JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Monte Carlo Free Ligand Diffusion with Markov State Model Analysis and Absolute Binding Free Energy Calculations

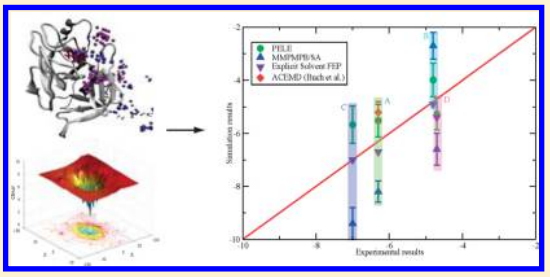
Ryoji Takahashi,[†] Víctor A. Gil,[†] and Victor Guallar*[,†,‡]

[†]Joint BSC-IRB Research Program in Computational Biology, Barcelona Supercomputing Center, c/Jordi Girona 29, 08034 Barcelona, Spain

[‡]Institució Catalana de Recerca i Estudis Avançats (ICREA), Passeig Lluís Companys 23, 08010 Barcelona, Spain

Supporting Information

ABSTRACT: Obtaining absolute binding free energies from unbiased ligand diffusion has attracted a significant amount of attention due to its implications in drug design. Several studies have used special purpose computers and software to achieve microsecond molecular dynamics which, combined with a Markov state model analysis, are capable of providing absolute binding free energies. We have recently developed a Monte Carlo based technique, PELE, capable of performing a dynamical exploration of the protein−ligand energy landscape including free ligand diffusion into the active site, at a fraction of the computational cost of molecular dynamics techniques. We demonstrate here the capabilities of our Monte Carlo technique in obtaining absolute binding free energies for a series of benzamidine like inhibitors into trypsin. Our results are in good agreement with experimental data and other molecular dynamics simulations, indicating that PELE can be a useful tool for quick estimates of binding free energies and mechanisms.



1. INTRODUCTION

Obtaining absolute (and relative) binding free energies from an atomic modeling of the protein−ligand conformational space is a challenging task in biophysics. From quick approximate docking calculations to more accurate slow molecular dynamics (MD) techniques, obtaining binding free energies estimations have centered a significant amount of work due to its implication in drug design.[1,2] Most biophysical all-atom modeling techniques have been applied to address ligand binding processes, such as, thermodynamics integration,[3] umbrella sampling,[4] steered molecular dynamics,[5−7] and metadynamics.[8] These techniques belong to the family of potential mean field (PMF) biased methods, explorations within (into/away from) the active site. Remarkably, using specialized hardware such as the Anton computer (http://www.deshawresearch.com/) or graphical processing units (GPUs), protein−ligand unbiased MD simulations have evolved rapidly in the recent years.[9−11] The pioneering studies from the Shaw research group, for example, showed that by running microsecond long MD it is possible to model the spontaneous binding of a ligand from the solvent to the active site.[12]

Combining these long MD simulations with Markov state models (MSMs),[13,14] where "Markovianity" stands for memoryless transitions between kinetic microstates and macrostates, absolute binding free energies from free diffusion simulations are now available.[10,11] MSM is now being used for the analyses of protein folding,[15−20] conformational transitions,[21−23] and protein−ligand binding.[10,24] The current state of the MSM and its validation are reported in ref 13.

In our group we have developed PELE (protein energy landscape exploration), a hybrid method combing Monte Carlo (MC) sampling with protein structure prediction techniques capable of quickly describing the protein and the protein−ligand conformational space. PELE, initially designed to map biased ligand entrance and exit pathways,[25] has been shown to accurately reproduce nonbiased ligand diffusion and active site search,[26] induced fit docking,[27,28] and overall protein dynamics,[29] at a fraction of the computational cost of molecular dynamics techniques.

Sufficient MC sampling combined with the right analytical tools can provide equilibrium states and free energy difference between them, as well as transition pathways.[30] In this paper, we report that PELE's random ligand exploration, when combined with a MSM analysis, is capable of accurately describing these stationary properties, in a similar fashion to extensive MD simulations. We test our methodology with trypsin−benzamidine binding, which has also been used as a benchmark for MD calculations by Buch et al.[10] Furthermore, we expand our studies to three more ligands: 1,4-diazamidine, 4-amino-benzamidine, and benzylamin, as shown in Figure 1, for which experimental binding free energies have been reported.[31−37] Our results show good agreement with experimental data and previous MD simulations, indicating that PELE can be a useful tool for quick estimates of binding free energies and mechanisms.

Received: July 30, 2013
Published: December 9, 2013

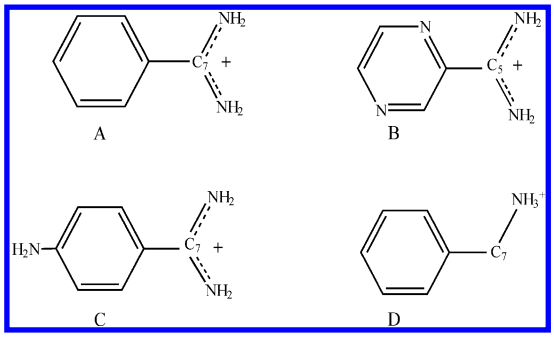


Figure 1. Chemical structures of trypsin ligands: (A) benzamidine, (B) 1,4-diazamidine, (C) 4-amino-benzamidine, and (D) benzylamine. The carbon atoms C_5 and C_7, used in the clustering processes, are underlined.

2. METHODS

The unbiased ligand's diffusion has been modeled with PELE, and the resulting trajectories have been analyzed with EMMA software.

2.1. PELE Sampling. PELE's method combines a stochastic approach with protein structure prediction tools. New trial configurations are produced with sequential ligand/protein perturbation, side chain prediction, and minimization steps.

Perturbation. The ligand is perturbed through random rotations and translations. Furthermore, ligand internal degrees of freedom are taken into account by building a ligand specific rotamer library (see the work of Borrelli et al. for more details[25]). In the case of the protein, the perturbation is based on the α-carbons anisotropic network model (ANM);[38] all atoms are displaced by a minimization where the α-carbons are forced to follow a randomly picked low eigenvector (or a combination of them) obtained in the ANM approach.

Side Chain Prediction. The side chain sampling step uses Xiang and Honig's rotamer libraries[39] to predict all local side chains to the ligand (within a defined distance), as well as those side chains with higher energy increase along the previous ANM step.

Minimization. Finally, the last stage involves the minimization of a region including, at least, all local residues to the atoms involved in the perturbation and side chain steps. The minimization algorithm uses a fast multiscale truncated Newton method[40,41] and includes a harmonic constraint in all α-carbons. These harmonic constraints enforce the (random) backbone sampling obtained in the initial perturbation step.

The combination of these three steps forms a *move* that is accepted (defining a new minimum) or rejected based on a Metropolis criterion. PELE uses an OPLS-AA force field (next version will include AMBER) with implicit surface-generalized Born (SGB).[42,43] The combination of ligand and protein perturbations explores effectively the energy landscape, reproducing large conformational changes along ligand migration. These perturbations, associated to random processes, constitute the major part of the sampling procedure. As explained above, they are maintained along the relaxation (minimization) step, enforcing the stochastic nature of the sampling and the maintenance of (an approximation to) detailed balance.

2.2. Building A Markov State Model with EMMA. Detailed descriptions of MSM analysis and the software, EMMA, can be found in refs 13 and 14 and on the Web site https://simtk.org/home/emma. Briefly, the analysis of PELE's trajectories involved:

(i) Clustering. The first step is to determine microstates using geometrical clustering. The Cartesian coordinates of the amidine's carbon (C_5 of 1,4-diazamidine and C_7 of benzamidine, 4-amino-benzamidine, and benzylamine) were subtracted from the PELE's trajectories and used as a ligand's locator. Then, the microsates were constructed with the *K*-means clustering method (Lloyd's algorithm[44]) using an Euclidean distance metric.

(ii) Connectivity. MSMs need to have a connected set of microstates. Using Tarjan's algorithm,[45] the connectivity between the microstates was examined. The largest set of connected microstates was chosen.

(iii) Implied time scale. For a model to be Markovian, a lagtime τ must be found during which its transition probabilities are time independent. The eigenvalues λ_i of the counting matrix $\mathbf{C}(\tau)$, where c_{ij} elements are the number of transitions from state i to j between time t and $t + \tau$, should decay exponentially in time. By testing an arbitrary τ, we are able to find an implied time scale ($t^*(\tau) = -\tau/\ln|\lambda_i(\tau)|$), from which the transition probability matrix $\mathbf{T}(\tau)$ can be approximated.

(iv) Stationary distribution. When the molecular system is in thermal equilibrium, the stationary states are given by the average of the microscopic Boltzmann distributions. The above eigenvalues represent the different stationary distributions for the (clustered) microstates, where $\lambda_1 = 1$ corresponds to the active site. The stationary distribution π can then be obtained as $\pi^T = \pi^T\mathbf{T}(\tau)$, with the PMF profile given by the Boltzmann inversion of the stationary distribution, $G_i = -k_BT \ln \pi_i$, where k_B is the Boltzmann constant and T is the temperature.

(v) Binding Free Energy. Binding free energies can be defined as $\Delta G_0 = -k_BT \ln(V_b/V_0) - \Delta\mathcal{W}$,[10] where V_b is the bounded volume of the G_{PMF} surface and $\Delta\mathcal{W}$ is a difference from the average value of the bulk to the minimum value of the G_{PMF} surface. For all calculations, the standard-state volume V_0 = 1661 Å^3 (1 M concentration)[46] and T = 300 K were used.

2.3. Simulation Setup and PELE Parameters. Initial coordinates for the bovine pancreatic trypsin were taken from the PDB file 3PTB.[47] The original structure was optimized with Schrödinger's protein wizard. In particular His40 and His91 were changed to Hie (epsilon protonated histidine), all other residues kept their standard protonation state. Ligand charges were extracted from the ab initio electrostatic potential, by means of a gas phase DFT-B3LYP optimization with the 6-31G^{**+} basis set. As shown in Figure 2, for each ligand six random different initial positions were produced with the ligand being away from the surface—the ligand's solvent accessible area (SASA) being 1. At each initial condition 140 PELE trajectories were started, for a total of 840 trajectories being produced for each ligand. Each trajectory was stopped after reaching 1000 MC steps of unbiased free ligand exploration combining long (7.0 Å) and short (0.5 Å) ligand's translation perturbations. Side chains within 5 Å from the ligand were included for resampling. The ANM backbone perturbation used the lowest 6 modes with a displacement of 1.2 Å. In order to avoid ligands diffusing away, a weak harmonic restrain of $k = 5 \times 10^{-4}$ kcal mol^{-1} Å^{-2} was applied to the ligand and the α-carbon of Pro198, which is located approximately at the center

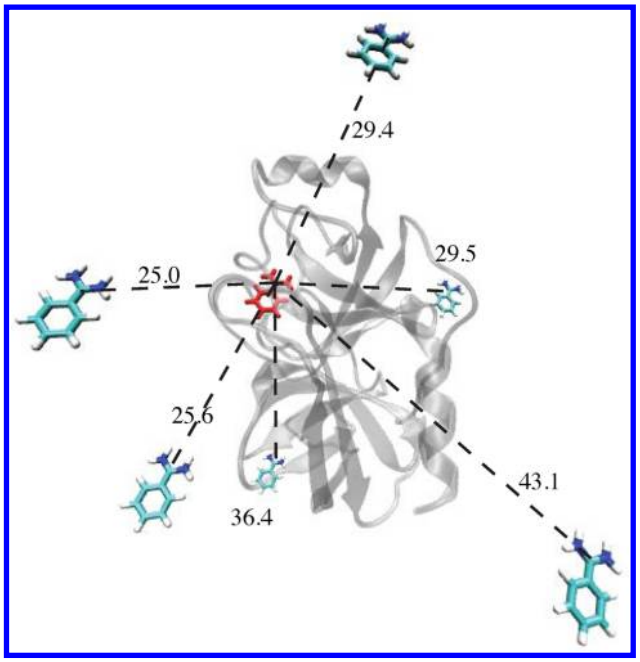


Figure 2. Initial positioning for the ligands before diffusion simulations. Distances to the crystal structure bound ligand (red color) are also shown. Distance units are in angstroms.

of the trypsin. Each PELE trajectory was obtained in about 12 CPU h using one core of a Xeon E5-2670 8C CPU at 2.600 GHz. Total simulation for each ligand required about 1 week using 64 cores (less than a couple of days using 256 cores).

3. RESULTS AND DISCUSSION

3.1. Ligand Diffusion. A set of 24 trajectories for benzamidine, shown in Figure 3, illustrates the behavior for the overall trajectories and ligands. Panel I shows the heavy atom ligand's RMSD to the native 3PTB structure along the accepted MC steps. Panel II indicates the ligand's interaction energy against its RMSD to the native (bound crystal structure, 3PTB). The interaction energy is the result of the complex energy minus the isolated protein and ligand energies at the current complex structure. Finally, panel III displays the selected benzamidine C7 positions along two selected trajectories (see below) after protein superposition (only one protein structure is shown in ribbons). Panel I (and II) clearly shows long excursions of the ligand into the bulk solvent indicating that the weak constraint only starts being effective at about 30 Å from the protein surface. By following the RMSD evolution in panel I, we can observe several trajectories finding (and very few leaving) the active site. More importantly, and in analogy to recent long molecular dynamics for the Src kinase by the Shaw research group,[12] we clearly recognize the active site by following the interaction energy (panel II). To underline the ligands' behavior, two different trajectories are highlighted with blue and magenta colors in all three panels. In the blue trajectory, the ligand spends significant time in the bulk solvent finding the active site at about the 250 MC accepted step. In the magenta trajectory, however, the ligand quickly finds the protein surface and performs a larger exploration of it before entering the active site at the ∼300 MC accepted step.

As shown below, notice that about 600−800 trajectories are necessary for absolute binding free energies to converge. However, it is a remarkable advance with respect to MD simulations, that by running 20−30 trajectories for 5−10 CPU h each we can identify the binding site producing several binding events. Such capabilities of the PELE software have already been shown in globin systems,[26,48] and additional examples (for example with addition of explicit waters) are displayed on our server http://www.pele.bsc.es where the software can be run for free.[49]

3.2. Ligand's Binding Pose. Figure 4 shows the active site distribution of bound complexes for benzamidine, comparing them with the crystal structure after α-carbon superposition. The heavy atom ligand's RMSD for most of the structures (accounting for ligand's rotation symmetries) is below 2 Å, with a good overall correlation between the interaction energy and the RMSD. The distribution of structures for the other ligands are very similar, indicating that for these systems an RMSD < 3 Å is enough to discriminate bound complexes. Using this value, the summary of binding events for each ligand is as follows: 16% for benzamidine, 16% for 1,4-diazamidine, 10% for 4-amino-benzamidne, and 12% for benzylamine.

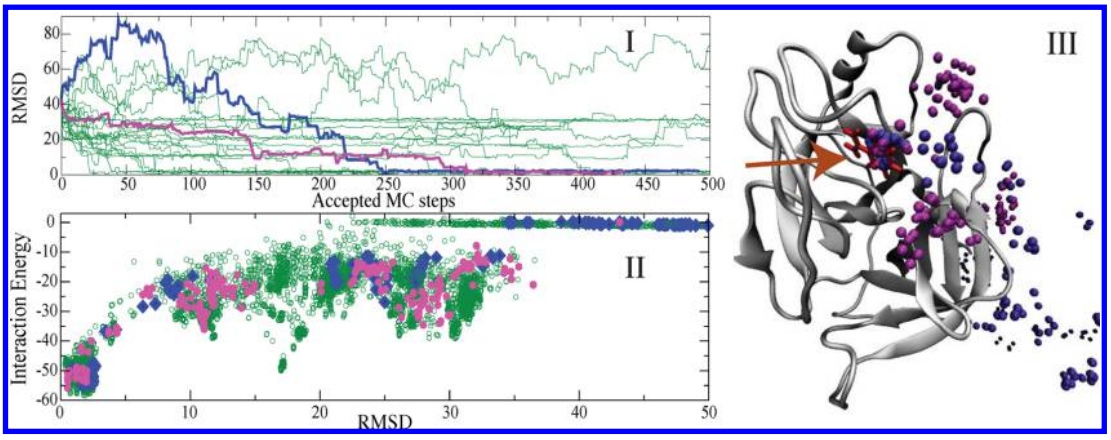


Figure 3. Twenty-four representative trajectories for the benzamidine migration. Panel I shows the ligand RMSD evolution versus the accepted MC steps. Panel II displays the protein−ligand interaction energy versus the ligand's RMSD. The RMSD was computed using the ligand's heavy atoms after superimposing all the α-carbons to the crystallographic structure. Panel III shows the ligand migration for the two selected trajectoires after protein superposition (only one protein frame is shown with ribbons). The arrow helps to locate the crystallographic binding site. RMSD units are angstroms, and the interaction energy is in kilocalories per mole.

dx.doi.org/10.1021/ct400678g | J. Chem. Theory Comput. 2014, 10, 282−288

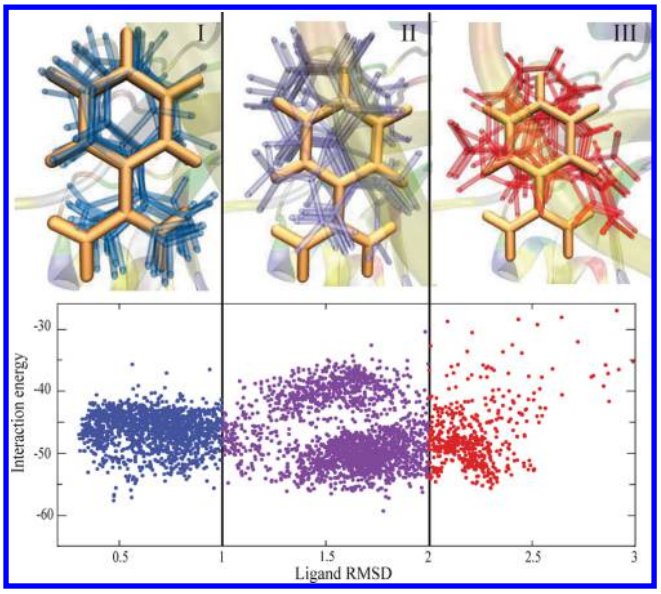


Figure 4. Distribution of benzamidine bounded structures in the active site, showing the interaction energy with respect the heavy atom ligand's RMSD to the native crystal structure. Representative structures in three different ranges: RMSD < 1 Å (blue), 1 < RMSD < 2 Å (purple), and 2 < RMSD < 3 Å (red) are compared to the superposed crystal (in orange). RMSD units are angstroms, and interaction energy in kilocalories per mole.

Supporting Information Figures S1 and S2 show a detailed comparison of the lowest interaction energy prediction (with RMSD < 1 Å) for benzamidine and benzylamine, with their corresponding crystallographic structures.

3.3. Free Energy Calculations. As described in the Methods section, constructing the MSM first involved building the 3D PMF derived from the ligands's coordinates along the PELE trajectories. We tried different sets of microstates (clusters), observing convergence in the 400−600 range. The transition matrix was constructed using $\tau \approx 120$ accepted MC steps (see Supporting Information Figure S3). Figure 5 shows the PMF surface and contour plot for benzamidine; additional contour plots are shown in Supporting Information Figure S5. In order to calculate V_b in an accurate manner, we first used a quadratic interpolation to increase the resolution around the bounded region, and then integrated the bound region with the same criteria used by Buch et al. in their MD studies, $G_{PMF} \leq 3.0$ (kcal/mol).[10] Table 1 summarizes the absolute binding free energies obtained for all ligands. Overall agreement between theoretical and experimental energies is good; for some compounds there are large fluctuations in experimental numbers, making direct comparison more difficult. Supporting Information Figure S4 shows the evolution of the free energy with respect to the number of trajectories, requiring >600 trajectories to achieve convergence.

Table 1. Absolute and Relative Binding Free Energies from the PELE Simulations and Experimental Results[a]

	A	B	C	D
PELE results	−5.50 (±0.61)	−3.99 (±0.63)	−5.68 (±0.70)	−5.26 (±0.60)
experiment	−6.3, −6.4, −7.3	−4.8	−7.0, −7.2	−3.8, −4.7
$\Delta\Delta G$ PELE results	0	1.52	−0.18	0.24
$\Delta\Delta G$ experimental	0	1.5	−0.6/−0.9	1.6

[a]Errors were computed by applying the bootstrapping method. Individual values at a given number of trajectories include also the standard deviation for the volume integration, see Supporting Information Figure S4. Relative binding energies were calculated with respect to benzamidine (A). All units are kilocalories per mole.

Figure 6 compares our results with recent free energy perturbation (FEP) studies and MMPB/SA results from ref 37, including the free energy results for the nonbiased MD by Buch et al. with benzamidine. Several points can be observed when comparing the different methodologies. First, when comparing absolute free energy techniques, we observe that the free ligand diffusion methods, using MD or MC, give analogous results, which are significantly better than the MMPB/SA calculations. We should notice here that PELE uses a GBSA implicit solvent model. Thus, it seems that the solvent model is not a limiting factor in MMPB/SA poor results. Certainly this is the case for this particular system, many others might require the inclusion of explicit waters. Second, the relative free energies obtained with FEP and PELE are in very good agreement. Of course FEP, being a relative free energy prediction, adjusts the value of

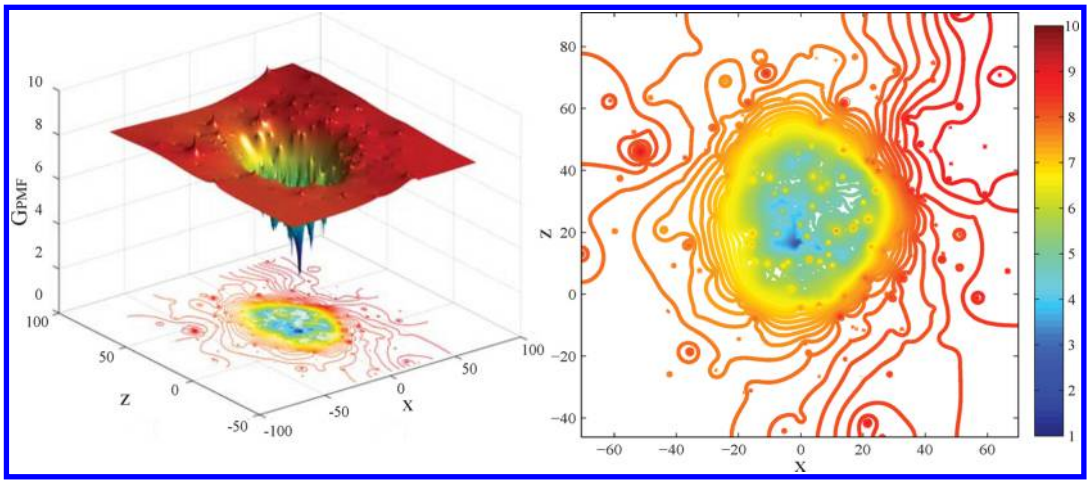


Figure 5. PMF surface in the *xz* plane. The active site can be seen as the global minimum of the PMF surface. Distance units are angstroms, and G_{PMF} units are kilocalories per mole.

dx.doi.org/10.1021/ct400678g | J. Chem. Theory Comput. 2014, 10, 282−288

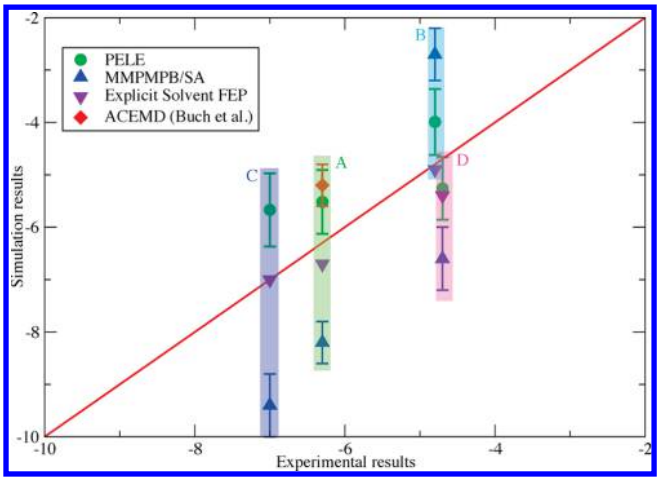


Figure 6. Comparison of experimental binding free energies to different prediction methods. Transparent color bars group together the comparison of the different simulations for the same ligand. Both units are kilocalories per mole.

one reference ligand (C: 4-amino-benzamidine in this case) to its experimental value. Interestingly, as seen in Figure 6, all methods (which also use different force fields) overestimate the binding energy of benzylamine (D). This discrepancy, together with the uncertainty in benzylamine experimental values, might indicate the necessity to revise the experimental data for this ligand.

3.4. Pathway Analysis. One of the advantages of performing nonbiased ligand diffusion simulations is the possibility of studying the binding mechanism by means of the ligand migration pathways and metastable states. Buch et al., for example, performed such an analysis by using the PMF Voronoi tessellation, which is a coarse-graining MSM. We have used here the Perron cluster−cluster analysis (PCCA$^+$)[50] method, which uses the eigenspectrum of the transition matrix to determine the partitioning with the slowest transitions between metastable states, together with transition path theory (TPT).[19] For benzamidine, the PCCA$^+$ analysis for the 840 PELE trajectories identified seven metastable states, the main five (*B*, *C*, *D*, *E*, and *G*) being shown in Figure 7 (states *A* and *F* are not shown). The overall results are very similar to those of Buch et al.:[10] *E* (≈−5.5 kcal/mol) showing the location of the active site, while *G* (≈−2.5 kcal/mol) and *C* (≈−1.7 kcal/mol) agree with states *S*2,*S*3 in the MD study. *G* (magenta color) is not as localized as in the MD results, possibly due to the constrained box simulation (reducing the total exploration space) used by Buch et al.[10] To prove the quality of the PCCA$^+$ analysis, the Chapman−Kolmogorov test was performed for the main five metastable states (Supporting Information Figure S6). The predicted and real data curves are in excellent agreement within 2σ, indicating the good performance of the nonbiased ligand diffusion in PELE simulations.

To examine the binding mechanism obtained with PELE, we calculated the reaction rates[19,24] (in 1/accepted MC steps) between the metastable states *B*, *C*, *D*, *E*, and *G*, panel II in Figure 7. Similarly to the MD kinetic analysis, we observe the largest transitions to the active site from the *C* and *G* states. While in both studies the relative rates are similar, PELE shows a slightly larger contribution from the *G* state. This is due most likely to the largest delocalization over the surface of our *G* state as a (possible) result of the full surface exploration in our simulations. Thus, our kinetic analysis indicates that the mechanism involves diffusion from the solvent into the surface states *C* and *G* followed by the entrance into the active site. In *C*, the ligand is involved in a $\pi-\pi$ interaction with Trp215. While it might access the binding site directly from *C*, it evolves twice as fast into *G* by a hydrogen bond between the NH_2 groups of benzamidine and Gln175. Thus, as seen in nonbiased MD simulations, active site binding is performed through metastable states.

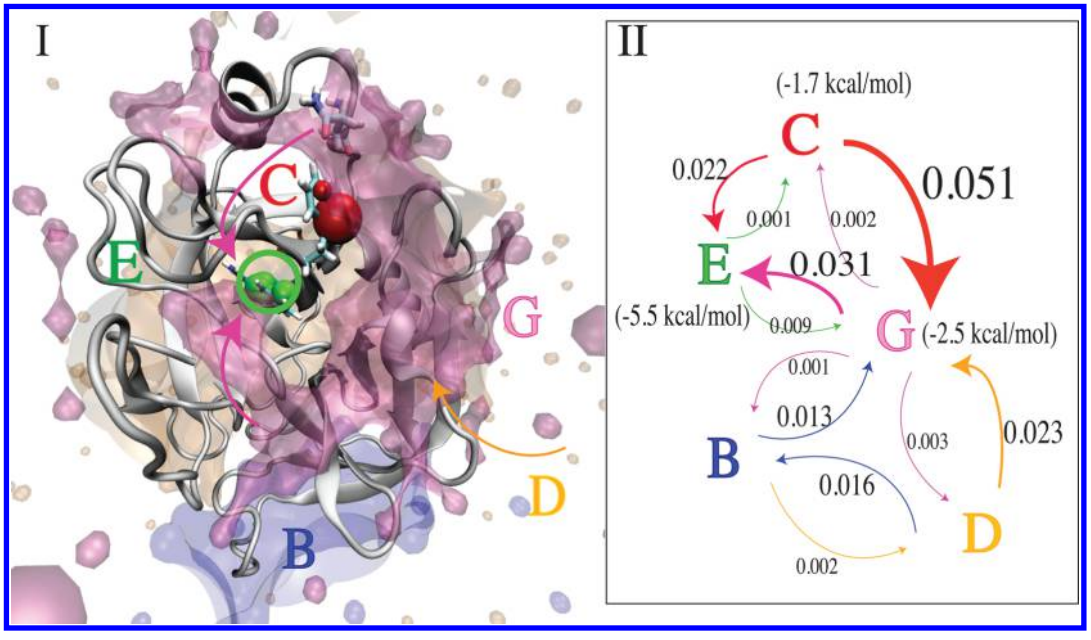


Figure 7. (I) Metastable states, *B*, *C*, *D*, *E*, and *G* (calculated by the PCCA$^+$ analysis), projecting them onto the crystallographic structure. (II) Reaction rates[24] between the different clusters, with arrows indicating the transition direction.

dx.doi.org/10.1021/ct400678g | J. Chem. Theory Comput. 2014, 10, 282−288

4. CONCLUSION

In this paper, we show the possibilities of PELE, a Monte Carlo based technique, in providing free ligand dynamics exploration, including active site binding mechanism and free energy calculations. With modest computational resources, 12−24 h in a 24 core cluster, we can identify the nonbiased ligand diffusion into the active site, obtaining bound poses in good agreement with crystallographic structures. Furthermore, by increasing the number of trajectories for constructing a MSM, we can obtain converged absolute binding free energies. A PCCA$^+$ analysis was used to determine the binding mechanism, indicating the importance of surface metastable states. Although these initial tests were conducted on relatively simple systems, the good agreement with experimental binding energies and molecular dynamics simulations show that PELE may be a useful tool for quick estimates of binding free energies and binding mechanisms.

■ ASSOCIATED CONTENT

Supporting Information

Figures S1 and S2 compare the lowest interaction energy poses with the crystallographic structures. Figure S3 shows the convergence of the implied time scales. Figure S4 shows the convergence of the free energies with the number of trajectories. Figure S5 shows additional PMF contour plots. Figure S6 shows the Chapman−Kolmogorov test. This material is available free of charge via the Internet at http://pubs.acs.org/.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: victor.guallar@bsc.es. Phone: +34 934137727. Fax: +34 934137721.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We acknowledge Martin Senne for providing the earlier version of EMMA, and Ignasi Buch and Gianni De Fabritiis for providing the details of their studies. This work was support by fundings from European Research Council 2009-Adg 25027-PELE grant. Calculations were performed at the Barcelona Supercomputing Center.

■ REFERENCES

(1) Reddy, M. R.; Erion, M. D. *Free Energy Calculations in Rational Drug Design*; Kluwer Academic/Premium Publisher: New York, 2001.
(2) Shirts, M. R. *Methods Mol. Biol.* **2012**, *819*, 425−467.
(3) Frenkel, D.; Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications*; Academic Press: San Diego, 2002.
(4) Doudou, S.; Burton, N. A.; Henchman, R. H. *J. Chem. Theory Comput.* **2009**, *5*, 909−918.
(5) Park, S.; Khalili-Araghi, F.; Tajkhorshid, E.; Schulten, K. *J. Chem. Phys.* **2003**, *119*, 3559.
(6) Park, S.; Schulten, K. *J. Chem. Phys.* **2004**, *120*, 5946.
(7) Nicolini, P.; Frezzato, D.; Gellini, C.; Bizzarri, M.; Chelli, R. *J. Comput. Chem.* **2013**, *34*, 1561−1576.
(8) Gervasio, F. L.; Laio, A.; Parrinello, M. *J. Am. Chem. Soc.* **2005**, *127*, 2600−2607.
(9) Pan, A. C.; Borhani, D. W.; Dror, R. O.; Shaw, D. E. *Drug Discov. Today* **2013**, *18*, 667−673.
(10) Buch, I.; Giorgino, T.; Fabritiis, G. D. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 10184−10189.
(11) Sadiqa, S. K.; Noé, F.; Fabritiis, G. D. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 20449−20454.
(12) Shan, Y.; Kim, E. T.; Eastwood, M. P.; Dror, R. O.; Seeliger, M. A.; Shaw, D. E. *J. Am. Chem. Soc.* **2011**, *133*, 9181−9183.
(13) Prinz, J.-H.; Wu, H.; Sarich, M.; Keller, B.; Senne, M.; Held, M.; Chodera, J. D.; Schütte, C.; Noé, F. *J. Chem. Phys.* **2011**, *134*, 174105.
(14) Senne, M.; Trendelkapm-Schroer, B.; Mey, A.; Schütte, C.; Noé, F. *J. Chem. Theory Comput.* **2012**, *8*, 2223−2238.
(15) Klimov, D.; Thirumalai, D. *Chem. Phys.* **2004**, *307*, 251−258.
(16) Wolynes, P. G.; Onuchic, J. N.; Thirumalai, D. *Science* **1995**, *267*, 1619−1620.
(17) Dill, K. A.; Chan, H. S. *Nat. Struct. Biol.* **1997**, *4*, 10−19.
(18) Bowman, G. R.; Beauchamp, K. A.; Boxer, G.; Pande, V. S. *J. Chem. Phys.* **2009**, *131*, 124101.
(19) Noé, F.; Schütte, C.; Vanden-Eijnden, E.; Reich, L.; Weikl, T. R. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 19011−19016.
(20) Beauchamp, K. A.; McGibbon, R.; Lin, Y.-S.; Pande, V. S. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 17807−13.
(21) Fischer, S.; Windshügel, B.; Horak, D.; Holmes, K. C.; Smith, J. C. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6873−6878.
(22) Noé, F.; Krachtus, D.; Smith, J. C.; Fischer, S. *J. Chem. Theory Comput.* **2006**, *2*, 840−857.
(23) Pan, A. C.; Roux, B. *J. Chem. Phys.* **2008**, *129*, 064107.
(24) Held, M.; Metzner, P.; Prinz, J.-H.; Noé, F. *Biophys. J.* **2011**, *100*, 701−710.
(25) Borrelli, K. W.; Vitalis, A.; Alcantara, R.; Guallar, V. *J. Chem. Theory Comput.* **2005**, *1*, 1304−1311.
(26) Lucas, M. F.; Guallar, V. *Biochim. Biophys. Acta* **2013**, *1834*, 1739−1743.
(27) Borrelli, K. W.; Cossins, B.; Guallar, V. *J. Comput. Chem.* **2010**, *31*, 1224−1235.
(28) Hosseini, A.; Espona-Fiedler, M.; Soto-Cerrato, V.; Quesada, R.; Pérez-Tomás, R.; Guallar, V. *PloS one* **2013**, *8*, e57562.
(29) Cossins, B. P.; Hosseini, A.; Guallar, V. *J. Chem. Theory Comput.* **2012**, *8*, 959−965.
(30) Klenin, K. V.; Wenzel, W. *J. Chem. Phys.* **2011**, *135*, 235105.
(31) Gräter, F.; Schwarzl, S. M.; Dejaegere, A.; Fischer, S.; Smith, J. C. *J. Phys. Chem. B* **2005**, *109*, 10474−10483.
(32) Talhout, R.; Engberts, J. B. F. N. *Eur. J. Biochem.* **2001**, *268*, 1554−1560.
(33) Katz, B. A.; et al. *J. Mol. Biol.* **2001**, *307*, 1451−1486.
(34) Schwarzl, S. M.; Tschopp, T. B.; Smith, J. C.; Fischer, S. *J. Comput. Chem.* **2002**, *23*, 1143−1149.
(35) Ota, N.; Stroupe, C.; da Silva, J. F.; Shah, S. A.; Mares-Guia, M.; Brunger, A. T. *Proteins: Struct., Funct., Bioinf.* **1999**, *37*, 641−653.
(36) Leiros, H.-K. S.; Brandsdal, B. O.; Andersen, O. A.; Os, V.; Leiros, I.; Helland, R.; Otlewski, J.; Willassen, N. P.; Smalås, A. O. *Protein Sci.* **2004**, *13*, 1056−1070.
(37) Jiao, D.; Zhang, J.; Duke, R. E.; Li, G.; Schnieders, M. J.; Ren, P. *J. Comput. Chem.* **2009**, *30*, 1701−1711.
(38) Atilgan, A. R.; Durell, S. R.; Jernigan, R. L.; Demirel, M. C.; Keskin, O.; Bahar, I. *Biophys. J.* **2001**, *80*, 505−515.
(39) Xiang, Z.; Honig, B. *J. Mol. Biol.* **2001**, *311*, 421−430.
(40) Jacobson, M. P.; Friesner, R. A.; Xiang, Z.; Honig, B. *J. Mol. Biol.* **2002**, *320*, 597−608.
(41) Zhu, K.; Shirts, M. R.; Friesner, R. A.; Jacobson, M. P. *J. Chem. Theory Comput.* **2007**, *3*, 640−648.
(42) Ghosh, A.; Rapp, C. S.; Friesner, R. A. *J. Phys. Chem. B* **1998**, *102*, 10983−10990.
(43) Yu, Z.; Jacobson, M. P.; Josovitz, J.; Rapp, C. S.; Friesner, R. A. *J. Phys. Chem. B* **2004**, *108*, 6643−6654.
(44) Lloyd, S. P. *IEEE Trans. Inf. Theory* **1982**, *28*, 129−137.
(45) Tarjan, R. E. *SIAM J. Comput.* **1972**, *1*, 146−160.
(46) General, I. J. *J. Chem. Theory Comput.* **2010**, *6*, 2520−2524.
(47) Marquart, M.; Walter, J.; Deisenhofer, J.; Bode, W.; Huber, R. *Acta Crystallogr.* **1983**, *B 39*, 480−490.
(48) Lucas, M. F.; Guallar, V. *Biophys. J.* **2012**, *102*, 887−896.
(49) Madadkar-Sobhani, A.; Guallar, V. *Nucleic Acids Res.* **2013**, *41*, W322−W328.

(50) Deuflhard, P.; Weber, M. *Lin. Alg. Appl.* **2005**, *398*, 161−184.