JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

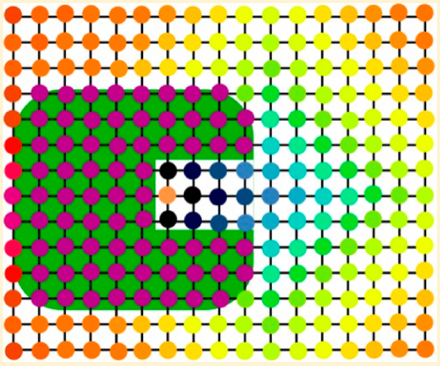
Protein−Ligand Binding from Distancefield Distances and Hamiltonian Replica Exchange Simulations

Anita de Ruiter and Chris Oostenbrink*

Institute for Molecular Modeling and Simulation, University of Natural Resources and Life Sciences (BOKU), Muthgasse 18, 1190 Vienna, Austria

Supporting Information

ABSTRACT: The calculation of protein−ligand binding free energies is an important goal in the field of computational chemistry. Applying path-sampling methods for this purpose involves calculating the associated potential of mean force (PMF) and gives insight into the binding free energy along the binding process. Without *a priori* knowledge about the binding path, sampling reversible binding can be difficult to achieve. To alleviate this problem, we introduce the distancefield (DF) as a reaction coordinate for such calculations. DF is a grid-based method in which the shortest distance between the binding site and a ligand is determined avoiding routes that pass through the protein. Combining this reaction coordinate with Hamiltonian replica exchange molecular dynamics (HREMD) allows for the reversible binding of the ligand to the protein. A comparison is made between umbrella sampling using regular distance restraints and HREMD with DF restraints to study aspirin binding to the protein phospholipase A_2. Although the free energies of binding are similar for both methods, the increased sampling with HREMD has a significant influence on the shape of the PMF. A remarkable agreement between the calculated binding free energies from the PMF and the experimental estimate is obtained.



INTRODUCTION

Molecular recognition, and in particular protein−ligand binding, plays a central role in many biological processes. Accurate calculation of the free energy associated with these binding processes is consequently of great importance in fields like drug discovery and biotechnology. In computational chemistry, the calculation of protein−ligand binding free energies remains one of the main challenges. Even with the ever-increasing computational power, the most accurate methods easily become too computationally demanding, because the most interesting biomolecular systems are often rather large. Simply simulating the protein and ligand and waiting until several binding and unbinding occasions are observed is therefore usually not feasible, although exceptions have been reported recently.[1,2] A wide range of methods has been developed over the years to tackle the problem of ligand binding computationally. On the one hand, there are methods like molecular docking that are very efficient and can be applied in a high-throughput manner. On the other hand, the more thermodynamically sound methods like thermodynamic integration[3] and free energy perturbation[4] can be expected to yield more accurate results, at the cost of a much reduced efficiency. So-called end point methods and alchemical methods seem most popular to calculate protein−ligand binding free energies. Using end point methods like linear interaction energy (LIE)[5] or molecular mechanics/Poisson−Boltzmann surface area (MM/PBSA),[6] one only simulates the ligand in the bound and free state, without any intermediate states. Obviously, this saves a lot of computational time, but also the accuracy that can be obtained is lower. Using alchemical methods, the free energy between two states is estimated via a number of discrete, possibly unphysical, intermediate states.[7,8] One can either calculate relative free energies of binding between several ligands or get an "absolute" binding free energy by using, e.g., the double decoupling method (DDM).[9] Higher accuracy can be expected from these methods than from the end point methods, but the intermediate states typically lead to a higher demand on computational power. The downside of all these methods is that they do not give any information about the kinetics of the actual binding process of the ligand. Especially for proteins with more buried binding sites, knowledge about which path the ligand takes and with which residues it interacts along this path may be of great importance for, e.g., drug design. This information can be obtained by applying methods that enhance or focus sampling along a certain path, which give rise to the free energy along this path, the potential of mean force (PMF). The well-known umbrella sampling method can be used for this purpose,[1,10−13] but several other methods have been developed over the years as well, including among others; smooth reaction path generation (SRPG),[14,15] local elevation umbrella sampling (LEUS),[16,17] metadynamics,[18−22] multidimensional adaptive biasing force (ABF),[23,24] steered molecular dynamics,[25,26] selective accelerated molecular dynamics (selective aMD),[27] and Hamiltonian replica exchange (HREMD, or REUS).[28,29] However, several

Received: November 6, 2012

   dx.doi.org/10.1021/ct300967a | *J. Chem. Theory Comput.* XXXX, XXX, XXX−XXX

methods require *a priori* knowledge of the dissociation path, which is often not available. Even in cases where the path is chosen during the simulation, assumptions are often made which might not be justified for the system under consideration. Examples include the consideration of just a single dissociation path or the assumption that the path is linear. Especially in larger systems, multiple paths might be accessible for the ligand, which may lead to different free energy profiles when properly considered. Also, only the dissociation path is sampled in most cases, while the molecular recognition that takes place during association is probably just as important and may lead to differently sampled paths. Sampling of all relevant dissociation and association paths is rather tedious. Consider, for example, a situation in which the ligand is only distance restrained to the binding site of the protein. As soon as the ligand is free in solution, it is free to move around the protein at the specified radius and can thus diffuse through a large part of the computational box. If the reference value of the distance restraint is subsequently decreased, to sample the association path, the ligand gets pulled to the surface of the protein but will be stuck there if the entrance is too far away.

In this study, a new type of reaction coordinate will be introduced, which will be referred to from now on as the distancefield (DF) distance. This is a grid-based method, where the distance between two points is in principle the sum of the grid lengths between them, while a penalty is added when one of the grid points coincides with the protein. In this way, the shortest path between the ligand and a reference point in the binding site is to go around the protein, as illustrated in Figure 1. It

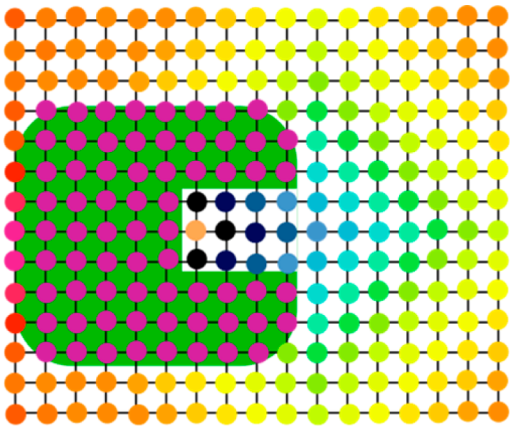


Figure 1. Schematic representation of the distancefield (DF). A protein (green) and a reference point in the binding site (orange dot) are shown on a grid. The colors of the grid points represent the DF distance, where black indicates a small distance and red a very large distance. The magenta dots represent the protein penalty.

is thus possible to find the entrance to the binding site even if the ligand diffused to the other side of the protein, without *a priori* knowledge about the association path. The distancefield reaction coordinate thus allows for the application of path sampling methods which include dissociation and association simultaneously, which is not possible with regular distance restraints. More details about DF and possible applications can be found in the Theory and Methodology section. We will combine this new reaction coordinate with Hamiltonian replica exchange molecular dynamics (HREMD),[30] which allows the ligand to associate to and dissociate from the protein reversibly. The results will be compared to those obtained with US with normal distance restraints, for which a replica exchange scheme leads to the issues outlined above. In addition, the free energy of binding is also calculated with the double decoupling method (DDM) to compare the path sampling methods with an alchemical method.

The binding of aspirin to phospholipase A_2 isoform 3 (PLA_2) will be taken as a case study. PLA_2 belongs to the secreted phospholipases and degrades phospholipids. One of the products of these reactions is arachidonic acid, which activates inflammatory reactions. The crystal structure of PLA_2 with bound aspirin was solved some years ago and confirmed that inflammatory reactions can be inhibited through PLA_2.[31] The PLA_2−aspirin complex is thus a very interesting target for rational drug design. The crystal structure offers important structural information, but more atomic details about the binding process and kinetic information may be of great value. The free energy calculations of the current study can be compared to the experimentally determined free energy of binding of −29.6 kJ mol^{-1}.[31]

THEORY AND METHODOLOGY

Distancefield. In the following, a description is given of the implementation of DF as briefly introduced in the Introduction. In the case of protein ligand binding, a DF restraint can be applied between (virtual) atom *i* and (virtual) atom *j*, representing the active site and the ligand, respectively. At the start of a simulation employing a DF restraint, a three-dimensional grid is created with the grid spacing, g_s, as set by the user. The minimal DF distance from a (virtual) atom *i* to each of the grid points is then assigned by applying Dijkstra's algorithm:[32]

(1) The DF distance at every grid point is initialized to a large value that cannot be reached during the simulation. Here, we have chosen this value to be $4 \times L_x \times L_y \times L_z$, where L_x, L_y, and L_z are the box lengths in the *x*, *y*, and *z* direction, respectively. The grid points that are positioned within a user-specified cutoff distance of any protein atom are flagged.

(2) All grid points are marked as unvisited. The initial node is the grid point closest to (virtual) atom *i*. This grid point is assigned as the current node and a DF distance of 0 is assigned to it.

(3) The unvisited nodes that are neighboring the current node are subsequently considered.

(a) A new DF distance for the neighboring grid point is initially calculated as the DF distance assigned to the current node plus the distance between the two grid points (g_s). If a neighboring node is flagged as protein, an additional user-specified protein penalty is added to the distance.

(b) When the new DF distance is less than the previously assigned preliminary distance, the latter is overwritten. Otherwise the distance does not change. In both cases, the neighbors are still marked "unvisited", and the assigned DF distances are still preliminary.

(4) When all neighbors are considered, the current node is marked as "visited". With this marking, the preliminary distance becomes final, and this node will not be considered anymore.

(5) The unvisited grid point with the lowest preliminary distance becomes the current node. Go back to step 3 and continue until all grid points are marked "visited".

Using this approach, periodic boundary conditions are taken fully into account, and grid points are not revisited. Once the updating step is completed, all grid points are assigned the shortest DF distance from atom *i*. The DF distances are updated (by applying steps 1−5) every UPDATE steps. Not updating every time step speeds up the simulation significantly and is allowed as long as the protein does not change its

dx.doi.org/10.1021/ct300967a | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

conformation too much between the updates, even though this will come at the expense of a slight loss of energy conservation. Using an UPDATE value of 100, the simulations are slowed down by ∼20% with respect to regular distant restraint simulations.

To avoid simulation artifacts, several extra features are implemented.

i. If (virtual) atom i is flagged as being within the protein, the DF distances of the grid points are not updated. It sometimes happens that this virtual atom temporarily gets buried in the active site due to the flexibility of the protein. If an update step would be performed, all DF distances are very large, because each of them has the additional protein penalty which was added to the first point. The forces induced by a DF restraint will thus get very large and may disrupt the structure. Also, because each grid point has the additional penalty, the DF distance converges to a normal radial distance with an initial offset.

ii. The user can specify the number of smoothening rounds that are applied after each updating step. In a smoothening round, we loop over all nonprotein grid points and check if one of its neighbors is flagged as protein. If this is the case, we are dealing with a grid point that is at the edge of the protein. In order to avoid large forces pointing away from the protein, the DF distance on the flagged protein grid point is determined again based on its direct neighbors, but now without the protein penalty. In this way, the large forces arising from the protein penalty on the grid are buried within the protein, and the regular van der Waals repulsion will ensure that the ligand never reaches these grid points. The DF distances of other grid points and the optimal route for atom j are not affected as this smoothening step is performed after the normal updating steps (1−5) have finished.

Once the DF distances are defined for all grid points, we can impose a DF restraint on the distance between atoms i and j. The forces and energies due to the DF restraint are calculated at each time step. We start this process by determining the eight grid points that are closest to atom j, to which DF distances have been determined in the updating step. In order to be able to calculate the forces, the derivatives of the DF distance in the x, y, and z directions have to be determined for each of the eight neighboring grid points. This is done using a finite differences approach as illustrated in eq 1 for grid point k, in the x direction.

$$\frac{dl}{dx_k} = \frac{l(x_k - 1, y_k, z_k) - l(x_k + 1, y_k, z_k)}{2^* g_s} \tag{1}$$

Here, $l(x_k - 1, y_k, z_k)$ is the DF distance assigned to the neighbor of grid point k with the smaller x coordinate, and g_s is the grid spacing. In order to interpolate the DF distance and its derivatives from the neighboring points to atom j, an assignment function of order 2 is used, similar to the one used for charges in the PPPM method.[33]

The potential energy associated with the DF restraint, V_{df}, can now be calculated using

$$\begin{aligned} V_{df} &= -k_{df}\left(l_{ij} - l_0 + \frac{1}{2}\Delta l^h\right)\Delta l^h && l_{ij} \le l_0 - \Delta l^h \\ V_{df} &= \frac{1}{2}k_{df}(l_{ij} - l_0)^2 && l_0 - \Delta l^h < l_{ij} < l_0 + \Delta l^h \\ V_{df} &= k_{df}\left(l_{ij} - l_0 - \frac{1}{2}\Delta l^h\right)\Delta l^h && l_{ij} \ge l_0 + \Delta l^h \end{aligned} \tag{2}$$

k_{df} is the force constant of the harmonic DF restraint, l_0 is the reference DF distance, and l_{ij} is the current DF distance between atoms i and j. The interaction term is linearized after a certain deviation Δl^h to prevent too large energies and forces for larger deviations. The forces on atoms i and j are calculated with

$$\begin{aligned} \vec{f}_j &= k_{df}\Delta l^h \vec{\nabla} l && l_{ij} \le l_0 - \Delta l^h \\ \vec{f}_j &= -k_{df}(l_{ij} - l_0)\vec{\nabla} l && l_0 - \Delta l^h < l_{ij} < l_0 + \Delta l^h \\ \vec{f}_j &= -k_{df}\Delta l^h \vec{\nabla} l && l_{ij} \ge l_0 + \Delta l^h \end{aligned} \tag{3}$$

and

$$\vec{f}_i = -\vec{f}_j \tag{4}$$

Here, the notation $\vec{\nabla} l$ is used for the derivative of the DF distance assigned to atom j.

The DF restraint can also be made dependent on a coupling parameter λ, making it more broadly applicable. In this case, the force constant and reference value of a DF restraint at a certain λ value can be obtained with

$$l_0^{AB}(\lambda) = (1 - \lambda)l_0^A + \lambda l_0^B \tag{5}$$

$$k_{df}^{AB}(\lambda) = (1 - \lambda)k_{df}^A + \lambda k_{df}^B \tag{6}$$

Here, l_0^A and l_0^B are the reference values in two states A and B, respectively. Similarly, k_{df}^A and k_{df}^B represent the force constants for the DF restraint in states A and B, respectively. The potential energy can be calculated with

$$V_{df}(\lambda) = 2^{n+m}\lambda^n(1 - \lambda)^m V_{df}^{AB}(\lambda) \tag{7}$$

where

$$\begin{aligned} V_{df}^{AB}(\lambda) &= -k_{df}^{AB}(\lambda)\left(l_{ij} - l_0^{AB}(\lambda) + \frac{1}{2}\Delta l^h\right) && \Delta l^h l_{ij} \le l_0^{AB}(\lambda) - \Delta l^h \\ V_{df}^{AB}(\lambda) &= \frac{1}{2}k_{df}^{AB}(\lambda)(l_{ij} - l_0^{AB}(\lambda))^2 && l_0^{AB}(\lambda) - \Delta l^h < l_{ij} < l_0^{AB}(\lambda) + \Delta l^h \\ V_{df}^{AB}(\lambda) &= k_{df}^{AB}(\lambda)\left(l_{ij} - l_0^{AB}(\lambda) - \frac{1}{2}\Delta l^h\right)\Delta l^h && l_{ij} \ge l_0^{AB}(\lambda) + \Delta l^h \end{aligned} \tag{8}$$

Here, we have introduced the prefactor term with variables m and n to be able to create so-called hidden restraints.[34] Depending on the choice of these variables, the restraint is not present in one or both of the end states.

The energy derivative can now be calculated using eqs 9 and 10.

$$\frac{\partial V_{df}}{\partial \lambda} = 2^{n+m}(n\lambda^{n-1}(1 - \lambda)^m - m\lambda^n(1 - \lambda)^{m-1})\cdot V_{df}^{AB}(\lambda) + 2^{n+m}\lambda^n(1 - \lambda)^m \frac{\partial V_{df}^{AB}(\lambda)}{\partial \lambda} \tag{9}$$

dx.doi.org/10.1021/ct300967a | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

$$\frac{\partial V_{\mathrm{df}}^{AB}(\lambda)}{\partial \lambda} = -(k^B - k^A)\left[l_{ij} - l_0^{AB}(\lambda) + \frac{1}{2}\Delta l^h\right]\frac{1}{2}\Delta l^h + k_{\mathrm{df}}^{AB}(\lambda)(l_0^B - l_0^A)\Delta l^h \qquad l_{ij} \le l_0^{AB}(\lambda) - \Delta l^h$$

$$\frac{\partial V_{\mathrm{df}}^{AB}(\lambda)}{\partial \lambda} = \frac{1}{2}(k^B - k^A)[l_{ij} - l_0^{AB}(\lambda)]^2 - k_{\mathrm{df}}^{AB}(\lambda)(l_{ij} - l_0^{AB}(\lambda))(l_0^B - l_0^A) \qquad l_0^{AB}(\lambda) - \Delta l^h < l_{ij} < l_0^{AB}(\lambda) + \Delta l^h$$

$$\frac{\partial V_{\mathrm{df}}^{AB}(\lambda)}{\partial \lambda} = (k^B - k^A)\left[l_{ij} - l_0^{AB}(\lambda) - \frac{1}{2}\Delta l^h\right]\frac{1}{2}\Delta l^h - k_{\mathrm{df}}^{AB}(\lambda)(l_0^B - l_0^A)\Delta l^h \qquad l_{ij} \ge l_0^{AB}(\lambda) + \Delta l^h \tag{10}$$

The forces are calculated similarly to that for the non-perturbed DF restraint:

$$\vec{f}_j(\lambda) = 2^{n+m}\lambda^n(1-\lambda)^m\vec{f}_j^{AB}(\lambda) \tag{11}$$

where

$$\vec{f}_j^{AB}(\lambda) = k_{\mathrm{df}}^{AB}(\lambda)\Delta l^h\vec{\nabla}l \qquad l_{ij} \le l_0^{AB}(\lambda) - \Delta l^h$$

$$\vec{f}_j^{AB}(\lambda) = -k_{\mathrm{df}}^{AB}(\lambda)(l_{ij} - l_0^{AB}(\lambda))\vec{\nabla}l \qquad l_0^{AB}(\lambda) - \Delta l^h < l_{ij} < l_0^{AB}(\lambda) + \Delta l^h$$

$$\vec{f}_j^{AB}(\lambda) = -k_{\mathrm{df}}^{AB}(\lambda)\Delta l^h\vec{\nabla}l \qquad l_{ij} \ge l_0^{AB}(\lambda) + \Delta l^h \tag{12}$$

and

$$\vec{f}_i^{AB}(\lambda) = -\vec{f}_j^{AB}(\lambda) \tag{13}$$

As will be shown in the Results and Discussion section, the current implementation allows for applications of DF in normal MD simulations, in free energy calculations, or in combination with enhanced sampling methods like local elevation (LE) or Hamiltonian replica exchange (HREMD).

Umbrella Sampling. Umbrella sampling (US) simulations are often used to focus sampling along a certain reaction coordinate.[10] In this study, we take the distance between the ligand and the active site of the protein as the reaction coordinate. Several independent simulations are performed with harmonic distance restraints on this distance, each centered at a different value. Using the weighted histogram analysis method (WHAM), the result can be unbiased, and the free energy along the reaction coordinate, the potential of mean force (PMF), can be obtained.[35] WHAM basically fits the distributions of the different windows to each other, so one has to be sure that sufficient overlap is present between neighboring windows.

Hamiltonian Replica Exchange Molecular Dynamics. A popular method to enhance sampling is replica exchange molecular dynamics (REMD). In this method, multiple replicas of the system are simulated simultaneously, but without any interactions between them. The replicas can either be simulated at different temperatures or with different Hamiltonians (HREMD), which is applied in this study. The key point of HREMD is that at certain time intervals, switching of the Hamiltonians for two neighboring configurations is attempted. The switching probability is determined on the basis of the energy differences by imposing the detailed balance condition. The switch is then accepted or rejected according to the Metropolis criterion. More detailed information on HREMD can be found elsewhere.[30,36] Here, the difference in the Hamiltonians between the replicas consists of a different reference value of the applied DF restraint. Upon configuration exchanges between the neighboring replicas, the ligand is restrained to DF distances closer or further from the binding site. Each replica can thus move in and out of the protein and is not restricted to a single route. However, with the use of a normal, radial distance restraint, it would be difficult to sample binding once a ligand has moved to a side or the back side of the protein, as was discussed above. The DF restraint is thus applied between the ligand and the binding site. At $\lambda = 0$ and $\lambda = 1$, it corresponds to reference values that will sample the bound and free states, respectively. Again, the PMF can be obtained by applying WHAM, or in the case of λ-dependent DF restraints, from a direct integration of eq 9 from the bound to the unbound state.

PMF. In the following, a brief description of the calculation of the PMF by using WHAM is given. As mentioned before, the PMF is the free energy along a certain reaction coordinate r and is defined as

$$\Delta G_{\mathrm{PMF}}(r) = -k_{\mathrm{B}}T\ln g(r) \tag{14}$$

Here, k_{B} is the Boltzmann constant, T is the temperature in Kelvin, and $g(r)$ is the radial distribution function in the case of a radial reaction coordinate. In general, one can define $g(r)$ as the ratio of the probability of finding a particle at position r in the simulations and the probability at position r for a homogeneous distribution, accounting for the Jacobian volume corrections.

Both US and HREMD result in several trajectories, which are different in the biasing potential applied to them. For each of these, the restraining energy distributions as a function of the reaction coordinate are obtained, which are then unbiased and combined using WHAM. This leads to a free energy as a function of r, $\Delta G_{\mathrm{WHAM}}(r)$, which can be transformed into a probability to find the ligand at distance r, $P(r)$:

$$P(r) \sim \mathrm{e}^{-\Delta G_{\mathrm{WHAM}}(r)/k_{\mathrm{B}}T} \tag{15}$$

In analogy to the radial distribution function, we define $g(r)$ as the ratio between $P(r)$ and the probability of finding the ligand at distance r in the case of a homogeneous distribution, $V(r)/V_{\mathrm{box}}$ where $V(r)$ is the volume accessible to the ligand at distance r and V_{box} is the volume of the simulation box. This gives

$$g(r) = \frac{P(r)V_{\mathrm{box}}}{V(r)} \tag{16}$$

and the PMF can be obtained from the WHAM results by using

$$\begin{aligned}\Delta G_{\mathrm{PMF}}(r) &= \Delta G_{\mathrm{WHAM}}(r) + k_{\mathrm{B}}T\ln V(r) - k_{\mathrm{B}}T\ln V_{\mathrm{box}} \\ &= \Delta G_{\mathrm{WHAM}}(r) + k_{\mathrm{B}}T\ln V(r) + C\end{aligned} \tag{17}$$

However, to get the free energy of binding, the concentration penalty as shown in eq 18 needs to be applied. This accounts for the free energy associated with bringing the ligand from V_{unb} to a distance r at which the volume $V(r)$ is available to the

dx.doi.org/10.1021/ct300967a | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

ligand, where V_{unb} is the volume sampled in the unbound, but restrained simulation.

$$\Delta G_{conc}(r) = -k_B T \ln \frac{V(r)}{V_{unb}} \tag{18}$$

Combining eqs 17 and 18 gives

$$\begin{aligned}\Delta G_{bind}(r) &= \Delta G_{PMF}(r) + \Delta G_{conc}(r) + C \\ &= \Delta G_{WHAM}(r) + k_B T \ln V(r) - k_B T \ln V(r) + C \\ &= \Delta G_{WHAM}(r) + C\end{aligned} \tag{19}$$

Finally, a standard state correction is added to $\Delta G_{bind}(r)$ to be able to compare to experimental data, which is independent of the reaction coordinate.

$$\Delta G_{std} = -k_B T \ln \frac{V_{unb}}{V_0} \tag{20}$$

Here, V_0 is the standard state volume of 1.661 nm^3. The binding free energy is calculated by choosing C in eq 19 such that the bound state corresponds to $\Delta G_{bind}(r) = 0$, and calculating the exponential average over an unbound region, e.g., where the protein−ligand interaction is zero.

$$\Delta G_{bind} = -k_B T \ln \int_{unb} e^{-\Delta G_{bind}(r)/k_B T} \, dr \tag{21}$$

Similarly, the unbound volume V_{unb} is calculated as

$$V_{unb} = \int_{unb} V_{unb}(r) e^{-\Delta G_{bind}(r)/k_B T} \, dr \Big/ \int_{unb} e^{-\Delta G_{bind}(r)/k_B T} \, dr \tag{22}$$

with $V_{unb}(r)$ as the sampled volume during a simulation at distance r, approximated from the observations of the ligand position on a grid around the protein. Once the $\Delta G_{PMF}(r)$ has reached a plateau (i.e., the protein and ligand no longer interact), the sum of ΔG_{bind} and ΔG_{std} will become independent of the exact definition of the unbound region.

Double Decoupling Method. The double decoupling method (DDM) is an alchemical approach to calculate binding free energies and is used here to compare efficiency and accuracy with respect to the path sampling techniques. DDM makes use of a thermodynamic cycle, where the free energy differences of two of the legs are determined with thermodynamic integration (TI) and one by an analytical formula. The two TI calculations correspond to the ligand being decoupled from the remaining of the system, once in solution and once when bound to the protein. During the TI within the protein, a (hidden) harmonic distance restraint is gradually introduced on the position of the ligand to prevent it from sampling the whole simulation box as it gets decoupled.[37] The free energy associated with the removal of this restraint can be calculated with the analytical formula

$$\Delta G_r = -k_B T \ln \frac{V_0}{(2\pi k_B T/K)^{3/2}} \tag{23}$$

Here, K is the force constant of the harmonic distance restraint, and the correction for the standard state is accounted for by the inclusion of V_0.

Simulation Settings. All molecular dynamics simulations were performed using the GROMOS11 simulation package[38] with the 54a7 force field parameter set.[39] All distance field functionality was implemented in release version 1.1.0 and will be made available in future releases of the program. Force field parameters for aspirin were derived on the basis of analogy with similar functional groups in the force field and are available in the Supporting Information.

Simulations of aspirin in solution were performed in a rectangular periodic box filled with 1438 SPC water molecules.[40] Initial coordinates for phospholipase A_2 (PLA_2) with bound aspirin were obtained from the crystal structure with PDB entry code 1OXR.[31] The system consists of 119 residues of PLA_2, one aspirin molecule and a calcium ion that is expected to play an important role in the binding process. The rectangular periodic box was filled with 9825 SPC water molecules and two Na^+ ions to neutralize the system. The resulting box has dimensions of 6.8 × 6.8 × 6.8 nm. Equilibration of the system was started by generating the velocities from a random Maxwell−Boltzmann distribution at 50 K. Position restraints were introduced on the solute with a force constant of 2.5 × 10^4 kJ mol^{-1} nm^{-2}. In six 20 ps simulations, the temperature was raised 50 K, until the final temperature of 298 K was reached. With every increase in temperature, the force constant of the position restraints was reduced by a factor of 10. The final equilibration step was a 40 ps simulation at 298 K, without position restraints. As the position restraints were switched off, rototranslational restraints were activated, thereby preventing rotation of the protein within the water box.[41]

The solvent and solute degrees of freedom were weakly coupled to separate temperature baths at 298 K with a relaxation time of 0.1 ps.[42] In simulations with normal distance restraints, the pressure was kept constant at 1 atm by isotropic weak coupling with a compressibility of 4.575 × 10^{-4} (kJ mol^{-1} nm^{-3})$^{-1}$ and a relaxation time of 0.5 ps.[42] In simulations where the distancefield was used, volume was kept constant. Simulations with constant pressure can in principle also be performed in combination with DF, although the grid must be updated every step due to the volume changes, making this setup rather computationally demanding. The SHAKE algorithm was used to constrain bond lengths,[43] allowing for a time step of 2 fs. A triple range cutoff scheme was used for the calculation of the nonbonded interactions. Within a distance of 0.8 nm, all interactions were calculated on the basis of a pairlist which was generated every fifth time step. Interactions at distances between 0.8 and 1.4 nm were calculated only with pairlist updates and kept constant at intermediate time steps. The interactions beyond 1.4 nm were accounted for by a reaction field contribution with a dielectric permittivity of 61.[44] It was shown that Ca^{2+} directly interacts with aspirin as well as other inhibitors.[31,45] However, in preliminary simulations, the Ca^{2+} drifted from its original position. Distance restraints with a force constant of 1500 kJ mol^{-1} nm^{-2} were therefore applied to Ca^{2+} with respect to the carbonyl and carboxylate oxygen atoms of four coordinating residues.

The reaction coordinate that is used throughout the manuscript is the distance between the center of geometry (cog) of C2 and C3 of aspirin (atom names according to the PDB entry) and the cog of the C_α atoms of six residues (L2, W19, A23, C29, D49, Y64) around the active site together with the Ca^{2+} ion. This can be calculated either as the regular Cartesian (DR) distance or as the distancefield (DF) distance. By using the cog of seven atoms, the force on the active site is distributed, thereby minimizing the chance of distorting the active site.

All simulations involving the DF restraints had a grid spacing, g_s of 0.2 nm, and the grid was updated every 100 time steps. The cutoff to determine if a grid point is lying in the protein was 0.2 nm, and one smoothening round was applied. A protein

dx.doi.org/10.1021/ct300967a | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

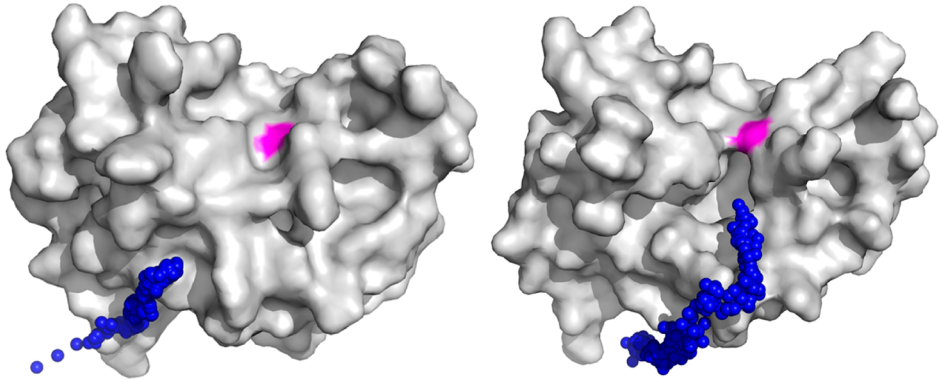


Figure 2. Trace of the aspirin molecule during slow-growth simulations using DR (left) and DF (right) restraints. The final configuration of Pla2 is shown as a gray surface, Ca^{2+} is shown in magenta, and the trace of the center of geometry of atoms C2 and C3 of aspirin is shown as blue spheres. The bound configuration is not reached using DR and distortions of the protein can be observed. Using DF restraints, aspirin moves around Pla2 and does reach the bound configuration, without distorting the protein.

penalty of 15 nm was added to the DF distance upon moving into the protein as described above. DF restraints were never "hidden", so in all cases $n = m = 0$ was used in eq 7. The DF restraint was linearized for deviations larger than $\Delta l^h = 1$ nm.

Slow-growth simulations were performed with DR as well as DF restraints. Two representing initial structures where aspirin was unbound were chosen from preliminary simulations. During the simulations, λ was changed from 0 to 1 by increasing it every time step with 4×10^{-6}, resulting in a total simulation length of 0.5 ns. A force constant of 500 kJ mol^{-1} nm^{-2} was kept constant throughout the simulation and was applied to the DR as well as the DF restraints.

Local elevation (LE) with the DF distance as reaction coordinate was used to simulate the binding and unbinding of aspirin to PLA_2. The biasing potentials were added every time step in the form of truncated polynomials.[46] The energy penalty of the DF reaction coordinate was increased with 5×10^{-3} kJ mol^{-1} nm^{-2} with each visit. The reaction coordinate was divided into 70 grid points with a width of 0.2 nm and was thus defined from 0 to 14 nm. In order to prevent unnecessary sampling of the unbound state at large DF distances, an additional half harmonic attractive distance restraint was applied with a high force constant (6000 kJ mol^{-1} nm^{-2}) at 1.75 nm.

The unbinding of aspirin from PLA_2 with umbrella sampling was performed at 31 evenly spaced windows, with restraining distances ranging from 0 to 3 nm. The first window (0 nm) was started from the equilibrated structure. Subsequent windows were started from the equilibrated structure at the previous window. The force constant on the harmonic distance restraints was 1500 kJ mol^{-1} nm^{-2}, and each window was simulated for 10 ns.

The Hamiltonian replica exchange simulation was performed with 24 replicas with nonequidistantly spaced λ points. The λ parameter is linearly coupled to the DF reference values, where $\lambda = 0$ corresponds to a DF reference value of 0 nm and $\lambda = 1$ to a DF reference value of 4 nm. The force constant applied to the DF restraints was set to 500 kJ mol^{-1} nm^{-2}. Initial coordinates for the replicas were obtained from a slow growth simulation. All replicas were simulated for a total of 10 ns with switching attempts every 2 ps.

Two thermodynamic integration (TI) simulations are performed for the double decoupling method. The TI of decoupling aspirin in the solvent is performed at λ points 0, 0.025, 0.05, 0.1, 0.125, 0.15, 0.2, 0.3, 0.4, 0.5, 0.55, 0.6, 0.625, 0.65, 0.7, 0.8, 0.85, 0.9, 0.95, and 1.0. The error estimates of $\partial H/\partial\lambda$ were determined by block averaging and extrapolation to infinite block lengths.[47] Initial simulation lengths at each λ point were 0.4 ns and were subsequently prolonged until the error estimates were smaller than 3 kJ mol^{-1}. This criterion was reached for all simulations, the maximum simulation time being 4 ns. For the decoupling of aspirin when bound to PLA_2, the same λ points were simulated, with the exception of 0.625. When λ is approaching 1, the ligand is almost decoupled from the rest of the system and is thus able to move away from the binding site. In order to reach convergence without the necessity to sample the whole simulation box, a harmonic distance restraint was gradually introduced during the TI. The restraint was defined between the cog of three atoms of the ligand and the cog of four atoms of neighboring residues such that the ligand remains in the binding cavity. The force constant was linearly coupled to the λ value, such that at $\lambda = 0$ and $\lambda = 1$ the force constants of the restraint corresponded to 0 and 1500 kJ mol^{-1} nm^{-2}, respectively. Here, the simulation length was at least 1 ns, and subsequent prolongations were again started when the error estimates were larger than 3 kJ mol^{-1}. In contrast to the TI in solution, not all simulations fulfilled this criterion before a maximum simulation length of 10 ns was reached.

RESULTS AND DISCUSSION

Use of Distancefield. In this section, the possibilities of using DF restraints as compared to DR restraints are described. As outlined in the Introduction, the use of DF restraints allows us to pull a ligand into the binding cavity starting from any position in the simulation box. This is demonstrated by making use of a slow-growth simulation, using the λ-dependent versions of the DF and DR restraints. In these simulations, aspirin was free in solution in the initial structures. As the simulations proceeded, the reference values of the restraints were slowly changed to be centered at the binding site (see simulation settings for more details). As shown in Figure 2A, aspirin gets stuck at the side of the protein when normal distance restraints are used. Further decreasing of the reference distance leads to deformation of the protein structure. Figure 2B shows that the ligand is guided around the protein and is able to find the entrance to the binding site, when a DF restraint was used.

Reversible binding in a single simulation can also be achieved when the DF restraints are used. This can be shown by using local elevation (LE) simulations, where sampling of previously visited states is made less favorable by the use of memory-based biasing potentials. Applying this to the DF distance between a

ligand and a protein, LE will enforce binding and unbinding, because the ligand cannot stay at a certain distance for a longer period of time. As discussed above, this approach does not work well with regular distance restraints, as the ligand will have difficulties moving back into the binding site as soon as it is fully solvated. Figure 3 shows the DF distance over time between

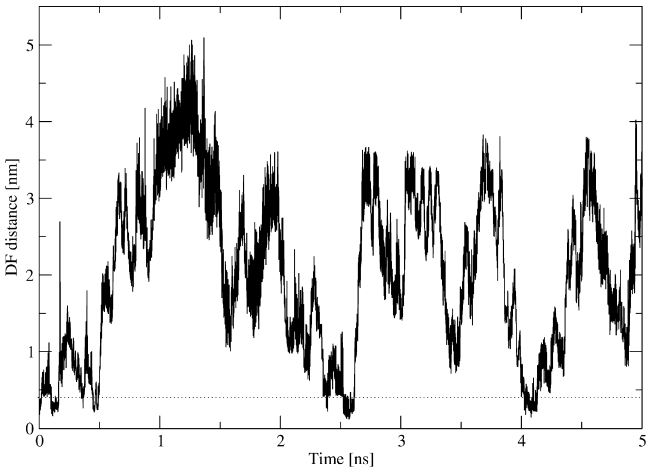


Figure 3. Time series of the DF distance during LE buildup simulation. A DF distance up to 0.4 nm can be considered as a bound state.

aspirin and PLA_2 during a LE simulation. Several dissociation and association events can be observed during the 5 ns simulation.

These examples involve nonequilibrium simulations in which the Hamiltonian changes with time, which makes a quantitative assessment of the binding affinity difficult. In the remainder of this work, we focus on equilibrium simulations using DR or DF restraints, which include thermodynamic integration (TI) and Hamiltonian replica exchange (HREMD).

PMF from US Using Distance Restraints. First, the results of the US simulations in combination with DR restraints will be evaluated. As described in the Theory and Methodology section, each of the 31 windows was sampled for 10 ns. Applying WHAM then yields the free energy as a function of the distance, as shown in Figure 4. It is immediately clear that the obtained curve is not very smooth. In principle, the free energy barriers encountered along the dissociation path can be related to interactions that aspirin has with PLA_2. However, it is also possible that these are artifacts, caused by a lack of sampling. Analysis of the trajectories shows that the first barrier, at 0.6 nm, corresponds to aspirin leaving the actual binding pocket. After aspirin leaves the pocket, it gets some more freedom to move, although the interaction with Ca^{2+} becomes more prominent. The large barrier in free energy starting from a DR distance of 1 nm (Figure 4) corresponds to configurations in which aspirin rotates around Ca^{2+} in order to keep the interaction with this ion as long as possible. Only when the plateau at around 1.4 nm is reached is the interaction with Ca^{2+} no longer retained, although aspirin is still very close to the surface of PLA_2. Up to this point, the results look reliable, at least for this particular dissociation path. The sudden decrease in free energy around 2 nm, however, is caused by a lack of sampling. During the first part of the simulation of the window centered at 1.9 nm, aspirin is sampling much more volume than at the smaller distances. However, after 4 ns, it moves to a slightly larger distance and subsequently remains trapped at a certain position, as shown in Figure 5. The next window (centered

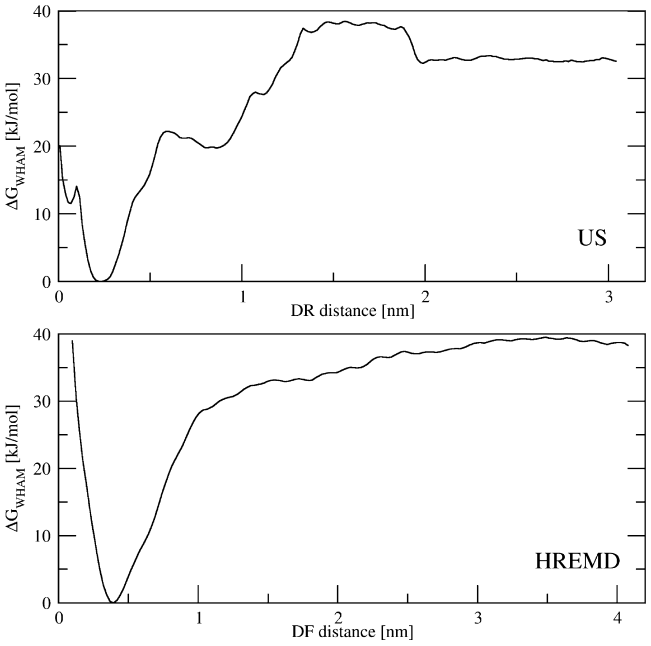


Figure 4. Free energies as a function of the reaction coordinates obtained from HREMD simulations with DF restraints (bottom) and from US with DR restraints (top).

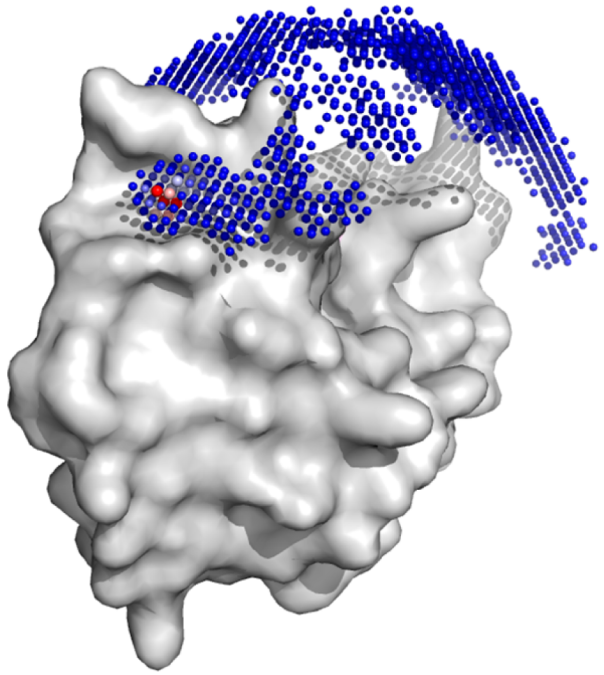


Figure 5. Representation of the sampling during US with restraints centered at 1.9 nm. Pla2 is shown as the surface (gray), and the center of geometry of atoms C2 and C3 of aspirin is shown as spheres on a grid with 0.1 nm spacing. Blue spheres indicate that aspirin was there only for a very short time, and red spheres indicate that aspirin occupied this position for a long time. For at least 1/3 of the simulation time, aspirin was stuck at the position of the colored spheres.

at 2 nm) does not show this behavior and is therefore more favorable. This explains the decrease in free energy around 2 nm. Note that the ligand would not get trapped when using the nonlinear DF distance.

Determining the PMF and the free energy of binding at several intermediate simulation lengths allows the evaluation of overall

dx.doi.org/10.1021/ct300967a | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

convergence (Figure S1). The largest changes are observed in the magnitude of the binding free energy during the first 6 ns. In the range of 6 to 10 ns, the PMF at distances up to 1.4 nm seems relatively converged. However, there are still significant changes at larger distances. The lack of sampling at the window centered at 1.9 nm can also be observed in Figure S1. As the ligand gets trapped after 4 ns, the drop in free energy at 2 nm can be observed for the first time after 5 ns of simulation.

We have thus shown that with the 10 ns US simulations, a single dissociation path is sampled and some sampling problems occurred. The free energy as a function of the DR distance is probably not the (only) physical one. Despite these issues, one can still determine the free energy of binding from these simulations because free energy is a state function and therefore independent of the chosen path. The only restriction here is that the bound and free states have to be well-defined. This seems to be the case here and the binding free energy can be determined to be −29.8 kJ mol^{-1}, where the correction for the standard state is already applied (see Table 1). The unbound state was averaged over distances larger than 2.1 nm.

Table 1. Binding Free Enegies in kJ mol^{-1} As Obtained by the Different Methods[a]

method		ΔG_{raw}	$\Delta G_{corr.}$
US (DR)	WHAM	−32.1	−29.8
HREMD (DF)	WHAM	−39.0	−32.2
HREMD (DF)	TI	−35.1	−30.8
DDM	TI	−20.0	−7.4
experiment	SPR		−29.6

[a]The raw free energy is obtained from the simulations, whereas the corrected free energy includes the standard state correction, eq 20. SPR: surface plasmon resonance.

PMF from HREMD Using Distancefield. The previous section showed that the quality of the PMF obtained with US was mainly hampered by the simulation of only a single dissociation path and lack of sampling. HREMD in combination with DF restraints should improve these issues as reversible binding and alternative paths can be sampled. However, this requires a reaction coordinate along which the binding process may also be sampled. Figure 4 also shows the free energy as a function of the DF distance as obtained from the HREMD simulations. It can be seen immediately that this curve is much smoother than the one obtained with US and has only a single well. This indicates that once the ligand gets close enough to the start of the association path, it is favorable to move into the binding site, and barrier-free paths are possible. Using the results shown in Figure 4 in combination with the standard state correction (eq 20) and averaging the unbound state after 3.2 nm, a free energy of binding of −32.2 kJ mol^{-1} is obtained (see Table 1).

The convergence and sampling of the simulations can be evaluated in several ways. Figure 6 shows the number of visited λ points over time for each of the 24 replicas. It shows that three replicas visited all λ points, indicating they sampled a complete association or dissociation path. Although these paths are helpful for the visualization of the paths that can be taken, it must be noted here that the convergence of the simulation is not dependent on the individual replicas visiting all λ points. Rather, a broad spectrum of configurations needs to be sampled at each λ point. On average, the replicas visit 16 different λ values. The free energy as a function of the DF distance shows very nice convergence (see Figure S2).

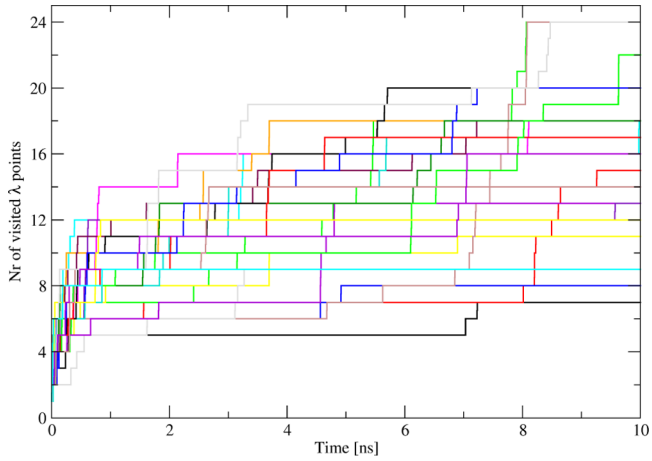


Figure 6. Time series of the number of λ points that were visited by each of the 24 replicas (different colors) during HREMD.

The enhanced sampling is optimal when the distributions of neighboring potentials overlap such that sufficient replica exchanges can occur. The exchanges can be visualized by plotting the λ values of the replicas over time, as shown in Figure S3. A more quantitative measure is to evaluate the acceptance probabilities of switching to the next λ value, as shown in Table 2. The average acceptance probability is ∼20%, with a minimum of 11%. A minimal value of 10% is required, and 20% is reported as the optimal probability as determined for standard REMD.[36,48]

Table 2. Average Properties for the 24 Replicas from HREMD[a]

id	λ	l_0 [nm]	V [nm^3]	$\langle P_{switch}\rangle$	$\langle \partial H/\partial\lambda\rangle$
1	0.000	0.00	0.049	0.38	−548.1
2	0.048	0.19	0.063	0.42	−283.9
3	0.093	0.37	0.083	0.33	−24.5
4	0.138	0.55	0.140	0.18	158.6
5	0.180	0.72	0.226	0.27	175.1
6	0.220	0.88	0.315	0.11	247.4
7	0.263	1.05	0.337	0.12	108.7
8	0.305	1.22	0.330	0.20	26.0
9	0.345	1.38	0.416	0.16	33.0
10	0.388	1.55	0.547	0.19	0.3
11	0.430	1.72	0.687	0.17	0.9
12	0.478	1.91	0.907	0.19	5.3
13	0.523	2.09	1.106	0.19	−2.2
14	0.573	2.29	1.360	0.22	12.7
15	0.618	2.47	1.597	0.25	3.6
16	0.660	2.64	1.582	0.22	22.9
17	0.703	2.81	1.977	0.22	30.3
18	0.745	2.98	2.296	0.19	27.1
19	0.788	3.15	2.513	0.22	3.6
20	0.830	3.32	2.715	0.22	2.2
21	0.873	3.49	2.767	0.21	−8.2
22	0.915	3.66	2.746	0.21	−10.0
23	0.958	3.83	2.697	0.24	1.2
24	1.000	4.00	2.599	-	13.5

[a]For each replica id, the corresponding λ value, reference DF distance (l_0), and the sampled volume (V) are shown. $\langle P_{switch}\rangle$ represents the average switching probability with the next λ point. The final column contains the ensemble average of the λ derivative (in kJ mol^{-1}) associated with each λ value. These are used for the TI calculation on the HREMD data, by integrating from the bound state (id 3) to the free state (id 24).

dx.doi.org/10.1021/ct300967a | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

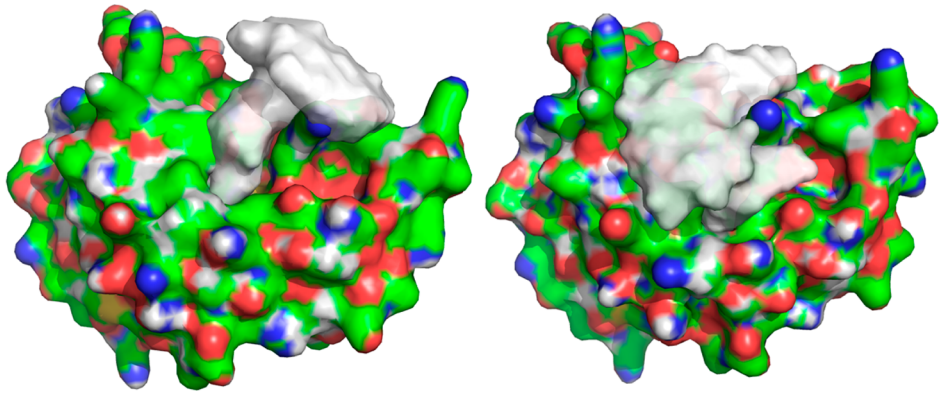


Figure 7. Volume sampled (white surface) during the US (left) and HREMD (right) simulations up to a DR distance of 1.5 nm and a DF distance of 1.55 nm, respectively. From the sampled volumes, one can see that with US, only a single dissociation path is sampled, whereas with HREMD, the sampled region is broader, indicating multiple paths being sampled.

The difference in sampling between regular US and HREMD can clearly be seen in Figure 7. The left panel shows the volume sampled by aspirin in the US DR windows that are centered from 0 up to 1.5 nm. Similarly, the right panel shows the volume sampled by aspirin during the simulations at λ points with reference DF distances from 0 up to 1.55 nm. The volume sampled with US clearly shows that only a single path is sampled. HREMD on the other hand shows a larger sampled volume, indicating that multiple paths are being sampled.

Binding Free Energies from Other Methods. In the previous section, the free energy of binding was determined from the PMF obtained with HREMD. However, since in HREMD the λ-dependent Hamiltonian is used and the derivative of the free energy with respect to λ at every λ point is known, these data can be used to perform thermodynamic integration. The integration should then take place starting from the bound configuration up to the unbound configuration. One should not confuse the bound configuration with $\lambda = 0$ in this case. After measuring the DF distance in the original bound structure and from evaluating the PMF, $\lambda = 0.093$ was chosen as the bound state. Integration from $\lambda = 0.093$ up to $\lambda = 1$ and applying the standard state correction with the sampled volume at $\lambda = 1$ yields a binding free energy of -30.8 kJ mol^{-1}.

In order to compare the path sampling methods with an alchemical method, DDM was also performed as described in the Theory and Methodology section. Two TI calculations and the application of one analytical formula are required to calculate the binding free energy with this alchemical method. The decoupling of aspirin in solution was, not surprisingly, relatively easy to perform and converge. However, the decoupling of aspirin when bound to PLA_2 was rather tedious. Several λ points were prolonged up to 10 ns and were still far from converged with remaining error estimates up to 16 kJ mol^{-1}. Figure S4 shows the free energy profiles for these calculations. These convergence issues turned out to be related to the rotation of aspirin within the binding cavity as its interactions with the protein were lowered. Although part of the convergence problems could thus have been prevented by applying an additional rotational restraint on aspirin,[37] this shows that alchemical methods to calculate the binding free energy are not an easy job either. The fact that aspirin is a charged molecule makes it even more difficult as a small difference is obtained from integration over very large free energies. In the end, a free energy of -7.4 kJ mol^{-1} was obtained (see Table 1). The deviation from the experimental data is thus quite large, but this can be expected, as the decoupling of aspirin bound to PLA_2 was ill converged.

CONCLUSIONS

Several methods are used to study free energies of protein ligand binding and in particular of aspirin binding to PLA_2, without any *a priori* knowledge of the association or dissociation path. Umbrella sampling simulations were used to determine the potential of mean force along the distance between aspirin and PLA_2. However, it was shown that this method only samples a single dissociation path and suffers from a lack of sampling. Hamiltonian replica exchange molecular dynamics (HREMD) would in principle be able to sample reversible binding and multiple paths but is in practice difficult as it involves pulling the ligand back into the protein. In cases where the shortest distance from the ligand to the binding site would be straight through the protein, normal distance restraints are usually not able to pull the ligand into the binding site. For this reason, we have introduced a new measure of distance, the distancefield (DF), on which restraints can be exerted. DF is a grid-based method in which the shortest distance between a protein and a ligand is determined without passing through the protein. In combination with HREMD it allows for reversible binding, and thus multiple dissociation and association paths can be taken into account when generating the potential of mean force (PMF).

As a case study, the above-mentioned methods were applied to aspirin binding to PLA_2. The binding free energies obtained with US (-29.8 kJ mol^{-1}) and HREMD (-32.2 kJ mol^{-1}) are both very close to the experimentally determined -29.6 kJ mol^{-1}, because these values do not depend on the path taken. However, the PMF obtained with HREMD includes more paths and is most likely more representative than the one obtained from US.

ASSOCIATED CONTENT

Supporting Information

Force-field parameters for aspirin, the binding free energy as function of r for different simulation lengths, the replica exchange switches, and free energy profiles for DDM. This information is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*Tel.: +43 1 47654 8302. Fax: +43 1 47654 8309. E-mail: chris.oostenbrink@boku.ac.at.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors gratefully acknowledge financial support of the Vienna Science and Technology Fund (WWTF) grant number LS08-QM03 and the European Research Council (ERC) grant number 260408.

■ REFERENCES

(1) Buch, I.; Giorgino, T.; De Fabritiis, G. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 10184−10189.
(2) Dror, R. O.; Pan, A. C.; Arlow, D. H.; Borhani, D. W.; Maragakis, P.; Shan, Y.; Xu, H.; Shaw, D. E. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 13118−13123.
(3) Kirkwood, J. G. *J. Chem. Phys.* **1935**, *3*, 300.
(4) Zwanzig, R. W. *J. Chem. Phys.* **1954**, *22*, 1420.
(5) Åqvist, J.; Medina, C.; Samuelsson, J.-E. *Protein Eng.* **1994**, *7*, 385−391.
(6) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E. *Acc. Chem. Res.* **2000**, *33*, 889−897.
(7) Beveridge, D. L.; DiCapua, F. M. *Annu. Rev. Biophys. Biophys. Chem.* **1989**, *18*, 431−492.
(8) Kollman, P. *Chem. Rev.* **1993**, *93*, 2395−2417.
(9) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. *Biophys. J.* **1997**, *72*, 1047−1069.
(10) Torrie, G. M.; Valleau, J. P. *J. Comput. Phys.* **1977**, *23*, 187−199.
(11) Woo, H.-J.; Roux, B. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 6825−6830.
(12) Hajjar, E.; Perahia, D.; Débat, H.; Nespoulous, C.; Robert, C. H. *J. Biol. Chem.* **2006**, *281*, 29929−29937.
(13) Doudou, S.; Burton, N. A.; Henchman, R. H. *J. Chem. Theory Comput.* **2009**, *5*, 909−918.
(14) Fukunishi, Y.; Mitomo, D.; Nakamura, H. *J. Chem. Inf. Model.* **2009**, *49*, 1944−1951.
(15) Mitomo, D.; Fukunishi, Y.; Higo, J.; Nakamura, H. *Genome Inf. Ser.* **2009**, *23*, 85−97.
(16) Huber, T.; Torda, A. E.; Van Gunsteren, W. F. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 695−708.
(17) Hansen, H. S.; Hünenberger, P. H. *J. Comput. Chem.* **2010**, *31*, 1−23.
(18) Branduardi, D.; Gervasio, F. L.; Cavalli, A.; Recanatini, M.; Parrinello, M. *J. Am. Chem. Soc.* **2005**, *127*, 9147−9155.
(19) Gervasio, F. L.; Laio, A.; Parrinello, M. *J. Am. Chem. Soc.* **2005**, *127*, 2600−2607.
(20) Masetti, M.; Cavalli, A.; Recanatini, M.; Gervasio, F. L. *J. Phys. Chem. B* **2009**, *113*, 4807−4816.
(21) Zhang, Y.; Voth, G. A. *J. Chem. Theory Comput.* **2011**, *7*, 2277−2283.
(22) Söderhjelm, P.; Tribello, G. A.; Parrinello, M. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 5170−5175.
(23) Darve, E.; Rodríguez-Gómez, D.; Pohorille, A. *J. Chem. Phys.* **2008**, *128*, 144120.
(24) Hénin, J.; Fiorin, G.; Chipot, C.; Klein, M. L. *J. Chem. Theory Comput.* **2010**, *6*, 35−47.
(25) Colizzi, F.; Perozzo, R.; Scapozza, L.; Recanatini, M.; Cavalli, A. *J. Am. Chem. Soc.* **2010**, *132*, 7361−7371.
(26) Xiong, H.; Crespo, A.; Marti, M.; Estrin, D.; Roitberg, A. *Theor. Chem. Acc.* **2006**, *116*, 338−346.
(27) Wereszczynski, J.; McCammon, J. A. *J. Chem. Theory Comput.* **2010**, *6*, 3285−3292.
(28) Kokubo, H.; Tanaka, T.; Okamoto, Y. *J. Comput. Chem.* **2011**, *32*, 2810−2821.
(29) Oostenbrink, C.; De Ruiter, A.; Hritz, J.; Vermeulen, N. *Curr. Drug Metab.* **2012**, *13*, 190−196.
(30) Sugita, Y.; Kitao, A.; Okamoto, Y. *J. Chem. Phys.* **2000**, *113*, 6042−6051.
(31) Singh, R. K.; Ethayathulla, A. S.; Jabeen, T.; Sharma, S.; Kaur, P.; Singh, T. P. *J. Drug Targeting* **2005**, *13*, 113−119.
(32) Dijkstra, E. W. *Numer. Math.* **1959**, *1*, 269−271.
(33) Hünenberger, P. H. *AIP Conf. Proc.* **1999**, *492*, 17−83.
(34) Christen, M.; Kunz, A.-P. E.; Van Gunsteren, W. F. *J. Phys. Chem. B* **2006**, *110*, 8488−8498.
(35) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13*, 1011−1021.
(36) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141−151.
(37) Boresch, S.; Tettinger, F.; Leitgeb, M.; Karplus, M. *J. Phys. Chem. B* **2003**, *107*, 9535−9551.
(38) Schmid, N.; Christ, C. D.; Christen, M.; Eichenberger, A. P.; Van Gunsteren, W. F. *Comput. Phys. Commun.* **2012**, *183*, 890−903.
(39) Schmid, N.; Eichenberger, A. P.; Choutko, A.; Riniker, S.; Winger, M.; Mark, A. E.; Van Gunsteren, W. F. *Eur. Biophys. J.* **2011**, *40*, 843−856.
(40) Berendsen, H. J. C.; Postma, J. P. M.; Van Gunsteren, W. F.; Hermans, J. In *Intermolecular Forces*; Reidel: Dordrecht, The Netherlands, 1981; Vol. *11*, pp 331−342.
(41) Bekker, H.; Van Den Berg, J. P.; Wassenaar, T. A. *J. Comput. Chem.* **2004**, *25*, 1037−1046.
(42) Berendsen, H. J. C.; Postma, J. P. M.; Van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684.
(43) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327−341.
(44) Heinz, T. N.; Van Gunsteren, W. F.; Hünenberger, P. H. *J. Chem. Phys.* **2001**, *115*, 1125.
(45) Schevitz, R. W.; Bach, N. J.; Carlson, D. G.; Chirgadze, N. Y.; Clawson, D. K.; Dillard, R. D.; Draheim, S. E.; Hartley, L. W.; Jones, N. D.; Mihelich, E. D. *Nat. Struct. Biol.* **1995**, *2*, 458−465.
(46) Hansen, H. S.; Daura, X.; Hünenberger, P. H. *J. Chem. Theory Comput.* **2010**, *6*, 2598−2621.
(47) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Oxford University Press: Oxford, U. K., 1987.
(48) Rathore, N.; Chopra, M.; De Pablo, J. J. *J. Chem. Phys.* **2005**, *122*, 024111.