bioRxiv preprint doi: https://doi.org/10.1101/169607; this version posted July 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Energy Barriers, Molecular Motions, and Residence Time in Ligand Dissociation: A Computational Study on Type II Inhibitors Binding to CDK8/CycC

Zhiye Tang and Chia-en A. Chang*

Department of Chemistry, University of California, Riverside, CA92521

Telephone: (951) 827-7263

Email: chiaenc@ucr.edu

Abstract:

It is important but unclear how protein-ligand system motions affect free energy profiles, create energy barriers, and lead to slow residence time. We computed residence time (RT) and potential of mean force (PMF) of the dissociations of five structure-kinetic relationship (SKRs) series inhibitors of CDK8/CycC using metadynamics and milestoning theory. By using a novel way to represent the reaction coordinate based on principal component analysis (PCA), we found one-to-one mapping of hydrogen bond (H-bond) breakage and protein domain motions with the energy barriers in the PMFs. This work provides a novel and well-defined approach to study the dissociation kinetics of large and flexible systems.

bioRxiv preprint doi: https://doi.org/10.1101/169607; this version posted July 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Binding kinetics has become an important topic in molecular recognition because of the growing awareness of the correlation between kinetics and the drug efficacy. [1-4] k_{off} and residence time (RT) are particularly important as determinants of the efficacy of a drug candidate. [5] It is desirable to optimize these properties for drug discovery but, binding kinetics and its determinants in the ligand-receptor structures are not yet fully understood. It is also important to predict these quantities computationally when experimental measurements are not available. Therefore, computational methods, which are able to provide an atomistic description of temporal and spatial dissociation pathway details of ligand-receptor, can be employed to unravel the secret behind the RT and kinetics.

Cyclin-dependent kinase 8 (CDK8) is a kinase protein from the CDK family. CDK8 forms a complex with cyclin C (CycC), Med12, Med13 and performs phosphorylation involved with positive and negative signaling of transcription and regulation of transcription activities. [6-7] Abnormal activities of CDK8 and its partner CycC are related to a variety of human cancers. [8] 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. [9] Series of CDK8 drug candidates were discovered by structure based drug design and virtual screening. [10-11] A structure-kinetic relationship (SKR) series compounds were developed and targeted at the allosteric binding site of CDK8. [12] Up to date, a few computational works were performed to study CDK8. Molecular dynamics (MD) simulation was used to study the structural stability of CDK8/CycC complex. [13] The RT of SKR candidate compounds series [12] has been studied by metadynamics stimulations [14] using reweighting technique, but the timescale is fast on nanosecond level. [15]

Milestoning theory is useful method for studying kinetics. [16] Using rates and transition time from unbiased samplings between milestones dividing the transitions, milestoning computes the lifetime of each milestone and the kernel matrix which includes the transition probabilities between milestones. The potential of mean force (PMF) and the time the system takes to travel between milestones can be computed from the kernel matrix and the lifetime. Milestoning has witnessed a few applications and succeeded in reproducing the free energy surface and transition time. [17] How to describe the reaction coordinate using collective variables (CVs) faithfully is a challenging problem for applying methods including milestoning, metadynamics, umbrella sampling [18] and other methods. It is critical to include degrees of freedom involved in the reaction coordinate as more as possible. With

bioRxiv preprint doi: https://doi.org/10.1101/169607; this version posted July 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

this concern in mind, it is nature to solve this issue by using principal component analysis (PCA), a mathematical method that extracts the major motions from a collection of data. Therefore, it is desirable to take the advantage of PCA to suggest a more straightforward but robust way to define reaction coordinate for calculation using milestoning theory.

In this work, we used metadynamics to provide dissociation pathways and computed PMF and RT for five ligands from the SKR compounds series [12] using milestoning theory with a novel way for defining the milestones. The details of methods are included in SI. Using PC1 and PC2 of the metadynamics trajectories, we unambiguously defined the dissociation pathways and set up milestoning calculations using this definition (Figure SI 1). The computed PMF plots indicate clear energy barriers that correspond to and originate from the hydrogen bond (H-bond) formations and important interactions between ligand and CDK8, and also protein motions featured by relative motions of N-lobe, C-lobe of CDK8 and CycC. The computed RTs are on the time scale of milliseconds and microseconds and correlate with the compute PMF reasonably well.

In our metadynamics trajectories, ligands completely dissociate at about 5 to 30 ns. Instead of reweighting the kinetics from the metadynamics trajectories, we used milestoning theory to more accurately evaluate RTs. RTs were computed as the time it takes to bring the ligand from the bound state milestone to the solvated state milestone (where the gray horizontal lines intersect with the PMFs in Figure 1). The computed RTs are listed in Table 1. The relative ranking distinguishes ligand1 and 2 that have much slower RTs than the rest ligands. Among the ligands, ligand 1 has a RT on the millisecond level which reproduces the experimental value the best. The RTs of other ligands are on the micro- or submicro-level. Interestingly, the computed RT of ligand 2, whose experimental RT is slightly slower, is hundred times faster than ligand 1. We noticed that ligand 2 forms half less H-bonds with CDK8 and requires minor protein motions for it to dissociate compared to ligand 1. Therefore, the faster kinetics may be an intrinsic property underlying in the ligand 2 dissociation pathway found by metadynamics. The same reversed rank of ligand 1 and 2 was reported previously. [15]

The PMF plots along milestones describing ligand dissociations was computed by using milestoning theory and are shown in Figure 1. In the figure, the binding affinities and the energy barriers are labeled correspondingly. The computed binding affinities agree with experimental values reasonably well with deviations smaller than 3 kcal/mol. The PMF of

bioRxiv preprint doi: https://doi.org/10.1101/169607; this version posted July 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

ligand 1 suggests a binding affinity of roughly -12 kcal/mol, which is comparable to the experimental value -10 kcal/mol (Table SI 1). Interestingly, the computed energy barrier is higher than the experimental value but the computed RT is faster. The energy barriers shown in PMF correspond to observable conformational changes (Figure 2) and trespassing the energy barriers requires significant amount of time (Table 2). β1-2 on the N-lobe of CDK8 moves up for one or two Å to open the ATP binding site, and the R-group of the ligand 1 is unlocked (L1-A). Then, the R-group of the ligand moves up towards the N-lobe of CDK8 (L1-B). The huge energy barrier C comes when the R-group of the ligand shrinks into the ATP by passing β1-2 binding site (L1-C). This energy barrier takes 15 ns to cross, and is slowest in all observed energy barriers. The energy barrier D comes from the motion when the R-group of ligand 1 passes LYS52. Between D and E, the R-group of ligand 1 finds a local comfortable position between β1-3 and αC helix. The bulky part of the ligand passes through the narrow tunnel between αC helix and activation loop and results in the energy barrier E. Finally, the ligand gets solvated on the surface of CDK8 (L1-F and L1-G).

The motions and interaction indicated by the PMF can be further quantified by measuring distances, angles and H-bonds along the dissociation milestones. We computed the distances between centers of mass of heavy atoms of β1-2 and β8, αC helix and β8, C-lobe of CDK8 and CycC, and angle between centers of mass of heavy atoms of N-Lobe, hinge region and C-lobe of CDK8 for ligand 1 system, and projected these profiles onto the corresponding path and milestones (Figure SI 2). We also computed the H-bond between ligand 1 and CDK8, and projected the existence of H-bond onto the milestones (Figure SI 3). These plots correlate with the energy barriers and system motions perfectly. At energy barrier A, the opening of the CDK8 binding site is indicated by the β1-2/β8 and αC helix/β8 distances (Figure SI 2 A and B). At energy barrier C, ligand 1 starts to move through the ATP and allosteric binding sites, where the N-lobe of CDK8 needs to move away to open up the binding site again (Figure SI 2 A and B), and H-bonds between ligand 1 and Val27, Arg29, Asp103, Ala155, Asn156 and Arg356 break right at this energy barrier (Figure SI 3). As ligand 1 moves further toward allosteric dissociation, H-bonds between ligand 1 and Glu66 and Asp173, which lock ligand 1 in the binding site, breaks right at energy barrier D and E (Figure SI 3). Meantime, the β1-2/β8 distance starts to restore to native state value, as ligand 1 has already leave this region (Figure SI 2 A). On the contrary, the αC helix/β8 and C-lobe/CycC distances do not restore until ligand 1 completely dissociates at energy barrier F (Figure SI 2 B and C). This is also the moment when all bound state and

bioRxiv preprint doi: https://doi.org/10.1101/169607; this version posted July 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

intermediate state H-bonds break (Figure SI 3). These observations indicate that the ligand-protein interactions and protein domain motions contribute specifically to the energy barriers shown in the PMF.

The ligand 2 binding affinity predicted by milestoning is roughly -10 kcal/mol. Although the computed binding affinity of ligand 2 is similar to experimental value (Table SI 1), the computed RT is orders faster. Unlike Ligand 1, the PMF of ligand 2 only suggests three major relevant barriers (Figures 1 and 2); the R-group of ligand 2 passes through α-C helix (L2-A); ligand 2 completely moves out of the allosteric binding site (L2-B); ligand 2 gets solvated and desolvated (L2-C, D, E, F). Moreover, ligand 2 starts to encounter a huge energy barrier (about 3 kcal/mol) at milestone 5, while ligand 1 starts to encounter the first major energy barrier at milestone 20. This is because the stretching motion and preparation for dissociation of ligand 1 can be clearly distinguished by PC1/PC2 space path, but for ligand 2 the motions in the first 20 ns metadynamics simulations are dominated by protein and ligand vibrations near the same bound state so that the track of these regions in PC1/PC2 space are convoluted and we cannot separate the motions using our approach (Figure SI 4). The PMF suggests that three major events are involved in the dissociation of ligand 2, and this is less complicated than ligand 1. The PMFs of the other ligands indicate roughly -6 kcal/mol binding affinities for ligand 5, 10 and 11, which are 2 to 3 kcal/mol shallower than the experimental values (Table SI 1). Similar to ligand 2, the tracks in PC1/PC2 space of these three ligands are also convoluted in the first 10-20 ns. Therefore, the dynamics included in these periods of simulations are projected onto only a few milestones (1-5), and the PMFs of these ligands also start to encounter the major energy barrier within 10 milestones. The first distinguishable energy barriers for these ligands are also from the motion when the ligand passes through αC helix. The other major energy barriers in the PMFs of these ligands are from desolvation of the ligands.

The average lifetime of the milestones for the systems are about 1 to 7 ps. In our calculation, the time resolution is 0.1 ps, and this means the timestep resolution is able to capture the transition between the milestones. The lifetime of the milestones is affected by two factors. First, although our milestones in the PC1/PC2 space is evenly distributed along the path, the areas divided into each milestone are not exactly identical. The first 20 milestones for ligand 1 system are located close to each other, so that the life time of these milestones can be as low as 0.5 ps (Figure SI 1 and Figure SI 5). The other factor is the nature of the population in that milestone. The lifetime of the milestones corresponding to the sloped

bioRxiv preprint doi: https://doi.org/10.1101/169607; this version posted July 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

regions near the energy barriers of the PMF is generally shorter than those corresponding to the flat regions on the PMF (Figure SI 5). The computed RT of the systems are several orders faster than experimental values and this is similar to other works. A millisecond-level substrate translocation study using milestoning showed the energy barrier was 4-5 kcal/mol and transition time was 1.6 μs. [19] Another milestoning study of CO entry and exit in myoglobin also suggested minimal energy barrier path with energy barrier and transition time on the same scale. [20] This discrepancy implies that MD simulations tend to overestimate the rate of slow motions.

Although we used distance of centers of mass, a naïve way to define the CVs, in metadynamics, we computed PMFs and RTs using a novel approach. This approach extracts the most remarkable motions from the metadynamics trajectory by choosing the first two PC modes, provides a 2D projection of the complex motions of the diffusion of ligand and ligand/protein internal motions. The path combines related motions from multiple degrees of freedom, not limiting to a few CVs like distance or angles, although the ligand-protein relative motions dominate in these degrees of freedom. The advantage of this approach is that it reduces the dimension of the spatial motions, but keeps the most important dimensions of the motions. This approach can be further improved. First, the 2D projection of the smoothed track must be able to suggest a clear path without convoluted loops in the important regions of the dissociation. For example, as shown in Figure SI 6, the conformations involved with dissociation of ligand 11 is projected onto the PC1/PC2 space as a folded curve indicated by the black arrow, and this region cannot be effectively distinguished by the 2D projection. For the same reason, the early period of the metadynamics trajectories for ligand 2, 5, 10 and 11 are more or less projected onto a small convoluted region on the PC1/PC2 space (Figure SI 4), and this partly results in the overestimated dissociation kinetics and underestimated PMF for these ligands. An optimization of the metadynamcis path using methods like string method [21] may further improve the performance by removing the convoluted regions in the PC1/PC2 space. Also, only the first two PC mode may not be able to describe all major motions in the system. As shown in Figure SI 7, PC1 and PC2 captures 85% (sum of eigenvalue of PC1 and PC2) of the motions in ligand 1 complex, and roughly 75% for the other ligand complexes. This implies that this approach can be expanded to 3-dimensional cases in which the first three PC modes can be used to generate a 3D PC space to include slightly more major motions of the system. A 3D curve in the space can be used to define milestones that resemble a series of Domino along a thread in the same way. Although the path will be 3D, but the

bioRxiv preprint doi: https://doi.org/10.1101/169607; this version posted July 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

milestones will still be one dimensional as in the 2D space.

We used metadynamics and milestoning theory to study the dissociations of five SKR series compound from CDK8/CycC complex and suggested a novel way based on PCA to define the reaction coordinate of the dissociation. By using this novel reaction coordinate definition, we yielded RTs on milli- or micro-second level, and explained the origins of energy barriers indicated on PMF by using H-bonds and protein motions that were projected to the same milestone coordinate.

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).

Abbreviations

RT residence time; PMF potential of mean force; SRK structure-kinetic relationship; PCA principal component analysis; H-bond hydrogen bond; MD Molecular dynamics; CV collective variable

bioRxiv preprint doi: https://doi.org/10.1101/169607; this version posted July 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

References

(1) Swinney, D. C., *Curr. Opin. Drug Discovery Dev.* **2009,** *12* (1), 31-39.

(2) Schreiber, G., et al., *Chem. Rev.* **2009,** *109* (3), 839-860.

(3) Pan, A. C., et al., *Drug Discov. Today* **2013,** *18* (13-14), 667-673.

(4) Guo, D., et al., *Br. J. Pharmacol.* **2012,** *166* (6), 1846-1859.

(5) Copeland, R. A., *Nat. Rev. Drug Discov.* **2016,** *15* (2), 87-95.

(6) Nemet, J., et al., *Biochimie* **2014,** *97*, 22-27.

(7) Knuesel, M. T., et al., *Genes Dev.* **2009,** *23* (4), 439-451.

(8) Xu, W.; Ji, J. Y., *Journal of Genetics and Genomics* **2011,** *38* (10), 439-452.

(9) Schneider, E. V., et al., *J. Mol. Biol.* **2011,** *412* (2), 251-266.

(10) Wang, T. J., et al., *Eur. J. Med. Chem.* **2017,** *129*, 275-286.

(11) Mallinger, A., et al., *J. Med. Chem.* **2016,** *59* (3), 1078-1101.

(12) Schneider, E. V., et al., *Proc. Natl. Acad. Sci. U.S.A.* **2013,** *110* (20), 8081-8086.

(13) Xu, W., et al., *Comput. Biol. Chem.* **2014,** *51*, 1-11.

(14) Barducci, A., et al., *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011,** *1* (5), 826-843.

(15) Callegari, D., et al., *J. Chem. Inf. Model.* **2017,** *57* (2), 159-169.

(16) Bello-Rivas, J. M.; Elber, R., *J. Chem. Phys.* **2015,** *142* (9).

(17) Elber, R., *Q. Rev. Biophys.* **2017,** *50*.

(18) Torrie, G. M.; Valleau, J. P., *J. Comput. Phys.* **1977,** *23* (2), 187-199.

(19) Ma, W.; Schulten, K., *J. Am. Chem. Soc.* **2015,** *137* (8), 3031-3040.

(20) Yu, T.-Q., et al., *J. Am. Chem. Soc.* **2015,** *137* (8), 3041-3050.

(21) Weinan, E., et al., *Phys. Rev. B* **2002,** *66* (5).

bioRxiv preprint doi: https://doi.org/10.1101/169607; this version posted July 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

	Expt. RT	Computed RT
Ligand1	1626 Min	34.7 ms
Ligand2	1944 Min	112.8 us
Ligand5	14 Min	1.2 us
Ligand10	7 Min	0.18 us
Ligand11	57 Min	0.25 us

Table 1. The experimental and computed residence time (RT) of the five ligands.

	Energy Barrier	Passing Energy Barrier	From Bound State
L1	A	35.6 ps	35.6 ps
	B	18.3 ps	69.0 ps
	C	15.4 ns	106 ns
	D	111 ps	482 ns
	E	6.48 ns	80.4 μs
	F	3.52 ns	13.5 ms
	G	77.1 ps	34.7 ms
L2	A	1.36 ns	1.36 ns
	B	4.19 ns	306 ns
	C	437 ps	4.89 μs
	D	158 ps	71 μs
	E	42.6 ps	122.8 μs

Table 2. The time required to cross energy barriers for ligand 1 and ligand 2. The time required to cross one energy barrier is in column 3, and the time required to travel from bound state to pass one energy barrier is in column 4.

bioRxiv preprint doi: https://doi.org/10.1101/169607; this version posted July 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

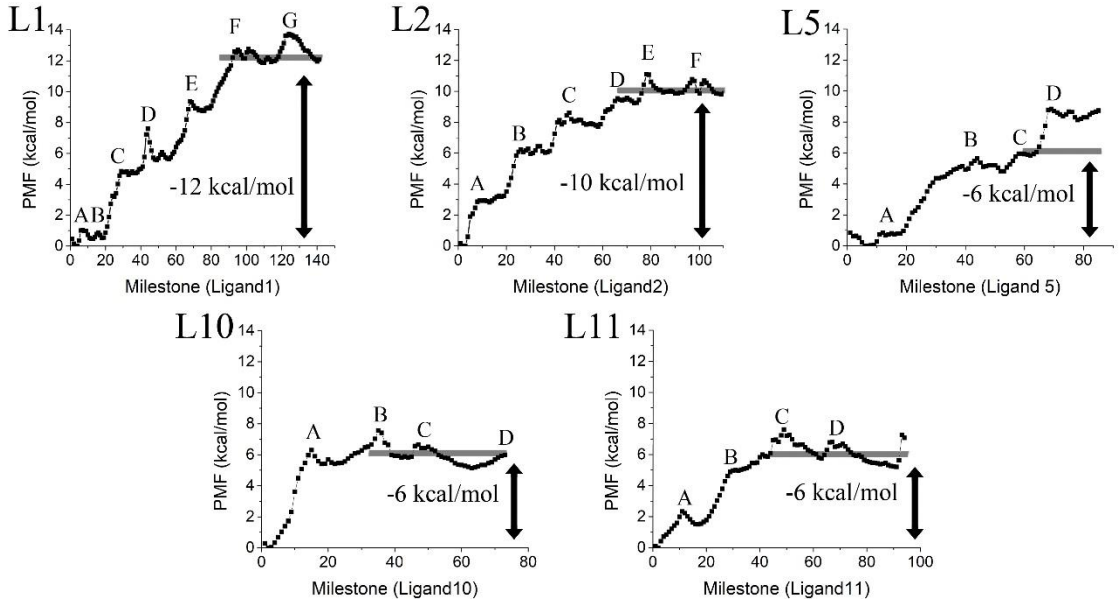

Figure 1. PMF plots of the five ligand dissociations along milestones. Important energy barriers are label by A-G. The desolvated states used for computing binding affinities and RTs are indicated by the gray horizontal line. Binding affinities are labeled next to the black double arrows.

bioRxiv preprint doi: https://doi.org/10.1101/169607; this version posted July 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

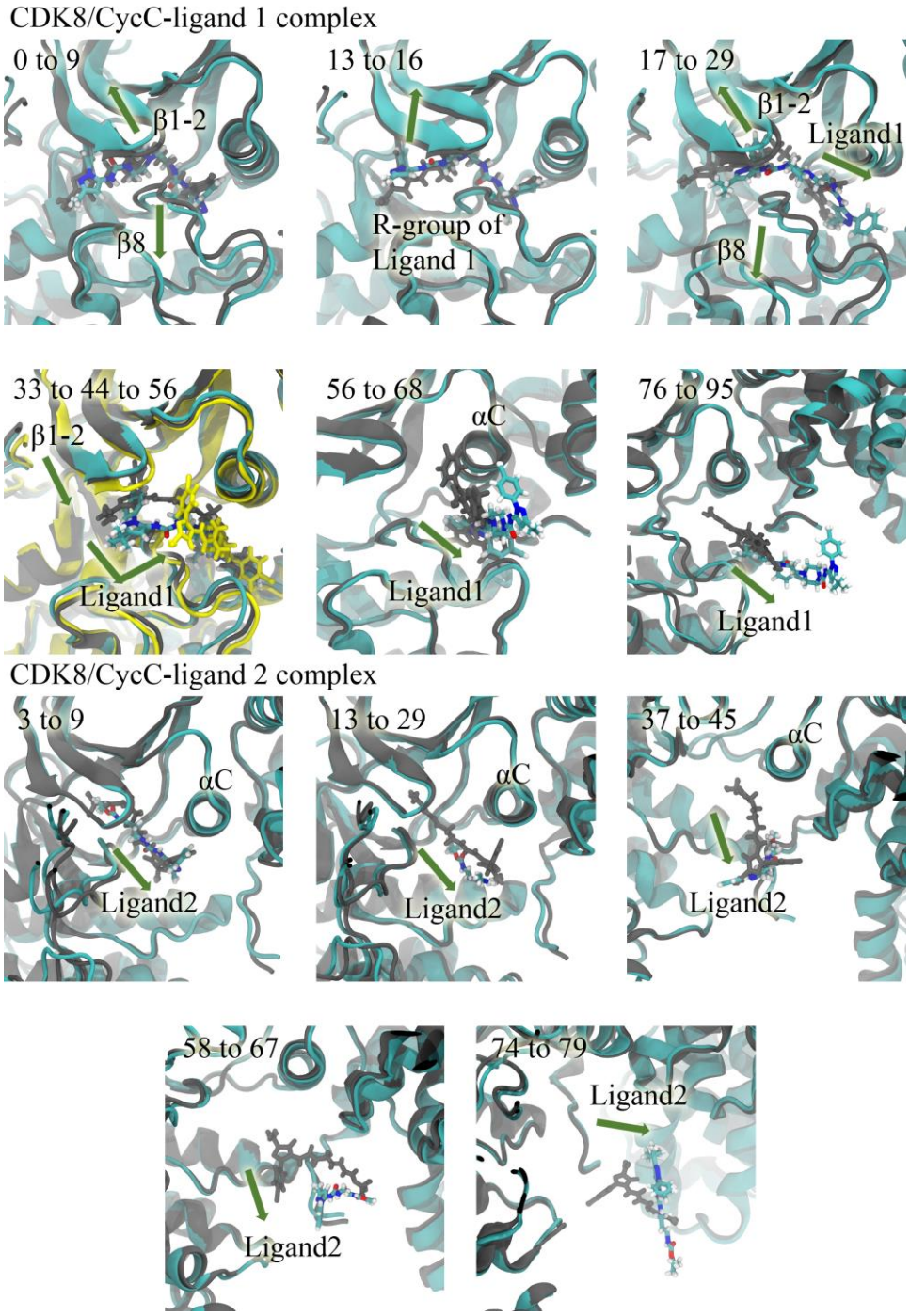

Figure 2. Conformational changes of CDK8/CycC-ligand 1 and 2 complexes along milestones. Ligands are shown in licorice and CDK8/CycC is shown in new cartoon. Indices of milestones are labeled on left top corner. First, second and third (if any) conformations are shown in gray, cyan and yellow. Conformational changes are indicated by green arrows.

bioRxiv preprint doi: https://doi.org/10.1101/169607; this version posted July 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Supporting Information for

Energy Barriers, Molecular Motions, and Residence Time in Ligand Dissociation: A Computational Study on Type II Inhibitors Binding to CDK8/CycC

Zhiye Tang and Chia-en A. Chang*

Department of Chemistry, University of California, Riverside, CA92521

Telephone: (951) 827-7263

Email: chiaenc@ucr.edu

bioRxiv preprint doi: https://doi.org/10.1101/169607; this version posted July 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Molecular Dynamics

We obtained the initial structures of CDK8/CycC-ligand 1, 2, 5, and 11 complexes from PDB database (PDB ID: 4F6W, 4F7L, 4F6U, 4F7N [1]), and manually modified PDB structure of CDK8/CycC-ligand 11 (4F7N) to get the initial structure of CDK8/CycC-ligand 10. We built the missing activation loop of the structures by using Swiss Model [2] based on p38 DFG-out crystals structure (PDB ID: 1W82 [3]) as a template. The other missing residues of CDK8 residue 1 to 360 and CycC residue 361 to 619 were also added by using Swiss-Model. We determined the protonation states of histidine residues in the CDK8/CycC-ligand complexes by using MCCE [4], and all histidine residues were in the single protonated state. Then we added missing hydrogen atoms to the structures by using tleap module from AmberTool14. [5]

AMBER FF14SB force field and GAFF [6] were used for CDK8/CycC complex and the ligands. We performed minimizations on the hydrogen atoms, sidechain atoms, and the whole structures to relax the complex conformations. Then 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 we minimized the water and the solvated systems, we equilibrated the water molecules with the solute fixed at 298K for 20 ns in NPT (isothermal-isobaric) ensemble. Then the systems were heated from 200 K to 298K gradually. 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. All MD simulations were performed using pmemd.cuda from AMBER 14 package. [5] The trajectories were resaved every 20 ps for post-analysis.

Metadynamics

We obtained the equilibrated conformations at 298K, and conformations after 100, 400, 500 ns MD simulations as initial structures, and performed twelve metadynamics simulations starting from these initial structures for each of the CDK8/CycC-ligand complexes. The distance between centers of mass of the heavy atoms on the ligands and the heavy atoms of residue 26 to 38 96 to 106 356 to 359 on CDK8 were chosen as the collective variables (CVs), and the forces from the Gaussian functions were only applied on the atoms in the ligands. A Gaussian of 0.02 kcal/mol was deposited on to the CV coordinate every 0.1 ps. A 1-fs time step were used in the simulations, and the metadynamics run were performed until the ligand dissociation event occurs (roughly 25-30 ns). The metadynamics simulations were performed using NAMD. [7] The metadynamics simulations were saved every 5 ps as initial conformations for milestoning calculations.

bioRxiv preprint doi: https://doi.org/10.1101/169607; this version posted July 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Milestoning

We selected the conformations from the metadynamics trajectory every 50 ps and obtained roughly 500 to 600 conformations as initial conformations to run 20 replicas of 100 ps conventional MD simulations with 2 fs time step at 298K. The replica trajectories were saved every 50 fs. Other settings were the same as the MD simulation section above. Then we used the principal component analysis (PCA) to define the milestones.

First, we computed PCA for the metadynamics trajectory. We selected the α-carbon atoms of CDK8/CycC and heavy atoms in the ligand, and computed covariance matrix of the Cartesian coordinates of these atoms using the first frame in each trajectory as references. We saved the eigenvectors of the covariance matrix, and use equation below to compute PC1 and PC2 for the metadynamics trajectory.

$$PC_i = R^T (X(t)-\langle X\rangle) \qquad \text{Eq. 1}$$

where R^T is the eigenvector with highest eigenvalue for PC1 and second highest for PC2, X(t) and <X> are the Cartesian coordinates of the selected atoms at time t and average over the trajectory. Then, the original metadynamics trajectory was smoothed by averaging forward and backward 100 frames, and PC1 and PC2 of averaged trajectory were also computed by using Eq. 1.

We use CDK8/CycC-ligand 1 system as an example to show how we set up the milestone using PC1 and PC2 values of the original and smoothed trajectories. The PC1 and PC2 values of the original and smoothed trajectories were plotted in the PC1/PC2 space (Figure SI 1 A). The track of smoothed trajectory in PC1/PC2 space indicates a clear dissociation path for CDK8/CycC-ligand 1 complexes. Along the track of the smoothed trajectory (red line in Figure SI 1 A), we manually built a path that represented the dissociation pathway. This circle region in the track of the smoothed trajectory at the negative of X-axis (in Figure SI 1 A) was intentionally neglected when manually building the path. This is because this region corresponds to the motion when ligand 1 moves towards the APT binding site and moves back towards dissociation via the allosteric binding site and this is not relevant to the dissociation of ligand 1. We smoothed the path by using Origin. A series of point was located along the path smoothed by Origin every 3.0 in the PC1/PC2 space (purple dots in Figure SI 1 B). Then short lines of length 20.0 were put perpendicularly to the path on the series of points and the slopes of the lines were optimized to minimize the overlapping of the lines and these lines were used as milestones (black lines in Figure SI 1 B). The first 30 ps of the 100 ps replica trajectories were also projected onto PC1/PC2 space using Eq. 1 and the projection points were divided by the milestones into cells between milestones

bioRxiv preprint doi: https://doi.org/10.1101/169607; this version posted July 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(Figure SI 1 C). We removed the data points of the 100 ps replica trajectories that fall onto the areas that were 10 unit away from the path in PC1/PC2 space (Figure SI 1 D). Using the data shown in Figure SI 1 D, we computed the duration time of each milestone and the transition counts between adjacent milestones. Finally, the PMF and residence time (RT) were computed following exact milestoning. [8]

Post-Analysis of Metadynamics Trajectories

We computed the hydrogen bonds (H-bonds) for the metadynamics trajectories. For a H-bond X-H…Y, the distance between H and Y was limited to 2.5 Å and the complimentary angle of X-H-Y was limited to 30 degrees. The h-bonds were computed for each frame of metadynamics trajectory. As described in Milestoning Section, each frame in metadynamics trajectory was assigned to a milestone, and by using the same assignment, we projected the H-bonds onto the milestone coordinate in the same way.

We computed the distances between centers of mass of heavy atoms of β1-2 (residue 26 to 36) and β8 (residue 170 to 173), αC helix (residue 61 to 71) and β8. We also computed the angle between centers of mass of heavy atoms of N-Lobe (residue 20 to 100), hinge region (residue 100 to 103) and C-lobe of CDK8 (residue 104 to 160 and 190 to 340). Similarly, we projected the distances and angle onto the milestones as described above.

The metadynamics dissociation trajectory were also projected onto the milestones. The conformations along the dissociation trajectories were clustered according to the milestones they belong to, and the conformations in one milestones were averaged to yield the average conformation for the milestone. Then a new trajectory along the milestones was generated by collecting the conformations for each milestone. The same procedure was repeated for the complexes with the five ligands.

bioRxiv preprint doi: https://doi.org/10.1101/169607; this version posted July 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

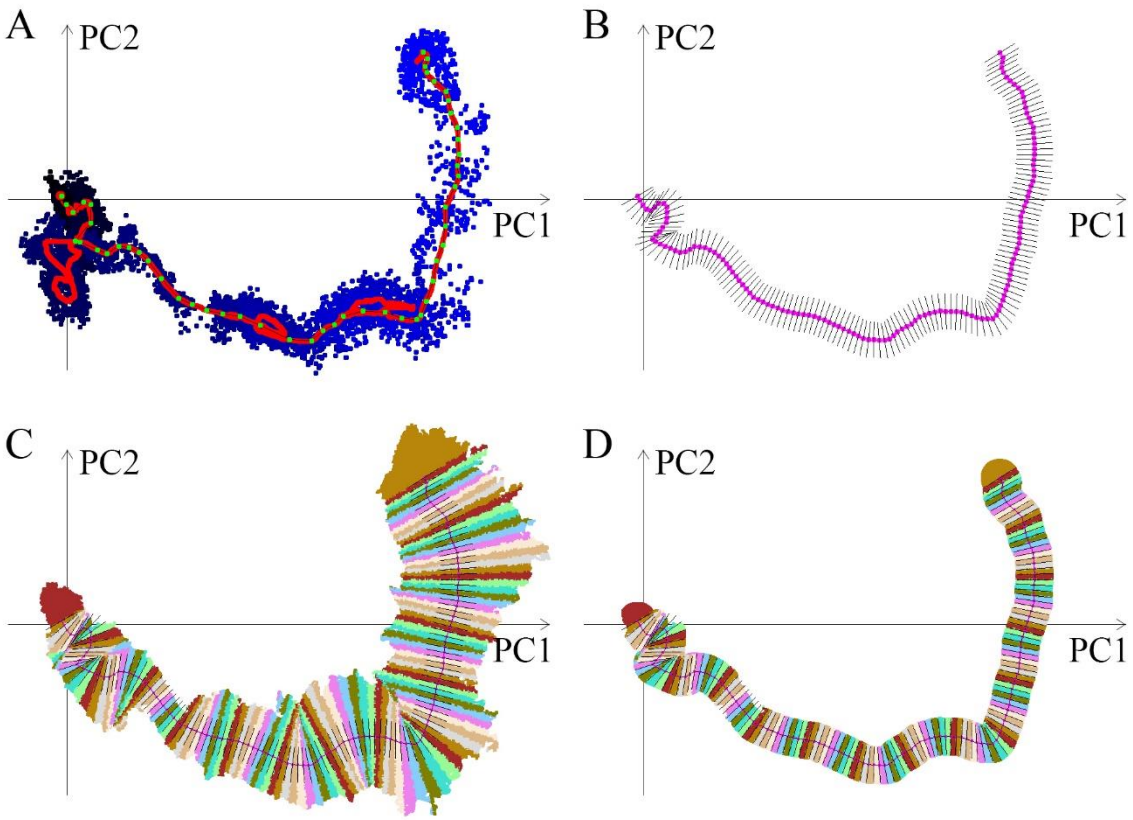

Figure SI 1. The projection of metadynamics trajectory of ligand 1 onto PC1/PC2 space and definition of milestones. A: the tracks of original trajectory (dark and light blue dots) and smoothed trajectory (red), and the manually defined path (green dots). B: the optimized path (purple) and milestones (black lines). C: the path and milestone (black lines) and all projection of short replicas of unbiased MD simulations used for milestoning theory (the color areas). D: the area of projections in C within 10 from the path (black line) used for computing PMF and RT.

bioRxiv preprint doi: https://doi.org/10.1101/169607; this version posted July 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Ligand	Residence Time (Min)	k_{on} (1/s/M)	k_{off} (1/s)	K_{eq} (1/M)	ΔG (kcal/mol)
1	1626	2.99×10^{2}	1.02×10^{-5}	2.9×10^{7}	-10.18
2	1944	5.73×10^{2}	8.57×10^{-6}	6.7×10^{7}	-10.67
5	14	1.68×10^{3}	1.18×10^{-3}	1.4×10^{6}	-8.39
10	7	1.85×10^{3}	2.41×10^{-3}	7.7×10^{5}	-8.02
11	57	3.65×10^{3}	2.90×10^{-4}	1.3×10^{7}	-9.68

Table SI 1. The experimental data of CDK8/CycC-ligand complexes taken from ref [1].

bioRxiv preprint doi: https://doi.org/10.1101/169607; this version posted July 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

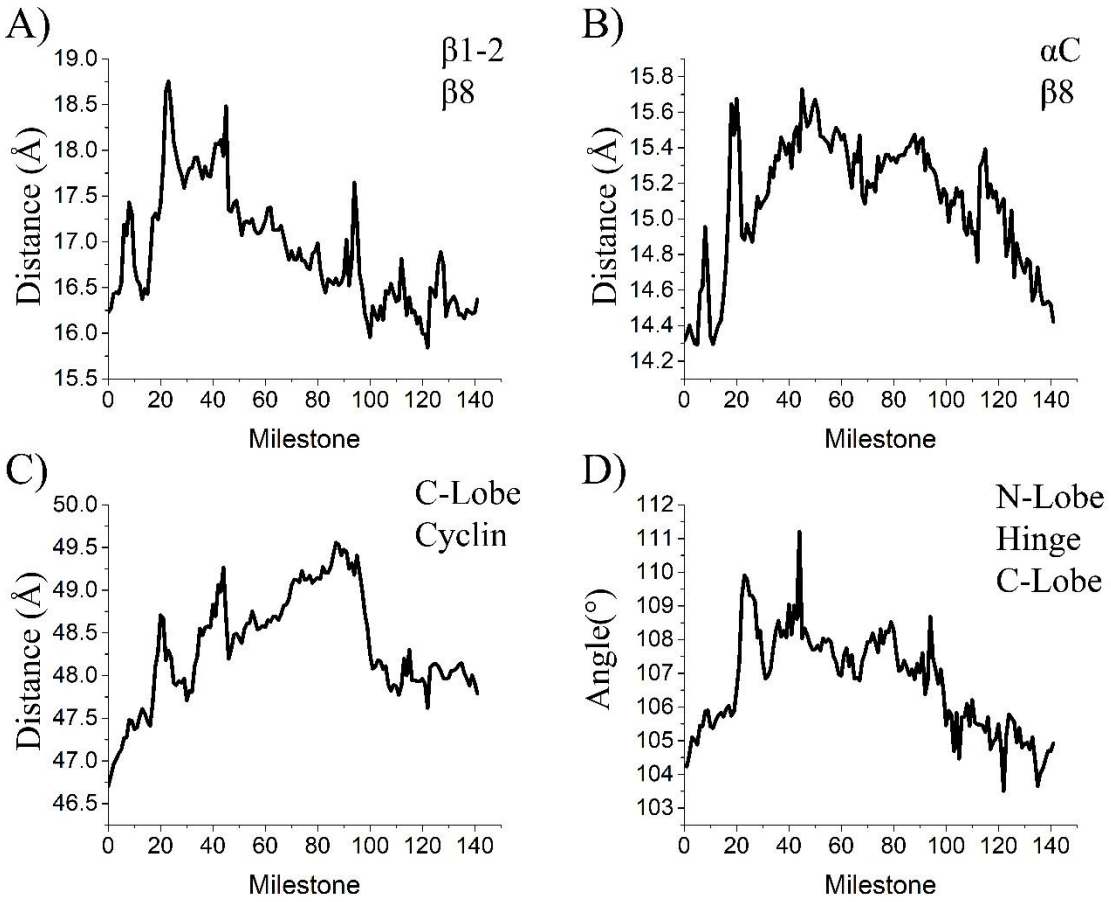

Figure SI 2. Distances and angles between atom groups on CDK8 and ligand 1 along milestones. A: distances between β1-2 (residue 26 to 36) and β8 (residue 170 to 173). B: distance between αC helix (residue 61 to 71) and β8. C: Distance between C-lobe of CDK8 and Cyclin C. D: angle between N-Lobe, hinge region (residue 100 to 103) and C-lobe of CDK8.

bioRxiv preprint doi: https://doi.org/10.1101/169607; this version posted July 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

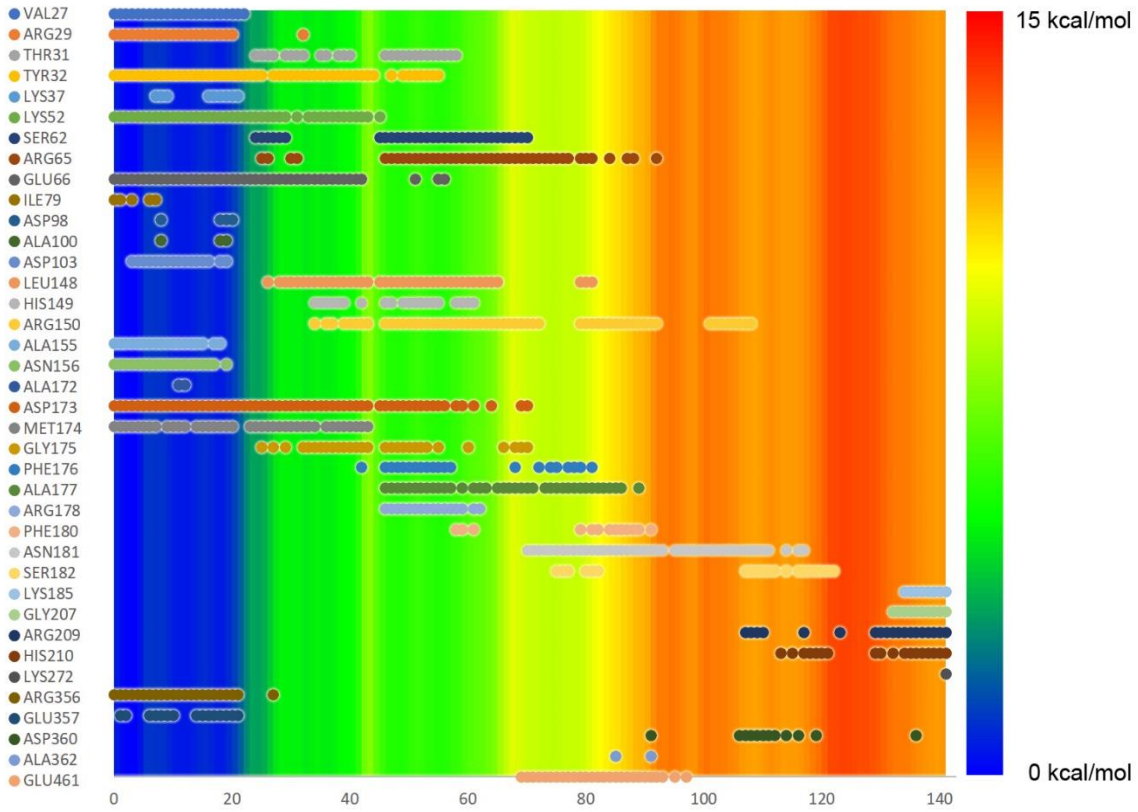

Figure SI 3. Plot of H-bond between ligand 1 and CDK8/CycC complex along milestones. The PMF of ligand 1 dissociation is colored for comparison and the color bar is shown on the right.

bioRxiv preprint doi: https://doi.org/10.1101/169607; this version posted July 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

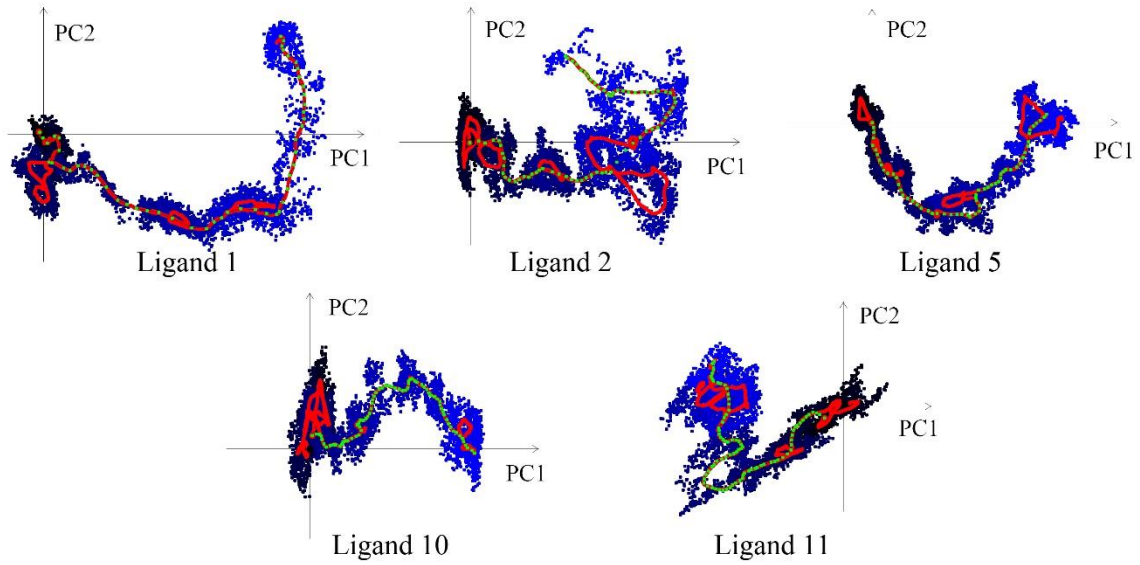

Figure SI 4. The paths of dissociations of the five ligands from metadynamics trajectories projected on PC1/PC2 space. The tracks of original trajectories are marked by dark and light blue points, and smoothed trajectories are marked by red points. The manually defined paths are marked by green points.

bioRxiv preprint doi: https://doi.org/10.1101/169607; this version posted July 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

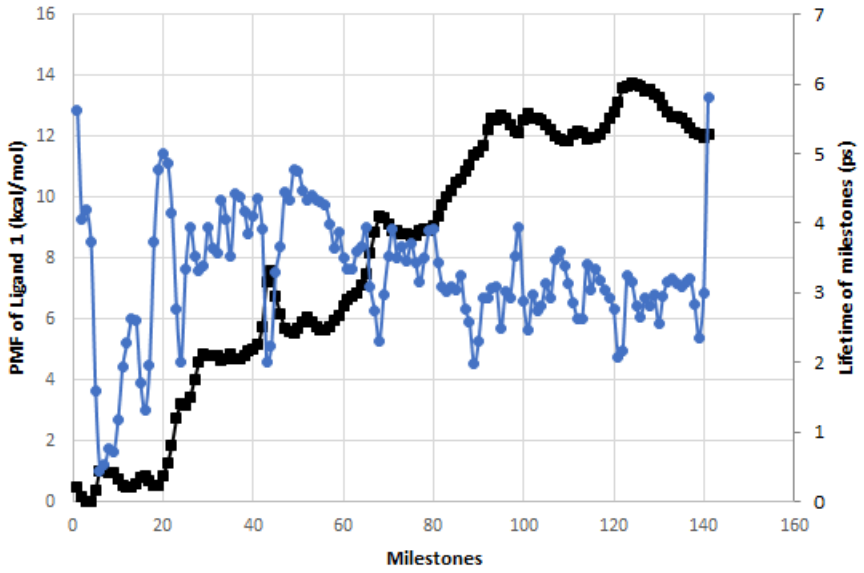

Figure SI 5. The PMF and lifetime of milestones. PMF is shown as black dotted line and the lifetime is shown as blue dotted line.

bioRxiv preprint doi: https://doi.org/10.1101/169607; this version posted July 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

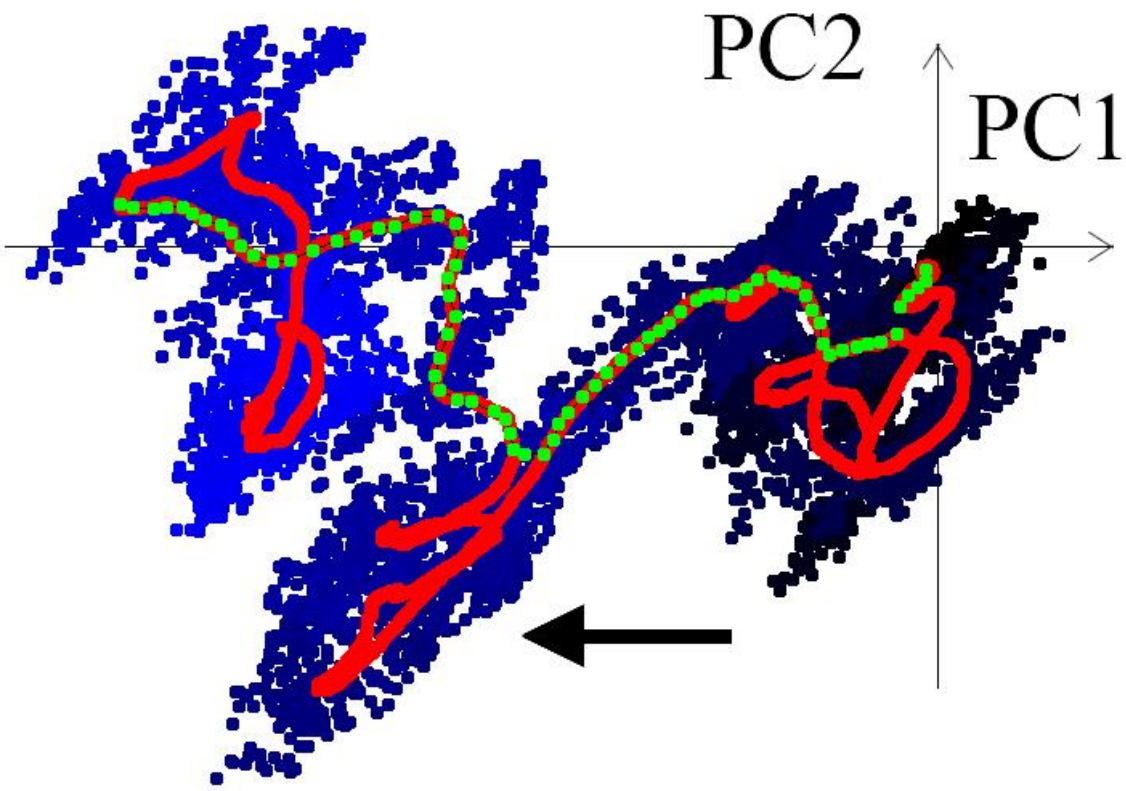

Figure SI 6. The original and smoothed track of an unused metadynamics trajectory of ligand 11 in PC1/PC2 space. The original track is shown as dark and light blue dots. The smoothed path is shown as red line. The path is defined as green dots.

bioRxiv preprint doi: https://doi.org/10.1101/169607; this version posted July 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

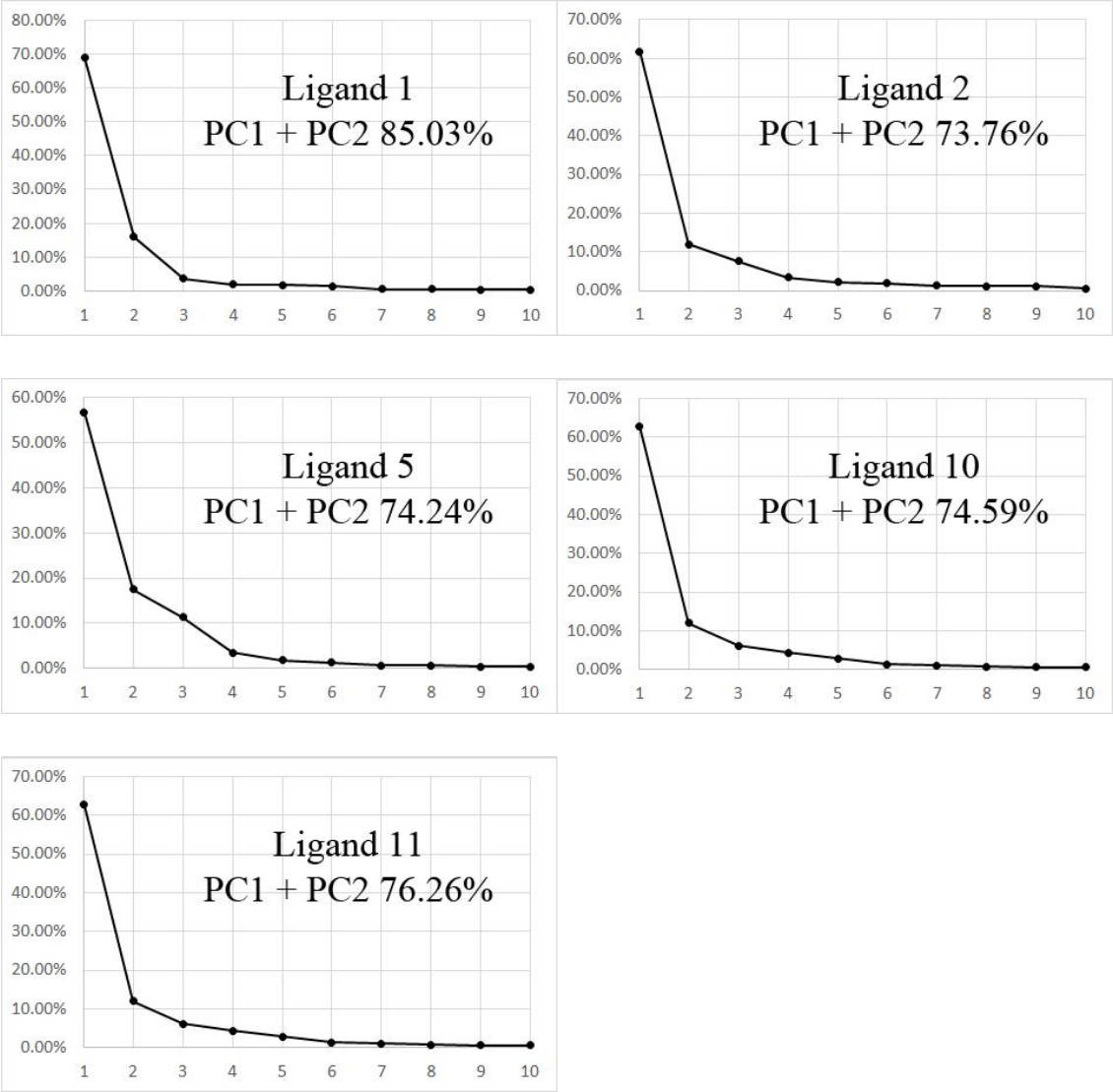

Figure SI 7. Plots of first 10 eigenvalues of the PCA of the metadynamics trajectories. Percentages were computed by dividing the eigenvalues by the sum of all eigenvalues. The sum of percentages of eigenvalues of PC1 and PC2 are labeled.

bioRxiv preprint doi: https://doi.org/10.1101/169607; this version posted July 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

References

(1) Schneider, E. V., et al., *Proc. Natl. Acad. Sci. U.S.A.* **2013,** *110* (20), 8081-8086.

(2) Biasini, M., et al., *Nucleic Acids Res.* **2014,** *42* (W1), W252-W258.

(3) Gill, A. L., et al., *J. Med. Chem.* **2005,** *48* (2), 414-426.

(4) Georgescu, R. E., et al., *Biophys. J.* **2002,** *83* (4), 1731-1748.

(5) Case, D. A., et al., Amber 2015. University of California, San Francisco: 2015.

(6) Wang, J. M., et al., *J. Comput. Chem.* **2004,** *25* (9), 1157-1174.

(7) Phillips, J. C., et al., *J. Comput. Chem.* **2005,** *26* (16), 1781-1802.

(8) Bello-Rivas, J. M.; Elber, R., *J. Chem. Phys.* **2015,** *142* (9).