

Free-Energy-Based Methods for Binding Profile Determination in a Congeneric Series of CDK2 Inhibitors

Jérémy Fidelak,^{†,||} Jarek Juraszek,^{‡,||} Davide Branduardi,[§] Marc Bianciotto,[†] and Francesco Luigi Gervasio^{*,‡}

Chemical and Analytical Sciences/In Silico Sciences, Sanofi-Aventis SA, 195 route d'Espagne, Toulouse, France, Computational Biophysics Group, Spanish National Cancer Research Centre (CNIO), calle Melchor Fernández Almagro 3, E-28029, Madrid, Spain, and Drug Discovery and Development (D3) Department, Italian Institute of Technology, Via Morego 30, I-16163, Genova, Italy

Received: December 10, 2009; Revised Manuscript Received: June 9, 2010

Free-energy pathway methods show great promise in computing the mode of action and the free energy profile associated with the binding of small molecules with proteins, but are generally very computationally demanding. Here we apply a novel approach based on metadynamics and path collective variables. We show that this combination is able to find an optimal reaction coordinate and the free energy profile of binding with explicit solvent and full flexibility, while minimizing human intervention and computational costs. We apply it to predict the binding affinity of a congeneric series of 5 CDK2 inhibitors. The predicted binding free energy profiles are in accordance with experiment.

Introduction

Predicting the energy profile of binding leading to reliable estimates of both thermodynamics and kinetics quantities is an ambitious objective of modern rational drug design, which would give a firm guidance both to the lead discovery and the lead optimization phases.

In structure-based drug design, several methods are nowadays available to obtain the structure–activity relationship (SAR), whereas few are able to predict the full free energy binding profile along the reaction coordinate.

Various docking methods^{1–3} can be used both for virtual screening of compound libraries and lead optimization. Evaluations of docking engines, which are being published regularly (see ref 4 for such a recent evaluation), focus mostly on pose prediction and/or on virtual screening capabilities.

The correlation of the raw docking score with the affinity is generally poor,^{5,6} even in congeneric series.³ Moreover, docking is particularly difficult on systems where buried interfacial water molecules are present in the cavity^{7,8} or where only the apo-structure of the target is known.⁹ In order to better take into account desolvation effects, MM-GBSA or MM-PBSA have been used for rescoring, with various degrees of success.^{10–12} More recently, approaches based on more accurate and more expensive QM/MM calculations,¹³ alone or together with PBSA,¹⁴ have been shown to improve the predictions of docking poses. Another rescoring approach, which leads to good performance in affinity prediction, is to tailor a scoring function from a carefully validated set of known actives and their docking poses.¹⁵

A different approach that has been used for many years is relying on free energy based methods. Free energy methods are based on conformational sampling and detailed force fields

generating thermodynamic averages on the full ligand plus target system with explicit or implicit solvent.^{3,16}

The use of conformational averaging removes sensitivity to the details of the single representative conformation. Moreover, the exploitation of the full flexibility of the target and the ligand along with a proper consideration of the solvent effect overcomes most of the concerns of fast-docking methods. This has been clearly shown in the design of HIV-integrase inhibitors. In this enzyme X-ray crystallography studies and traditional docking methods pointed to the existence of only one site where inhibitors could bind. Simulations by McCammon group revealed that a second “trench” opens intermittently close to the catalytic site,¹⁷ suggesting the way to design stronger inhibitors binding on both sides.¹⁸

The many methods available have been sometimes distinguished in two general classes: end-point methods, which only sample conformations of the free and bound species in order to compute the binding free energy, and pathway methods, which sample the conformations all along a multistep pathway connecting the initial and final states.

A detailed description of the end-point methods can be found in several reviews.^{3,19} Pathway methods that proceed through nonphysical decoupling of the interactions between the ligand and its receptor have been shown to provide reliable binding free energies, unless the solvation free energy is very large.¹⁶ But, as the end-points methods, they do not provide any information on transition states (TS) nor metastable minima that could play a fundamental role in the lead-optimization phase. Optimizing the TS and the underlying ligand binding kinetics could potentially provide an additional mechanism to achieve therapeutically safe and differentiated response.²⁰

Calculating the free energy profile along a physical pathway approximating the reaction coordinate does not suffer from the above limitations as it also provides TS and metastable states and can be used to estimate the kinetics by application of the standard Eyring equation or more advanced approximations.²¹

Several free energy methods have been applied to molecular recognition, only to name a few: thermodynamic integration,^{22–24}

* Corresponding author. E-mail: flgervasio@cnio.es.

[†] Sanofi Aventis.

[‡] Spanish National Cancer Research Center (CNIO).

[§] Italian Institute of Technology.

^{||} The first two authors contributed equally to this work.

Jarzynski method,²⁵ adaptive force bias,²⁶ umbrella sampling/weighted histogram analysis,²⁷ force probe MD,²⁸ transition path sampling,²⁹ and metadynamics.^{30–32} So far, umbrella sampling has been one of the most used to study binding in solutions.^{33–37} Among the emerging techniques, metadynamics and its major derivatives—parallel tempering metadynamics (PTmetaD)³⁸ and Bias-exchange metadynamics²¹—have been shown to be useful and efficient in studying molecular recognition.^{39–42}

However, metadynamics shares with umbrella sampling and other collective variable (CV) based methods two fundamental drawbacks that make it scarcely practical for routine docking tasks in the pharmaceutical industry work-flow. It requires a careful choice of the chosen CVs along which the free energy is sampled, and it is very computationally expensive. In order to reconstruct the FES, the CVs must describe all of the slow events that are relevant to the ligand binding/unbinding. As it was shown in refs 32, 41, and 43, when a relevant CV is neglected, the reconstruction of the FES fails. This often implies the need to run multiple metadynamics simulations to understand the optimal set of CVs that is needed for each ligand/protein complex or to use a metadynamics derivative as PTmetaD and bias-exchange MetaD (BEMD) that are able to reconstruct the FES even when relevant CVs are neglected.^{38,44} However, even when the optimal CVs or BEMD are employed, the need to explore the free energy of the ligand in bulk solution makes these methods extremely computationally demanding, depending on which approach is used to calculate the free energy corresponding to the “external area”. In Figure 1 the different possible approaches are presented. In Figure 1, scheme 1, the volume that must be filled with bias by a metadynamics simulation using r and θ as CVs is shown. After fully exploring the internal cavity A and filling it with bias, metadynamics must completely fill the “outside” space B, which in this case is delimited by a restraining the ligand within a conical area. This approach is inefficient due to the long time necessary to fill B and to the multiple recrossing of the narrow gate G needed to reach convergence. Moreover, the relative depths of area A and B of the free energy surface depend on the volume accessible to the drug in the area B and must be reweighted according to the standard volume used in the experiments. In a recent paper describing the use of bias-exchange metadynamics the authors were able to fully reconstruct the FES of docking and undocking of a peptide to the HIV-protease dimer. However, given the necessity of fully filling the external area with bias, the computational resources that were needed to complete the task are out of the reach of a typical pharmaceutical work-flow, limiting the applicability of this approach to selected cases.²¹

A different approach (“coarse metadynamics”) was proposed in ref 41 (Figure 1, scheme 2) where only the internal area of the cavity is explored by metadynamics. The relative binding free energy of similar inhibitors can be calculated by making the assumption that once the drugs reach the “outside area” their free energy is approximately the same. This seemingly drastic approximation gave surprisingly good results in different protein/ligand complexes.

A more accurate approach was used in ref 43, where again only the internal area of the cavity was explored by metadynamics (Figure 1, scheme 3). The absolute free energy of binding was obtained by performing an additional metadynamics simulation on a straight line connecting the gate G to a point several angstroms far away from the protein and correcting for the standard-state dependence.^{3,16}

In the light of the above considerations, here we propose a protocol that minimizes human intervention on the choice of

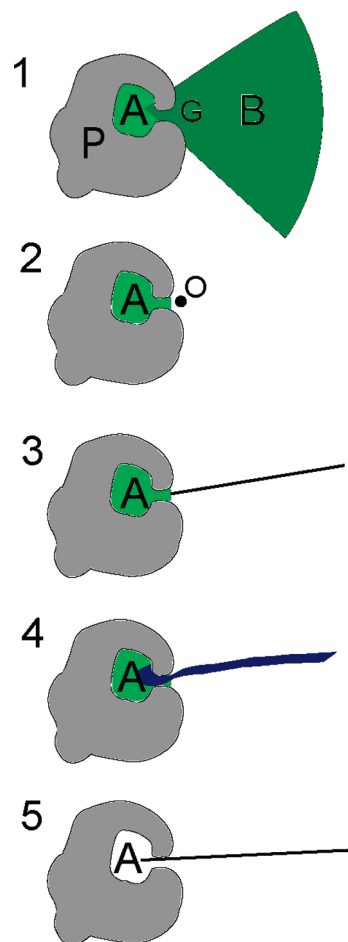


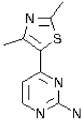
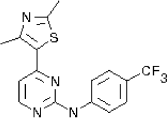
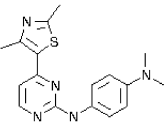
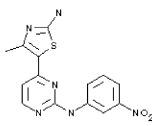
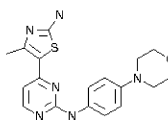
Figure 1. Different approaches to calculate binding free energies with metadynamics or umbrella sampling on a path. In the scheme P = protein; A = binding cavity; G = narrow gate; B = external area; O = a point just outside the internal cavity. The schemes in the figure correspond to (1) running metadynamics with r and θ as CVs; (2) the approach used in ref 41; (3) the approach used in ref 43; (4) metadynamics using the path as CV; (5) metadynamics or umbrella sampling on a straight line, respectively. The different approaches are discussed in the text.

the CV, drastically decreases the computational resources needed to calculate the binding-unbinding free energy, but still provides an optimal binding reaction coordinate and the free energy profile along it, without renouncing to full flexibility and explicit solvent. Our protocol uses the approach of ref 41 (Figure 1, scheme 2) to explore the internal cavity and find a pathway of docking or undocking and a variant of the path collective variables (PCV)⁴⁵ to optimize the path in free energy space and to calculate the free energy profile (Figure 1, scheme 4). As the free energy is sampled only inside a tube around an optimal exit path, the sampling time is minimized.

The protocol was applied to study the binding profile and the mode of inhibition of an homologous series of 2-anilino-4(hetero) aryl-pyrimidine derivatives to cyclin dependent kinase CDK2. The choice of this system was guided by three relevant characteristics: its pharmaceutical relevance, the great flexibility of the target and the large availability of experimental data, including X-ray structures of all the complexes and IC_{50} inhibition constants.⁴⁶

CDK2 is a member of the CDK family which performs numerous functions in the regulation of the cell proliferation cycle and RNA polymerase II transcription cycle.⁴⁷ The kinase complex CDK2-cyclin A (CDK2A) operates primarily in the

TABLE 1: Comparison between Experimental and Calculated Relative and Absolute $\Delta G_{\text{binding}}^a$

inhibitor					
$\Delta\Delta G_b^{\text{exp}}$	0.0	-2.0	-2.0	-4.0	-1.5
$\Delta\Delta G_b^{\text{coarse}}$	0.0	-1.5	-2.7	-2.2	-1.8
$\Delta\Delta G_b^{\text{pathCV}}$	0.0	-2.0	-1.4	-4.7	-1.4
ΔG_b^{exp}	-7.0	-9.0	-9.0	-11.0	-8.5
$\Delta G_b^{\text{pathCV}}$	-7.5 (-13.1)	-9.5 (-14.9)	-8.9 (-15.9)	-12.2 (-18.2)	-8.9 (-13.5)

^a The naming convention for the inhibitors and the experimental values for the $\Delta\Delta G_b$ have been obtained from ref 46. Coarse and pathCV are used to refer to the results obtained with “coarse metadynamics” and pathCV metadynamics, respectively. The absolute binding free energies were obtained after correcting for standard state dependence (see the Supporting Information). The uncorrected values corresponding to the profiles of Figure 4 are reported in parentheses.

DNA synthesis phase (S phase) of the cell cycle. It has been recognized and validated that pharmacological inhibition of CDK2A leads to E2F-induced apoptosis.⁴⁸ Cells with constitutively deregulated E2F activity, that is, most cancer cells, are especially sensitive to this inhibition, and CDK2A inhibition therefore represents a promising tumor-selective therapeutic strategy.⁴⁷

CDK2 is known to be flexible, as crystal structures revealed the existence of an active (open) conformation and at least one inactive (closed) conformation. Interaction with cyclin A and phosphorylation induces a conformational change from the closed to the open conformation. An interesting feature of the ligands studied here is that they all bind to the open conformation.

Anticipating our results, enhanced sampling methods together with the path collective variables (PCV) approach were able to predict an optimal reaction coordinate, the absolute binding free energies ($\Delta\Delta G_{\text{binding}}$) of all the ligands, and to reconstruct the full docking FE profile including the transition states and metastable minima. The approach proposed provides both the SAR and an estimate of the kinetics at a much lower computational cost when compared to a full multidimensional umbrella sampling, metadynamics, bias-exchange metadynamics, or transition path sampling. The added bonus of our protocol is that it does not require an a priori choice of the CV while still providing a full and in-depth understanding of the docking mechanism which could turn out to be crucial in cases where the improvement of the drug binding kinetics is a concern.²⁰ In the specific case of CDK2, the knowledge of the metastable states and of all of the residues involved in the access of the ligands to the binding cavity could lead to the design of novel and more selective drugs for this important oncological target.

Methods

System Setup. For each inhibitor, the protein–ligand complex was obtained from the Protein Data Bank⁴⁹ (2C5X; X = O,V,N; 2WEV; 2UUE). From now on we shall name the inhibitors inh1...inh5 as shown in Table 1. We did not consider the congeneric structure 2C5Y since, at difference with the other structures, the CDK2 is in a fully closed conformation. The ligand bonded parameters were assigned by Antechamber/GAFF force field.^{50,51} The point charges were derived with a RESP fit on ab initio calculations.⁵² Each structure was relaxed with a 4 ns long MD simulation at 300 K and 1 atm (NPT ensemble) followed by 1 ns in NVT ensemble with NAMD MD code using Langevin thermostat and Berendsen’s barostat.⁵³ Simulations

were performed using the TIP3P water model⁵⁴ and the AMBER parm99⁵⁵ force field. All of the simulations were performed in an orthorhombic box with periodic boundary conditions. The dimensions of the box were chosen to be at least 8 Å larger than the solute in every direction. Long-range electrostatics was computed with the PME method.⁵⁶ Short range nonbonded interactions were calculated using a cutoff radius of 12 Å for both Coulomb and van der Waals potentials. The temperature was kept constant at 300 K by a Langevin thermostat with a collision coefficient of 5 ps⁻¹. The time step was set to 2 fs. Bond lengths involving hydrogens were restrained by the SHAKE⁵⁷ algorithm.

Metadynamics Simulations. We used the well-tempered formulation of metadynamics,⁵⁸ which we implemented in the NAMD MD code.⁵³ To perform a metadynamics simulation, a set of functions of the system coordinates $s(R)$ able to describe the process of interest have to be chosen (i.e., the collective variables, CVs). The CVs must be able to distinguish between the different states visited by the system, which in the present context were all of the docked and undocked conformations as well as the stable intermediates.

The dynamics in the space of the CVs is then guided by the free energy of the system plus a history-dependent potential having the functional form of a sum of Gaussians centered along the CVs space up to the current time.³² As with other methods that sample free energy in generalized coordinates, the reliability of the reconstructed free energy is strongly dependent upon the choice of the CVs. They should include the main “slow” variables that change during the transformation, such as large scale motions of the protein or changes in the solvation shell of the ligand. Failure to take into account changes in the solvation shell may adversely affects both binding free energies calculated both with FEP and metadynamics.^{16,21}

Since we used metadynamics to find an approximate docking path and not to obtain a fully converged FES, we did not try to find the best set of CVs for the system. Following the protocol defined in ref 41 (coarse metadynamics), we used a distance (r) and a dihedral angle (ω), where (1) r is the distance between a carefully chosen reference point on the protein pocket and the center of mass (CM) of a rigid moiety of the ligand and (2) ω is the dihedral angle between two reference points on the protein, and the CM and the major inertia axis of the rigid moiety of the ligand chosen to define r .

Our reference point was the center of mass of the two α helices of the C-terminal domain lying below the active site.

The rigidity of these protein portions was assessed during the unconstrained preliminary MD runs. An additional dihedral angle was used to constrain the ligands inside a conical region having an apex at the binding site and pointing toward the entrance of the cavity (see results and discussion). The metadynamics simulations were started from the pre-equilibrated complexes. The hills width was set to 0.3 Å for the distance and 0.15 rad for the dihedral. The height was set to 0.3 kcal mol⁻¹. The deposition interval was set to 1000 fs. An upper bound of 20 Å was set on r .

Path Collective Variables. The molecular recognition was studied using a variant of the path method of ref 45, whereby the path is represented in the space of the mean square deviation of the ligand coordinates, while the alignment is made on the two rigid α helices of the protein (differential alignment).

The path method is used to study a transition between states A and B. Here as state A we chose the relaxed crystallographic pose of the ligands. The choice of state B is more arbitrary as in principle one should take a point at a very large distance (ideally infinite) from the target. Here we took a point ≈ 8 Å away from the mouth of the enzyme cavity. We do correct the free energy for this choice.^{3,16,36,59}

As preliminary metadynamics simulations have shown a strong tendency of CDK2 to assume a closed structure upon unbinding of the ligand, we kept the CDK2 in the open conformation by applying a weak harmonic restraint (3 kcal mol⁻¹ Å⁻²) to the positions of the C_α of the kinase domain. This is justified by the fact that the experimental IC₅₀ were obtained with the CDK2 bound to cyclin A, which keeps the binding site in an open conformation.

If on the contrary we would be interested in the standard absolute free energy of binding to free CDK2, we should have included the free energy change due to the active-to-inactive conformational plasticity, which can be calculated with the path-like CVs as shown in ref 60.

We described the transition with 20 frames $S(1) \dots S(20)$, with the $S(1)$ and $S(20)$ representing the bound (state A) and the unbound (state B), respectively. Care must be taken that all of the frames were equally spaced, relative to the metric used. If the choice of S is appropriate, we would expect the reactive trajectories to be bundled in a tube around the path when the ligand is inside the cavity and assume a funnel shape as the ligand escapes to the bulk solution. To trace the path, we follow the procedure of ref 45 introducing the two variables $s(R)$ and $z(R)$

$$s(\mathbf{R}) = \frac{1}{P-1} \left(\frac{\sum_{l=1}^P e^{-\lambda \|\mathbf{S}(\mathbf{R}) - \mathbf{S}(l)\|^2}}{\sum_{l=1}^P e^{-\lambda \|\mathbf{S}(\mathbf{R}) - \mathbf{S}(l)\|^2}} - 1 \right) \quad (1)$$

and

$$z(\mathbf{R}) = -\frac{1}{\lambda} \ln \left(\sum_{l=1}^P e^{-\lambda \|\mathbf{S}(\mathbf{R}) - \mathbf{S}(l)\|^2} \right) \quad (2)$$

where P is the number of frames and for any microscopic configuration R and $s(R)$ and $z(R)$ measure its intercept and distance from the path $S(l)$, respectively. In practice, while the $s(R)$ CV calculates which, among the S frames, is the closest to the system position, $z(R)$ measures the distance from the

reference path. This allows visiting conformations far from the original path. The λ value was set to 0.9 Å⁻². The square distance, $\|\bullet\bullet\bullet\|^2$, between a generic state $s(R)$ and a point along the path described by $s(l)$ is measured as the MSD between the atoms of the 2-anilino-4(hetero) aryl-pyrimidine scaffold. The alignment of the structures was performed with the Kearsley⁶¹ algorithm using the rigid helices as the reference.

These reference paths were obtained in the following way. The first 18 frames of the guess path have been extracted from the first fraction (2 ns) of the metadynamics undocking simulation of each inhibitor. This first fraction of the metadynamics simulation was sufficient to undock the drug but was far from providing a converged FES. The last two frames were obtained by pulling the inhibitors to a point ≈ 8 Å away from the mouth of the enzyme cavity.

The initial path was optimized in free energy space separately for each inhibitor according to the procedure described in ref 45. In practice the free energy surface in the proximity of each node of the path is sampled by umbrella sampling²⁷ and the nodes are moved perpendicularly to the path in the direction of the lowest free energy.

To obtain the FES of the docking-undocking transition we performed well-tempered metadynamics in the space of $s(R)$ and $z(R)$ with an hills height of 0.2 kcal mol⁻¹ and a deposition time of 2 ps and an energy threshold of 2400 K. $z(R)$ was constrained to be less than 6 Å². This gives the possibility to the system to explore conformations different from the original path however maintaining at the same time the system reasonably close to the chosen intermediate frames.

Results and Discussion

Here, we study the molecular recognition of the 2-anilino-4(hetero) aryl-pyrimidine derivatives to the active open-state cyclin dependent kinase CDK2 and their docking/undocking pathways. CDK2 exhibits the canonical bilobal kinase fold, where N-terminal domain is composed mainly of the β sheet and one α helix (see Figure 2). The C-terminal part is mostly composed by α -helices and is linked to the N-terminal by a flexible hinge (residues 81–84). The ATP-binding site consists in a deep cleft and is located between the two domains.

All of the inhibitors bind with the 2-anilino-4(hetero) aryl-pyrimidine scaffold to the hinge region of CDK2 and make two H-bonds with L83. The detailed structural features of the bound ligands have been described in refs 46 and 62. The differences in the structural features of the binding interface between the protein and inhibitors 2–5 are not remarkable: only inhibitors 4 and 5 have a supplementary interaction between their amino group located on thiazole ring and carboxylate function of Asp145. According to X-ray structures and our MD simulations, aryl-pyrimidine R groups (CF₃, N(CH₃)₂, NO₂ and morpholine) do not interact with CDK2, so relying on interactions which are visible on structures could not lead to a meaningful analysis. Thus, the SAR could not be explained only with the structural analysis of protein ligand interactions from the crystal structure.

Metadynamics Undocking Simulations. We performed undocking simulations of the inhibitors by metadynamics using a set of distance and dihedral angle collective coordinates.

Driven by metadynamics the five inhibitors escaped from the ATP-binding site. In agreement with the results of ref 41, the deepest minima in the reconstructed 2D FESs corresponded in all cases to X-ray structures. However, the free-energy surfaces in Figure 3 clearly show the presence of additional local minima. These are possibly important minima⁴³ and would not have been found with most of the other pathway methods. As in ref 41,

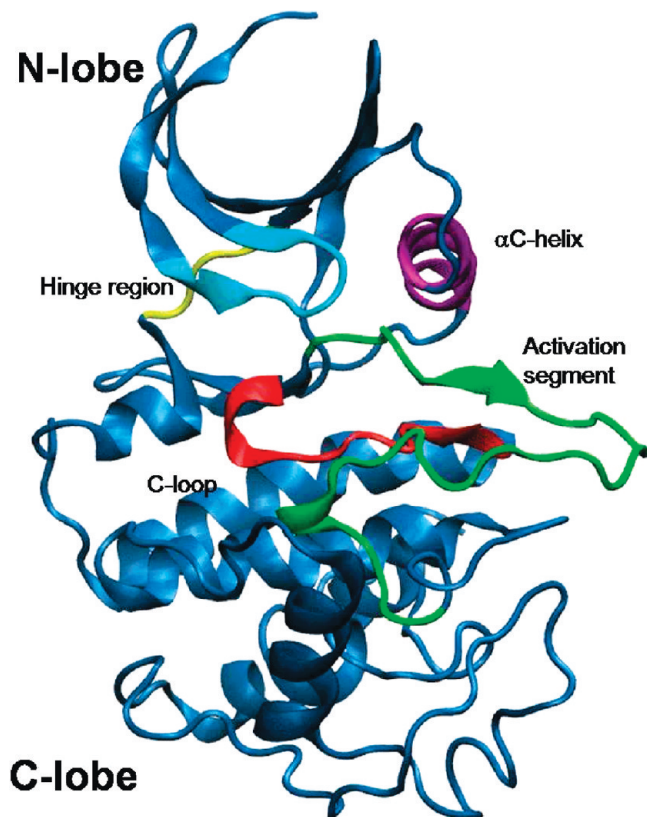


Figure 2. Three-dimensional structure of CDK2 in the open state. (cyan) The G-loop; (violet) the RC-helix; (yellow) the hinge region between the N- and C-lobes (aa 81–84); (red) the α C-loop; (green) the activation segment.

the $\Delta\Delta G_{\text{binding}}$ obtained by measuring the relative free energy difference from the deepest minima to the transition state is often a good approximation to the experimental $\Delta\Delta G_{\text{binding}}$ (see Table 1 and Figure 3). This is not always true, as can be seen in the case of inhibitor 4 where the internal cavity is separated by two barriers of increasing energy from the bulk solution.

It was not possible to converge with only two CVs the full FES, i.e., the FES ranging from the binding site all the way to the external water solution.

This is due to the fact that after exiting from the binding pocket, the ligands interact nonspecifically with the external surface of CDK2, finding several metastable minima with degenerate r and θ (see the Supporting Information). As shown elsewhere if the CVs do not clearly distinguish between different minima converging the FES impossible without resorting to special methods as PTmetaD.³⁸ Here, in order to obtain a full and fully converged FES a conical restraint that keeps the ligand inside a cone having an apex at the binding site and pointing toward the entrance of the cavity had to be added.

During the undocking we also noticed that CDK2 has a tendency to change its conformation toward the inactive closed form. In some cases the final conformation of the metadynamics runs was very similar to the closed X-ray structure, while in other cases the conformational change was more subtle. Thus, as described in the previous section, we added a weak positional restraint on the C-alpha of the kinase domain.

The convergence of the free energy surface was monitored by plotting the hills height as a function of simulation time and observing it decrease to zero as the FE profile converges.⁵⁸ During this time, each metadynamics run sampled several docking and undocking events (recrossing). With the conical

restraints and positional restraints the residual height of the hills was less than 10% of the initial height after 100–250 ns of MD-metadynamics sampling, depending on the inhibitor. This threshold and a minimum of two recrossings was used to establish the convergence of the FES. For inh1 and inh3 about 100 ns was enough to converge the metadynamics, whereas for inh2 and inh4, 200 ns were necessary. The metadynamics run of inh5 did not experience any recrossing within 250 ns and was terminated before converging. The “converged” metadynamics result in a relative binding free energy that is not as good as that obtained from “coarse metadynamics” (see the Supporting Information). This is not surprising in the light of refs 21 and 41, where it was shown that, for similar cases, more than 4 CVs are needed to get free energy differences in agreement with the experiment. Indeed, the chosen CVs are not able to fully discriminate the different minima present in the internal cavity A and the “outside” space B.

Unlike the convergence of the FES, the undocking events were fast. Indeed, the first undocking event of inh2 took only about 2 ns. The guess paths for the subsequent PCV runs was extracted from the metadynamics exit trajectories (see Figure 3).

Free Energy Profiles with Path Collective Variables. In a typical lead optimization process, given the huge computational resources needed to converge the FES and the human intervention needed to carefully select the CVs, a standard metadynamics simulation as shown above would be rather impractical. Thus here we adapted the recently developed PCV method to study molecular recognition.

The details of the procedures are described in the methods section. First, we obtained an optimal reaction coordinate for the docking and undocking of each inhibitor, and subsequently, we used it as a CV for a well-tempered metadynamics to reconstruct the FE profile along it as a function of S and Z . The reference path used to define S and Z is different for each inhibitor.

The convergence of the free energy surface was monitored by plotting the hills height as a function of simulation time and observing it decrease to zero as the FE profile converges.⁵⁸ During each run, we observe multiple docking and undocking events as the ligands exit and re-enter in the binding site multiple times. This is a further indication that the coordinate is optimal and the runs are fully converged. The full reconstruction of the FES for each ligand was reached after 40–200 ns corresponding to 1–2 weeks of calculations with NAMD on a standard 16-cores workstation or up to one week of calculations on a dual GPU-equipped workstation with the ACEMD code.⁶³ The relatively short time necessary for convergence is exclusively due to the optimized nature of the path used as CV. For comparison, in ref 37, umbrella sampling calculations were performed on a straight line (see Figure 1, scheme 5). This approach is similar to performing the sampling with path CVs. However, the nonoptimal nature of the undocking coordinate and the single undocking umbrella sampling simulation performed, results in the necessity of using an extremely long sampling time to converge the free energy profile. For instance, in ref 37, it was shown that a total simulation time of 20.5 μ s is necessary to converge the binding free energy profile of the Src SH2 domain/pYEEI ligand complex.

The FES resulting from our S , Z metadynamics runs, projected on one dimension for clarity, are shown in Figure 4. The deepest minimum of the free energy projection was set to zero.

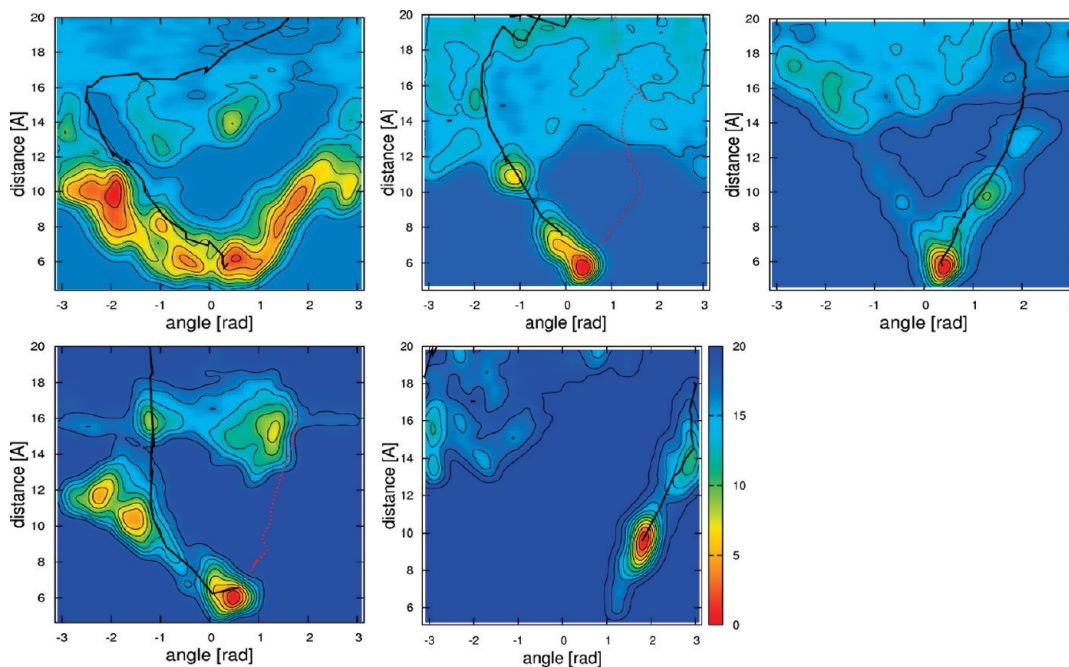


Figure 3. Free energy surfaces as a function of r and θ of the inh1–inh5 binding. The iso-surfaces are one per 2 kcal/mol. The global minimum free-energy paths are plotted as black lines. The paths plotted as red dotted lines (inh2 and inh4) were obtained from the equilibration of a suboptimal initial path, but did not fall into the global minimum (see text).

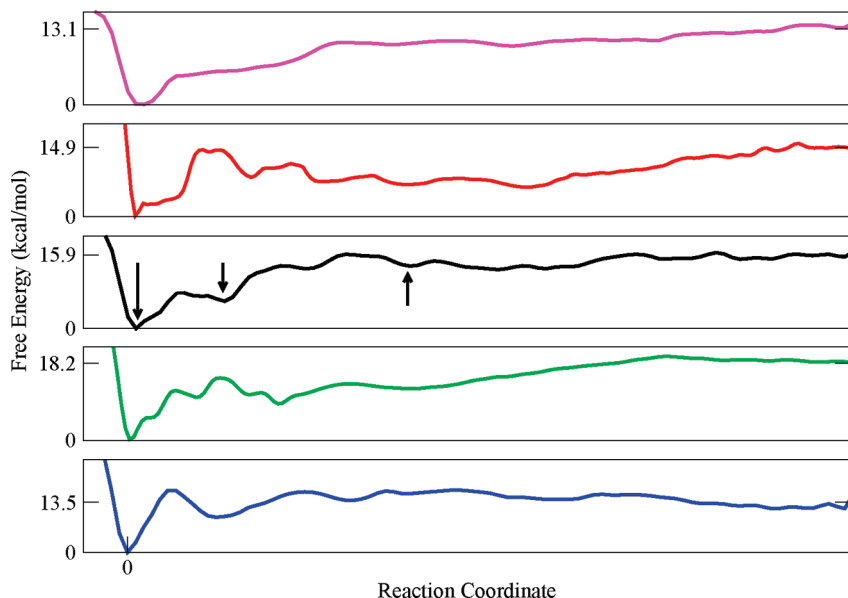


Figure 4. One-dimensional projection of free energy surfaces obtained by path collective variables metadynamics simulations. From top to bottom: inh1 to inh5. Inhibitor structures are reported in Table 1. The arrows in the central panel show the value of the reaction coordinate for the structures reported in Figure 5.

We calculated the $\Delta G_{\text{binding}}$ by projecting the converged 2D FES on S and taking the difference of the free energy between the deepest minima around the docked pose ($S = 1$) and the flat plateau around the undocked state. The lowest free energy difference (inh1) was set to 0. The standard absolute free energy of binding was obtained from the FES by correcting for the standard-state dependence (see the Supporting Information).^{3,16,36,59}

The calculated $\Delta G_{\text{binding}}$ are in excellent agreement with experimental data, leading to a correct ranking of the inhibitors (see Table 1).

As we obtained the reaction coordinates of binding, we have a good insight of the common and peculiar features of the docking mechanism of the various inhibitors. This leads to a biological insight that is not available neither with classical

docking nor with more accurate end-point methods. What is more, when multiple iso-energetic binding modes for the ligand exist, as it was recently shown to be the case for a selective COX2 inhibitor,⁴³ docking, end-point methods, as well as some path based methods would fail in detecting the two sites. This failure would lead to the wrong free energy of binding as well as to the underestimation of the kinetics of unbinding.

On the contrary, the approach that we propose here, by exploring the cavity with metadynamics, would be able to account for both sites.

Inhibitor 1, which is the smallest, is the only one undergoing a full 180 degrees rotation inside the cavity before starting to exit from the ATP-binding site. In this case, the thiazole group

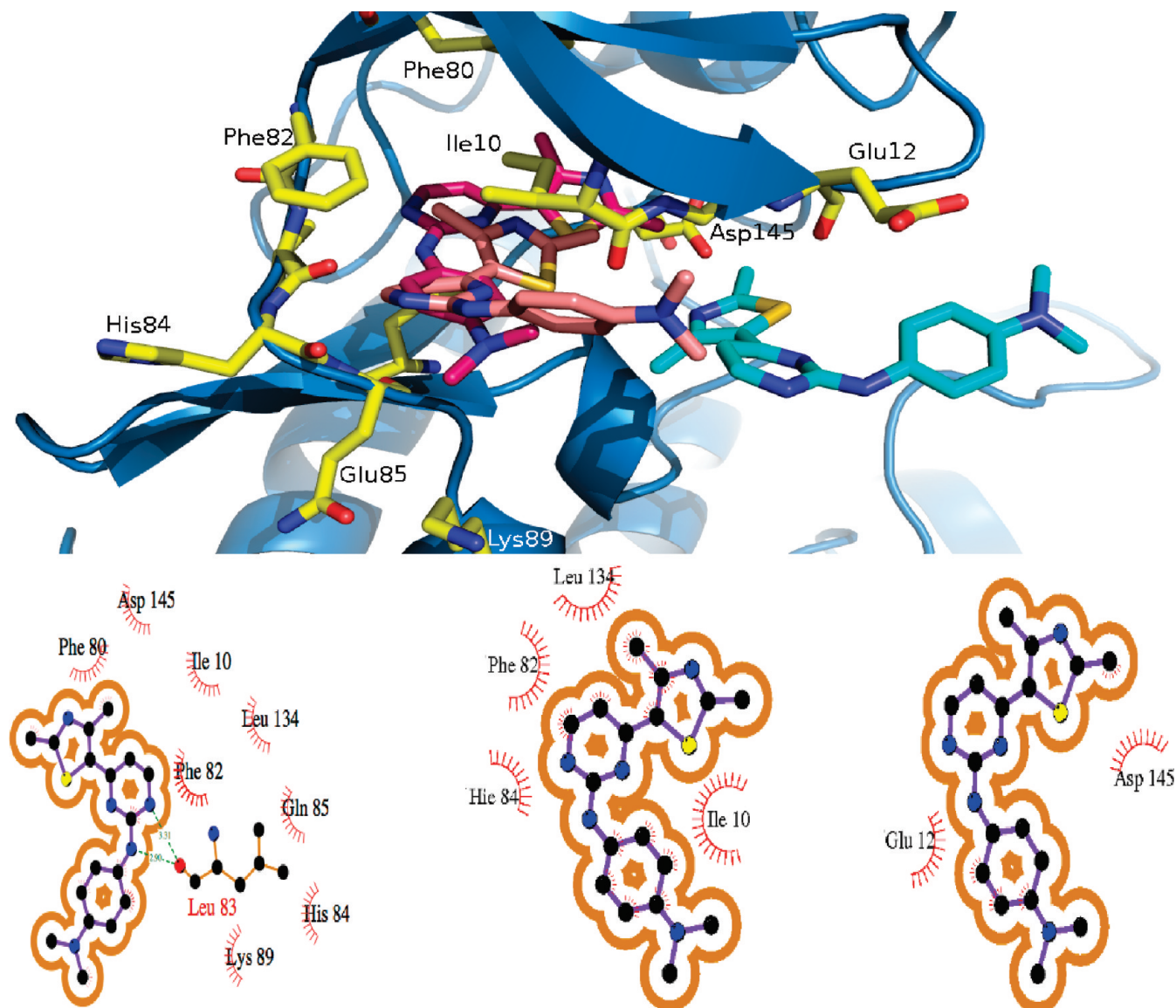


Figure 5. Top: three-dimensional representation of the exit path of Inh3. As the inhibitor leaves the cavity the color changes from magenta to cyan (the corresponding positions on the FES projection are reported in Figure 4 central panel). Bottom: from left to right Ligplot⁶⁴ schemes corresponding to the three-dimensional plots of the top panel. The residues in contact with the inhibitor are highlighted.

exits first. At the same time, Gln131 comes closer to the mouth of the pocket and interacts via H-bond with Glu12.

The other inhibitors do not have sufficient space to rotate inside the pocket, due to their size, and exit from the binding site with the R-amino group first. The thiazole group exits last (see Figure 5). At difference with inh1, the Gln131 moves away from the mouth of the pocket for inh2, 3, 4, and 5. This residue also plays an important role when the inhibitors re-enter the cavity as it interacts with their amino group, facilitating the entrance.

We can also distinguish inhibitors 2, 3, 4, and 5 into two groups according to their exit pathways: inh2 and inh3 for which amino group interacts with Asp86 through H-bond when exiting from the binding site and inh4 and inh5 which progressively exit while thiazole amino group still interacts with Asp145 (an important residue involved in CDK5 opening-closure mechanism⁶⁰). This latter residue acts as an anchor point which breaks at the end of the undocking process and releases the inhibitor into the solvent.

The two Phe80 and Phe83 which are involved in ligand recognition^{62,46} play a key role during ligand exit/entering: they

dynamically interact with the methyl group(s) of all inhibitors and help to position them in the binding site. As inhibitors 2 and 3 have two methyl groups both phenylalanine residues interact during docking/undocking pathway. Concerning inhibitors 4 and 5, each phenylalanine residue interacts one by one with the methyl group located on the thiazole ring. Ligand exit/entering is always assisted by Asp86 in our path collective variables simulations as it interacts with the amino group through an H bond. All of these detailed informations can be used to optimize the ligand-target interactions both in the cavity and on the path to the cavity to increase the binding energy and tune the entrance barrier for the ligands studied.

Importance of Using an Optimal Path. The sensitivity of the free energy profile on the path used to run the PCV simulations can be tuned to some extent by changing the distance from the path available to the system (the upper wall on the Z-path variable). Since we perform a 2-D metadynamics, the system is allowed to explore the Z-path direction and thus, even if the PCV is not chosen optimally, the metadynamics will relax to the free-energy minimum path. Of course, if the initial path is very far away from the lowest free-energy channel, and

the upper wall on Z is too close to the path, we will not explore it. To check the effect of these conditions on the reconstructed free energy profiles, we repeated our PCV simulations in two cases (inh2 and inh4) with a suboptimal path. The final $\Delta\Delta G$ turned out to be the same as compared to the previous runs. This is because even though we pass through higher transition states, the free energy difference between the solvated state and the docked state is the same.

Conclusions

Pathway free-energy-based methods can be very useful in establishing a quantitative structure–activity relationship which do not rely upon a training set. They are able to account both for ligand and target flexibility and explicit solvation effect generally predicting accurate free energies of binding. Moreover, they provide information on barriers and intermediate states that is not readily available with classical docking methods nor with more accurate end-point methods. However, most of the pathway methods require very significant computational resources and/or human intervention. We have introduced a protocol using metadynamics and path collective variables, which is able to find an optimal binding reaction coordinate, to reconstruct the full binding-unbinding free energy profile. We have also shown that it gives the $\Delta G_{\text{binding}}$ for a congeneric series of ligands binding to CDK2 in very good agreement with experiment. The protocol requires reduced human intervention with respect to metadynamics and converges the free energy profile with reasonable computational resources, highlighting its relevance for structure-based lead optimization purposes. The proposed approach, as other pathway methods, by predicting metastable and transition states, provides a biological insight that is not available neither with docking nor with end-point methods. Indeed optimizing the TS and the underlying ligand binding kinetics could potentially provide an additional mechanism to achieve therapeutically safe and differentiated response.²⁰ Finally, the use of metadynamics to explore the binding cavity allows the discovery of additional relevant binding modes, if they are present, something that is not possible even with most of the other pathway methods.

Acknowledgment. We thank Anna Berteotti and Matteo Masetti for useful discussions. F.L.G. and D.B. are grateful to the Barcelona Supercomputing Center for providing computer time. M.B. and J.F. thank the group members of s.a.'s “Protein-Ligand Interaction and Protein Flexibility” and “Virtual Screening” Special Interest Groups for insightful discussions.

Supporting Information Available: Additional figures and experimental details. This material is available free of charge via the Internet at <http://pubs.acs.org>.

References and Notes

- Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. *Proteins* **2006**, *65*, 15–26.
- Halperin, I.; Ma, B.; Nussinov, R. *Proteins* **2002**, *47*, 409–443.
- Gilson, M. K.; Zhou, H.-X. *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 21–42.
- Cross, J. B.; Thompson, D. C.; Rai, B. K.; Baber, J. C.; Fan, K. Y.; Hu, Y.; Humblet, C. J. *Chem. Inf. Model.* **2009**, *49*, 1455–1474.
- Kim, R.; Skolnick, J. J. *Comput. Chem.* **2008**, *29*, 1316–1331.
- Kroemer, R. T. *Curr. Protein Pept. Sci.* **2007**, *8*, 312–328.
- Ladbury, J. E. *Chem. Biol.* **1996**, *3*, 973–980.
- Moitessier, N.; Englebienne, P.; Lee, D.; Lawandi, J.; Corbeil, C. R. *Br. J. Pharmacol.* **2008**, *153*, S7–S26.
- McGovern, S. L.; Shoichet, B. K. *J. Med. Chem.* **2003**, *46*, 2895–2907.
- Pearlman, D. E. *J. Med. Chem.* **2005**, *48*, 7796–7807.
- Lyne, P. D.; Lamb, M. L.; Saeh, J. C. *J. Med. Chem.* **2006**, *49*, 4805–4808.
- Guimaraes, C. R. W.; Cardozo, M. J. *Chem. Inf. Model.* **2008**, *48*, 958–970.
- Cho, A. E.; Guallar, V.; Berne, B. J.; Friesner, R. J. *Comput. Chem.* **2005**, *26*, 915–931.
- Kaukonen, M.; Soderhjelm, P.; Heimdal, J.; Ryde, U. *J. Phys. Chem. B* **2008**, *112*, 12537–12548.
- Silber, K.; Heidler, P.; Kurz, T.; Klebe, G. *J. Med. Chem.* **2005**, *48*, 3547–3563.
- Deng, Y.; Roux, B. *J. Phys. Chem. B* **2009**, *113*, 2234–2246.
- Schames, J. R.; Henchman, R. H.; Siegel, J. S.; Sottriffer, C. A.; Ni, H. H.; McCammon, J. A. *J. Med. Chem.* **2004**, *47*, 1879–1881.
- Hazuda, D.; et al. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 11233–11238.
- Alonso, H.; Bliznyuk, A. B.; Gready, J. E. *Med. Res. Rev.* **2006**, *5*, 531–568.
- Swinney, D. C. *Curr. Opin. Drug Discovery Dev.* **2009**, *12*, 31–39.
- Pietrucci, F.; Marinelli, F.; Carloni, P.; Laio, A. *J. Am. Chem. Soc.* **2009**, *131*, 11811–11818.
- Sneddon, S. F.; Tobias, D. J.; Brooks III, C. L. *J. Mol. Biol.* **1989**, *209*, 817–820.
- Straatsma, T. P.; McCammon, J. A. *Annu. Rev. Phys. Chem.* **1992**, *43*, 407–435.
- Kong, X. J.; Brooks, C. L., III. *J. Chem. Phys.* **1996**, *105*, 2414–2423.
- Jarzynski, C. *Phys. Rev. Lett.* **1997**, *78*, 2690–2693.
- Darve, E.; Pohorille, A. *J. Chem. Phys.* **2001**, *115*, 9169–9183.
- Patey, G. N.; Valleau, J. P. *J. Chem. Phys.* **1975**, *63*, 2334–2339.
- Heymann, B.; Grubmüller, H. *Biophys. J.* **2001**, *61*, 1295–1313.
- Dellago, C.; Bolhuis, P. G.; Csajka, F. S.; Chandler, D. *J. Chem. Phys.* **1998**, *108*, 1964–1977.
- Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562–12566.
- Laio, A.; Rodriguez-Fortea, A.; Gervasio, F. L.; Ceccarelli, M.; Parrinello, M. *J. Phys. Chem. B* **2005**, *109*, 6714–6721.
- Laio, A.; Gervasio, F. L. *Rep. Prog. Phys.* **2008**, *71*, 126601.
- Dang, L.; Kollman, P. J. *J. Am. Chem. Soc.* **1990**, *112*, 503–507.
- Bui, J. M.; Henchman, R. H.; McCammon, J. A. *Biophys. J.* **2003**, *85*, 2267–2272.
- Deng, Y.; Roux, B. *J. Chem. Theory Comp.* **2006**, *2*, 1255–1273.
- Lee, M. S.; Olson, M. A. *Biophys. J.* **2006**, *90*, 864–877.
- Buch, I.; Harvey, M. J.; Giorgino, T.; Anderson, D. P.; Fabritiis, G. D. *J. Chem. Inf. Model.* **2010**, *50*, 397–403.
- Bussi, G.; Gervasio, F. L.; Laio, A.; Parrinello, M. *J. Am. Chem. Soc.* **2006**, *128*, 13435–13441.
- Gervasio, F. L.; Laio, A.; Parrinello, M. *J. Am. Chem. Soc.* **2005**, *127*, 2600–2607.
- Branduardi, D.; Gervasio, F. L.; Cavalli, A.; Recanatini, M.; Parrinello, M. *J. Am. Chem. Soc.* **2005**, *127*, 9147–9155.
- Masetti, M.; Cavalli, A.; Recanatini, M.; Gervasio, F. L. *J. Chem. Phys.* **2009**, *113*, 4807–4816.
- Domene, C.; Klein, M. L.; Branduardi, D.; Gervasio, F. L.; Parrinello, M. *J. Am. Chem. Soc.* **2008**, *130*, 9474–9480.
- Limongelli, V.; Bonomi, M.; Marinelli, L.; Gervasio, F. L.; Cavalli, A.; Novellino, E.; Parrinello, M. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 5411–5416.
- Piana, S.; Laio, A. *J. Phys. Chem. B* **2007**, *111*, 4553–4559.
- Branduardi, D.; Gervasio, F.; Parrinello, M. *J. Phys. Chem. B* **2007**, *116*, 054103.
- Kontopidis, G.; McInnes, C.; Pandalaneni, S. R.; McNae, I.; Gibson, D.; Mezna, M.; Thomas, M.; Wood, G.; Wang, S.; Walkinshaw, M. D.; Fischer, P. M. *Chem. Biol.* **2006**, *13*, 201–211.
- Malumbres, M.; Barbacid, M. *Nat. Rev. Cancer* **2009**, *9*, 153–166.
- Chen, Y.-N. P.; Sharma, S. K.; Ramsey, T. M.; Jiang, L.; Martin, M. S.; Baker, K.; Adams, P. D.; Bair, K. W.; Kaelin, W. G. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 4325–4329.
- Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235–242.
- Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157–1174.
- Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.
- Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollman, P. A. *J. Am. Chem. Soc.* **1993**, *115*, 9620–9631.
- Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781–1802.
- Jorgensen, M. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, L. M. *J. Chem. Phys.* **1983**, *79*, 926–935.

- (55) Wang, J. M.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049–1074.
- (56) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577–8593.
- (57) Ryckaert, L. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327–341.
- (58) Barducci, A.; Bussi, G.; Parrinello, M. *Phys. Rev. Lett.* **2008**, *100*, 020603.
- (59) Doudou, S.; Burton, N.; Henchman, R. *J. Chem. Theory Comp.* **2009**, *5*, 909–918.
- (60) Berteotti, A.; Cavalli, A.; Branduardi, D.; Gervasio, F. L.; Recanatini, M.; Parrinello, M. *J. Am. Chem. Soc.* **2009**, *131*, 244–250.
- (61) Kearsley, S. K. *Acta Cryst. A* **1989**, *45*, 208–210.
- (62) Vulpetti, A.; Pevarello, P. *Curr. Med. Chem. Anti-cancer agents* **2005**, *5*, 561–573.
- (63) Harvey, M. J.; Giupponi; Fabritiis, G. D. *J. Chem. Theory Comp.* **2009**, *5*, 1632–1639.
- (64) Wallace, A. C.; Laskowski, R. A.; Thornton, J. M. *Protein Eng.* **1995**, *8*, 127–134.

JP911689R