JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Relationship between Protein Flexibility and Binding: Lessons for Structure-Based Drug Design

Daniel Alvarez-Garcia[†,‡] and Xavier Barril[†,‡,§,*]

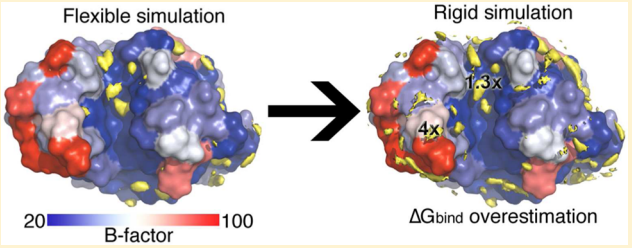
[†]Departament de Fisicoquímica, Facultat de Farmàcia, Universitat de Barcelona, Av. Joan XXIII s/n, 08028 Barcelona, Spain

[‡]Institut de Biomedicina de la Universitat de Barcelona (IBUB), Barcelona, Spain

[§]Catalan Institution for Research and Advanced Studies (ICREA), Passeig Lluís Companys 23, 08010 Barcelona, Spain

Supporting Information

ABSTRACT: Conceptually, the simplistic lock and key model has been superseded by more realistic views of molecular recognition that take into account the intrinsic dynamics of biological macromolecules. However, it is still common for structure-based drug discovery methods to represent the receptor as static structures. The practical advantages of this approximation, the notable success attained over the past few decades with such simple models and the absence of clear guidelines for weighing the pros and cons of accounting for flexibility may prompt some investigators to stretch the rigid model beyond its scope. Here, we investigate the relationship between protein flexibility and binding free energy and present some useful hints for understanding when, and to what extent, flexibility should be considered. Using molecular dynamics simulations of hen egg-white lysozyme (HEWL) with explicit aqueous/organic solvent mixtures and a range of restraint conditions, we find out how artificially restricted mobility affects binding hot spots. Barring sampling errors or an inappropriate choice of reference structure, we find that decreased mobility (measured as B-factors) leads to artifactually more negative binding free energies, but a logarithmic relationship between both terms attenuates the errors. Consequently, ignoring flexibility may be an acceptable approximation for intrinsically rigid regions (such as the active site of enzymes) but may lead to larger errors elsewhere. For the same reason, local conformational sampling yields very accurate predictions and, owing to its practical advantages, may be preferable to full conformational sampling for many applications.



INTRODUCTION

Predicting the binding free energy of noncovalent complexes (ΔG_{bind}) is the primary objective of computer-aided drug design, as it enables the identification of new bioactive compounds and rational optimization of initial hits.[1,2] However, ΔG_{bind} is an elusive property that can only be attained if the molecular system (including the solvent) and the bound−unbound transitions are accurately replicated in silico. Since this is generally not feasible, numerous approximations have been developed to reduce the complexity of the system (e.g., implicit solvent), decrease the degrees of freedom (e.g., limited conformational sampling), consider the bound and unbound states only (e.g., linear response theory), bypass physics with empirical relationships (e.g., docking scoring functions), or settle for relative binding free energies ($\Delta\Delta G_{bind}$) to a reference state (e.g., thermodynamic integration of alchemical transformations). Several such approximations are often used simultaneously, making it difficult to evaluate their individual merits and limitations. Sampling the conformational flexibility of the protein is particularly challenging and, though well recognized as a necessary consideration in drug discovery,[3,4] most often it is considered to a minimal extent or simply neglected. The consequences of such approximation are not well-understood and there have been conflicting reports, with some authors suggesting that protein rigidity can cause loss of binding affinity on some sites and the opposite effect on others,[5] a finding that has been challenged by other authors.[6]

With new software and hardware solutions, it has become possible to bypass most approximations and to use unbiased molecular dynamics (MD) simulations to identify the association process of protein−ligand complexes, retrieving the correct binding mode as well as their binding affinity and kinetic constants.[7−9] Such simulations are still limited by the molecular mechanics framework, but at least on the protein side, molecular force fields have proven as accurate as experimental methodologies.[10] Unfortunately, even with the most advanced software and hardware solutions only binding events in the microsecond scale have been replicated in silico, while many drugs have slower association rates and the typical residence time ranges from minutes to hours.[11] But proteins also display affinity for small organic solvents, a feature that has proven useful in experimental and computational hot spot and binding site identification.[12,13] In this case, the whole binding and unbinding event takes only a few nanoseconds and can be fully reproduced by MD simulations.[14] This provides an ideal test bed to investigate the

Received: March 2, 2014
Published: May 20, 2014

dx.doi.org/10.1021/ct500182z | J. Chem. Theory Comput. 2014, 10, 2608−2614

relationship between protein flexibility and binding affinity, which is the main goal of this study.

We have used MD simulations with explicit aqueous/organic solvent mixtures (MDmix method)[15] to identify binding hot spots on hen egg white lysozyme (HEWL) and to investigate the consequences of limiting protein flexibility. Comparison with existing experimental data and with previous computational results allows us to quantify the consequences of using a rigid representation of the protein and to unravel the problems that arise from including full flexibility in the simulations, namely convergence and transferability of results.

MATERIALS AND METHODS

MD simulations of HEWL in different aqueous/organic solvent mixtures (MDmix method) were used to identify the high binding affinity spots (hot spots) over the surface of the protein and to estimate their binding free energies using different restraining schemes with varying strengths. The protein was simulated in ethanol (EtOH), acetonitrile (ACN), methanol (MeOH), and isopropyl alcohol (IPA) aqueous mixtures. EtOH was chosen as the simplest representation of a druglike molecule with fastest diffusion rate and is the basis for this study. The latter three solvents where chosen for availability of experimental binding data.[13] Moreover ACN has been widely used in theoretical studies on HEWL in hot-spot identification and is used for comparison and cross-validation.[5,6,16] After simulation, binding free-energy grids are calculated and lower points identified and quantified as hereafter described.

MDmix Simulations. HEWL protein structure was downloaded from the Protein Databank (PDB: 2lyo).[16] All molecular dynamics simulations were prepared and run following the same protocol: the protein was protonated and termini-capped using MOE.[17] Using AMBER's tLeap,[18] the protein was placed in a truncated octahedral box spanning 13 Å further from the solute in each direction and solvated using pre-equilibrated solvent mixture boxes (details on the preparation of the different solvent boxes can be found in the Supporting Information). AMBER force field 99SB was used to parametrize the system. The protein and solvent were minimized for 3000 steps, initial velocities assigned to obtain a distribution at 150 K and gradually warmed up to 300 K in 800 ps in the NPT ensemble. The system was further equilibrated for 1 ns at 300 K in the NPT ensemble. The production part of the simulation was run in the NVT ensemble, storing the coordinates every picosecond. During all equilibration and production steps, a Langevin thermostat with a collision frequency of 4 was used; the cutoff for van der Waals interactions was set to 9 Å, and the time step to 2 ps.

Different restraining schemes were used: the protein was allowed to move freely (FREE simulations) or positional restraints on all nonhydrogen atoms were applied with varying intensity. Restraints are implemented as harmonic potentials centered on the crystallographic coordinates with a force constant of 1 kcal/mol·Å^2 (HA1 simulations), 0.1 kcal/mol·Å^2 (HA01 simulations), or 0.01 kcal/mol·Å^2 (HA001 simulations).

For each condition, several independent replicas were run with different seed numbers to ensure better sampling. A summary of simulations run can be found in Table 1.

Binding Free-Energy Grids. AMBER's ptraj program was used to align the trajectories over a single reference structure (the crystal coordinates). The density grids were calculated using a custom program that allows tracking the center-of-mass of the solvent molecules instead of particular atoms. A grid spacing of 0.5 Å was used to construct the density map. The resulting count units were converted to free energies using the inverse Boltzmann relationship, shown in eq 1:

$$\Delta G = -k_b T \ln \frac{N_i}{N_o} \quad (1)$$

where the expected density (N_0) was calculated dividing the number of molecules in the equilibrated mixture box by the equilibrated box volume (e.g., the equilibrated mixture of EtOH 20% had 17 EtOH molecules and an average volume of 7984 Å^3; $N_0 = 2.13 \times 10^{-3}$ counts/Å^3 per snapshot). Since the change in free energy of the organic solvent molecules is caused by, and measured at, the protein surface, from now on we refer to this quantity as binding free energy.

Standard State Correction. The aqueous/organic mixtures were prepared to a final concentration of 20% (v/v). In order to compare the calculated binding free energies with published experimental affinities, a standard state correction was analytically introduced to estimate the binding free energy at 1 M concentration. In accordance with the literature,[19] this cost can be estimated as follows:

$$\Delta G^\circ = \Delta G + k_b T \ln \frac{V_{sim}}{V^0} \quad (2)$$

where V_{sim} is the simulation volume accessible by each solvent molecule and V^0 is the standard volume corresponding to a 1 M concentration (i.e., 1661 Å^3). All binding free-energy values reported in this work were corrected accordingly.

Hot Spot Binding Free-Energy Estimation. To provide an estimate of binding free energy to a particular hot spot, the grid point containing the minimum is identified and values 2 Å around (considering the average radii of an atom) are averaged to reduce possible artifacts introduced by the grid spacing. The total hot spot binding free energy will be

$$\Delta G^0_{hotspot} = -k_b T \ln\langle e^{-\beta \Delta G_i^\circ} \rangle \quad (3)$$

where <···> indicates an arithmetic average over the values surrounding the minimum point and β the thermal factor ($1/k_b T$).

Hot Spot *B*-Factor Calculation. The *B*-factor value for a particular hot spot is calculated from the trajectory using an in-house program following the standard *B*-factor formula ($B = 8\pi^2 \langle x^2 \rangle$). Only protein atoms within 6 Å of the center of the hot spot are considered.

Table 1. Summary of Simulations Run in This Work

solvent	restraint scheme	replicas × simulation time
water	free	1 × 20 ns
MeOH (methanol 20% water)	free	3 × 20 ns
IPA (isopropanol 20% water)	free	3 × 50 ns
ACN (acetonitrile 20% water)	free	3 × 20 ns
	heavy atoms (HA1, $k = 1^a$)	3 × 20 ns
EtOH (ethanol 20% water)	free	3 × 200 ns
	heavy atoms (HA1, $k = 1^a$)	3 × 20 ns
	heavy atoms (HA01, $k = 0.1^a$)	3 × 20 ns
	heavy atoms (HA001, $k = 0.01^a$)	3 × 200 ns

^{a}Force constant in kcal/(mol Å^2) units.

dx.doi.org/10.1021/ct500182z | J. Chem. Theory Comput. 2014, 10, 2608−2614

Solvent-Exchange Plots. All solvent molecules visiting the C-site along the trajectory were tracked using custom software and results plotted using the python Matplotlib package.[20] The C-site was defined as the coordinate inside the cavity with minimum binding free energy and added a tolerance of 2 Å around.

RESULTS

Unbiased MDmix Simulations Reproduce Experimental Binding Data. Before proceeding to investigate the effect of protein flexibility, we must confirm that MDmix simulations are capable of correctly reproducing binding hot spots. HEWL was the first system on which NMR solvent screening was described, and the binding affinity of several organic molecules for the catalytic site (C-site) was reported.[13] The predicted ΔG_{bind} for ACN in this site is −1.25 ± 0.1 kcal/mol (see Methods for calculation details), in good agreement with the experimental value of −0.8 ± 0.1 kcal/mol. In the case of MeOH, the results differ by at least 1.5 kcal/mol from the experimental data (−0.75 ± 0.1 kcal/mol calculated vs +0.7 ± 0.1 kcal/mol or +1.4 ± 0.2 kcal/mol from the NMR data, depending on the observed chemical shift[13]). A similar level of error is found for IPA: −2.0 ± 0.4 kcal/mol calculated vs 0.0 ± 0.3 kcal/mol, according to the NMR experiments. With consideration of the nonobvious relationship between the ensemble of molecular structures provided by MD and the experimental observables (chemical shift differences of one particular atom, W108 NεH, in this case), the agreement is remarkable. MD approaches remain limited by the quality of the force field, but consideration of protein flexibility and explicit solvation bring about a major improvement relative to standard methods: the corresponding values obtained with a rigid description of the protein and an implicit solvent method are −10.2 kcal/mol (ACN), −7.2 kcal/mol (MeOH), and −11.6 kcal/mol (IPA).[21]

Simulations with all the organic solvents investigated confirm that the C-site is the most favorable binding site (i.e., highest occupancy and lowest binding free energy), but other putative hot spots are also identified on the surface of HEWL, as shown in Figure 1a for EtOH. Numerous crystal structures of HEWL with organic solvents, larger ligands, or protein complexes confirm that, besides the C-site, other sites are recurrently used to form noncovalent complexes. As shown in Figure S1 of the Supporting Information, these tend to overlap with hot spots derived from the simulations, providing confidence on the quality of the predictions. This allows us to investigate the effect of protein rigidity on binding free energy, binding kinetics, and convergence of results in the following sections.

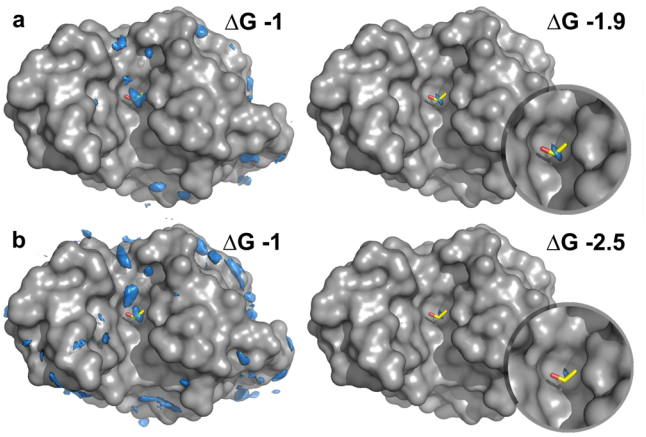


Figure 1. Binding free energy of ethanol on the surface of HEWL simulated: (a) without restraints (FREE) and (b) with strong restraints (HA1). Left column: general hot spots overview at same isocontour level (−1 kcal/mol). Right column: global minima at the level indicated (in kcal/mol).

Rigidity Favors Binding, With Exceptions. Protein rigidity is generally associated with increased binding. This becomes evident observing the distribution of binding hot spots over the protein surface, as they tend to localize on rigid areas (Figure S2 of the Supporting Information). A recent analysis of hot spots occurring in protein−protein interfaces also suggests that hot spots are located in "stability islands", that is, regions of the protein with lesser degree of mobility.[22] In trying to explain the avidity of hot spots to engage in noncovalent complexes, many authors have pointed toward a tendency to become desolvated,[23,24] but reduced mobility of the site in the apo state is a logical contributing factor, as it implies a lower entropic cost of conformational selection. This is consistent with the fact that structured proteins can achieve strong binding while intrinsically disordered proteins display lower affinities.[25] Introducing artificial restraints has the same effect and, as previously noted by Lexa and Carlson,[5] when the protein is kept rigid, existing hot spots acquire more negative ΔG_{bind} values and a myriad of additional hot spots emerge (Figure 1b).

However, protein rigidity can also cause the opposite effect. This is the case of ligand-induced binding sites, as illustrated in Figure S3 of the Supporting Information, where a modest shift of position of the lateral chain of Asn93 is necessary to make a small hydrophobic pocket accessible to EtOH molecules. Mutual, and sometimes substantial, adaptation is a natural part of the binding process (see the animations in the work by Giorgino, T. et al.[26] for a visual example), and a rigid structure may conceal part of its binding surface. This probably explains why apo structures yield significantly worse results than holo structures in rigid receptor docking.[27,28] The same can happen if the initial structure contains inaccuracies, as in the case of homology models generated from low identity templates.[27,29] In summary, rigidity can cause both gain or loss of binding affinity, depending on whether the reference structure is suitable for binding or not, but the first situation is far more common on well-defined structures and will be better-characterized in the next section.

Logarithmic Relationship between Protein Flexibility and ΔG_{bind}. In order to quantify the effect of protein rigidity on binding affinity, it is useful to consider the thermodynamic cycle depicted in Figure 2, according to which, the free energy of binding to a constrained protein ($\Delta G_{bind}'$) can be related to the intrinsic binding free energy (ΔG_{bind}) as

$$\Delta G'_{bind} = \Delta G_{bind} + \Delta G^{PL}_{fix} - \Delta G^{P}_{fix} \quad (4)$$

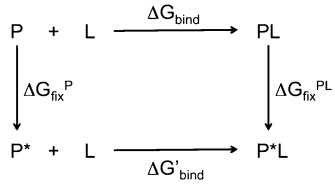


Figure 2. Thermodynamic cycle showing the relationship between free energy of binding and free energy of restraining the protein conformational space.

dx.doi.org/10.1021/ct500182z | J. Chem. Theory Comput. 2014, 10, 2608−2614

Where ΔG^{PL}_{fix} and ΔG^{P}_{fix} are the free-energy costs of rigidifying the protein in the presence or absence of the ligand, respectively. It is useful to consider the ideal situation where the protein atoms surrounding the binding site behave as a harmonic oscillator. In such a case, the cost of changing the mobility from a state i to a reference state ref can be expressed as

$$\Delta G_{fix} = -RT \ln \frac{Z_{ref}}{Z_i} = -\frac{1}{2}RT \ln \frac{k_i}{k_{ref}} = -\frac{1}{2}RT \ln \frac{\langle \Delta x^2 \rangle_{ref}}{\langle \Delta x^2 \rangle_i} \quad (5)$$

where Z is the partition function, k the spring constant and $\langle \Delta x^2 \rangle$ the mean squared displacement (see the Supporting Information for details). Introducing the last part of eq 5 in eq 4 and using a common reference state, $\Delta\Delta G_{bind}$ ($\Delta G_{bind} - \Delta G'_{bind}$) can be expressed as

$$\Delta\Delta G_{bind} = \Delta G^{P}_{fix} - \Delta G^{PL}_{fix} = -\frac{1}{2}RT\left(\ln \frac{\langle \Delta x^2 \rangle_{ref}}{\langle \Delta x^2 \rangle_P} - \ln \frac{\langle \Delta x^2 \rangle_{ref}}{\langle \Delta x^2 \rangle_{PL}}\right) = \frac{1}{2}RT \ln \frac{\langle \Delta x^2 \rangle_P}{\langle \Delta x^2 \rangle_{PL}} \quad (6)$$

This expression suggests that the change in binding free energy upon protein rigidification should bear a logarithmic relationship with the change of atomic mobility upon complex formation. In other words, the binding free energy derived from restrained or unrestrained simulations should be very similar for intrinsically rigid sites, where ligand binding barely alters the mobility of the site. If, on the other hand, ligand binding in the unrestrained state causes a significant reduction of mobility, the difference in calculated binding free energies will be larger but still limited by the logarithmic relationship. This simplified model provides an initial guess of the type of relationship that one should be looking for.

Empirically, we have measured the binding free energy of EtOH at 17 different sites on the surface of HEWL (Figure S4 of the Supporting Information) in the free state (ΔG_{bind}) and after applying strong harmonic restraints (k = 1 kcal/mol Å^2) on the nonhydrogen protein atoms ($\Delta G'_{bind}$). The sites were selected for having an estimated binding free energy below −0.3 kcal/mol, and they include the C site (named hp1), sites validated by superposition with crystal structures (Figure S1 of the Supporting Information: names hp3, hp4, hp6, hp9, hp12, hp13, and hp14) and others randomly chosen. We have also calculated the average B-factor ($B = 8\pi^2 \langle \Delta x^2 \rangle$) of the protein atoms surrounding the hot spots in the free and restrained simulations (B and B', respectively). Plotting the relative change of binding free energy ($\Delta G'_{bind}/\Delta G_{bind}$) versus the logarithm of the B-factors (Figure 3), we can see that most sites have a similar slope (−0.4 ± 0.15) and therefore the change in ΔG_{bind} caused by protein rigidification can be anticipated from the change in mobility by the following relationship:

$$\frac{\Delta G'_{bind}}{\Delta G_{bind}} = 1 - 0.4(\ln B' - \ln B) \quad (7)$$

$$\frac{\Delta G'_{bind}}{\Delta G_{bind}} = 1 - 0.4\left(\ln \frac{B'}{B}\right) \quad (8)$$

Which can be rewritten as

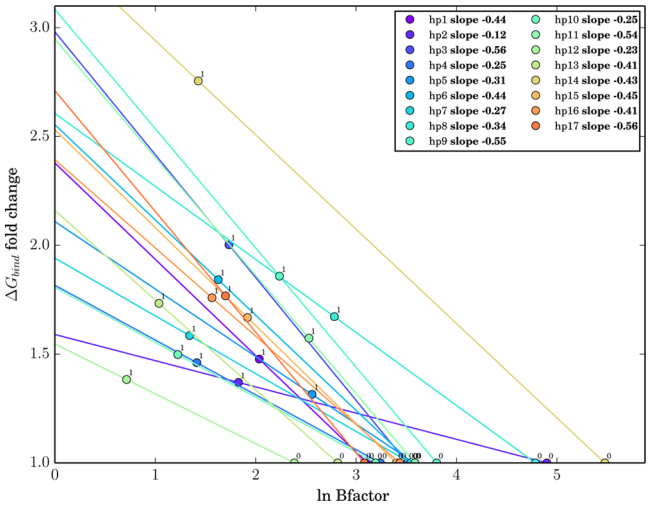


Figure 3. Binding free energies (relative to the free simulation: $\Delta G'_{bind}/\Delta G_{bind}$) plotted against the mobility of the binding site (in logarithmic scale) corresponding to eq 7. The binding free energy of EtOH and the B-factor of the binding site was calculated at 17 different positions from free (ΔG_{bind}; B) and restrained ($\Delta G'_{bind}$; B') MD simulations. Restricted mobility has similar consequences for most binding sites, with an average slope of −0.4 ± 0.15.

$$\Delta\Delta G_{bind} = \Delta G_{bind} 0.4\left(\ln \frac{B'}{B}\right) \quad (9)$$

It is noteworthy that binding sites with very different structural and chemical properties conform to the same relationship. For instance, the C site is mostly hydrophobic, relatively rigid and occluded, while hp14 (shown in Figure S5 of the Supporting Information), is very exposed, flexible, and positively charged. Though the theoretical relationship (6) and the empirical one (9) have similar forms, they are not equivalent, but it is comforting to see that in both cases the change in binding free energy has a logarithmic relationship with the change in mobility of the binding site. In order to gain confidence on the empirically derived equation, we have carried out additional simulations with intermediate restraining forces [0.1 kcal/(mol Å^2) and 0.01 kcal/(mol Å^2)]. As shown in Figure S6 of the Supporting Information, a majority of the binding sites (11 out of 17) adapt to the previously described empirical relationship with high correlation coefficients ($r^2 > 0.8$). This shows that, in many cases, the change in binding free energy depends on, and can be predicted from, the initial mobility of the site. Indeed, enforced rigidification has a relatively mild effect on intrinsically rigid sites such the C-site (overestimated by ∼1.5-fold) but leads to much larger errors on flexible regions, where ΔG_{bind} can be overestimated by a factor of 3. This finding indicates that the same approximation (e.g., rigid receptor docking) may be acceptable or not depending on the nature of the binding site.

Protein Rigidity Alters the Ligand Exchange Pathway. As discussed, protein rigidity generally leads to more favorable binding free energies. However, Lexa and Carlson reported that the hot spot corresponding to the C site of HEWL essentially disappeared when the protein was kept rigid.[5] We find that the C-site adopts the same conformation in the apo and holo states, discarding the possibility of a ligand-induced effect. Further, in our simulations with EtOH, binding affinity increases upon protein rigidification, as expected (Figure 1b). Since Lexa and Carlson used ACN as organic solvent, we have replicated those experiments in order to understand the causes of such conflicting

results. Surprisingly, ACN does not visit the C site in the course of 60 ns, while in the flexible simulation, we observe several binding events for the same period of time, but we also noted that if an ACN molecule is placed in the binding site in the beginning of the restrained simulations, it remains there for the whole period (Figure S7 of the Supporting Information). This indicates that ACN molecules cannot diffuse in and out of the binding site when the protein is held rigid, which is explained by the tight fit of the complex (Figure S8 of the Supporting Information) and the rigid nature of ACN. Entering and leaving the binding site likely requires a small breathing movement on the protein side, which cannot be accomplished if the protein atoms are restrained. In spite of the similarity between EtOH and ACN, both in terms of the chemical structure and in the way they bind to the C site,[30] the more compact shape of EtOH allows it to exchange in all simulations (Figure S9 of the Supporting Information). Even when the protein adopts the same conformation in the apo and holo states, as it is the case here, transient fluctuations may be necessary to facilitate ligand exchange. Not surprisingly then, one of the consequences of making the protein rigid is an alteration of the association pathway and a concomitant increment in the energy of the transition state, leading to slower on- and off-rates. In itself, a change in the association pathway should not affect ΔG_{bind} because it is a state function. But in the case of ACN, protein rigidity creates an insurmountable barrier, making it impossible to recover the bound state. Naturally, thermodynamic parameters can only be inferred if sampling is correct and solvent exchange should be closely monitored in MDmix simulations. End-state methods, on the other hand, simply ignore the binding process, and it is therefore not surprising that binding of ACN to the C-site is correctly predicted by FTMAP and GCMC, even if the protein is represented as a rigid body.[6,31]

Protein Rigidity Facilitates Convergence of Calculated ΔG_{bind}. Sampling the protein conformational space requires extremely long simulations, and obtaining an equilibrated distribution of the solvent (and ligands) around the protein for each one of the main conformational states makes convergence even harder to achieve. This is exemplified in Figure 4, where the predicted binding free energy of EtOH for the C-site of HEWL is calculated along three independent 200 ns MD simulations. Starting from the same initial structure, each replica reaches different conformational states with characteristic binding free energies and solvent-exchange properties (Figures S10−S11 of the Supporting Information). The macroscopic binding affinity corresponds to the average between these and other states, with exchange rates between them far surpassing the time that is amenable to simulation. In such circumstances, limiting protein flexibility provides a significant advantage in terms of convergence, but this should be balanced against the error in ΔG_{bind} caused by the artificial restraints. This is illustrated in Figure 5, where the average ΔG_{bind} of EtOH for the C-site is plotted against the average B-factor obtained under different restraining schemes. As indicated by the error bars, unrestrained simulations span a wide range both in the B-factor and the ΔG_{bind} axes. Expanding the simulation time changes the average value but retains the variability. At the other end, strong restraints produce simulations that are invariable in the B-factor coordinate and have very small standard deviation of the ΔG_{bind} value even after very short simulation times (20 ns). But, as discussed in the previous section, this leads to a significant error in the estimated value. Weak restraints offer a good compromise between precision and accuracy since, contrary to unrestrained simulations, predicted ΔG_{bind} from independent replicas converge

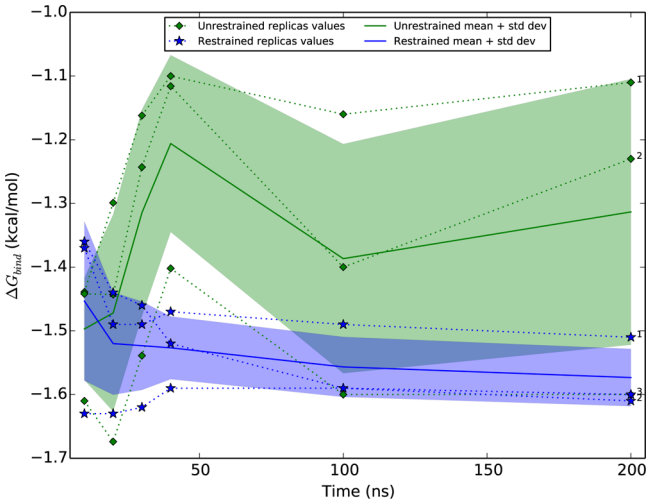


Figure 4. Convergence of binding free-energy predictions. Predicted ΔG_{bind} for the hot spot located at the C-site is represented at various time points along the 200 ns trajectories. Three independent replicas (dotted lines), their average value (solid lines), and the standard deviation (colored areas) are represented for the FREE (green) and HA001 (i.e., soft restraints) simulations (blue). Data can be found in Table S1 of the Supporting Information.

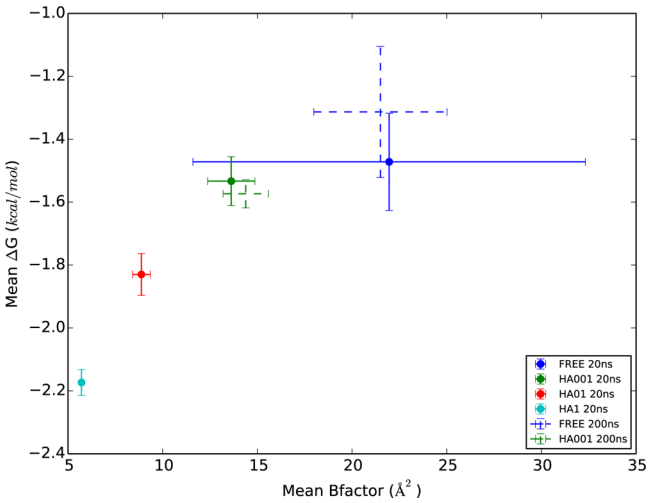


Figure 5. B-factor and binding free-energy variability in the C-site of HEWL, as a function of time and restraining scheme. Mean values (dots) and standard deviations (error bars) are calculated from 3 independent replicas of EtOH MDmix simulations under different restraining forces [FREE: no restraints; HA001: $k = 0.01$; HA01: $k = 0.1$; HA1: $k = 1$ kcal/(mol Å^2)] and simulation times (20 and 200 ns).

rapidly and continuously (Figure 4). Such mild restraints prevent large conformational changes but do allow for local conformational sampling. Accordingly, the B-factors are reduced by a factor of approximately 2. On virtue of the logarithmic relationship between mobility and ΔG_{bind}, this causes only a small overestimate of this property. Using mild restraints, one desists from exploring the whole conformational space available to the protein, but this has a small impact on accuracy and sampling becomes only limited by the diffusion rate of the ligand. For all the organic solvents investigated here, 20−50 ns are sufficient to obtain consistent results under these circumstances. Even for an intrinsically rigid site, such as the C-site of HEWL, unrestrained simulations would require several orders of

dx.doi.org/10.1021/ct500182z | J. Chem. Theory Comput. 2014, 10, 2608−2614

magnitude longer MD simulation to reach a similar level of confidence.

Protein Rigidity Facilitates Interpretation and Transferability of Calculated ΔG_{bind}. Often, proteins are probed with specific atom types or chemical moieties, and the results are extrapolated to larger ligands. This is the case in hot spot mapping,[32] identification of potential pharmacophores,[33] or in precalculated interaction potentials used in docking programs.[34] In such cases, it is important that interaction patterns can be visually interpreted and that ΔG_{bind} values obtained with a probe are transferable to larger ligands. Full protein flexibility may not be advisible in such cases because some protein regions adopt conformations with very different molecular recognition patterns that, when averaged out, become uninterpretable. This is illustrated on 3 EtOH binding hot spots that appear next to the main binding site (Figure S12 of the Supporting Information). In the absence of restraints, residues Trp62, Trp63, Arg73, Leu75, and Asp101, explore two different states: an open conformation corresponding to the crystal structure (Figure S12a of the Supporting Information) and a closed conformation in which Arg73 bends and closely interacts with Asp101, rendering two of those spots partly inaccessible (Figure S12b of the Supporting Information). The closed conformation was also observed during the simulation in pure water, excluding the possibility of an artifact or denaturation due to the cosolvent. The subset of the flexible trajectory adopting conformations similar to the crystallographic structure provides equivalent results to the restrained simulation, while in the fully flexible simulation, the two states become entangled. Large ligands will preselect either one of those two conformations, hence it would be more meaningful to use ΔG_{bind} values derived from the relevant conformation rather than from the scrambled states explored in the free simulation. Additionally, in the absence of restraints, the protein may adopt solvent-specific conformations or even denature,[35] in which case the conformational sampling obtained with the probe would be irrelevant for the ligand-bound state.

CONCLUSIONS

Proteins are very flexible entities, and their conformational freedom has major consequences for binding. Taking protein flexibility fully into account would seem the ideal solution, but it presents major methodological as well as practical challenges. In contrast, using a single receptor structure is a very crude approximation but its simplicity is very appealing and has been used with reasonable success for many years. Striking the right balance between those two extreme situations is key to achieve satisfactory results. Here we have studied the relationship between protein flexibility and binding, obtaining clues to choose the necessary level of flexibility for a given application. Proteins should never be constrained when the goal of the simulation is to recapitulate the binding event and/or the kinetic parameters because, as experimentally observed in viscous environments,[36] dampening protein mobility causes alteration of the ligand exchange pathways. In extreme cases, ligand exchange may be seriously hampered, preventing the system from reaching thermodynamic equilibrium. This effect explains the lack of ACN density observed in the C-site of HEWL, rather than an intrinsic limitation of rigid structures to identify proper minima, as was suggested by Lexa and Carlson.[5] Besides inadequate sampling, neglecting protein flexibility will only fail to find true minima if the selected conformation is not suitable for binding (e.g., cryptic sites, lack of ligand fit). In all other cases, protein rigidity favors binding. This is true both for natural and simulated systems: hot spots locate on intrinsically rigid parts of the protein,[22] while artificial rigidity leads to exaggerated binding affinities. We have found a logarithmic relationship between those two variables, meaning that small changes in protein flexibility have minimal consequences on ΔG_{bind}, but absolute rigidity may cause errors of several fold in the predicted value, particularly on very flexible sites. This indicates that using a single structure may be acceptable for binding sites that are intrinsically and homogeneously rigid but is a poor approximation on binding sites with a heterogeneous distribution of flexibility because interactions with the more mobile parts will be artificially exaggerated relative to others. Interestingly, a large majority of structure-based drug discovery efforts have focused on enzyme active sites, which display a much lower level of mobility than the average protein surface.[37] Although formally incorrect, the lock-and-key approximation may be approximately valid in those pockets. Our results warn that the same should not be expected on more dynamic sites such as those found in GPCRs, protein−protein interfaces, or allosteric pockets. We have also shown that most hot spots display a change in binding free energy that can be predicted from the mobility of the protein atoms in the free state. This information could be used to take flexibility into account in an implicit manner, a possibility that we will investigate in the future. Finally, we have shown how unrestrained protein mobility hampers convergence, transferability, and interpretability of the results. Soft restraints appear as an attractive option to allow local conformational sampling with a minimal impact on predicted ΔG_{bind}, affording significant gain in computational efficiency and preserving the univocal relationship between a given protein structure and the calculated interaction potentials, which facilitates visual interpretation. Taking larger conformational changes into account may require full protein flexibility, but it is also possible that, as is the case for docking, using a few carefully selected experimental structures may be a more realistic and productive alternative.[38,39] Ultimately, choosing the right level of flexibility will depend on the goal of each individual investigation and the nature of each particular system. We trust this study provides useful hints to make the right decision.

ASSOCIATED CONTENT

Supporting Information

Additional information on the preparation of the solvent mixtures for simulation, eq 5 derivation, and additional images, tables, and figures. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

* Address: Departament de Fisicoquímica, Facultat de Farmàcia, Av. Joan XXIII s/n, 08028 Barcelona, Spain. E-mail: xbarril@ub.edu. Tel: +34 934031304. Fax: +34 934035987.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank the Barcelona Supercomputing Center for access to computational resources. This work was financed by the Spanish Ministerio de Economia (SAF2009-08811 and SAF2012-33481). D.A.G. is funded by the University of Barcelona. We thank the reviewers for their very useful comments and Andreana M. Rosnik for editing the abstract.

REFERENCES

(1) Gohlke, H.; Klebe, G. Approaches to the Description and Prediction of the Binding Affinity of Small-Molecule Ligands to Macromolecular Receptors. *Angew. Chem., Int. Ed.* **2002**, *41*, 2644−2676.

(2) Jorgensen, W. L. The Many Roles of Computation in Drug Discovery. *Science* **2004**, *303*, 1813−1818.

(3) Cozzini, P.; Kellogg, G. E.; Spyrakis, F.; Abraham, D. J.; Costantino, G.; Emerson, A.; Fanelli, F.; Gohlke, H.; Kuhn, L. A.; Morris, G. M.; Orozco, M.; Pertinhez, T. A.; Rizzi, M.; Sotriffer, C. A. Target Flexibility: An Emerging Consideration in Drug Discovery and Design. *J. Med. Chem.* **2008**, *51*, 6237−6255.

(4) Spyrakis, F.; BidonChanal, A.; Barril, X.; Luque, F. J. Protein Flexibility and Ligand Recognition: Challenges for Molecular Modeling. *Curr. Top. Med. Chem.* **2011**, *11*, 192−210.

(5) Lexa, K. W.; Carlson, H. A. Full Protein Flexibility Is Essential for Proper Hot-Spot Mapping. *J. Am. Chem. Soc.* **2011**, *133*, 200−202.

(6) Hall, D. R.; Grove, L. E.; Yueh, C.; Ngan, C. H.; Kozakov, D.; Vajda, S. Robust Identification of Binding Hot Spots Using Continuum Electrostatics: Application to Hen Egg-White Lysozyme. *J. Am. Chem. Soc.* **2011**, *133*, 20668−20671.

(7) Buch, I.; Giorgino, T.; Fabritiis, G. De. Complete Reconstruction of an Enzyme-Inhibitor Binding Process by Molecular Dynamics Simulations. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 10184−10189.

(8) Shan, Y.; Kim, E. T.; Eastwood, M. P.; Dror, R. O.; Seeliger, M. a; Shaw, D. E. How Does a Drug Molecule Find Its Target Binding Site? *J. Am. Chem. Soc.* **2011**, *133*, 9181−9183.

(9) Dror, R. O.; Pan, A. C.; Arlow, D. H.; Borhani, D. W.; Maragakis, P.; Shan, Y.; Xu, H.; Shaw, D. E. Pathway and Mechanism of Drug Binding to G-Protein-Coupled Receptors. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 13118−13123.

(10) Beauchamp, K. A.; Lin, Y.-S.; Das, R.; Pande, V. S. Are Protein Force Fields Getting Better? A Systematic Benchmark on 524 Diverse NMR Measurements. *J. Chem. Theory Comput.* **2012**, *8*, 1409−1414.

(11) Swinney, D. C. The Role of Binding Kinetics in Therapeutically Useful Drug Action. *Curr. Opin. Drug Discovery Dev.* **2009**, *12*, 31−39.

(12) Mattos, C.; Bellamacina, C. R.; Peisach, E.; Pereira, A.; Vitkup, D.; Petsko, G. A.; Ringe, D. Multiple Solvent Crystal Structures: Probing Binding Sites, Plasticity and Hydration. *J. Mol. Biol.* **2006**, *357*, 1471−1482.

(13) Liepinsh, E.; Otting, G. Organic Solvents Identify Specific Ligand Binding Sites on Protein Surfaces. *Nat. Biotechnol.* **1997**, *15*, 264−268.

(14) Huang, D.; Caflisch, A. The Free Energy Landscape of Small Molecule Unbinding. *PLoS Comput. Biol.* **2011**, *7*, e1002002.

(15) Seco, J.; Luque, F. J.; Barril, X. Binding Site Detection and Druggability Index from First Principles. *J. Med. Chem.* **2009**, *52*, 2363−2371.

(16) Wang, Z.; Zhu, G.; Huang, Q.; Qian, M.; Shao, M.; Jia, Y.; Tang, Y. X-Ray Studies on Cross-Linked Lysozyme Crystals in Acetonitrile-Water Mixture. *Biochim. Biophys. Acta* **1998**, *1384*, 335−344.

(17) *Molecular Operating Environment (MOE)*, version 2013.8; Chemical Computing Group Inc.: Montreal, Canada, 2013.

(18) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber Biomolecular Simulation Programs. *J. Comput. Chem.* **2005**, *26*, 1668−1688.

(19) General, I. J. A Note on the Standard State's Binding Free Energy. *J. Chem. Theory Comput.* **2010**, *6*, 2520−2524.

(20) Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **2007**, *9*, 90−95.

(21) Dennis, S.; Kortvelyesi, T.; Vajda, S. Computational Mapping Identifies the Binding Sites of Organic Solvents on Proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 4290−4295.

(22) Kuttner, Y. Y.; Engel, S. Protein Hot Spots: The Islands of Stability. *J. Mol. Biol.* **2012**, *415*, 419−428.

(23) Bogan, A. A.; Thorn, K. S. Anatomy of Hot Spots in Protein Interfaces. *J. Mol. Biol.* **1998**, *280*, 1−9.

(24) Keskin, O.; Ma, B.; Nussinov, R. Hot Regions in Protein−Protein Interactions: The Organization and Contribution of Structurally Conserved Hot Spot Residues. *J. Mol. Biol.* **2005**, *345*, 1281−1294.

(25) Liu, J.; Faeder, J. R.; Camacho, C. J. Toward a Quantitative Theory of Intrinsically Disordered Proteins and Their Function. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 19819−19823.

(26) Giorgino, T.; Buch, I.; De Fabritiis, G. Visualizing the Induced Binding of SH2-Phosphopeptide. *J. Chem. Theory Comput.* **2012**, *8*, 1171−1175.

(27) McGovern, S. L.; Shoichet, B. K. Information Decay in Molecular Docking Screens against Holo, Apo, and Modeled Conformations of Enzymes. *J. Med. Chem.* **2003**, *46*, 2895−2907.

(28) Verdonk, M. L.; Mortenson, P. N.; Hall, R. J.; Hartshorn, M. J.; Murray, C. W. Protein-Ligand Docking against Non-Native Protein Conformers. *J. Chem. Inf. Model.* **2008**, *48*, 2214−2225.

(29) Novoa, E. M.; Pouplana, L. R.; De Barril, X.; Orozco, M. Ensemble Docking from Homology Models. *J. Chem. Theory Comput.* **2010**, *6*, 2547−2557.

(30) Deshpande, A.; Nimsadkar, S.; Mande, S. C. Effect of Alcohols on Protein Hydration: Crystallographic Analysis of Hen Egg-White Lysozyme in the Presence of Alcohols. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2005**, *61*, 1005−1008.

(31) Kulp, J. L.; Pompliano, D. L.; Guarnieri, F. Diverse Fragment Clustering and Water Exclusion Identify Protein Hot Spots. *J. Am. Chem. Soc.* **2011**, *133*, 10740−10743.

(32) Landon, M. R.; Lancia, D. R.; Yu, J.; Thiel, S. C.; Vajda, S. Identification of Hot Spots within Druggable Binding Regions by Computational Solvent Mapping of Proteins. *J. Med. Chem.* **2007**, *50*, 1231−1240.

(33) Sotriffer, C.; Klebe, G. Identification and Mapping of Small-Molecule Binding Sites in Proteins: Computational Tools for Structure-Based Drug Design. *Farm* **2002**, *57*, 243−251.

(34) Brooijmans, N.; Kuntz, I. D. Molecular Recognition and Docking Algorithms. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 335−373.

(35) Foster, T. J.; MacKerell, A. D.; Guvench, O. Balancing Target Flexibility and Target Denaturation in Computational Fragment-Based Inhibitor Discovery. *J. Comput. Chem.* **2012**, *33*, 1880−1891.

(36) Beece, D.; Eisenstein, L.; Frauenfelder, H.; Good, D.; Marden, M. C.; Reinisch, L.; Reynolds, A. H.; Sorensen, L. B.; Yue, K. T. Solvent Viscosity and Protein Dynamics. *Biochemistry* **1980**, *19*, 5147−5157.

(37) Yuan, Z.; Zhao, J.; Wang, Z.-X. Flexibility Analysis of Enzyme Active Sites by Crystallographic Temperature Factors. *Protein Eng.* **2003**, *16*, 109−114.

(38) Barril, X.; Morley, S. D. Unveiling the Full Potential of Flexible Receptor Docking Using Multiple Crystallographic Structures. *J. Med. Chem.* **2005**, *48*, 4432−4443.

(39) Rueda, M.; Bottegoni, G.; Abagyan, R. Recipes for the Selection of Experimental Protein Conformations for Virtual Screening. *J. Chem. Inf. Model.* **2010**, *50*, 186−193.