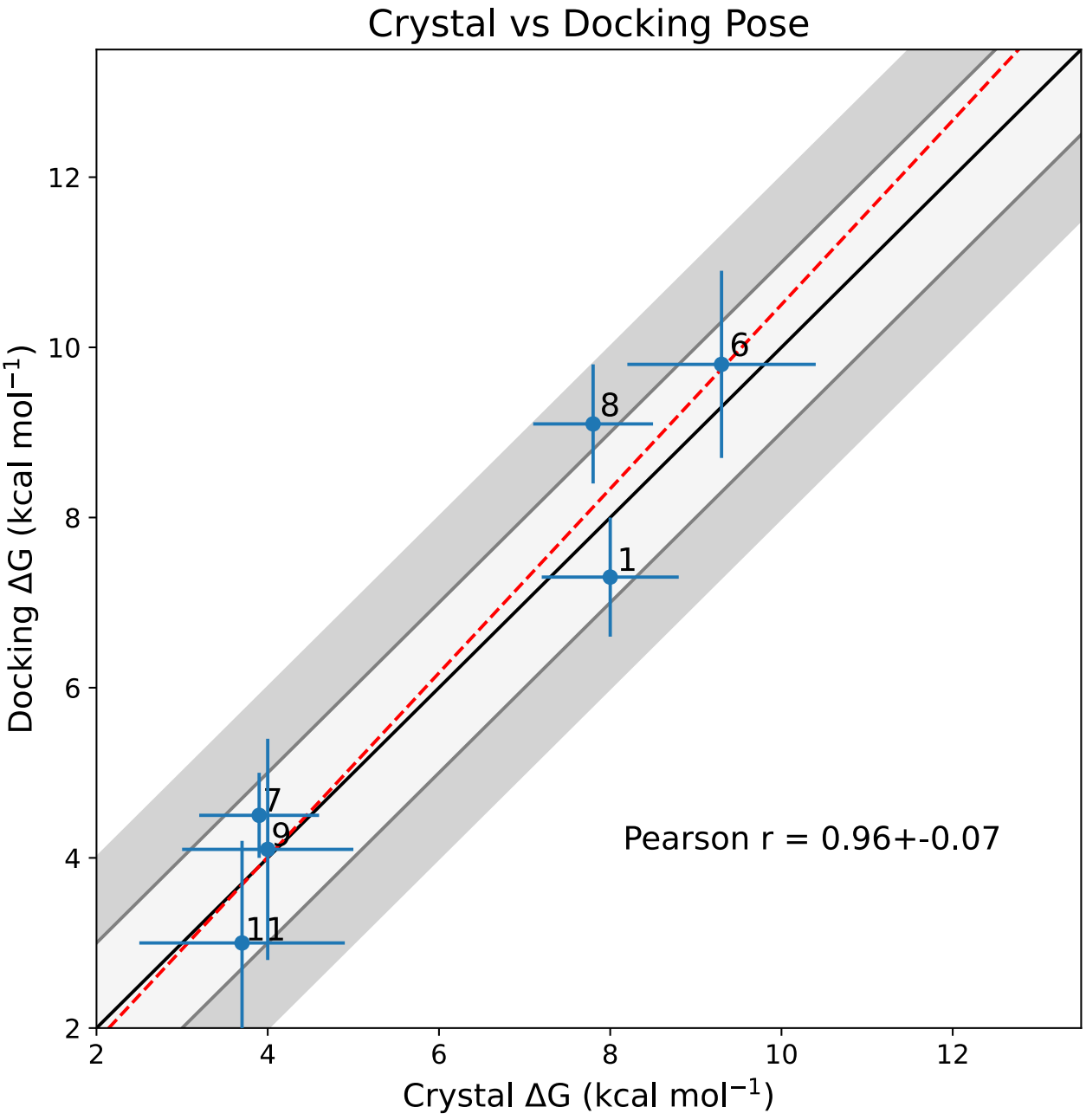

Figure 6: Comparison between the free energies obtained from the crystallographic and the docked poses.

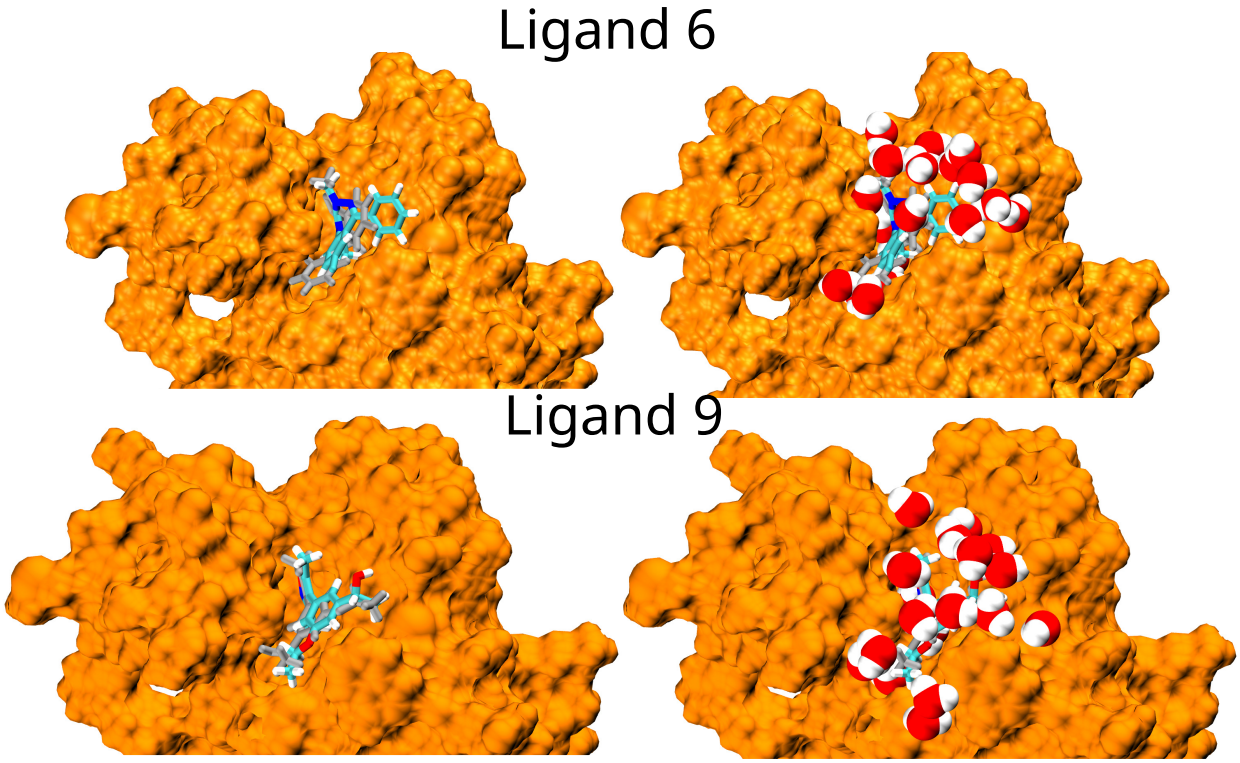

Figure 7: The crystallographic (colored) and docked (silver) poses of ligands 6 and 9. For both ligands, both the crystallographic and the docked poses are well enclosed by the protein surface. Therefore, for the ligand to go from one pose to the another, it cannot just rotate but it has to slightly exit and re-enter the pocket. Throughout the process, neighboring water molecules have to rearrange. The water molecules shown are those within a distance of 0.35nm from the ligands.

of ligand 6 appears to be different from the experimental one. In fact, the HREM simulation starting from the docked pose has converged to a stable probability distribution where the docked pose is the dominant one, while the HREM starting from the crystallographic pose is slowly but steadily shifting towards the docked pose (see Figure 4, SI Figure 1). In this docked pose, shown in Figure 7, the ligand is slightly deeper in the binding pocket and fills better the pocket than the crystal pose.

As shown in SI Figure 1, the HREM of ligand 6 starting from the crystallographic pose is the only one that does not reach a converged conformational distribution. An analysis of the HREM trajectories shows that, for the the ligand to pass from one pose to another, it is forced to slightly exit the binding pocket, rotate, and then re-enter while a number of neighboring water molecules rearrange. The rather large size of the ligand 6 and the presence of slow degrees of freedom that are not accelerated with our HREM setup such as the movement of the water molecules can explain the difficulty for HREM simulations to equilibrate. Because of this, in the case where one starts from the crystallographic pose, most replicas need to flip the pose to reach the equilibrium Boltzmann distribution and the process is difficult and time consuming. On the other hand, the case starting from the seemingly more stable docked pose seems to be able to reach equilibrium faster.

An analogous line of reasoning can be applied to ligand 9 where the crystallographic and docked poses are rotated respect to each other, but in this case ligand 9 is less bulky than ligand 6 and can flip in an easier way. Therefore, 300ns of HREM equilibration are enough to reasonably converge to equivalent conformational probability distributions starting from both binding initial poses.

These examples showcase one of the well known limits of molecular dynamics sampling: the presence of unknown slow degrees of freedom. HREM is helpful in alleviating this problem by enhancing the sampling of a predefined set of known physical elements such as torsional angles, but it is not built to pinpoint and accelerate each one of these degrees of freedom. Their comprehensive identification often requires extensive analysis and the use of

advanced techniques[78–81] while CV-based techniques are being developed to offer a solution to the sampling problem.[82–85] Notwithstanding its limitations, the fact that our method produces comparable and correlated results starting from largely different binding poses for all ligands, including the problematic ligand 6, is testament to the fact that it does manage to effectively accelerate relevant slow degrees of freedom.

Sampling slow degrees of freedom

A degree of freedom that plays a relevant role in our target is the opening of the ZA-loop between the HOLO and APO structures.[77,86] We estimate its opening and closing by evaluating the RMSD of residue ASP88. The HOLO protein tends to be in its closed state while the APO protein tends to be in its open state. In such a situation, bidirectional transformations are more robust to conformational changes between the APO and HOLO protein.[23] To verify this, we compute monodirectional free energies on SI Table 2 and observe that they are indeed systematically shifted towards higher values compared to the bidirectional ones. Nevertheless, also the convergence of bidirectional transformations can be negatively affected by complex APO/HOLO conformational changes and the presence of multiple water configurations.[14]

In this work, HREM simulations on both on the APO and HOLO structures allow to improve the situation by sampling rare fluctuations that are crucial to recover good quality free energy values.[87] In fact, as can be seen in SI Figures 10, 11, 12, we are able to visit the rare cases where the ligand gets annihilated from an open (APO-like) protein, and created in the closed (HOLO-like) protein. This is especially noticeable in SI Figure 12 where one can observe the RMSD of the ASP88 residue during the 200 non-equilibrium alchemical transformations. While for both the annihilation and the creation transformations the mean RMSD value is constant, it presents a rather large variance throughout the whole transformations, indicating that a variety of configurations is visited and a number of paths are taken.

As a consequence, our approach can better sample the tails of the forward and backward probability distributions of the alchemical works, allowing to recover less biased and more robustly converged results.[87] The use of HREM, indirectly, also allows us to sample different water configurations around the ligands, in analogy with the observations for the ZA-loop motion, as shown in SI Figures 13, 14.

Therefore the free energy values obtained with our method are expected to be less influenced by the convergence problems showcased in Ref.[14] than results extracted with more standard approaches.

Limits of applicability

Two aspects that may limit the application of our approach and affect its accuracy emerge from the results presented. First, the quality of the ligand and protein force field as well as the choice of the right protomer are crucial. The HREM step extensively explores the ligand binding pose space. Thus, if the force field favours a pose that is very different from the experimental one, the successively calculated free energy cannot be expected to agree with the experimental one, as seen in the case of ligand 6.

The second limitation involves the slow degrees of freedom that are not explicitly accelerated in the HREM setup. This can lead to a sub-optimal exploration of the binding pose space, especially in the case of large and flexible ligands. The HREM phase requires more time to converge the distributions as the ligand complexity and flexibility increases, potentially making the technique expensive. The case of ligand 6 exemplifies this scenario, where a transition between poses was achieved during an extensive HREM phase in which the ligand had first to partially exit the binding pocket, consequently displacing surrounding water molecules.

The fact that we did not observe a direct correlation between the error in the calculated free energy and the ligand size or the RMSD of the docked pose in our test set is an indication of the effectiveness of the HREM phase in sampling ligand binding poses. However, we would

not recommend using our method blindly especially in the case of larger and more flexible ligands where the exploration of multiple and/or vastly differing binding poses can prove essential. In this case, we would suggest to carefully inspect the binding poses and check the convergence of the HREM exploration.

Conclusions

In this work we have presented an approach combining HREM with non-equilibrium alchemical transformations which is able to provide good estimates of absolute ligand binding free energies, regardless of the quality of the initial binding pose. Decoupling the computed free energy from the starting pose is important for real-world drug discovery campaigns where the protein ligand binding poses are often obtained by docking to crystallographic poses, and increasingly to predicted protein structures.[88–91]

Given that the starting pose is typically kept fixed with tight restrains in most alchemical approaches, the uncertainty on the starting pose is a problem. One possibility is to repeat the free energy calculations on various poses obtained by docking and use the average of those free energy estimates (e.g. ligand 11 in Refs.[22,25]). For this approach to be successful, at least one of the selected poses must belong to the lowest energy basin in the binding site, while averaging between results obtained with different poses is only meaningful if these poses have similar free energy values, as the probability of encountering a conformation decays exponentially with its energy.

Our approach allows to directly use sub-optimal poses in alchemical calculations, as the extensive conformation pre-sampling with HREM would explore the relevant binding poses. Obtaining a set of initial conformations this way takes into account the highly dynamic nature of binding, eliminating the dependence of the result on the choice of the initial binding pose.

The algorithm proposed in this paper can be directly applied to protein targets of interest with new ligands in the absence of crystallographic data, and thanks to the automation

provided by our HPC_Drug package can be easily integrated into computational pipelines, making it a valuable tool for drug discovery.

Data and Software Availability

Supporting informations: additional computational details, free energy results, data analysis, and figures. HPC_Drug is available at:

`https://github.com/MauriceKarrenbrock/HPC_Drug`

The Gromacs input files used, the docked structures, and the modified versions of the partial tempering script are available on Zenodo at:

`https://doi.org/10.5281/zenodo.8377371`

Acknowledgements

We acknowledge PRACE and the Swiss National Supercomputing Centre (CSCS) for large supercomputer time allocations on Piz Daint, project IDs: pr126, s1107, s1169, s1228. FLG acknowledges the Swiss National Science Foundation and Bridge for financial support (projects number: 200021_204795, CRSII5_216587 and 40B2-0_203628). PP acknowledges the National Recovery and Resilience Plan, Mission 4 Component 2 - Investment 1.4

References

(1) McCammon, J. A.; Roux, B.; Voth, G.; Yang, W. Special Issue on Free Energy. *Journal of Chemical Theory and Computation* **2014**, *10*, 2631–2631.

(2) Chodera, J. D.; Mobley, D. L.; Shirts, M. R.; Dixon, R. W.; Branson, K.; Pande, V. S. Alchemical free energy methods for drug discovery: progress and challenges. *Current Opinion in Structural Biology* **2011**, *21*, 150–160.

(3) Laio, A.; Parrinello, M. Escaping free-energy minima. *Proceedings of the National Academy of Sciences* **2002**, *99*, 12562–12566.

(4) Cavalli, A.; Spitaleri, A.; Saladino, G.; Gervasio, F. L. Investigating DrugTarget Association and Dissociation Mechanisms Using Metadynamics-Based Algorithms. *Accounts of Chemical Research* **2015**, *48*, 277–285, PMID: 25496113.

(5) Pohorille, A.; Jarzynski, C.; Chipot, C. Good Practices in Free-Energy Calculations. *The Journal of Physical Chemistry B* **2010**, *114*, 10235–10253, PMID: 20701361.

(6) Cournia, Z.; Allen, B. K.; Beuming, T.; Pearlman, D. A.; Radak, B. K.; Sherman, W. Rigorous Free Energy Simulations in Virtual Screening. *Journal of Chemical Information and Modeling* **2020**, *60*, 4153–4169, PMID: 32539386.

(7) Wu, J. Z.; Azimi, S.; Khuttan, S.; Deng, N.; Gallicchio, E. Alchemical Transfer Approach to Absolute Binding Free Energy Estimation. *Journal of Chemical Theory and Computation* **2021**, *17*, 3309–3319, PMID: 33983730.

(8) Cournia, Z.; Chipot, C.; Roux, B.; York, D. M.; Sherman, W. *Free Energy Methods in Drug Discovery—Introduction*; ACS Symposium Series; American Chemical Society, 2021; Vol. 1397; pp 1–38, 0.

(9) Kuhn, M.; Firth-Clark, S.; Tosco, P.; Mey, A. S. J. S.; Mackey, M.; Michel, J. Assessment of Binding Affinity via Alchemical Free-Energy Calculations. *Journal of Chemical Information and Modeling* **2020**, *60*, 3120–3130, PMID: 32437145.

(10) Clark, F.; Robb, G.; Cole, D. J.; Michel, J. Comparison of ReceptorLigand Restraint Schemes for Alchemical Absolute Binding Free Energy Calculations. *Journal of Chemical Theory and Computation* **2023**, *19*, 3686–3704, PMID: 37285579.

(11) Rizzi, A.; Grinaway, P.; Parton, D.; Shirts, M.; Wang, K.; Eastman, P.; Friedrichs, M.; Pande, V.; Branson, K.; Mobley, D.; Chodera, J. YANK: a GPU-accelerated platform

for alchemical free energy calculations. `http://getyank.org/latest/references.html`, 2020.

(12) Gilson, M.; Given, J.; Bush, B.; McCammon, J. The statistical-thermodynamic basis for computation of binding affinities: a critical review. *Biophysical Journal* **1997**, *72*, 1047–1069.

(13) Chen, W.; Cui, D.; Jerome, S. V.; Michino, M.; Lenselink, E. B.; Huggins, D. J.; Beautrait, A.; Vendome, J.; Abel, R.; Friesner, R. A.; Wang, L. Enhancing Hit Discovery in Virtual Screening through Absolute ProteinLigand Binding Free-Energy Calculations. *Journal of Chemical Information and Modeling* **2023**, *63*, 3171–3185, PMID: 37167486.

(14) Baumann, H. M.; Gapsys, V.; de Groot, B. L.; Mobley, D. L. Challenges Encountered Applying Equilibrium and Nonequilibrium Binding Free Energy Calculations. *The Journal of Physical Chemistry B* **2021**, *125*, 4241–4261, PMID: 33905257.

(15) Pal, R. K.; Gallicchio, E. Perturbation potentials to overcome order/disorder transitions in alchemical binding free energy calculations. *The Journal of Chemical Physics* **2019**, *151*, 124116.

(16) Gervasio, F. L.; Laio, A.; Parrinello, M. Flexible Docking in Solution Using Metadynamics. *Journal of the American Chemical Society* **2005**, *127*, 2600–2607, PMID: 15725015.

(17) Evans, R.; Hovan, L.; Tribello, G. A.; Cossins, B. P.; Estarellas, C.; Gervasio, F. L. Combining Machine Learning and Enhanced Sampling Techniques for Efficient and Accurate Calculation of Absolute Binding Free Energies. *Journal of Chemical Theory and Computation* **2020**, *16*, 4641–4654, PMID: 32427471.

(18) Ansari, N.; Rizzi, V.; Carloni, P.; Parrinello, M. Water-Triggered, Irreversible Conformational Change of SARS-CoV-2 Main Protease on Passing from the Solid State to Aqueous Solution. *Journal of the American Chemical Society* **2021**, *143*, 12930–12934.

(19) Lukauskis, D.; Samways, M. L.; Aureli, S.; Cossins, B. P.; Taylor, R. D.; Gervasio, F. L. Open Binding Pose Metadynamics: An Effective Approach for the Ranking of ProteinLigand Binding Poses. *Journal of Chemical Information and Modeling* **2022**, *62*, 6209–6216, PMID: 36401553.

(20) Rizzi, A.; Murkli, S.; McNeill, J. N.; Yao, W.; Sullivan, M.; Gilson, M. K.; Chiu, M. W.; Isaacs, L.; Gibb, B. C.; Mobley, D. L.; Chodera, J. D. Overview of the SAMPL6 host–guest binding affinity prediction challenge. *Journal of Computer-Aided Molecular Design* **2018**, *32*, 937–963.

(21) Rizzi, A.; Jensen, T.; Slochower, D. R.; Aldeghi, M.; Gapsys, V.; Ntekoumes, D.; Bosisio, S.; Papadourakis, M.; Henriksen, N. M.; de Groot, B. L.; Cournia, Z.; Dickson, A.; Michel, J.; Gilson, M. K.; Shirts, M. R.; Mobley, D. L.; Chodera, J. D. The SAMPL6 SAMPLing challenge: assessing the reliability and efficiency of binding free energy calculations. *Journal of Computer-Aided Molecular Design* **2020**, *34*, 601–633.

(22) Gapsys, V.; Yildirim, A.; Aldeghi, M.; Khalak, Y.; van der Spoel, D.; de Groot, B. L. Accurate absolute free energies for ligand–protein binding based on non-equilibrium approaches. *Communications Chemistry* **2021**, *4*, 61.

(23) Khalak, Y.; Tresadern, G.; Aldeghi, M.; Baumann, H. M.; Mobley, D. L.; de Groot, B. L.; Gapsys, V. Alchemical absolute proteinligand binding free energies for drug design. *Chemical Science* **2021**, *12*, 13958–13971.

(24) Zwanzig, R. W. HighTemperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *The Journal of Chemical Physics* **1954**, *22*, 1420–1426.

(25) Aldeghi, M.; Heifetz, A.; Bodkin, M. J.; Knapp, S.; Biggin, P. C. Accurate calculation of the absolute free energy of binding for drug molecules. *Chemical Science* **2016**, *7*, 207–218.

(26) Laio, A.; Gervasio, F. L. Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. *Reports on Progress in Physics* **2008**, *71*, 126601.

(27) Masetti, M.; Cavalli, A.; Recanatini, M.; Gervasio, F. L. Exploring Complex Protein-Ligand Recognition Mechanisms with Coarse Metadynamics. *The journal of physical chemistry B* **2009**, *113*, 4807–4816.

(28) 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? *Journal of the American Chemical Society* **2011**, *133*, 9181–9183, PMID: 21545110.

(29) Henriksen, N. M.; Fenley, A. T.; Gilson, M. K. Computational Calorimetry: High-Precision Calculation of HostGuest Binding Thermodynamics. *Journal of Chemical Theory and Computation* **2015**, *11*, 4377–4394, PMID: 26523125.

(30) Heinzelmann, G.; Gilson, M. K. Automation of absolute protein-ligand binding free energy calculations for docking refinement and compound evaluation. *Scientific Reports* **2021**, *11*, 1116.

(31) Procacci, P. Methodological uncertainties in drug-receptor binding free energy predictions based on classical molecular dynamics. *Current Opinion in Structural Biology* **2021**, *67*, 127–134, Theory and Simulation/Computational Methods Macromolecular Assemblies.

(32) Pagadala, N. S.; Syed, K.; Tuszynski, J. Software for molecular docking: a review. *Biophysical Reviews* **2017**, *9*, 91–102.

(33) Cross, J. B.; Thompson, D. C.; Rai, B. K.; Baber, J. C.; Fan, K. Y.; Hu, Y.; Humblet, C. Comparison of Several Molecular Docking Programs: Pose Prediction and Virtual Screening Accuracy. *Journal of Chemical Information and Modeling* **2009**, *49*, 1455–1474, PMID: 19476350.

(34) Guo, J.; Janet, J. P.; Bauer, M. R.; Nittinger, E.; Giblin, K. A.; Papadopoulos, K.; Voronov, A.; Patronov, A.; Engkvist, O.; Margreitter, C. DockStream: a docking wrapper to enhance de novo molecular design. *Journal of Cheminformatics* **2021**, *13*, 89.

(35) Moore, J. H.; Margreitter, C.; Janet, J. P.; Engkvist, O.; de Groot, B. L.; Gapsys, V. Automated relative binding free energy calculations from SMILES to $\Delta\Delta$G. *Communications Chemistry* **2023**, *6*, 82.

(36) Cappel, D.; Jerome, S.; Hessler, G.; Matter, H. Impact of Different Automated Binding Pose Generation Approaches on Relative Binding Free Energy Simulations. *Journal of Chemical Information and Modeling* **2020**, *60*, 1432–1444, PMID: 31986249.

(37) Macchiagodena, M.; Pagliai, M.; Karrenbrock, M.; Guarnieri, G.; Iannone, F.; Procacci, P. Virtual Double-System Single-Box: A Nonequilibrium Alchemical Technique for Absolute Binding Free Energy Calculations: Application to Ligands of the SARS-CoV-2 Main Protease. *Journal of Chemical Theory and Computation* **2020**, *16*, 7160–7172, PMID: 33090785.

(38) Macchiagodena, M.; Karrenbrock, M.; Pagliai, M.; Procacci, P. Virtual Double-System Single-Box for Absolute Dissociation Free Energy Calculations in GROMACS. *Journal of Chemical Information and Modeling* **2021**, *61*, 5320–5326, PMID: 34723516.

(39) Bussi, G. Hamiltonian replica exchange in GROMACS: a flexible implementation. *Molecular Physics* **2014**, *112*, 379–384.

(40) Marsili, S.; Signorini, G. F.; Chelli, R.; Marchi, M.; Procacci, P. ORAC: A molecular dynamics simulation program to explore free energy surfaces in biomolecular systems at the atomistic level. *Journal of Computational Chemistry* **2010**, *31*, 1106–1116.

(41) Fukunishi, H.; Watanabe, O.; Takada, S. On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction. *The Journal of Chemical Physics* **2002**, *116*, 9058–9067.

(42) Procacci, P.; Macchiagodena, M.; Pagliai, M.; Guarnieri, G.; Iannone, F. Interaction of hydroxychloroquine with SARS-CoV2 functional proteins using all-atoms non-equilibrium alchemical simulations. *Chemical Communications* **2020**, *56*, 8854–8856.

(43) Procacci, P.; Cardelli, C. Fast Switching Alchemical Transformations in Molecular Dynamics Simulations. *Journal of Chemical Theory and Computation* **2014**, *10*, 2813–2823, PMID: 26586508.

(44) Sandberg, R. B.; Banchelli, M.; Guardiani, C.; Menichetti, S.; Caminati, G.; Procacci, P. Efficient Nonequilibrium Method for Binding Free Energy Calculations in Molecular Dynamics Simulations. *Journal of Chemical Theory and Computation* **2015**, *11*, 423–435, PMID: 26580905.

(45) Procacci, P. I. Dissociation free energies of drugreceptor systems via non-equilibrium alchemical simulations: a theoretical framework. *Physical Chemistry Chemical Physics* **2016**, *18*, 14991–15004.

(46) Nerattini, F.; Chelli, R.; Procacci, P. II. Dissociation free energies in drugreceptor systems via nonequilibrium alchemical simulations: application to the FK506-related immunophilin ligands. *Physical Chemistry Chemical Physics* **2016**, *18*, 15005–15018.

(47) Crooks, G. E. Nonequilibrium Measurements of Free Energy Differences for Microscopically Reversible Markovian Systems. *Journal of Statistical Physics* **1998**, *90*, 1481–1487.

(48) Crooks, G. E. Path-ensemble averages in systems driven far from equilibrium. *Physical Review E* **2000**, *61*, 2361–2366.

(49) Nicolini, P.; Chelli, R. Improving fast-switching free energy estimates by dynamical freezing. *Physical Review E* **2009**, *80*, 041124.

(50) Boresch, S.; Tettinger, F.; Leitgeb, M.; Karplus, M. Absolute Binding Free Energies: A Quantitative Approach for Their Calculation. *The Journal of Physical Chemistry B* **2003**, *107*, 9535–9551.

(51) Vakali, V.; Papadourakis, M.; Georgiou, N.; Zoupanou, N.; Diamantis, D. A.; Javornik, U.; Papakyriakopoulou, P.; Plavec, J.; Valsami, G.; Tzakos, A. G.; Tzeli, D.; Cournia, Z.; Mauromoustakos, T. Comparative Interaction Studies of Quercetin with 2-Hydroxyl-propyl-β-cyclodextrin and 2,6-Methylated-β-cyclodextrin. *Molecules* **2022**, *27*, 5490.

(52) Shirts, M. R.; Chodera, J. D. Statistically optimal analysis of samples from multiple equilibrium states. *The Journal of Chemical Physics* **2008**, *129*, 124105.

(53) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *Journal of Computational Chemistry* **1998**, *19*, 1639–1662.

(54) Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry* **2009**, *30*, 2785–2791.

(55) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Structure, Function, and Bioinformatics* **2010**, *78*, 1950–1958.

(56) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **1983**, *79*, 926–935.

(57) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *Journal of Computational Chemistry* **2004**, *25*, 1157–1174.

(58) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *Journal of Computational Chemistry* **2002**, *23*, 1623–1641.

(59) Devereux, C.; Smith, J. S.; Huddleston, K. K.; Barros, K.; Zubatyuk, R.; Isayev, O.; Roitberg, A. E. Extending the Applicability of the ANI Deep Learning Molecular Potential to Sulfur and Halogens. *Journal of Chemical Theory and Computation* **2020**, *16*, 4192–4202, PMID: 32543858.

(60) Galvelis, R.; Doerr, S.; Damas, J. M.; Harvey, M. J.; De Fabritiis, G. A Scalable Molecular Force Field Parameterization Method Based on Density Functional Theory and Quantum-Level Machine Learning. *Journal of Chemical Information and Modeling* **2019**, *59*, 3485–3493, PMID: 31322877.

(61) Pll, S.; Zhmurov, A.; Bauer, P.; Abraham, M.; Lundborg, M.; Gray, A.; Hess, B.; Lindahl, E. Heterogeneous parallelization and acceleration of molecular dynamics simulations in GROMACS. *The Journal of Chemical Physics* **2020**, *153*, 134110.

(62) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1*, 19–25.

(63) Páll, S.; Abraham, M. J.; Kutzner, C.; Hess, B.; Lindahl, E. Tackling Exascale Software Challenges in Molecular Dynamics Simulations with GROMACS. Solving Software Challenges for Exascale. Cham, 2015; pp 3–27.

(64) Pronk, S.; Pll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **2013**, *29*, 845–854.

(65) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *Journal of Chemical Theory and Computation* **2008**, *4*, 435–447, PMID: 26620784.

(66) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, flexible, and free. *Journal of Computational Chemistry* **2005**, *26*, 1701–1718.

(67) Lindahl, E.; Hess, B.; van der Spoel, D. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *Molecular modeling annual* **2001**, *7*, 306–317.

(68) Berendsen, H.; van der Spoel, D.; van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications* **1995**, *91*, 43–56.

(69) Bonomi, M.; Bussi, G.; Camilloni, C.; Tribello, G. A.; Banáš, P.; Barducci, A.; Bernetti, M.; Bolhuis, P. G.; Bottaro, S.; Branduardi, D.; Capelli, R.; Carloni, P.; Ceriotti, M.; Cesari, A.; Chen, H.; Chen, W.; Colizzi, F.; De, S.; De La Pierre, M.; Donadio, D.; Drobot, V.; Ensing, B.; Ferguson, A. L.; Filizola, M.; Fraser, J. S.; Fu, H.; Gasparotto, P.; Gervasio, F. L.; Giberti, F.; Gil-Ley, A.; Giorgino, T.; Heller, G. T.; Hocky, G. M.; Iannuzzi, M.; Invernizzi, M.; Jelfs, K. E.; Jussupow, A.; Kirilin, E.; Laio, A.; Limongelli, V.; Lindorff-Larsen, K.; Löhr, T.; Marinelli, F.; Martin-Samos, L.; Masetti, M.; Meyer, R.; Michaelides, A.; Molteni, C.; Morishita, T.; Nava, M.; Paissoni, C.; Papaleo, E.; Parrinello, M.; Pfaendtner, J.; Piaggi, P.; Piccini, G.; Pietropaolo, A.; Pietrucci, F.; Pipolo, S.; Provasi, D.; Quigley, D.; Raiteri, P.; Raniolo, S.; Rydzewski, J.; Salvalaglio, M.; Sosso, G. C.; Spiwok, V.; Šponer, J.; Swenson, D. W. H.; Tiwary, P.; Valsson, O.; Vendruscolo, M.; Voth, G. A.; White, A.; consortium, T. P. Promoting transparency and reproducibility in enhanced molecular simulations. *Nature Methods* **2019**, *16*, 670–673.

(70) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New feathers for an old bird. *Computer Physics Communications* **2014**, *185*, 604–613.

(71) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R. A.; Parrinello, M. PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Computer Physics Communications* **2009**, *180*, 1961–1972.

(72) Ciceri, P.; Müller, S.; O'Mahony, A.; Fedorov, O.; Filippakopoulos, P.; Hunt, J. P.; Lasater, E. A.; Pallares, G.; Picaud, S.; Wells, C.; Martin, S.; Wodicka, L. M.; Shah, N. P.; Treiber, D. K.; Knapp, S. Dual kinase-bromodomain inhibitors for rationally designed polypharmacology. *Nature Chemical Biology* **2014**, *10*, 305–312.

(73) Filippakopoulos, P.; Picaud, S.; Fedorov, O.; Keller, M.; Wrobel, M.; Morgenstern, O.; Bracher, F.; Knapp, S. Benzodiazepines and benzotriazepines as protein interaction inhibitors targeting bromodomains of the BET family. *Bioorganic & Medicinal Chemistry* **2012**, *20*, 1878–1886, The Chemistry Biology Interface.

(74) Picaud, S.; Wells, C.; Felletar, I.; Brotherton, D.; Martin, S.; Savitsky, P.; Diez-Dacal, B.; Philpott, M.; Bountra, C.; Lingard, H.; Fedorov, O.; Müller, S.; Brennan, P. E.; Knapp, S.; Filippakopoulos, P. RVX-208, an inhibitor of BET transcriptional regulators with selectivity for the second bromodomain. *Proceedings of the National Academy of Sciences* **2013**, *110*, 19754–19759.

(75) Hewings, D. S.; Fedorov, O.; Filippakopoulos, P.; Martin, S.; Picaud, S.; Tumber, A.; Wells, C.; Olcina, M. M.; Freeman, K.; Gill, A.; Ritchie, A. J.; Sheppard, D. W.; Russell, A. J.; Hammond, E. M.; Knapp, S.; Brennan, P. E.; Conway, S. J. Optimization of 3,5-Dimethylisoxazole Derivatives as Potent Bromodomain Ligands. *Journal of Medicinal Chemistry* **2013**, *56*, 3217–3227, PMID: 23517011.

(76) Vidler, L. R.; Filippakopoulos, P.; Fedorov, O.; Picaud, S.; Martin, S.; Tomsett, M.;

Woodward, H.; Brown, N.; Knapp, S.; Hoelder, S. Discovery of Novel Small-Molecule Inhibitors of BRD4 Using Structure-Based Virtual Screening. *Journal of Medicinal Chemistry* **2013**, *56*, 8073–8088, PMID: 24090311.

(77) Heinzelmann, G.; Henriksen, N. M.; Gilson, M. K. Attach-Pull-Release Calculations of Ligand Binding and Conformational Changes on the First BRD4 Bromodomain. *Journal of Chemical Theory and Computation* **2017**, *13*, 3260–3275, PMID: 28564537.

(78) Pérez-Hernández, G.; Paul, F.; Giorgino, T.; De Fabritiis, G.; Noé, F. Identification of slow molecular order parameters for Markov model construction. *The Journal of Chemical Physics* **2013**, *139*, 015102.

(79) Tiwary, P.; Berne, B. J. Spectral gap optimization of order parameters for sampling complex molecular systems. *Proceedings of the National Academy of Sciences* **2016**, *113*, 2839–2844.

(80) Wehmeyer, C.; No, F. Time-lagged autoencoders: Deep learning of slow collective variables for molecular kinetics. *The Journal of Chemical Physics* **2018**, *148*, 241703.

(81) Bonati, L.; Piccini, G.; Parrinello, M. Deep learning the slow modes for rare events sampling. *Proceedings of the National Academy of Sciences* **2021**, *118*, e2113533118.

(82) Bussi, G.; Gervasio, F. L.; Laio, A.; Parrinello, M. Free-Energy Landscape for β Hairpin Folding from Combined Parallel Tempering and Metadynamics. *Journal of the American Chemical Society* **2006**, *128*, 13435–13441.

(83) Gil-Ley, A.; Bussi, G. Enhanced Conformational Sampling Using Replica Exchange with Collective-Variable Tempering. *Journal of Chemical Theory and Computation* **2015**, *11*, 1077–1085.

(84) Rahimi, K.; Piaggi, P. M.; Zerze, G. H. Comparison of On-the-Fly Probability Enhanced Sampling and Parallel Tempering Combined with Metadynamics for Atomistic

Simulations of RNA Tetraloop Folding. *The Journal of Physical Chemistry B* **2023**, *127*, 4722–4732.

(85) Rizzi, V.; Aureli, S.; Ansari, N.; Gervasio, F. L. OneOPES, a Combined Enhanced Sampling Method to Rule Them All. *Journal of Chemical Theory and Computation* **2023**, *19*, 5731–5742, PMID: 37603295.

(86) Kuang, M.; Zhou, J.; Wang, L.; Liu, Z.; Guo, J.; Wu, R. Binding Kinetics versus Affinities in BRD4 Inhibition. *Journal of Chemical Information and Modeling* **2015**, *55*, 1926–1935, PMID: 26263125.

(87) Jarzynski, C. Rare events and the convergence of exponentially averaged work values. *Physical Review E* **2006**, *73*, 046105.

(88) Scardino, V.; Di Filippo, J. I.; Cavasotto, C. N. How good are AlphaFold models for docking-based virtual screening? *iScience* **2023**, *26*, 105920.

(89) Zhang, Y.; Vass, M.; Shi, D.; Abualrous, E.; Chambers, J. M.; Chopra, N.; Higgs, C.; Kasavajhala, K.; Li, H.; Nandekar, P.; Sato, H.; Miller, E. B.; Repasky, M. P.; Jerome, S. V. Benchmarking Refined and Unrefined AlphaFold2 Structures for Hit Discovery. *Journal of Chemical Information and Modeling* **2023**, *63*, 1656–1667, PMID: 36897766.

(90) Daz-Rovira, A. M.; Martn, H.; Beuming, T.; Daz, L.; Guallar, V.; Ray, S. S. Are Deep Learning Structural Models Sufficiently Accurate for Virtual Screening? Application of Docking Algorithms to AlphaFold2 Predicted Structures. *Journal of Chemical Information and Modeling* **2023**, *63*, 1668–1674, PMID: 36892986.

(91) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.;

Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589.

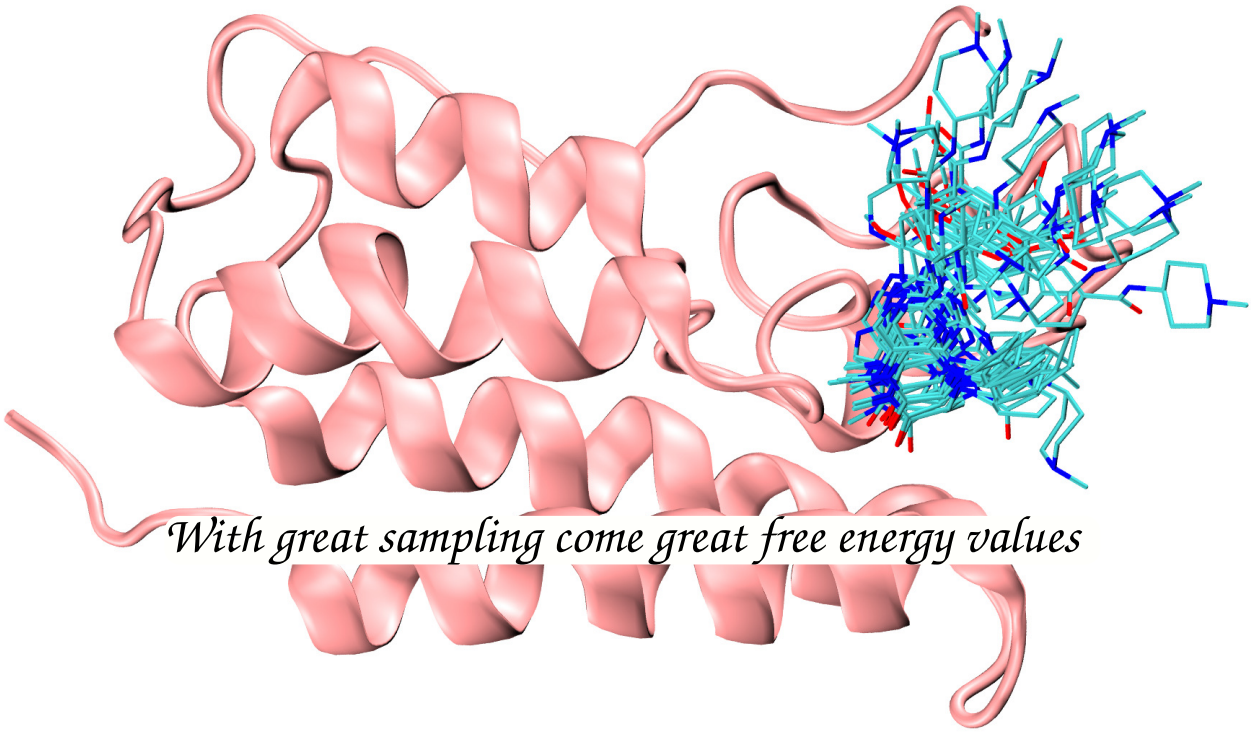

Figure 8: TOC Graphic