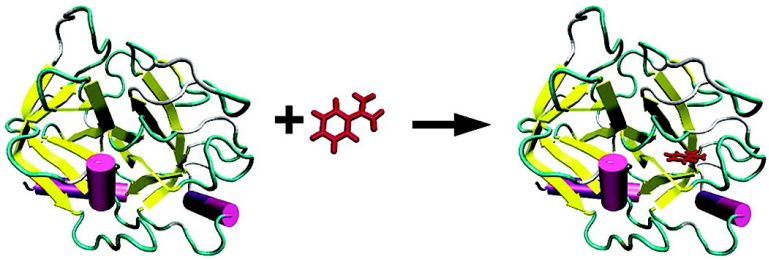
Article

Flexible Docking in Solution Using Metadynamics

Francesco Luigi Gervasio, Alessandro Laio, and Michele Parrinello

J. Am. Chem. Soc., **2005**, 127 (8), 2600-2607• DOI: 10.1021/ja0445950 • Publication Date (Web): 05 February 2005



J|A|C|S
ARTICLES
Published on Web 02/05/2005

Flexible Docking in Solution Using Metadynamics

Francesco Luigi Gervasio,* Alessandro Laio, and Michele Parrinello

Contribution from the Computational Science, Department of Chemistry and Applied Biosciences, ETH Zürich, USI Campus, Via Giuseppe Buffi 13, CH-6900 Lugano, Switzerland

Received September 7, 2004; E-mail: fgervasi@phys.chem.ethz.ch.

Abstract: We apply our recently developed metadynamics method to the docking of ligands on flexible receptors in water solution. This method mimics the real dynamics of a ligand exiting or entering an enzyme and in so doing reconstructs the free energy surface. We apply it to four docking cases: β-trypsin/benzamidine, β-trypsin/chlorobenzamidine, immunoglobulin McPC-603/phosphocholine, and cyclin-dependent kinase 2/staurosporine. In every case studied, the method is able to predict the docked geometry and the free energy of docking. Its added value with respect to many other available methods is that it reconstructs the complete free energy surface, including all the relevant minima and the barriers between them.

1. Introduction

Understanding the mechanism of the recognition process between a ligand and its receptor (docking) presents a fundamental theoretical challenge[1] and has significant importance in the drug discovery process.[2] The number of algorithms available to assess and rationalize ligand−receptor docking is large and steadily increasing.[3] A distinction can be made between fast and simplified methods and computationally intensive and more accurate methods. Methods in the former group are often based on effective potentials that model the ligand−receptor interactions and often assume a rigid receptor and/or ligand. With such methods it is now possible to screen thousands of ligands in a time scale useful to the pharmaceutical industry[4] with varying degrees of success.[3] The second group of methods uses more complex interaction Hamiltonians, flexible ligands and receptors, and explicit or implicit water solvation and therefore is much more expensive from a computational point of view. Both kinds of method can contribute to the discovery process: the first group can be used to perform a virtual screening of large libraries,[5] while the time-consuming methods of the second group are useful in the optimization phase to predict more reliable binding energies and/or to gain a better understanding of the docking process. Some of the most successful docking schemes are based on a combination of the two approaches in a multistage process, where fast methods are used to screen a large number of ligands and accurate methods to refine the docking geometry and binding energy for the most promising candidates.[3]

The efficiency of the search and optimization methods used to find the global minimum of the ligand−receptor conformation energy is important for the success of the first group of methods and fundamental for the second group in which, given the greater complexity of the Hamiltonian, each energy evaluation is computationally expensive.

Most of the successful search methods developed for the accurate ligand−receptor docking applications are based on Monte Carlo (MC)[6] and molecular dynamics (MD).[7] In principle, a straightforward MC or MD simulation, when based on a reasonably accurate force field and including solvation effects, should be able to find the docked geometry and the binding affinity. Unfortunately, most of the time this turns out to be not feasible, since the time spans that can be simulated are much shorter than the time necessary for the real binding process. For this reason, in the field of docking, MD and MC are used in connection with some other method to sample the conformational space efficiently. Such methods include parallel tempering,[8] stochastic tunneling,[9] taboo search,[10] multicanonical MD,[11] umbrella sampling/weighted histogram analysis method,[12] force probe MD,[13−15] and molecular dynamics docking.[16] When the docked geometry is known, a different class of methods based on MD or MC can be used to predict the binding affinities. These use free energy perturbation, thermodynamic integration,[17−21] MM/PBSA,[22] and their variations.

(1) McCammon, J. A. *Curr. Opin. Struct. Biol.* **1998**, *8*, 245−249.
(2) Gane, P. G.; Dean, P. M. *Curr. Opin. Struct. Biol.* **2000**, *10*, 401−404.
(3) Taylor, P. D.; Jewsbury, P. J.; Essex, J. W. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 151−166.
(4) Abagyan, R.; Totrov, M. *Curr. Opin. Chem. Biol.* **2001**, *5*, 375−382.
(5) Bissantz, C.; Folkers, G.; Rognan, D. *J. Med. Chem.* **2000**, *43*, 4759−4767.
(6) Apostolakis, J.; Plückthun, A.; Caflish, A. *J. Comput. Chem.* **1998**, *19*, 21−37.
(7) Wang, J.; Dixon, R.; Kollman, P. A. *Proteins: Struct., Funct., Genet.* **1999**, *34*, 69−81.
(8) Merlitz, H.; Wenzel, W. *Chem. Phys. Lett.* **2002**, *362*, 271−277.
(9) Merlitz, H.; Burghardt, B.; Wenzel, W. *Chem. Phys. Lett.* **2003**, *370*, 68−73.
(10) Cvijovic, D.; Klinowski, J. *Science* **1995**, *267*, 664−666.
(11) Nakajima, N.; Higo, J.; Kidera, A.; Nakamura, H. *Chem. Phys. Lett.* **1997**, *278*, 297−301.
(12) Kumar, S.; Payne, P. W.; Vásquez, M. *J. Comput. Chem.* **1996**, *17*, 1269−1275.
(13) Heymann, B. A.; Grubmüller, H. *Biophys. J.* **2001**, *61*, 1295−1313.
(14) Heymann, B. A.; Grubmüller, H. *Phys. Rev. Lett.* **2000**, *84*, 6126−6129.
(15) Grubmüller, H.; Heymann, B. A.; Tavan, P. *Science* **1996**, *271*, 997−999.
(16) Mangoni, M.; Roccatano, D.; Di Nola, A. *Proteins: Struct., Funct., Genet.* **1999**, *35*, 153−162.
(17) Bash, P. A.; Singh, U. C.; Brown, F. K.; Langridge, R.; Kollman, P. A. *Science* **1987**, *235*, 574−576.

10.1021/ja0445950 CCC: $30.25

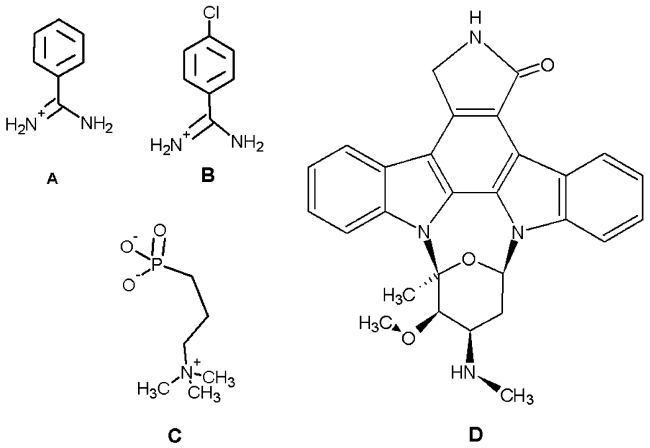


Figure 1. Schematic representation of the ligand structures used for docking: (A) benzamidine, (B) *p*-chlorobenzamidine, (C) phosphocoline, and (D) CDK2 inhibitor staurosporine.

Recently our group has developed a sampling method (metadynamics)[23] that has proved to be useful in various fields, including physics,[24] biophysics,[25] and statistical mechanics.[26] In this article, we show that this method can also be successfully adapted to docking, where it is able not only to find the docked geometry and predict the binding affinity ($\Delta G_{\text{binding}}$) but also to explore the whole docking process from the solution to the docking cavity, including barriers and intermediate minima.

Our test set encompasses four docking cases: β-trypsin/benzamidine, β-trypsin/chlorobenzamidine, immunoglobulin McPC-603/phosphocholine (PC), and cyclin-dependent kinase 2 (CDK2)/staurosporine (Figure 1). In each case, we predict the correct docking geometry and the experimental binding energy within 1 kcal/mol, and we calculate the full free energy surface (FES) of docking as a function of two coordinates: the angle between the major inertia axis of the ligand and the line connecting the ligand and the centroid of the receptor and the distance of the ligand from an atom in the binding site or the distance of the ligand from the centroid of the receptor.

The accuracy of our method in calculating the two-dimensional FES is tested through extensive comparison with an umbrella sampling[27]/weighted histogram[28] calculation. We find that the efficiency of our method is superior, while the accuracy is comparable. We also discuss similarities and differences between our approach and methods based on the taboo search, such as the local elevation method[29] or the filling potential method,[30] which share some common traits.

2. Methods

2.1. Metadynamics. We briefly review here the basic features of the metadynamics method introduced in refs 23 and 24 and report the necessary adaptation for an optimal application to the docking problem. The method is a dynamics in the space of collective coordinates that are evolved with a standard restrained molecular dynamics supplemented by a history-dependent potential. Hence the name *metadynamics*, which in the sense described above, refers to a dynamics of a dynamics. More precisely, a metadynamics run consists of a standard MD run in which we impose harmonic restraints on a set of collective variables ($S_\alpha(r)$) and evolve the values imposed on these collective variables using a modified Lagrangian of the form:

$$H + \sum_\alpha \left[\frac{1}{2} M_\alpha \left(\frac{ds_\alpha}{dt} \right)^2 + \frac{1}{2} k_\alpha (s_\alpha - S_\alpha(r))^2 \right] + V_G(s, t) \qquad (1)$$

where s_α is an auxiliary variable (defining the value of the restraints), and the fictitious masses M_α and coupling constants k_α are free parameters, the optimal choice of which has been discussed in ref 25. The potential $V_G(s, t)$ disfavors configurations in s_α space that have already been visited and is constructed as a sum of Gaussians centered around the values of the $s_\alpha(t)$ explored during the dynamics:

$$V_G(s, t) = w \sum_{t'=T,2T,3T,\ldots t'<t} \exp\left\{ -\frac{[s_\alpha - s_\alpha(t')]^2}{2\delta s^2} \right\} \qquad (2)$$

where w and δs are the height and the width of the Gaussians and T is the time interval after which a new Gaussian is added.

The use of Gaussian repulsive potentials is not new in docking algorithms. It has been used in methods based on taboo search, such as the local elevation method[29] or the filling potential method.[30] These methods are a type of umbrella sampling that uses the sum of many Gaussians as the repulsive potential to avoid revisiting the same minima. Two features make our method novel and different from the taboo-based searches: the reduced dimensionality of the search and the possibility to obtain the free energy directly from the sum of the added Gaussians. The reduction of dimensionality is obtained by choosing the collective variables S_α of the metadynamics. This is an important difference relative to taboo search methods, where the dimensionality of the space to be explored is large, being approximately three times the number of atoms relevant for the docking.

The second difference is the possibility of obtaining the FES from the added Gaussians. As shown in refs 23 and 24 in the limit of a long metadynamics, the sum of V_G and FES tends to become flat as a function of the S_α. When this happens, the V_G gives the negative image of the FES.

To obtain a flat V_G + FES, the metadynamics has to be terminated at the appropriate time. In a system that has two main minima in the FES, which in this case correspond to the docked and solvated ligand, the metadynamics first fills one well, then the second, and then re-enters the first well. To obtain the correct relative height of the two wells, the run is best terminated when the ligand exits from the second well before re-entering the first, that is, immediately after a *recrossing* event. At this stage, V_G + FES is approximately flat. Continuing the run carries the risk of pushing the system outside the basin of interest into higher free energy regions, for instance by a partial unfolding of the protein. Figure 2 shows the exact point where the metadynamics simulations were truncated and Figure 7 shows the effect of overfilling the FES.

2.2. Computational Details. As has been shown elsewhere,[23,24] accurate results can be obtained in a relatively short simulation time if

(18) Bash, P. A.; Singh, U. C.; Langridge, R.; Kollman, P. A. *Science* **1987**, *236*, 564−568.
(19) Sneddon, S. F.; Tobias, D. J.; Brooks, C. L., III. *J. Mol. Biol.* **1989**, *209*, 817−820.
(20) Straatsma, T. P.; McCammon, J. A. *Annu. Rev. Phys. Chem.* **1992**, *43*, 407−435.
(21) Kong, X.; Brooks, C. L. B., III. *J. Chem. Phys.* **1996**, *105*, 2414−2423.
(22) Swanson, J. M. J.; Henchman, R. H.; McCammon, J. A. *Biophys. J.* **2004**, *86*, 67−74.
(23) Laio, A.; Parrinello, M. *Proc Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562−12566.
(24) Iannuzzi, M.; Laio, A.; Parrinello, M. *Phys. Rev. Lett.* **2003**, *90*, 238302.
(25) Ceccarelli, M.; Danelon, C.; Laio, A.; Parrinello, M. *Biophys J.* **2004**, *87*, 58−64.
(26) Micheletti, C.; Laio, A.; Parrinello, M. *Phys. Rev. Lett.* **2004**, 170601.
(27) Patey, G. N.; Valleau, J. P. *J. Chem. Phys.* **1975**, *63*, 2334−2339.
(28) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. *J. Comput. Chem.* **1995**, *16*, 1339−1350.
(29) Huber, T.; Torda, A. E.; van Gunsteren, W. F. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 695−708.
(30) Fukunishi, Y.; Mikami, Y.; Nakamura, H. *J. Phys. Chem. B* **2003**, *107*, 13201−13210.

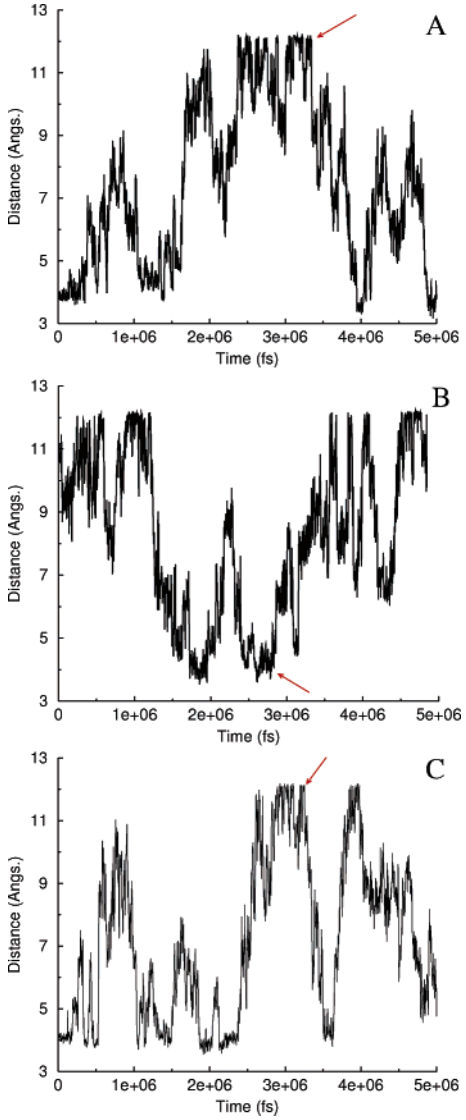


Figure 2. Ligand−protein distance as a function of the total simulation time in the case of: (A) β-trypsin/benzamidine starting with the ligand inside, (B) β-trypsin/benzamidine starting with the ligand in the solution, and (C) McPC-603/phosphocholine. The arrows indicate the point where the metadynamics was stopped in order to calculate the FES.

the S_α are able to discriminate between the initial and final states and include all the modes relevant to the reaction that cannot be sampled within the time scale of an MD run. In the present case, the choice of the metavariables was guided by two requirements: their general applicability in the study of most small ligand−receptor docking and keeping their number low, since the efficiency of the method decreases with the number of metavariables. In the choice of metavariables, one could draw inspiration from the available docking algorithms.[31] Typically variables defining the translation, rotation, and conformation of the ligand are used. This would lead to a metadynamics space that is too large to be efficiently explored. However, this is not necessary, since in the metadynamics approach we exploit the ability of MD to equilibrate the fast degrees of freedom. We found appropriate to use the angle between the line connecting the centroid of receptor and ligand and the principal axis of inertia of the ligand and a ligand-active site distance or a ligand-centroid of the receptor distance. Among the discarded variables were the number of ligand-enzyme hydrogen bonds, hydrophobic contacts, coordinating water molecules, etc. Some of these coordinates were discarded because of their lack of general applicability, others because they were not able to distinguish between different basins.

An important class of docking situations not covered by the present choice of metavariables is the one in which the protein has to undergo a nonlocal rearrangement with a high activation energy in order to accommodate the ligand; in this case the collective coordinate describing the rearrangement has to be included explicitly in the calculation. On the contrary, as shown in the present article, in the case of the McPC-603/phosphocholine and in ref 25 the internal flexibility of the ligand does not need to be explicitly included, unless its activation energies are significantly higher than the average thermal energy.

Since we are not interested in sampling the free energy of the ligand solvated in water, a maximum was enforced on the ligand−protein distance by means of a reflective wall. Whenever the protein−ligand distance hits the reflective wall, the velocity of the corresponding metavariable is reversed. To ensure that at the maximum distance from the binding site the ligand is well-solvated in water and its free energy is comparable to the free energy of a ligand in a diluted solution, we performed a metadynamics on the freely fluctuating ligand. At long distances the FES is both flat within 1 kcal/mol and unstructured, as is to be expected for a water-solvated ligand (Figure 1, SI). This calculation also proves that the reference unbound state that we use to calculate the $\Delta G_{\text{binding}}$ is roughly equivalent to having the unbound ligand and the protein in a dilute solution.

The choice of the metadynamics parameters, namely the height w and the width of the Gaussians, the number of molecular dynamics steps between each metadynamics step (T), as well as the total length of the metadynamics trajectory, affects the accuracy of the FES reconstruction.

Many different combinations of parameters were tried, and in agreement with the results obtained in ref 26, the accuracy for a given value of T is crucially influenced by the height of the added Gaussians w. For high accuracy a small value of w is advisable; however, if small Gaussians are used, it takes longer to fill the wells. A more efficient strategy is to have first a rapid but coarser estimation of the FES, followed when needed by a refinement with a smaller w.[26] After much testing we found that, if the length T of the metadynamics step is set to 3 ps, convenient values of w for this class of problems are 0.48 kcal/mol for the coarse exploration and 0.17 kcal/mol for the refinement. However, provided that the ratio w/T is kept constant, T can be decreased to ca. 0.5 ps without influencing the results. All other parameters were kept constant. The width of the Gaussians, the coupling constant, and the mass M were set to 0.1 rad, 1600 kJ mol^{-1} rad^{-2}, 400 kJ fs^2 mol^{-1} rad $^{-2}$, and 0.4 Å, 600 kJ mol^{-1} Å^{-2}, 150 uma for the angle and the distance, respectively. The rationale that guided the choice of the Gaussian width is that it should be between 1/4 and 1/3 of the average fluctuations of a metacoordinate during a free molecular dynamics run.

In the following, we shall measure the length of the metadynamics run in seconds. However, the reader should note that this is NOT the physical time over which in reality the process takes place, but only a measure of the sampling length. When w is set at 0.48 the recrossing happens in most of the test systems after approximately 3 ns of MD. At this stage, the calculated docked geometry and the $\Delta G_{\text{binding}}$ are already in good agreement with experiment. The refinement with $w =$ 0.17 kcal/mol takes an additional 3 ns. A run which from the start used $w = 0.17$ kcal/mol took about 9 ns to reach the same level of accuracy.

All our simulations were performed with the Amber all-atom force field,[32] the molecules were solvated in a box of explicit water molecules, and sodium or chloride counterions were added to neutralize the charge. The NaCl concentration was 0.1 M. The point charges of the ligands

(31) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W.; Belew, R.; Olson, A. *J. Comput. Chem.* **1998**, *19*, 1639−1662.

(32) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollmann, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.

were obtained using a RESP fit to a HF/6-31G* calculation[32] and are reported in the Supporting Information (Figure 2, SI and Table 1, SI). Electrostatic interactions were evaluated using the PME schemes with 64 × 64 × 64 grid points, $k = 0.40$ Å^{-1}, and a cutoff of 10 Å, the same used for the Lennard-Jones energy terms. The starting structures were obtained from the PDB database, and the highest resolution wild-type structure complexed with the ligand was selected. The hydrogen atoms, not resolved in the crystal structure, were added, and the resulting structure was thermalized in the isobaric−isothermal (NPT) ensemble at $T = 300$ K and $P = 0.1$ MPa for 1 ns in explicit TIP3P[33] water solution. To keep the pressure constant, the tetragonal MD cell was allowed only uniform deformation in a Parrinello−Rahman-like scheme.[34] The temperature was controlled by a Nosè thermostat.[35] Once equilibrated at constant pressure, the run was continued in the canonical ensemble. The metadynamics was started after an additional 5 ns of thermalization. All the simulations were performed with the program ORAC.[36] Figures 4, 10, and 12 were realized using the program VMD.[37]

As stated in the Introduction, our aim is to test the viability of the method with respect to four different goals of increasing difficulty, namely to find the docked geometry, to reproduce a reasonable $\Delta G_{\text{binding}}$, to predict differences in $\Delta G_{\text{binding}}$ from different ligands, and finally to reconstruct the whole docking FES.

To check if the method is able to find the docked geometry, we compare the coordinates of the ligand in the global minimum of the FES with the crystallographic structure, assuming that the docked geometry corresponds to the most stable structure. Moreover, we calculate the average root-mean-square deviation (rmsd) of the heavy atoms of the ligand and of the residues within 6 Å from the ligand between the experimental geometry and all the configurations within the deepest well of the metadynamics.

The $\Delta G_{\text{binding}}$ was calculated as the difference between the free energy of the minimum and the deeper free energy of the unbound state. The $\Delta\Delta G_{\text{binding}}$ was calculated for the complexes β-trypsin/benzamidine and β-trypsin/chlorobenzamidine.

Finally, our ability to calculate the whole docking FES was tested in the case of the β-trypsin/benzamidine complex using the FES obtained by a bi-dimensional umbrella sampling. We performed the bi-dimensional umbrella sampling using the same variables (angle and distance). The umbrella potential used was parabolic on the two variables: $0.5[K_x(X - X_0)^2 + K_y(Y - Y_0)^2]$ with $K_x = 5.73$ kcal/mol Å^{-2} and $K_y = 5.73$ kcal/mol rad^{-2}. In bi-dimensional umbrella sampling, the space spanned by the two variables is divided into many overlapping windows, and constraints in the form of the umbrella potential are imposed on each of them to keep the value of the variables inside the window. The original FES is reconstructed by re-weighting the distribution of the variables in order to eliminate the umbrella bias and by connecting all the windows together using a weighted histogram analysis method.[28,38] It is crucial for the success of the procedure that a significant fraction of the trajectory of each window has values of the variables close to the desired one. If this is not the case, the method fails. The space was divided into 10 × 10 windows. Two different strategies were tried in order to obtain the starting configurations. In the first case, the starting configuration was obtained by restarting from the last configuration of the preceding run and applying the new values of the restraints. A zigzag path on the two-dimensional space of the variables angle and distance was followed. Because of the high barriers in the free energy surface, the restraint potential was in many cases not strong enough to obtain a starting configuration useful for the window, leaving many regions of the FES unexplored (Figure 3). Therefore, to obtain the starting configurations we used the trajectory obtained by metadynamics and chose the configurations with the variables closest to the umbrella sampling windows. This approach enabled us to use very short thermalization times, considerably increasing the computational efficiency of the umbrella sampling. In each window, the MD run was 150-ps long, and thus the total length of the umbrella sampling simulation was 15 ns. It goes without saying that other methods are available to obtain a good initial configuration for each window.

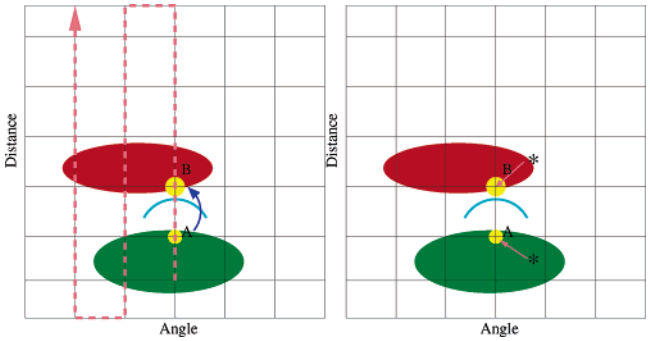


Figure 3. Schematic representation of the two strategies adopted to obtain the starting configurations for the two-dimensional umbrella sampling. In the first case, (left picture), not assuming any prior knowledge of the FES topology, we followed a zigzag path and obtained the starting point of each window starting from the last point of the preceding window and applying the new constraints (red dashed line). In some cases (as here passing from A to B), the two points lie in different energy basins (red and green ellipses in the scheme) and are separated by high barriers. In these cases, the umbrella constraints might not be strong enough to overcome the barrier and carry the system to the new basin of attraction. Applying a higher umbrella potential does not always help: if the barrier is too high, forcing the system through it might unfold the protein. The second strategy adopted uses the trajectory obtained in the metadynamics run and applies the umbrella potential to points in the phase space already close to the grid points. In this case, it is always possible to find a suitable starting point for each window of the umbrella sampling.

The two-dimensional FES was then reconstructed using the weighted histogram analysis method[28,38] as implemented in the program WHAM-2d.[39]

3. Results and Discussion

3.1. β-Trypsin/Benzamidine. The first protein−ligand system studied is the β-trypsin/benzamidine complex. In this case, the interaction is dominated by electrostatic interactions and hydrogen bonding (Figure 4A). The experimental free energy of binding is −6.5 kcal/mol.[40] This is a classic docking test case, where the ligand and the binding pocket are scarcely flexible. However, even here, predicting the correct binding free energy and geometry is far from trivial, since we use a generic force field (i.e., not specifically parametrized to obtain the docking free energy); both the ligand and the protein were treated as fully flexible, and the system was immersed in a solution of explicit water molecules and counterions.

The initial structure was obtained starting from the crystal structure obtained from the Brookhaven Protein Data Bank,[41] code: 1J8A, resolved at 1.21 Å at 105 K.[42] The calcium cation

(33) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.
(34) Parrinello, M.; Rahman, A. *J. Appl. Phys.* **1981**, *52*, 7182.
(35) Nose, S. *Mol. Phys.* **1984**, *52*, 255−268.
(36) Procacci, P.; Paci, E.; Darden, T.; Marchi, M. *J. Comput. Chem.* **1997**, *18*, 1848−1862.
(37) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33−38.
(38) Roux, B. *Comput. Phys. Commun.* **1995**, *91*, 275−282.
(39) Grossfield, A. *An Implementation of WHAM: The Weighted Histogram Analysis Method*; 2004. http://dasher.wustl.edu/alan/wham/doc.pdf. Accessed Sept. 2004.
(40) Mares-Guia, M.; Nelson, D. L.; Rogana, E. *J. Am. Chem. Soc.* **1977**, *99*, 2331−2336.
(41) Berman, H.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I.; Bourne, P. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(42) Cuesta-Seijo, J.; Garcia-Granda, S. *Bol. R. Soc. Esp. Hist. Nat.* (*Sec. Geol.*) **2002**, *97*, 123−129.

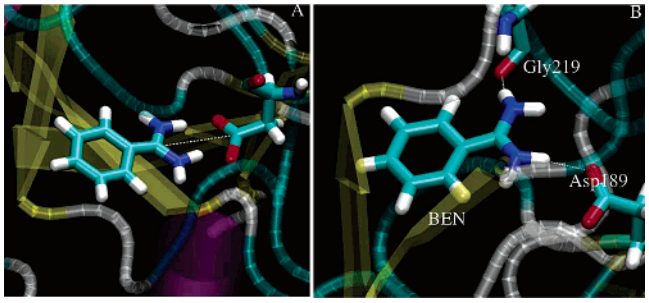


Figure 4. (A) β-Trypsin/benzamidine metadynamics docking geometry. The interaction is stabilized by two hydrogen bonds and a salt bridge with the aspartate 189 and additional hydrogen bonds to backbone oxygens. The white dashed line represents the distance chosen as one of the metacoordinates. (B) First intermediate state following the breaking of a hydrogen bond and preceding the exit of the benzamidine from the binding pocket. This state is stabilized by the formation of a new hydrogen bond with a backbone oxygen.

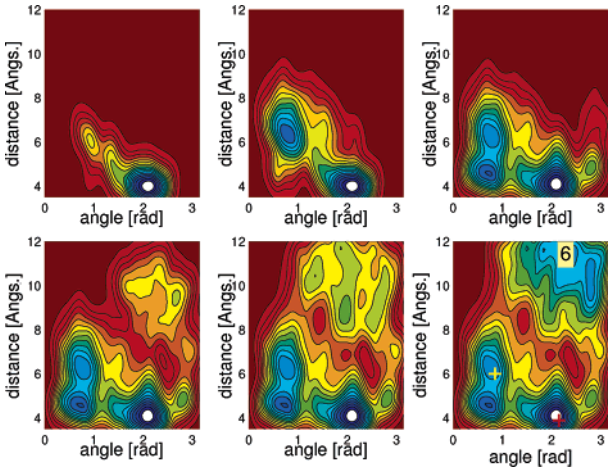


Figure 5. Exit of the ligand. Free energy surface reconstructed using metadynamics as a function of the angle between the centroid of the enzyme and the atoms C_7 and C_4 of the benzamidine and of the distance between C_7 of the benzamidine and the C_γ of the β-trypsin aspartate 189. The isosurfaces are one per kcal/mol. The total metadynamics trajectory (3 ns, 1000 steps) was divided into six snapshots, each adding 166 Gaussians to the previous integrated surface to show the dependence of the reconstructed surface on the simulation time. The number in the yellow box shows the free energy difference of the underlying isocontour from the docked geometry. The purple cross in the last snapshot indicates the position of the crystallographic docked geometry; the yellow cross indicates the position of the metastable state.

and the crystallographic water molecules were kept fixed only during the NPT equilibration. The charge of the calcium atom (which is relatively far away from the active site) was assumed to be +2. After the constant pressure equilibration, the simulation box had dimensions of 43.7 × 42.5 × 49.1 Å, and the protein was solvated with 1939 water molecules. The amidine moiety of the benzamidine was treated as being protonated. The first metacoordinate was chosen to be the distance between C_7 of the benzamidine and the C_γ of the β-trypsin aspartate 189, which is approximately in the center of the binding pocket (Figure 4). The maximum ligand−protein distance explored by the metadynamics was set at 13 Å. The angle was measured between the centroid of the enzyme and the atoms C_7 and C_4 of the benzamidine, which corresponds roughly to the principal axis of inertia of the ligand.

In Figure 5, the FES is reported as a function of the total metadynamics simulation time using $w = 0.48$. Already after

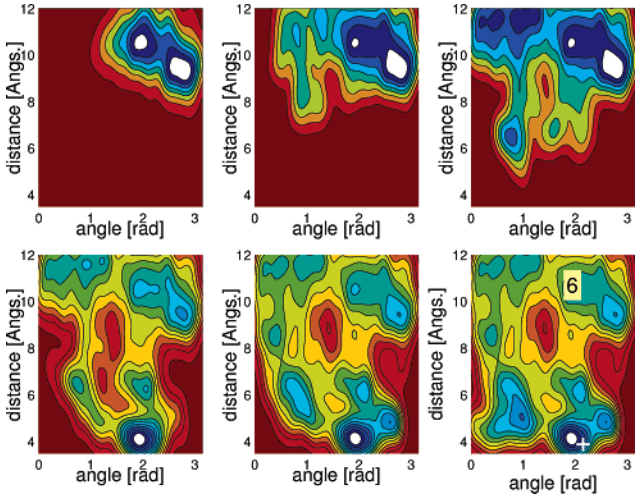


Figure 6. Entrance of the ligand. Free energy surface reconstructed using metadynamics as a function of the angle between the centroid of the enzyme and the atoms C_7 and C_4 of the benzamidine and of the distance between C_7 of the benzamidine and the C_γ of the β-trypsin aspartate 189. The isosurfaces are one per kcal/mol. The total metadynamics trajectory (1000 steps, 3 ns) was divided into six snapshots, each adding 166 Gaussians to the previous integrated surface to show the dependence of the reconstructed surface on the simulation time. The number in the yellow box shows the free energy difference of the underlying isocontour from the docked geometry. The purple cross in the last snapshot indicates the position of the crystallographic docked geometry.

the deposition of 166 Gaussians (0.5 ns), the deepest minimum is close to the experimental geometry (purple cross in the figure). This proves that the experimental geometry corresponds to the FES minimum. In addition, the method provides a wealth of additional information. For instance, we find an intermediate metastable state inside the cavity. This state is visited by the ligand on its way out of the binding pocket and is stabilized by the formation of a new hydrogen bond between the benzamidine and a backbone oxygen of glycine 219 (Figure 4).

Continuing to fill the free energy basin by means of metadynamics pushes the ligand out of the protein. Since during the metadynamics we are adding energy to the system and we are not imposing constraints to the protein backbone, it is important to monitor the rmsd of the protein backbone (C_α) at the end of the metadynamics. In this case, the average rmsd of the α carbons was around 0.6 Å as in all the other metadynamics runs of this complex, unless the FES was overfilled. In the Supporting Information, a picture showing the backbones of the β-trypsin in the crystal structure and after the metadynamics can be found (Figure 3, SI).

In this first calculation, we started our metadynamics from the docked geometry. A more challenging test for the method is to try to reconstruct the FES starting with the ligand in the external solution. This means looking for a physical path leading the ligand inside the protein to docked geometry. In our second simulation we did just that, and the outcome is shown in Figure 6. We started from a geometry having an angle around 1 rad and a distance of ≃12 Å and already after 2.67 ns found the docked geometry. This corresponds, on a modern Pentium 4 workstation, to about 60 CPU hours. The overall reconstruction of the FES is in excellent agreement with that obtained by starting the metadynamics from the docked geometry (Figure 5). The minimum of our FES crystallographic is in this case somewhat shifted from the docked conformation (indicated by

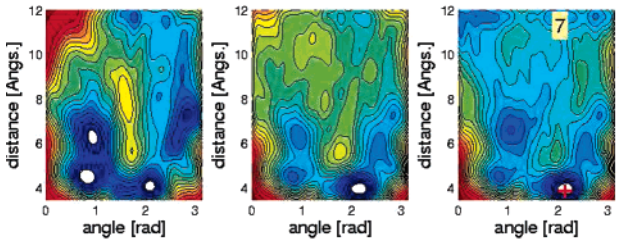


Figure 7. Effect of overfilling. Free energy surface reconstructed using metadynamics as a function of the angle between the centroid of the enzyme and the atoms C_7 and C_4 of the benzamidine and of the distance between C_7 of the benzamidine and the C_γ of the β-trypsin aspartate 189. The isosurfaces are one per kcal/mol. The total metadynamics trajectory (1000 steps added on top of the 1000 steps of Figure 5) was divided into three snapshots, each adding 333 Gaussians (1 ns) to the previous integrated surface to show the dependence of the reconstructed surface on the simulation time. The effect of overfilling is clear in the appearance of spurious minima and is because if we add too much energy to the system (too many Gaussians) after filling all the relevant minima we start exploring regions of the space that are too high in energy and the protein is partially unfolded. The number in the yellow box shows the free energy difference of the underlying isocontour from the docked geometry.

a purplecross in the figure), but this is mainly due to a slight shift of the center of mass of the enzyme with respect to the crystal structure, which slightly changes the value of the angle. Overall, the predicted docking geometry is in very good agreement with the crystallographic, and the average rmsd is 0.3 Å.

The second question that we wish to address is whether the method is able to reproduce the $\Delta G_{\text{binding}}$, the free energy difference between the docked state and the unbound state. Remarkably, both simulations gave a very consistent value of $\simeq -6$ kcal/mol, quite close to the experimental -6.5 kcal/mol.

After 3 ns of the $w = 0.48$ run, the FES is satisfactory. The minima and their relative height are correctly predicted (compare Figures 5 and 6 with Figure 8). This is the classical goal of docking algorithms. However, if one wishes to go beyond this and obtain an accurate estimate of the whole FES, including the barriers, the run has to be refined. As discussed in the Methods section, this can be achieved by continuing the run with a smaller w. The beneficial effect of refining is apparent from Figure 8; the metadynamics FES is smoother, and the difference from the umbrella sampling FES in the region of relevant minima and transition states is less than 1 kcal/mol. The only exception is the region around angle $= 3$ rad and distance $= 7$ Å. We tracked this anomaly down to the use of Gaussians that are too large (0.4 Å) to accurately reproduce this thin minimum. This result is even more encouraging when one considers that the best scenario, umbrella sampling, took 15 ns to complete.

3.2. β-Trypsin/Chlorobenzamidine. This complex was selected to test if the method is able to reproduce small variations in the binding energy due to slight changes of the ligand. We chose to dock the chlorobenzamidine on the β-trypsin since its experimental binding energy is $\simeq$1 kcal/mol less favorable than that of the benzamidine. The setup of the simulation was the same as in the β-trypsin/benzamidine case apart from the *p*-chloro substitution on the benzamidine ligand. In Figure 9, we report the refined FES. The metadynamics was run for a total of 6 ns, of which 3 ns with Gaussians of height set to 0.48 kcal/mol were followed by 3 ns with Gaussians of height 0.17 kcal/mol. The computed $\Delta\Delta G$ is $\simeq$1 kcal/mol, to be compared with the experimental value of 0.7 kcal/mol.[40]

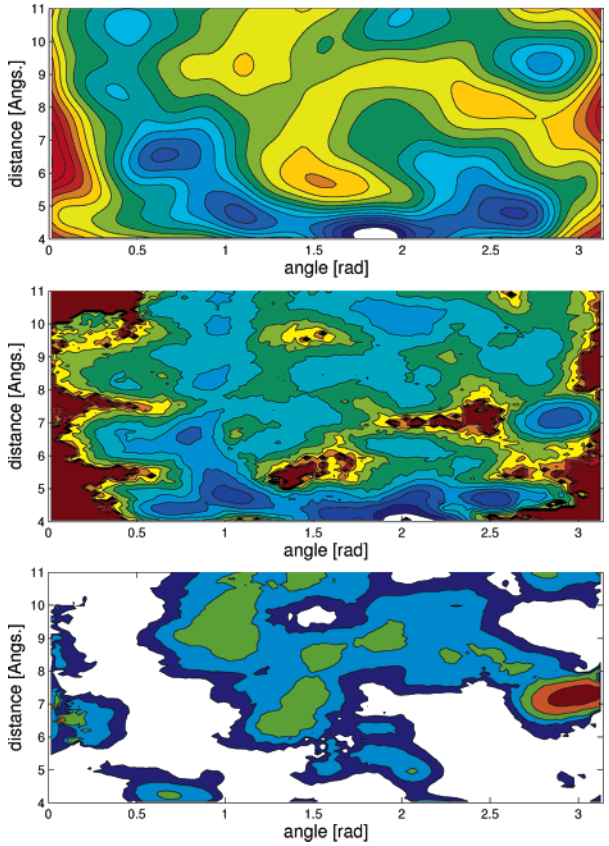


Figure 8. (A) Free energy surface reconstructed using the $w = 0.48 + w = 0.17$ setup. (B) Free energy surface reconstructed using umbrella sampling/WHAM as a function of the angle between the active-site ligand and the major axis of inertia and of the active-site ligand distance. The brown areas are regions where the sampling was too scarce. (C) Difference between the (A) and (B) FES. The isosurfaces are one per kcal/mol.

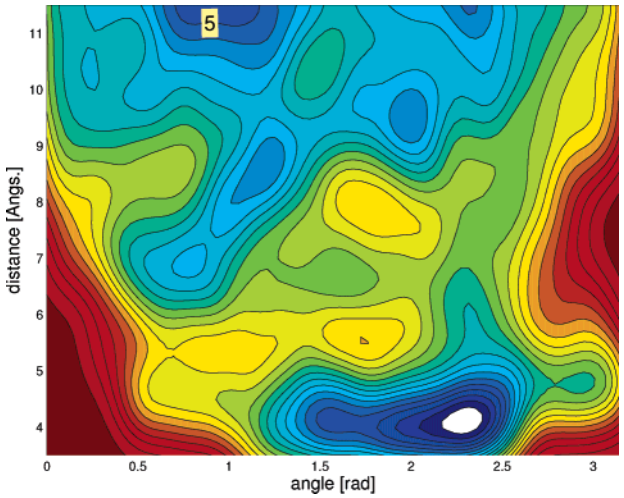


Figure 9. Free energy surface reconstructed using metadynamics as a function of the angle between the centroid of the enzyme and the atoms C_7 and C_4 of the chlorobenzamidine and of the distance between C_7 of the benzamidine and the C_γ of the β-trypsin aspartate 189. The isosurfaces are one per kcal/mol. The number in the yellow box shows the free energy difference of the underlying isocontour from the docked geometry.

The difference arises mostly from a deeper solvation state and is due to the higher dipole moment (12 D vs 6 D of the unsubstituted benzamidine), an observation that is in agreement with the trend observed by Essex et al. when performing Monte

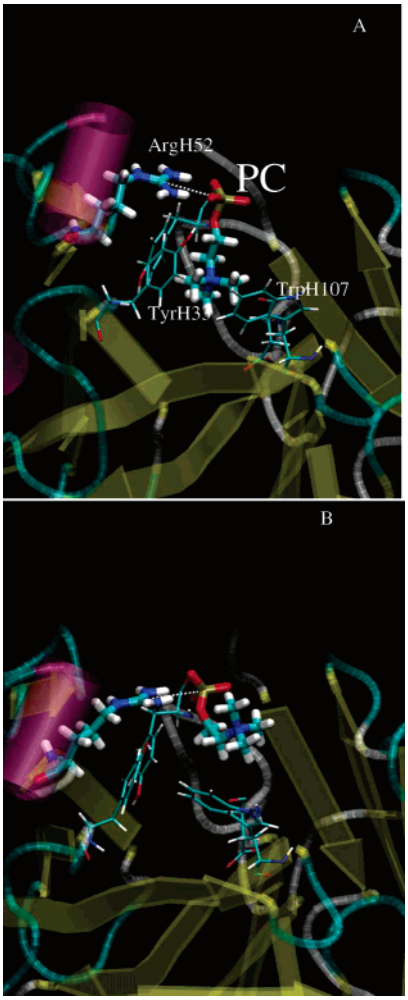


Figure 10. (A) Immunoglobulin McPC-603/phosphocholine metadynamics docking geometry. The interaction is stabilized by electrostatic interaction with Arg H52 as well as by van der Waals and hydrogen bonding with other residues of the cavity (represented as sticks in the picture). The white dashed line represents the distance chosen as a metacoordinate. (B) Transition state preceding the exit of PC from the pocket. Notice the change of the P−C−C−O dihedral angle.

Carlo simulations on substituted benzamidines/trypsin complexes.[43]

3.3. McPC-603/Phosphocholine. The third ligand−protein complex, immunoglobulin McPC-603/phosphocholine, is also a classic docking test case[16,31] and is generally considered harder since the flexibility of the protein plays an important role. The recognition is predominantly electrostatic in character because of the influence of Arg H52, although van der Waals and hydrogen bonding with other residues of the cavity also play an important role (Figure 10). The experimental free energy of binding is −7.1 kcal/mol.[44] The initial structure was obtained by adding hydrogen atoms, water, and counterions to the crystal structure obtained from the Brookhaven Protein Data Bank repository,[41] code: 2MCP.[45] The protein was solvated with 4449 water molecules, and the equilibrated simulation box had dimensions of 51.5 × 50.6 × 76.3 Å. The distance metacoordinate was chosen to be the distance between C_ζ of the Arg

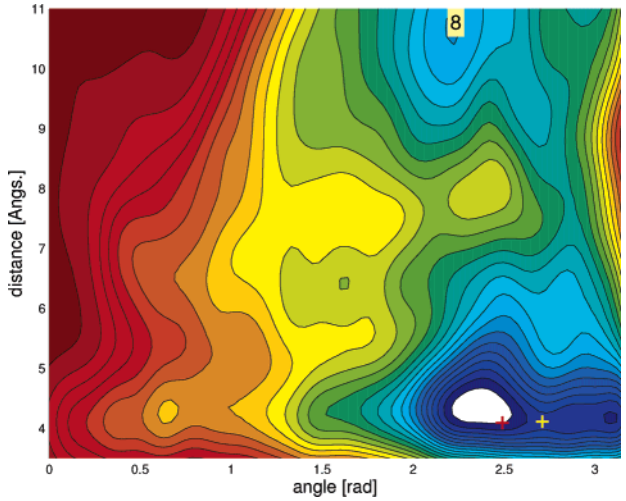


Figure 11. Free energy surface reconstructed using metadynamics as a function of the angle between the centroid of the enzyme and the atoms P and N of the PC and of the active-site ligand distance. The isosurfaces are one per kcal/mol. The total metadynamics trajectory was 7.2 ns long. The number in the yellow box shows the free energy difference of the underlying isocontour from the docked geometry. The purple cross indicates the position of the crystallographic docked geometry; the yellow cross indicates the position of the transition state.

H52 of the McPC-603 antibody and the P of PC (Figure 10). The angle was measured between the centroid of the enzyme and the atoms P and N of the PC. During the equilibration phase at room temperature, the ligand remains bound in its docked geometry. In Figure 11, the final FES is reported. The deepest minimum is to be found around the docked ligand geometry, and the average rmsd is 0.4 Å. We find a transition state preceding the exit of the ligand that involves the change in the P−C−C−O dihedral angle of the PC (Figure 11). This change is aided by the interaction with the Trp H107 and makes the PC more compact (Figure 10B). In this case, it takes a long time to fill the *external* portion of the FES and reach the converged value of $\Delta\Delta G$. Using the $w = 0.48$ + $w = 0.17$ setup, after 7.2 ns we obtain a value of −8 kcal/mol, in agreement with the experimental value. The average rmsd of the α carbons between the initial equilibrated structure and the post metadynamics structure was around 0.9 Å. This value is larger than that obtained in the case of the β-trypsin and confirms the role of the flexibility of the protein in the docking process. A closer analysis shows that it is due to the opening of the mouth of the enzyme and not to a partial unfolding.

3.4. Cyclin-Dependent Kinase 2/Staurosporine. The last ligand−protein complex, cyclin-dependent kinase 2/staurosporine inhibitor, is not a classic docking test case and was included to test the performance of the method on a more complex inhibitor. The CDKs regulate the eukaryotic cell cycle, and their aberrant activity is a common defect in human tumors.[46] For this reason, designing ligands that inhibit the CDK activity is of great pharmacological interest.[47] The inhibitor studied here (staurosporine) is nonselective and too toxic for use in therapy; nonetheless, it has become the lead structure for development of novel CDK inhibitors.[46] The recognition is due to four hydrogen bonds as well as extensive van der Waals contacts in

(43) Essex, J. W.; Severance, D. L.; Tiraldo-Rives, J.; Jorgensen, W. L. *J. Phys. Chem. B* **1997**, *101*, 9663−9669.
(44) Böhm, H.-J. *J. Comput.-Aided Mol. Des.* **1992**, *8*, 593−606.
(45) Satow, Y.; Cohen, G. H.; Padlan, E. A.; Davies, D. R. *J. Mol. Biol.* **1986**, *190*, 593−604.
(46) Huwe, A.; Mazitschek, R.; Giannis, A. *Angew. Chem., Int. Ed.* **2003**, *42*, 2122−2138.
(47) Davies, T. G.; Pratt, D. J.; Endicott, J. A.; Johnson, L. N.; Noble M. E. *Pharmacol. Ther.* **2002**, *93*, 125−133.

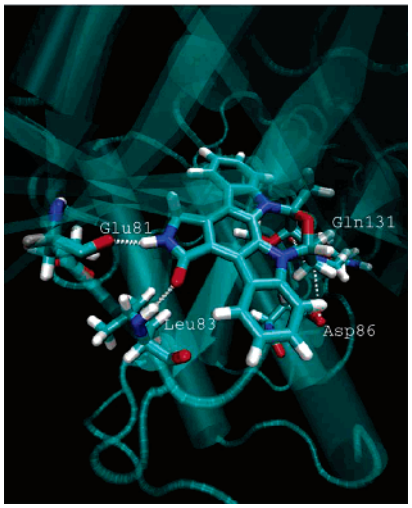


Figure 12. CDK2/staurosporine docking geometry. The interaction is stabilized by hydrogen bonding with Glu81, Leu83, Asp86, and Gln131 as well as by extensive van der Waals interactions. The white dashed line represent the H-bonds.

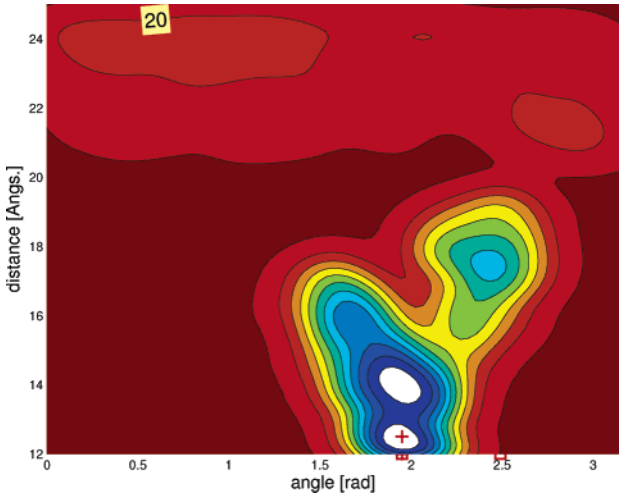


Figure 13. Free energy surface reconstructed using metadynamics as a function of the angle between the centroid of the enzyme and the atoms O4 and N1 of the staurosporine and of the receptor-centroid ligand distance. In this figure, the isosurfaces are one per 2 kcal/mol. The total metadynamics trajectory was 8 ns long. The number in the yellow box shows the free energy difference of the underlying isocontour from the docked geometry. The purple cross indicates the position of the crystallographic docked geometry.

the active site[48] (Figure 12). The experimental IC50 is 7 nM when the CDK2 is complexed with cyclin A,[49] setting an upper bound to the experimental free energy of binding of −11 kcal/mol. The initial structure was obtained by adding hydrogen atoms, water, and counterions to the crystal structure obtained from the Brookhaven Protein Data Bank repository, code: 1AQ1.[50] The nonresolved residues of the crystal structure (residues 36−43 and 149−161) were reconstructed, and the resulting structure was equilibrated. The protein was solvated with 5782 water molecules, and the equilibrated simulation box had dimensions of 50.0 × 70.0 × 60.0 Å. In this case, we did not use the information available on the binding site and of the interactions between the ligand and the receptor: the distance metacoordinate was chosen to be the distance between the centroid of the enzyme and the centroid of the ligand. The angle was measured between the centroid of the enzyme and the atoms O4 and N1 of the staurosporine. In Figure 13, the reconstructed FES is reported. As expected, the deepest minimum is around the docked ligand geometry with an average rmsd of 0.4 Å. Using the $w = 0.48 + w = 0.17$ setup, after 8 ns we obtain a value of −20 kcal/mol. The average rmsd of the α carbons between the initial equilibrated structure and the post-metadynamics structure was around 1.0 Å.

4. Conclusions

In this article, we have described the application of metadynamics to the flexible docking of ligands in explicit water solution. This method mimics the real dynamics of a ligand in exiting or entering an enzyme and in doing so reconstructs the FES of the event. We have shown that in the case of the complexes studied the method is able to predict the docked geometry even without assuming previous knowledge of the docked geometry. In this latter case, the metadynamics finds a path leading the ligand from the external solution to the binding cavity. Moreover, it predicts in a quantitative way the $\Delta G_{\text{binding}}$, the free energy of docking for different ligand−protein complexes, and the $\Delta\Delta G$ of binding arising from slight modifications of the ligand.

Its added value with respect to many other methods available is that it reconstructs the complete FES, including all the relevant minima and the barriers between them. We have shown that this FES is in quantitative agreement with that obtained by running a much longer two-dimensional umbrella sampling.

The information given by the FES could be used to further optimize the binding path and provide a tool for more powerful rational drug design.

Acknowledgment. We thank A. Cavalli for useful discussions.

Supporting Information Available: A figure showing the flatness of the reconstructed FES of the ligand in solution, table of the RESP charges used for the ligands, and a figure showing the effect of the metadynamics on the backbone of the β-trypsin. This material is available free of charge via the Internet at http://pubs.acs.org.

JA0445950

(48) Johnson, L. N.; Moliner, E. D.; Brown, N. R.; Song, H.; Barford, D.; Endicott, J. A.; Noble M. E. *Pharmacol. Ther.* **2002**, *93*, 113−124.
(49) Meijer, L. *Trends Cell Biol.* **1996**, *6*, 396−397.
(50) Lawrie, A. M.; Noble, M. E.; Tunnah, P.; Brown, N. R.; Johnson, L. N.; Endicott, J. A. *Nat. Struct. Biol.* **1997**, *4*, 796−801.