Dissection of ligand–CDK8/CycC unbinding free energy barriers and kinetics by molecular simulations

Zhiye Tang and Chia-en A. Chang*

Department of Chemistry, University of California, Riverside, CA92521

Telephone: (951) 827-7263

Email: chiaenc@ucr.edu

Abstract

Understanding the inhibitor dissociation is critical for rational drug discovery. This study used metadynamics and a pathway search method to sample five pyrazolourea ligand dissociation pathways from CDK8/CycC, then applied principal component analysis and milestoning theory to compute a ligand-unbinding free energy profile along the path and estimate the ligand binding residence time. The unbinding paths and free energy barriers provide atomistic details of the unbinding processes to characterize and explain important molecular rearrangements, solvent effects and breakage of key interactions during ligand dissociation. The theory uses interfaces (milestones) without the need to *a priori* identify important states during ligand dissociation. The calculations also provide the lower limit of the residence time, on a time scale of milliseconds and microseconds. This work provides a novel and robust approach to investigate dissociation kinetics of large and flexible systems and to assist in designing new drugs with desired unbinding kinetics.

Introduction

Binding kinetics has become an important topic in molecular recognition because of the growing awareness of the correlation between kinetics and drug efficacy[1-4]. Drug binding residence time, which can be estimated by a dissociation rate constant, $1/k_{off}$, is particularly important for determining the efficacy and selectivity of drug candidates. Experiments provide measured binding affinities (ΔG), rate constants (k_{on} and k_{off}) and molecular structures. However, details are not fully presented by the experimental values and static conformations. As well, why a compound can bind/unbind fast or slowly is not fully understood. Molecular simulations, which are able to provide atomistic descriptions of temporal and spatial details of ligand–protein association and dissociation processes, become an important tool to characterize mechanistic features of binding kinetics and further assist drug development[5]. Features that govern binding kinetics are system-dependent and include ligand properties, conformational fluctuations, intermolecular interactions, and solvent effects[6-10]. However, the determinants to adjust for optimizing kinetic properties for a drug discovery project are not well understood.

All-atom molecular dynamics (MD) simulation in explicit solvent has been widely used to investigate protein dynamics and function as well as ligand–protein binding affinity. However, ligand binding/unbinding processes can be excessively longer than microsecond simulation lengths using computer hardware available to most scientists[11,12]. Various methods such as aMD, metadynamics, scale-MD, and PaCS-MD have been used to accelerate sampling the ligand binding/unbinding processes[13-18]. In terms of the binding paths found, various algorithms such as umbrella sampling[19] and milestoning[20,21] have been used to estimate kinetic rates and free energy profile. Multiple states may be identified

from the trajectories, and the Markov state model (MSM) can be applied to estimate the transition rates[22] (Noe paper). Using a reaction coordinate to accurately present ligand-unbinding free energy barriers provides invaluable information to understand binding kinetics and the mechanism. However, the task becomes daunting when a ligand–protein system under study is large and flexible. All important degrees of freedom involved during the unbinding processes must be included, but the use of high dimensionality to construct the ligand-unbinding free energy plot is impractical with umbrella sampling or milestoning theory. Therefore, this work developed strategies to obtain variables that cover important degrees of freedom for constructing a ligand-unbinding free energy profile with milestoning theory.

Cyclin-dependent kinase 8 (CDK8) is a promising cancer drug target because of its vital role in regulation[23]. CDK8 forms a complex with cyclin C (CycC), Med12, and Med13 for phosphorylation involved in positive and negative signaling of transcription and regulation of transcription activities. [24,25] Abnormal activities of CDK8 and its partner CycC are implicated in various human cancers. [26] CDK8 has an allosteric binding site adjacent to the ATP binding controlled by a DMG motif (Asp-Met-Gly) that characterizes the DFG-in/DFG-out conformations as in other protein kinases. [27] A series of CDK8 drug candidates was discovered by structure-based drug design and virtual screening. [28,29] A series of pyrazolourea ligands (PLs) were developed and targeted at the allosteric binding site of CDK8, [30] and a few studies used MD simulation to examine the structural stability and ligand binding of the CDK8/CycC complex[31,32] A recent work used a metadynamics-based protocol to successfully rank the experimental residence time of CDK8 inhibitors[33]. The screening method provides a tool to identify inhibitors with relative short or long residence

time; however, further investigation to understand determinants with atomistic details that govern the binding kinetics is needed.

In this work, we sampled dissociation pathways by using metadynamics and the newly developed pathway search guided by internal motions (PSIM) [34]. Free energy profile and residence time for five PLs shown in Figure 1 were computed by using the milestoning theory with a novel way to define the milestones. The work developed a new strategy involving principal component analysis (PCA), a mathematical method that can be used to extract the major motions from a collection of data, to define unbinding coordinates for constructing unbinding free energy barriers by using milestoning theory. By projecting the metadynamics trajectories onto the first two principal component (PC) modes, we were able to provide a reaction coordinate for our milestones that captures the most important degrees of freedom involved during ligand unbinding. Unlike MSM, which usually considers a handful of key states, we used more than 100 milestones to reveal detailed molecular motions and interactions corresponding to the unbinding free energy barriers. The computed free energy profile for ligand dissociation clearly indicates and explains where and why the energy barriers occur, such as the hydrogen bond (H-bond) and important interaction formations/breakages between the ligand and CDK8, motions of CDK8 and CycC, and solvation of the ligands and CDK8. The calculations also provide the lower limit of the residence time, on a time scale of milliseconds and microseconds, and the trend agreed with experiments as well. The binding free energy agreed with experimental measurements, which further supports the computation with the milestoning theory.

Results and Discussion

Uncovering energy barriers during dissociation using metadynamics and milestoning theory

Sampling the ligand dissociation pathways is the first critical step in understanding kinetic mechanisms of ligand dissociation. We used metadynamics, an enhanced molecular dynamics simulation method, to obtain multiple ligand dissociation pathways for each compound. We also used the PSIM method to sample dissociation pathways. From our dissociation trajectories, we observed several kinetically relevant features such as loss of intermolecular interactions and conformational rearrangement of both the ligand and protein. Then the milestoning method was used to further quantify the free energy barriers associated with ligand unbinding. Because the dissociation process is complicated, manually selected states or milestones can create artificial bias. Therefore, we developed a novel strategy that uses natural molecular motions illustrated by principal component modes from a dissociation trajectory to represent the reaction coordinate of the milestones along a dissociation pathway. Here we used dissociation pathways obtained by metadynamics, and the strategy can be used for any trajectory with other modeling methods. As shown in Figure S1, the first two PC modes can represent more than 73% of protein motions. Several frames that present conformations along the first two PC modes still cannot be included in the milestones; therefore, it is inevitable that some small energy barriers may be missed in our dissociation free energy profile. However, because the PC modes capture major motions during ligand dissociation, all the important energy barriers should be included in the free energy profile.

The milestoning method reveals that ligands PL1 and PL2 have a more rugged dissociation

free energy profile and more high energy barrier to overcome during unbinding (Figures 2-4 and S2). The computed residence time reflects the trends from experiments, and the computed residence time of PL1 and PL2 are orders of magnitude slower than the other ligands. In addition, the average time required to cross each notable barrier can be estimated (Figures 2-3). Of note, when a ligand is unbinding from the cavity, it can visit various regions on the protein surface before finally dissociating into the solvent. Because the surface diffusion step is not critical in determining a fast or slow binder, it is not a focus in the current study. The free energy difference between the first and last few milestones may be used to approximate the absolute binding free energy, ΔG. The last a few milestones do not appear to be the completely unbound states, so, a highly accurate ΔG is not anticipated. However, the approximated ΔG agrees with the experimental data, with errors < 2 kcal/mol for all ligands, which further validates our computed free energy profile (Table S1).

Identifying the initial dissociation step and the structure–kinetic relationship

The energy barriers illustrated by our free energy profiles reveal protein conformational rearrangements, loss of intermolecular attractions, and changes of H-bond network during ligand unbinding. For unbinding a ligand, we expect that key interactions formed in the stable bound states must be broken; however, this step may need various protein rearrangements for different ligands that are unseen in the static experimental structures. For example, intermolecular H-bonds between the urea linker with Glu66 and Asp173 must be broken for unbinding this series of PLs (Figure 1); however, this step does not yield the same free energy cost for every ligand. Although the most important stage in molecular dissociation is the initial step, during which a compound mostly breaks crucial intermolecular interactions in the bound state and starts to leave the binding pocket, it is

not the whole picture of binding kinetics. The intermediates affected by the structure and properties of a compound may largely contribute to the free energy barriers and binding kinetics as well. Notably, the energy barrier associated with this important initial stage is not related to binding free energy either (Figures 2-4, S2). PL1 has a second set of polar linkers, which needs larger energy and takes significantly longer to overcome the first barrier (first two peaks of Figure 4A). The H-bond donor and acceptor of PL1 also display multiple formation and breakage of complex H-bond networks with CDK8 and/or the bridge water molecules, as illustrated in Figure S3. The interactions contribute to stable intermediate states and multiple energy barriers to leave these local energy minima (Figure 4A). In contrast, the morpholine group of PL4 is not able to form strong attractions with CDK8, which results in less rugged free energy barriers during PL4 unbinding (Figure 3).

PL2, the largest of the five ligands, shows similar behavior to PL1: stepwise H-bond formation and breakage, and multiple energy barriers during unbinding (Figure 2, S2 and S4). The compound rearranges itself a couple of times before further breaking the H-bonds between Glu66 and Asp173, which results in two small energy barriers before heading to a major energy barrier (Figure 2A/B). Notably, although unbinding free energy profiles computed from different dissociation pathways are not identical, the key interactions are the same. For example, the first major peaks of Figure 2 and S2 denote the free energy cost for breaking the H-bonds between Glu66 and Asp173, and the energy barrier from both pathways is ~5 kcal/mol. However, PL2 can have slightly different rearrangements before breaking this major interaction, which results in minor variation in the free energy plots. PL3 and PL5 are the two smallest compounds, with a linear 5-hydroxypentyl and 4-hydroxybutyl group, respectively, and the intermolecular interactions between CDK8 and

the linear groups break immediately when the compounds begin to dissociate. Interestingly, during PL3 and PL5 unbinding, the αC helix (Figure 1) moves with the compounds to keep the H-bonds between the compounds and Glu66 and Asp173 while the compounds are unbinding (first major peaks of Figure 4 B/C). The free energy continues to increase, but H-bonds are retained until the αC helix stops moving in concert with the compounds, which results in large energy barriers for H-bond breakage. After overcoming the major energy barriers, PL3 and PL5 do not form highly stable intermediates, as shown in a smoother free energy profile.

Mechanism of fast/slow dissociation

Here we investigated the unbinding paths and their intermediates in greater detail for one slow, PL2, and one relatively fast, PL4, binding/unbinding compounds. We selected these two compounds because the selected milestones cover most motions appearing in their first two PC modes, with the natural motion PC modes presenting 85% for PL2 and 74% for PL4 conformational fluctuations sampled by a metadynamics run (Figure S1). The free energy profile accurately reproduces experimental binding data as well (Table S1).

PL2 is the largest compound in this study and is anchored by the H-bonds between Glu66 and Asp173; the other functional groups all form nice contacts with the kinase front and deep pocket (Figures 1 and 5). At energy barriers A and B, van der Waals interactions start to loosen between Tyr32 in the β1-2 loop connecting two beta-sheets, β1 and β2, and Gly185 in the activation loop. Also, β1 and β2 sheets move upward to create space for further movement (Figures 2A/B, 5B, S5B). While PL2 is jiggling in this intermediate state, the direct H-bond with Arg356 of the front pocket also breaks. We also observed the same

molecular motions and disruption of interactions from other unbinding pathways for PL2 (Figures S5 and S6), although not all local energy minima are identical during unbinding (Figure S2). Note that the milestoning theory also allows for estimating the time required to cross each energy barrier, and PL2 only needs < 50 ps to go across barriers A or B. The major time-limiting step during dissociation is the energy barrier C, where the stable H-bonds between Glu66 and Asp173 and PL2 break (Figure 5C/D), which allows the ligand to leave the binding site. During this step, the piperazine ring located in the center of PL2 is passing the cleft formed by β1/β2 sheets, β8 and the activation loop, which requires protein arrangement for opening of the cleft (Figure 2C, 5C/D, S2 and S5C/D). This step takes significantly longer, ~15 ns, than passing the first two energy barriers shown in Figure 2. Because Glu66 no longer forms H-bond interactions with PL2, the kinome-wide conserved salt bridge between Glu66 and Lys52 reforms.

At energy barrier D, the second pyrazol and benzene rings of PL2, which formed interactions with residues in the front pocket, are passing the cleft. During this process, multiple H-bond interactions between PL2 and transient bridge water molecules need to break, which contributes to the free energy cost (Figure S4), and the αC helix is moving outwards as well (Figure S7). After overcoming the unfavorable step, a new H-bond network between the compound, protein and bridge water molecules forms, which leads to an intermediate local energy minimum after energy barrier D. The outward movement of αC helix is more pronounced for unbinding PL2 because the αC helix can form more contacts with this large ligand and result in concurrent motion with PL2. Breaking the interactions between αC helix and PL2 and the H-bond network leads to energy barrier E, which takes longer than 6 ns and is another critical step for other compound dissociations.

After leaving the ligand binding site, PL2 continues forming and losing interactions with CDK8 and CycC (Figure 2 F/G) until it is fully solved by water molecules. The simulations also show the breathing motion between the N- and C-lobes of CDK8, whereby the two lobes move apart to open the binding sites for ligand dissociation and then shift to come closer after PL2 completely leaves CDK8 (Figure S7 Milestone Index 140). The major motions all appeared in our metadynamics and PSIM runs (for example, Figures 5, S5 and S6). We observed the same breathing motion during classical MD simulations for apo CDK8/CycC proteins[32]. The breathing motion was also found closely related to the ligand binding/unbinding from p38 MAP kinase[6]. The milestoning theory yielded a lower-limit residence time of 314.1 ms when PL2 dissociates from the protein along the path estimated by the first two PC modes.

The computed free energy profiles explain why inhibitors such as PL4 demonstrate fast unbinding kinetics by revealing the heights and numbers of the free energy barriers. As for all other inhibitors in this series, the bound state of PL4 is stabilized by the two H-bonds via the urea linker of type-II ligands with Glu66 on the αC helix and Asp173 on β8. The terminal [3-(morpholine-4-yl)propyl] group also forms favorable intermolecular interactions with residues Ala100 and Asp98. Of note, the first barrier in Figure 3A shows < 1 kcal/mol energy barrier, so these H-bonds are easy to break. This finding is consist with our previous MD simulations revealing that these H-bonds had fewer occurrences, as compared with PL1, for less stable polar attractions[32]. The major energy barrier is breaking the interactions with the αC helix and the activation loop (Figure 3B), and it takes ~100 ns to overcome the energy barrier. Flipping the morpholine group after barrier A is barrier-less, whereby the oxygen atom of morpholine begins to expose to the solvent (milestone

index 12 to 18 in Figure 3). However, PL4 cannot leave the pocket unless it pushes the αC helix to move upwards, thus creating room for further dissociation (Figure 3B). The activation loop does not directly involve ligand binding affinity[32], but the loop can affect binding kinetics. While leaving the pocket, the methylphenyl group of PL4 interacts with CycC and creates the last major energy barrier before completely unbinding. Although the free energy barrier heights of the two major energy barriers, B with 4 kcal/mol and D with 3 kcal/mol, are comparable with the energy barrier heights of C and F in PL2, lacking multiple small energy barriers results in a much faster estimated residence time than that in PL2.

Challenge in computing absolute residence time for complex biomolecular systems and insights into drug design

A free energy profile along the dissociation pathway provides information not seen in solely the bound state. The unbinding free energy barriers suggest how significant a feature influences the unbinding kinetics, which may come from interactions, protein/ligand rearrangement, solvent effects, etc., and assist in drug design of the desired kinetic property. Constructing an unbinding free energy profile needs a coordinate to present the unbinding pathway. A simple coordinate may be the distance between the center of mass of a ligand and protein pocket. However, during dissociation, the distance may remain the same, but the protein is rearranging its side-chains and/or backbone, which contribute to the free energy. One may add a couple of dihedral rotations in the coordinate. However, during the transient unbinding states, key dihedrals are not identical, and explicitly selecting certain degrees of freedom inevitably ignores motions that contribute to the free energy barriers. Therefore, we used molecular motions presented by the first two PCA modes, PC1 and

PC2, to present the unbinding coordinate. The first two modes cover more than 73% of the major motions for all five ligands in this study (Figure S1). Using this new quantitative approach to choose the unbinding coordinate for milestoning theory, we successfully obtained the free energy barriers and timescales of crossing these barriers during the dissociation pathway.

Figures 6 and S1 illustrate that the unbinding pathway presented by PC1 and PC2 combines related motions from multiple degrees of freedom, not limited to a few collective variables such as the distance between two centers of mass or selected dihedral angles. Each dot denotes a frame from the unbinding trajectory. The red curve is the smoothed path, and we divided the space along the path into compartments separated by milestones (black line in Figure 6B). The milestoning theory that follows statistical mechanics theory to estimate the transitions between compartments (Figures 2-4) along the unbinding pathways allows for using short trajectories initiated within each compartment. The trajectories are analyzed to produce free energy profiles and accurately estimate the time a ligand needs to pass each energy barrier and unbind. Notably, Figure S1 shows complex conformational changes; for example, PL2 slightly adjusted its position and orientation multiple times before it began to leave the binding pocket (red circles in the lower left of Figure 6A). Because of the complicated molecular rearrangements, not all natural motions shown in PC1 and PC2 can be included in the compartments. Therefore, some fluctuations that contribute to unbinding free energy barriers are ignored, which results in faster kinetics. The transition matrix constructed by multiple short unbiased MD trajectories provides the transition probability between nearby milestones; however, the frequency of a ligand moving back and forth between numerous major or tiny energy barriers may not be accurately captured. The

computed residence time may be considered the time required to smoothly travel from the bound to unbound state without excess backward and forward swing.

As compared with bound states, which have relatively few low-energy native conformations, in ligand binding/unbinding processes, a ligand and its target protein may have access to many more possible binding/unbinding routes that all contribute to the binding kinetics. For tight binders with only one dissociation direction, the ligands must break important and conserved interactions before unbinding. Therefore, the initial movement during the unbinding process is highly similar from multiple pathways, and the initial steps largely determine the binding residence time estimated by $1/k_{off}$, as shown in Figures 2 and S2 for two dissociation pathways of PL2. In this study, we obtained several unbinding pathways for each ligand by using metadynamics and PSIM, and all showed the breakage of interactions appearing in the crystal structures of the bound complexes, thereby resulting in one major dissociation pathway for each ligand when leaving the pocket. As a result, milestoning calculations along the first two PC modes from one metadynamics trajectory yielded good approximations for unbinding free energy profiles to provide valuable information for drug development. The energy barriers associated with particular movements and/or interactions during dissociation inform rational drug design for desired kinetic properties. For example, Figures 2 and S2 show that passing the second pyrazol and benzene rings of PL2 produces a high energy barrier. Modifying a functional group in this region, such as adding an H-bond to strengthen intermolecular interactions in this intermediate state, may further slow the unbinding process. For protein systems with a wide-open binding site, such as HIV-1 protease, instead of one well-defined binding/unbinding channel, molecular modeling may sample significantly different

dissociation pathways; thus, different unbinding directions need to be considered when constructing the unbinding free energy profiles.

Methods

Molecular simulations

We obtained the initial structures of CDK8/CycC–PL complexes from the PDB database (PDB ID: 4F6W, 4F7L, 4F6U, 4F7N [30]) and manually modified PDB structure of CDK8/CycC–PL3 (4F7N) to obtain the initial structure of CDK8/CycC-PL5. We added missing residues and built the missing activation loop of the structures by using Swiss Model [35] based on the p38 DFG-out crystal structure (PDB ID: 1W82 [36]) as a template. We determined the protonation states of histidine residues in the CDK8/CycC–ligand complexes by using the MCCE package [37]. AMBER FF14SB force field and GAFF [38] were used for the CDK8/CycC complex and PLs. We solvated the five complexes with TIP3P and a water buffer size of 12 Å and added 6 Cl^- ions to neutralize the formal charges of the system. After the standard setup detailed in SI, production runs of the five systems were performed at 298K for 500 ns in NPT ensemble and saved every 2 ps with a 2-fs time step by using pmemd.cuda from the AMBER 14 package. For each CDK8/CycC–PL complex, we chose the equilibrated conformations at 298K and conformations after 100-, 400-, 500-ns MD simulations as initial structures to run 12 metadynamics simulations. Details of the collective variables chosen for metadynamics are in SI.

Computing free energy and unbinding time by using milestoning theory

To construct unbinding free energy and compute the kinetics properties with milestoning theory, we needed to define milestones along reaction coordinates and estimate the

probability of a transition between two milestones by using many short classical MD simulations. The PCA obtained from a metadynamics trajectory of a complex was used to define the milestones. To obtain many short classical MD runs for each complex, we resaved a frame every 50 ps from a metadynamics trajectory. As a result, about 500 to 600 conformations were used as initial conformations, and we ran 20 replicas of 100-ps classical MD simulations for each initial conformation with a 2-fs time step at 298K. Frames were saved every 50 fs for each 100-ps MD trajectory, and the settings were the same as the MD simulation mentioned previously.

To compute PCA for a metadynamics trajectory, we selected the α-carbon atoms of CDK8/CycC and heavy atoms of PLs, and computed the covariance matrix of the Cartesian coordinates of these atoms by using the first frame in each trajectory as references. We saved the eigenvectors of the covariance matrix and used the equation $PC_i = R^T (X(t)-\langle X\rangle)$ to project frames from the metadynamics trajectory onto PC1 and PC2 space, where R^T is the eigenvector with the highest eigenvalue for PC1 and second highest for PC2, and X(t) and $\langle X\rangle$ are the Cartesian coordinates of the selected atoms at time t and the average over the trajectory, respectively. Frames from a metadynamics trajectory were projected onto PC1 and PC2 space, as exemplified in Figure 6A. The original metadynamics trajectory was further smoothed by averaging forward and backward 100 frames, and the frames were also projected onto the PC space, as shown by the red line in Figure 6A. The smoothed trajectory can more clearly represent the dissociation path, and multiple short lines of 20.0 eigenvalue units in length were placed perpendicular to the path. Each line is 3.0 eigenvalue units apart, and the lines were optimized to minimize the overlapping of the lines, as illustrated in Figure 5B. Therefore, each line serves as one milestone. Frames saved from

the first 30 ps of each 100-ps short MD run were then projected onto the PC1/PC2 space and fall into one of the milestones (Figure S8A). We used frames from only the first 30 ps because the protein or ligand may cover several milestones, which are not useful in estimating the transitions between nearby milestones. We also removed the data points that fall onto the areas that were 10 units away from the path in PC1/PC2 space (Figure S7B). The data points remaining in Figure 7B were used to compute the duration of each milestone and the transition counts between adjacent milestones. Finally, the transition kernel (matrix) K, free energy profile and residence time were computed following the exact milestoning theory. [20]

Acknowledgments

This study was supported by the US National Institutes of Health (GM-109045), US National Science Foundation (MCB-1350401), and NSF national supercomputer centers (TG-CHE130009). We thank Dr. Ron Elber for discussion relating to milestone theory and Timothy Cholko for making the CDK8 figure.

Abbreviations

SRK structure-kinetic relation; PCA principal component analysis; H-bond hydrogen bond; MD molecular dynamics; PL pyrazolourea ligand; CDK8 cyclin-dependent kinase 8

References

1 Schuetz, D. A. *et al.* Kinetics for Drug Discovery: an industry-driven effort to target drug residence time. *Drug Discovery Today* **22**, 896-911, doi:10.1016/j.drudis.2017.02.002 (2017).

2 Copeland, R. A. The drug-target residence time model: a 10-year retrospective. *Nature Reviews Drug Discovery* **15**, 87-95, doi:10.1038/nrd.2015.18 (2016).

3 Ferruz, N. & De Fabritiis, G. Binding Kinetics in Drug Discovery. *Molecular Informatics* **35**, 216-226, doi:10.1002/minf.201501018 (2016).

4 Bernetti, M., Cavalli, A. & Mollica, L. Protein-ligand (un)binding kinetics as a new paradigm for drug discovery at the crossroad between experiments and modelling. *Medchemcomm* **8**, 534-550, doi:10.1039/c6md00581k (2017).

5 Bruce, N. J., Ganotra, G. K., Kokh, D. B., Sadiq, S. K. & Wade, R. C. New approaches for computing ligand-receptor binding kinetics. *Current Opinion in Structural Biology* **49**, 1-10 (2018).

6 You, W. L. & Chang, C. E. A. Role of Molecular Interactions and Protein Rearrangement in the Dissociation Kinetics of p38 alpha MAP Kinase Type-I/II/III Inhibitors. *Journal of Chemical Information and Modeling* **58**, 968-981, doi:10.1021/acs.jcim.7b00640 (2018).

7 Tiwary, P., Mondal, J. & Berne, B. J. How and when does an anticancer drug leave its binding site? *Science Advances* **3**, doi:10.1126/sciadv.1700014 (2017).

8 Weiss, R. G., Setny, P. & Dzubiella, J. Principles for Tuning Hydrophobic Ligand-Receptor Binding Kinetics. *Journal of Chemical Theory and Computation* **13**, 3012-3019, doi:10.1021/acs.jctc.7b00216 (2017).

9 Huang, Y. M. M., Raymundo, M. A. V., Chen, W. & Chang, C. E. A. Mechanism of the Association Pathways for a Pair of Fast and Slow Binding Ligands of HIV-1 Protease. *Biochemistry* **56**, 1311-1323, doi:10.1021/acs.biochem.6b01112 (2017).

10 Wang, Y., Martins, J. M. & Lindorff-Larsen, K. Biomolecular conformational changes and ligand binding: from kinetics to thermodynamics. *Chemical Science* **8**, 6466-6473, doi:10.1039/c7sc01627a (2017).

11 Pan, A. C., Xu, H. F., Palpant, T. & Shaw, D. E. Quantitative Characterization of the Binding and Unbinding of Millimolar Drug Fragments with Molecular Dynamics Simulations. *Journal of Chemical Theory and Computation* **13**, 3372-3377, doi:10.1021/acs.jctc.7b00172 (2017).

12 Tang, Z. Y., Roberts, C. C. & Chang, C. E. A. Understanding ligand-receptor non-covalent binding kinetics using molecular modeling. *Frontiers in Bioscience-Landmark* **22**, 960-981, doi:10.2741/4527 (2017).

13 Teo, I., Mayne, C. G., Schulten, K. & Lelievre, T. Adaptive Multilevel Splitting Method for Molecular Dynamics Calculation of Benzamidine-Trypsin Dissociation Time. *J. Chem. Theory Comput.* **12**, 2983-2989, doi:10.1021/acs.jctc.6b00277 (2016).

14 Tiwary, P., Limongelli, V., Salvalaglio, M. & Parrinello, M. Kinetics of protein-ligand unbinding: Predicting pathways, rates, and rate-limiting steps. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E386-E391, doi:10.1073/pnas.1424461112 (2015).

15 Mollica, L. *et al.* Kinetics of protein-ligand unbinding via smoothed potential molecular dynamics simulations. *Sci. Rep.* **5**, doi:10.1038/srep11539 (2015).

16 Rydzewski, J., Jakubowski, R., Nowak, W. & Grubrmuller, H. Kinetics of Huperzine A Dissociation from Acetylcholinesterase via Multiple Unbinding Pathways. *Journal of Chemical Theory and Computation* **14**, 2843-2851, doi:10.1021/acs.jctc.8b00173 (2018).

17 Tran, D. P., Takemura, K., Kuwata, K. & Kitao, A. Protein-Ligand Dissociation Simulated by Parallel Cascade Selection Molecular Dynamics. *Journal of Chemical Theory and Computation* **14**, 404-417, doi:10.1021/acs.jctc.7b00504 (2018).

18 Casasnovas, R., Limongelli, V., Tiwary, P., Carloni, P. & Parrinello, M. Unbinding Kinetics of a p38 MAP Kinase Type II Inhibitor from Metadynamics Simulations. *Journal of the American Chemical Society* **139**, 4780-4788, doi:10.1021/jacs.6b12950 (2017).

19 Torrie, G. M. & Valleau, J. P. NON-PHYSICAL SAMPLING DISTRIBUTIONS IN MONTE-CARLO FREE-ENERGY ESTIMATION - UMBRELLA SAMPLING. *J. Comput. Phys.* **23**, 187-199, doi:10.1016/0021-9991(77)90121-8 (1977).

20 Bello-Rivas, J. M. & Elber, R. Exact milestoning. *J. Chem. Phys.* **142**, doi:10.1063/1.4913399 (2015).

21 Votapka, L. W., Jagger, B. R., Heyneman, A. L. & Amaro, R. E. SEEKR: Simulation Enabled Estimation of Kinetic Rates, A Computational Tool to Estimate Molecular Kinetics and Its Application to Trypsin-Benzamidine Binding. *Journal of Physical Chemistry B* **121**, 3597-3606, doi:10.1021/acs.jpcb.6b09388 (2017).

22 Husic, B. E. & Pande, V. S. Markov State Models: From an Art to a Science. *Journal of the American Chemical Society* **140**, 2386-2396, doi:10.1021/jacs.7b12191 (2018).

23 Philip, S., Kumarasiri, M., Teo, T., Yu, M. & Wang, S. Cyclin-Dependent Kinase 8: A New Hope in Targeted Cancer Therapy? *Journal of Medicinal Chemistry*, 5073-5092 (2018).

24 Nemet, J., Jelicic, B., Rubelj, I. & Sopta, M. The two faces of CDK8, a positive/negative regulator of transcription. *Biochimie* **97**, 22-27, doi:10.1016/j.biochi.2013.10.004 (2014).

25 Knuesel, M. T., Meyer, K. D., Bernecky, C. & Taatjes, D. J. The human CDK8 subcomplex is a molecular switch that controls Mediator coactivator function. *Genes Dev.* **23**, 439-451, doi:10.1101/gad.1767009 (2009).

26 Xu, W. & Ji, J. Y. Dysregulation of CDK8 and Cyclin C in tumorigenesis. *Journal of Genetics and Genomics* **38**, 439-452, doi:10.1016/j.jgg.2011.09.002 (2011).

27 Schneider, E. V. *et al.* The Structure of CDK8/CycC Implicates Specificity in the CDK/Cyclin Family and Reveals Interaction with a Deep Pocket Binder. *J. Mol. Biol.* **412**, 251-266, doi:10.1016/j.jmb.2011.07.020 (2011).

28 Wang, T. J. *et al.* Discovery of novel CDK8 inhibitors using multiple crystal structures in docking-based virtual screening. *Eur. J. Med. Chem.* **129**, 275-286, doi:10.1016/j.ejmech.2017.02.020 (2017).

29 Mallinger, A. *et al.* Discovery of Potent, Selective, and Orally Bioavailable Small-Molecule Modulators of the Mediator Complex-Associated Kinases CDK8 and CDK19. *J. Med. Chem.* **59**, 1078-1101, doi:10.1021/acs.jmedchem.5b01685 (2016).

30 Schneider, E. V., Boettcher, J., Huber, R., Maskos, K. & Neumann, L. Structure-kinetic relationship study of CDK8/CycC specific compounds. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 8081-8086, doi:10.1073/pnas.1305378110 (2013).

31 Xu, W., Amire-Brahimi, B., Xie, X.-J., Huang, L. & Ji, J.-Y. All-atomic molecular dynamic studies of human CDK8: Insight into the A-loop, point mutations and binding with its partner CycC. *Comput. Biol. Chem.* **51**, 1-11, doi:10.1016/j.compbiolchem.2014.03.003 (2014).

32 Cholko, T., Chen, W., Tang, Z. Y. & Chang, C. E. A. A molecular dynamics investigation of CDK8/CycC and ligand binding: conformational flexibility and implication in drug discovery. *Journal of Computer-Aided Molecular Design* **32**, 671-685, doi:10.1007/s10822-018-0120-3 (2018).

33 Callegari, D. *et al.* Metadynamics Simulations Distinguish Short- and Long-Residence-Time Inhibitors of Cyclin-Dependent Kinase 8. *J. Chem. Inf. Model.* **57**, 159-169, doi:10.1021/acs.jcim.6b00679 (2017).

34 Tang, Z. Y. & Chang, C. E. A. Systematic Dissociation Pathway Searches Guided by Principal Component Modes. *Journal of Chemical Theory and Computation* **13**, 2230-2244, doi:10.1021/acs.jctc.6b01204 (2017).

35 Biasini, M. *et al.* SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* **42**, W252-W258, doi:10.1093/nar/gku340 (2014).

36 Gill, A. L. *et al.* Identification of novel p38 alpha MAP kinase inhibitors using fragment-based lead generation. *J. Med. Chem.* **48**, 414-426, doi:10.1021/jm049575m (2005).

37 Georgescu, R. E., Alexov, E. G. & Gunner, M. R. Combining conformational flexibility and continuum electrostatics for calculating pK(a)s in proteins. *Biophys. J.* **83**, 1731-1748 (2002).

38 Wang, J. M., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development

and testing of a general amber force field. *J. Comput. Chem.* 25, 1157-1174, doi:10.1002/jcc.20035 (2004).

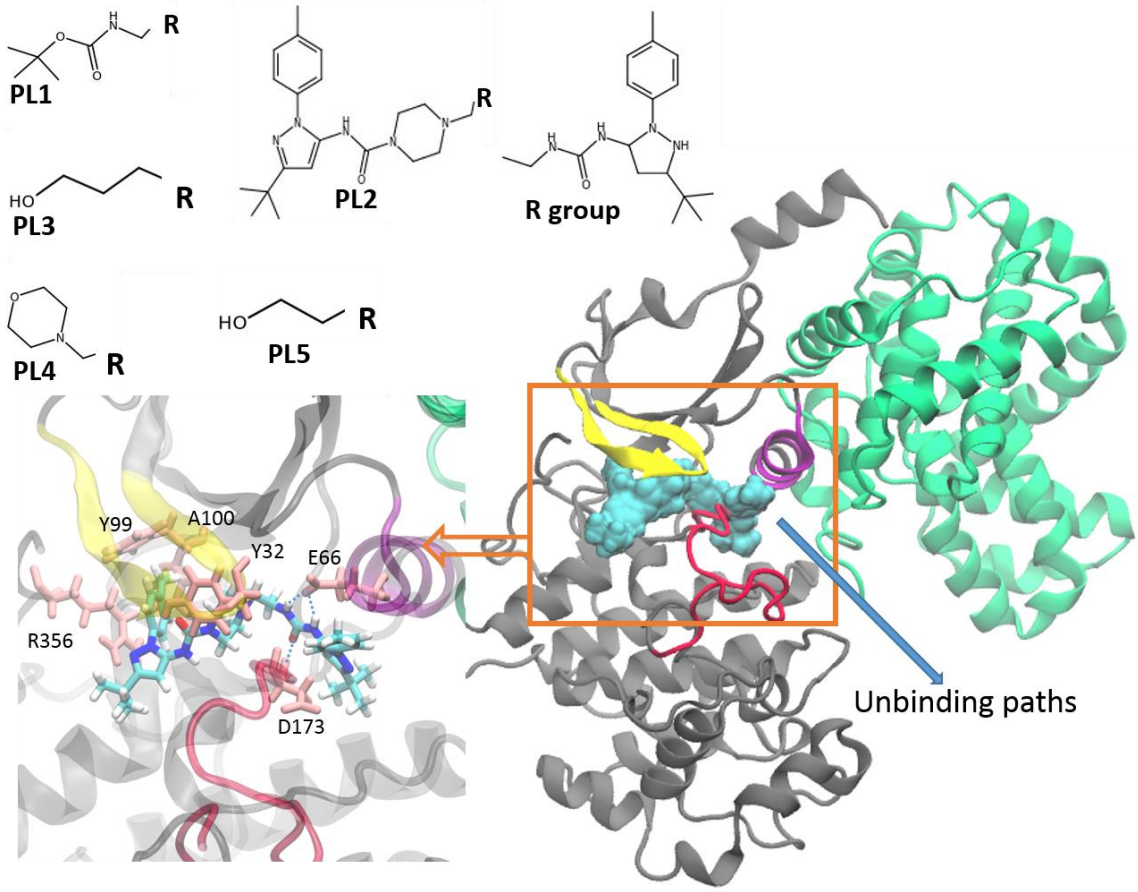

Figure 1. Protein structure and five PL compounds used in this study. Right: CDK8 (grey) and Cyclin C (green); left: a close-up view of ligand PL2 shown in licorice bound to CDK8. Conserved H-bonds between the urea linker of all five compounds with Glu66 and Asp173 are shown in blue. Regions with significant motions during ligand unbinding are presented by different color: αC helix (purple), β1 and β2 sheets (yellow) and activation loop (red). Residues engage in important interactions with PLs in the bound complex are labeled with one-letter amino acid codes (pink).

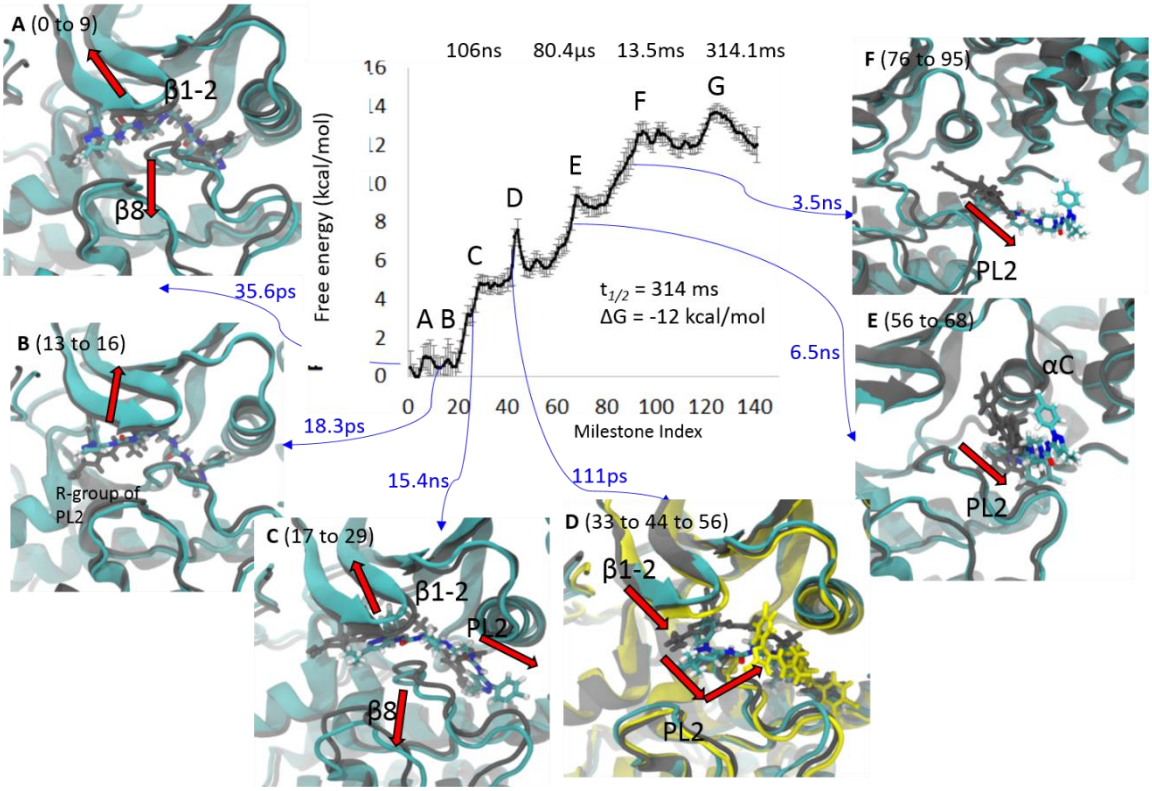

Figure 2. Free energy profiles along milestones and important conformational changes associated with each major energy barrier during PL2 dissociation. X-axis: milestone index to illustrate dissociation represented by PC modes shown in Figure 5B. $t_{1/2}$ and ΔG are computed residence time and absolute binding free energy, respectively. The transition time required for passing an energy barrier is in blue, and red arrows show major protein motions when PL2 crosses a peak labeled in letters. A-F: conformational changes of CDK8/CycC-PL2 when PL2 passes the labeled energy barriers. PL2 is shown in licorice and CDK8/CycC in new cartoon. First, second and third (if any) conformations are shown in gray, cyan and yellow. Conformational changes are indicated by red arrows.

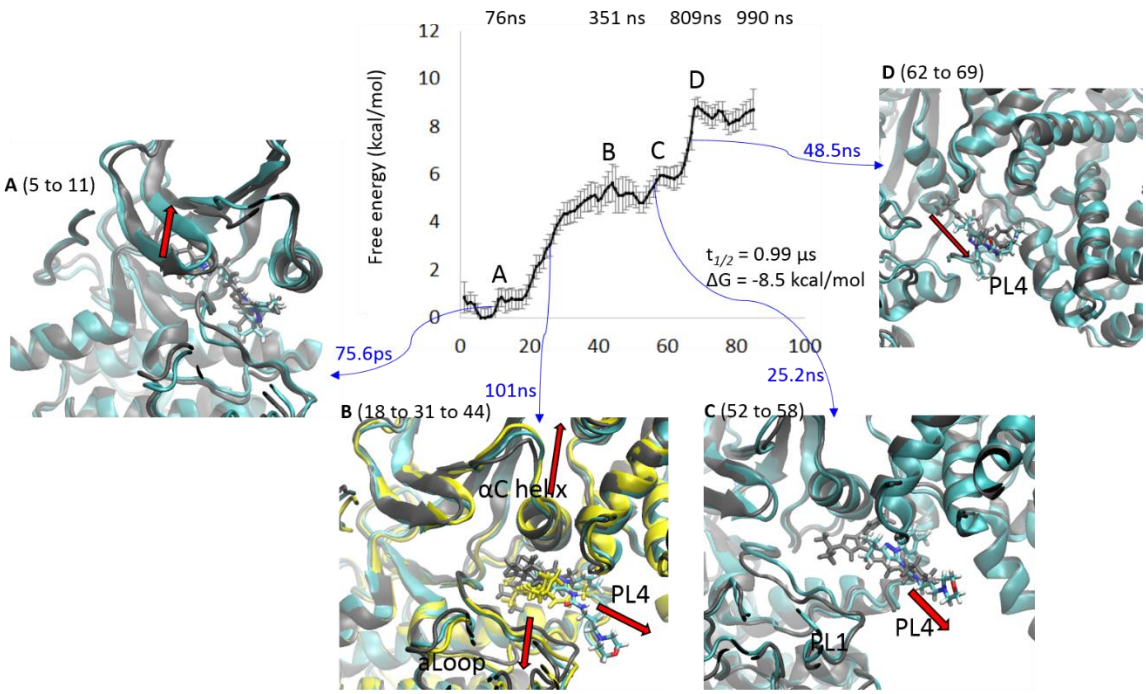

Figure 3. Free energy profiles along milestones and important conformational changes associated with each major energy barrier during PL4 dissociation. See Figure 2 legend for details.

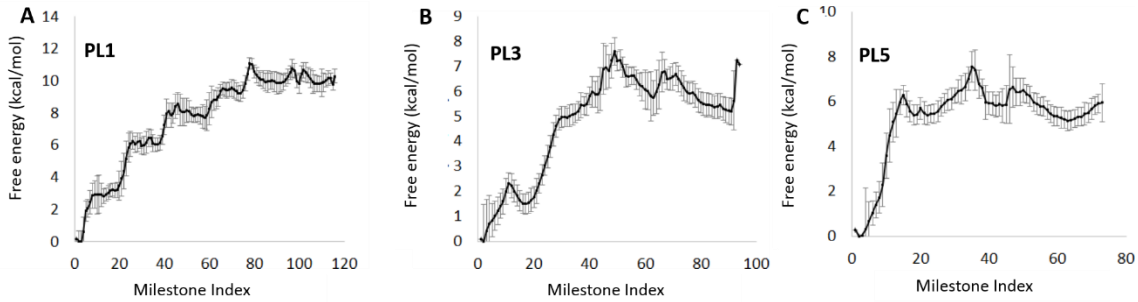

Figure 4. Free energy profiles along milestones and important conformational changes associated with each major energy barrier during PL1 (A), PL3 (B), and PL5 (C) dissociation.

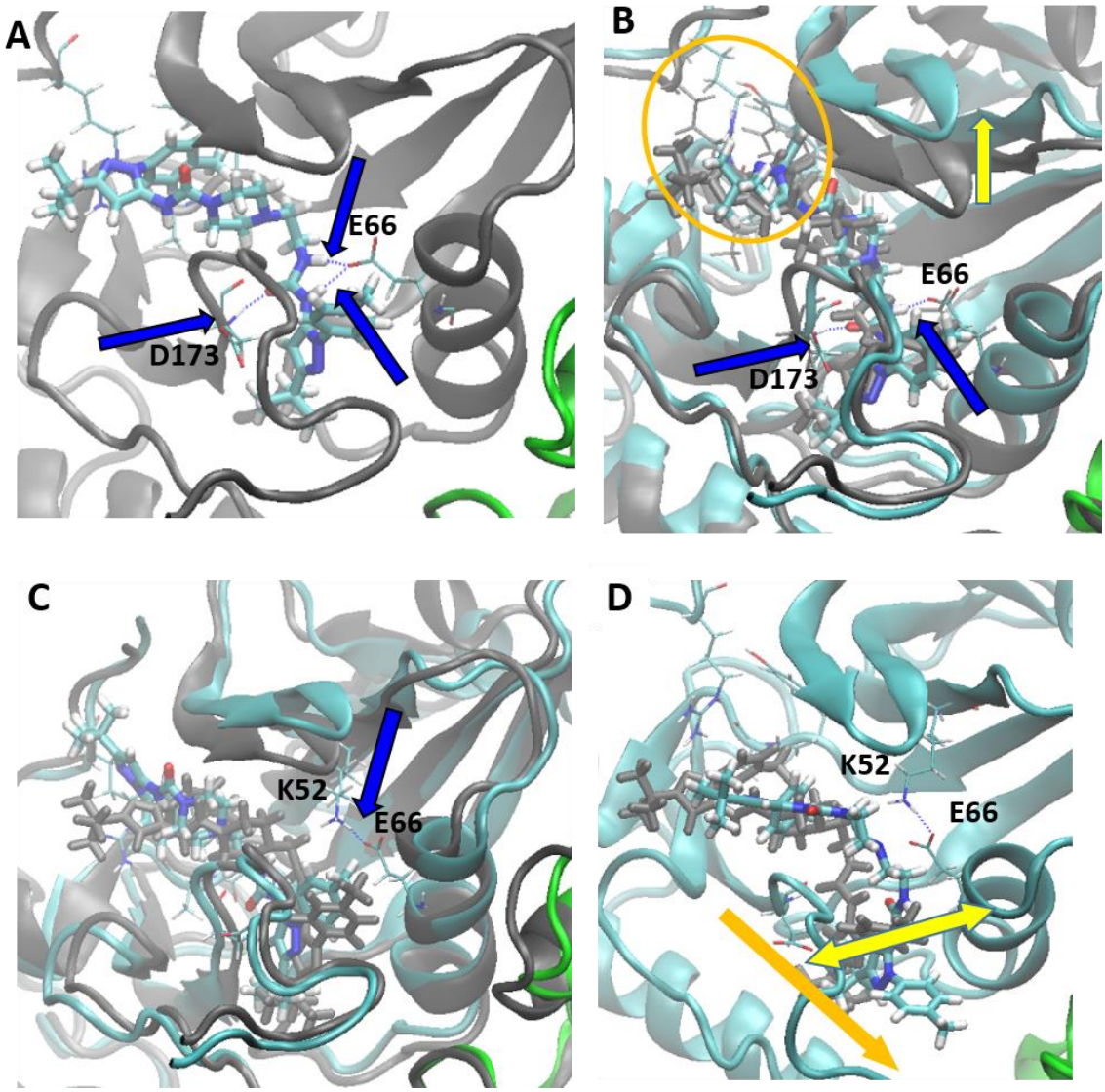

Figure 5. Key steps during PL2 dissociating from the binding pocket of CDK8. **A**. stable bound state seen in crystal structure (PDB: SKR1). Conserved H-bonds between all five compounds are indicated by blue dash line and arrows. **B**. β1 and β2 sheets that cover the top of the binding pocket open up. Interactions in the hinge area and front pocket (orange circle) are loosen but H-bonds between Glu66 and Asp 173 hold. Stable bound state shown in **A** is superimposed for comparison. **C**. H-bonds between Glu66 and Asp 173 break, and the conserved H-bond between Lys52 and Glu66 reforms. PL2 no longer forms stable H-bond with residues in the pocket, and starts to leave the pocket. The free energy continues

raising in this state, as shown in Figure 2 before energy barrier C. Stable bound state shown in **A** is superimposed for comparison. **D**. CDK8 arrangement for opening of the cleft to allow PL2 to leave the pocket, creating an energy barrier C in Figure 2.

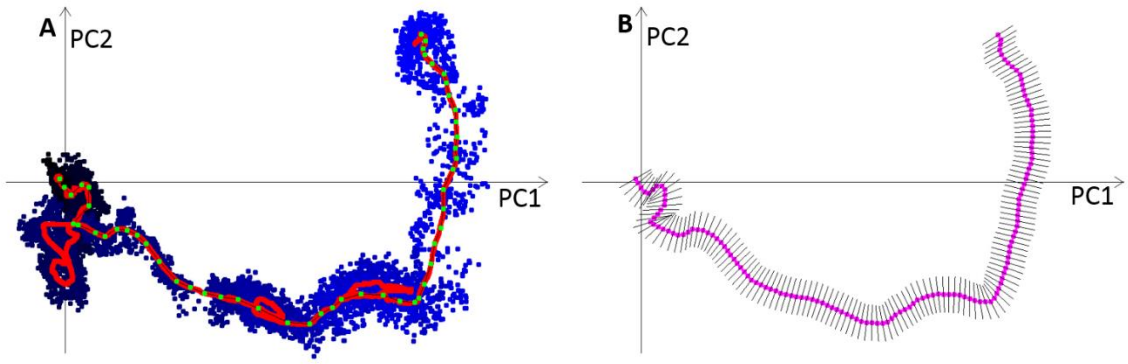

Figure 6. The projection of frames from a metadynamics trajectory of PL2 onto PC1/PC2 coordinates and spatial definition of milestones. A: the dissociation pathway projected on PC1/PC2 space (dots) and smoothed trajectory (red). Black and blue dots present the ligand position when it was inside or outside the protein binding pocket, respectively. The averaged position is illustrated in the red line as a smoothed trajectory. Green dots present the manually defined path based on the smoothed dissociation trajectory. B: the smoothed trajectory in A was optimized to remove the frames that do not lead to PL2 dissociation. The optimized path is in purple and the milestones are shown as black lines.