HHS Public Access

Author manuscript

J Phys Chem A. Author manuscript; available in PMC 2023 July 19.

Published in final edited form as:

J Phys Chem A. 2022 November 24; 126(46): 8761–8770. doi:10.1021/acs.jpca.2c05499.

A Binding Kinetics Toolkit for Analyzing Transient Molecular Conformations and Computing Free Energy Landscapes

Talant Ruzmetov,

Ruben Montes,

Jianan Sun,

Si-Han Chen[#],

Zhiye Tang[¶],

Chia-en A. Chang[*]

Department of Chemistry, University of California at Riverside, Riverside, CA 92521, USA

Abstract

Understanding ligand binding kinetics and thermodynamics, which involves investigating the free, transient and final complex conformations, is important in fundamental studies and applications for chemical and biomedical systems. Examining the important but transient ligand–protein-bound conformations, in addition to experimentally determined structures, also provides a more accurate estimation for drug efficacy and selectivity. Moreover, obtaining the entire picture of the free energy landscape during ligand binding/unbinding processes is critical in understanding binding mechanisms. Here, we present a Binding Kinetics Toolkit (BKiT) that includes several utilities to analyze trajectories and compute a free energy and kinetics profile. BKiT uses principal component space to generate approximated unbinding or conformational transition coordinates for accurately describing and easily visualizing the molecular motions. We implemented a new partitioning approach to assign indexes along the approximated coordinates that can be used as milestones or microstates. The program can generate input files to run many short classical molecular dynamics simulations and uses milestoning theory to construct the free energy profile and estimate binding residence time. We first validated the method with a host–guest system, aspirin unbinding from β-cyclodextrin, and then applied the protocol to pyrazolourea compounds and cyclin-dependent kinase 8 and cyclin C complexes, a kinase system of pharmacological interest. Overall, our approaches yielded good agreement with published results and suggest ligand design strategies. The computed unbinding free energy landscape also provides a more complete picture of ligand–receptor binding barriers and stable local minima for deepening our understanding of molecular recognition. BKiT is easy to use and has extensible features for future expansion of utilities for post-analysis and calculations.

[*]Corresponding authors: Chia-en A. Chang chiaenc@ucr.edu.

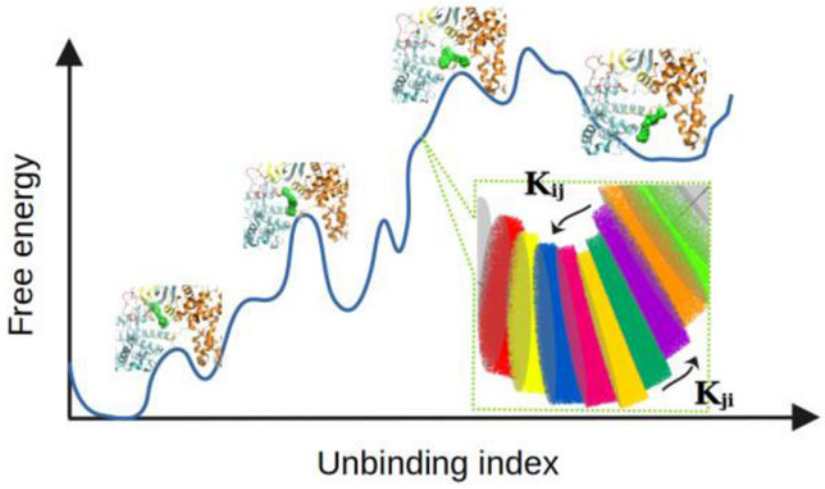
[#]Present address: VantAI, 33 Irving Pl, New York, NY 10003, USA

[¶]Present address: Institute for Molecular Science, Myodaiji, Okazaki, Aichi 444-8585, Japan

Supporting Information

Details of unbinding path construction, path smoothing, disk optimization, short MD simulation length validation and initial point distributions.

Graphical Abstract



Introduction

Transient states during ligand–protein binding/unbinding not seen in experiments contribute to ligand–receptor binding kinetics and thermodynamics, inform screening and lead to optimizing potential drug candidates. Although experiments provide measurements for molecular binding such as binding free energy (ΔG) and association (k_{on}) and dissociation (k_{off}) rate constants and determine molecular structures, the dynamic nature is critical for protein function and molecular recognition. Molecular simulations bridge the gap and have become increasingly appealing because the results provide valuable insights into structural, dynamical, and mechanistic properties of molecular systems[1,2]. Computational methods also predict ligand–receptor binding free energy and assist in molecular design[3–6]. Historically, studies of drug development and binding mechanisms have focused on equilibrium metrics such as binding affinity and the two end-point states, experimentally determined free and final bound structures. Investigation of transient states can help alter the drug's residence time to achieve desired binding kinetics and provide insights in mechanistic studies[7–10].

All-atom molecular dynamics (MD) simulations in explicit or implicit solvents are widely used to examine molecular motions. However, ligand binding/unbinding processes usually need much longer than a microsecond (μs) time scale. Although use of special computer hardware may achieve μs-to-ms simulation lengths, such computer resources are inaccessible to most scientists[11,12]. Many unbinding processes and/or protein conformational changes take longer than seconds. Previous work utilized natural dynamics to generate unbinding pathways for systems that took minutes to dissociation [13,14]; however, it can require significant computation time. Therefore, statistical mechanics-based enhanced sampling techniques, such as metadynamics, weighted ensemble, accelerated MD, steered MD, Tau-MD, and scale-MD, have been used to access long-timescale events at less computational cost[15–26]. These methods are efficient; however, because of the modified

potential during simulations, additional analysis steps are needed to extract thermodynamic and kinetic properties.

For known binding/unbinding pathways or large-scale protein conformational changes, various techniques, including umbrella sampling[27–29] and milestoning[30–33], have been used to predict kinetic rates and free energy profiles. The Markov state model is widely used to estimate the transition rates between different identified states [34–36]. Milestoning theory permits the subdivision of transition space into smaller sections, with boundaries called milestones, and uses many short trajectories at different positions to obtain a more complete kinetic picture. Coordinates to present the states are needed, such as reaction coordinates, ligand unbinding coordinates and protein conformational states; ideally, the coordinates can capture various motions to accurately estimate the population of these intermediate states. Center-of-mass distance between two molecules is widely used to assign milestones or states to estimate ligand–protein residence time. However, capturing motions of conformational rearrangements for most realistic ligand–receptor systems is a challenge for the simplified coordinate. Therefore, post-analysis strategies using physics-based or machine learning have been developed to select the coordinates that more accurately describe the above motions for free energy and kinetics studies[37–44].

Here, we introduce a Binding Kinetics Toolkit (BKiT) developed to post-analyze simulation trajectories and accurately compute the free energy profile and kinetics properties for given trajectories by using milestoning theory. The toolkit includes newly developed strategies to approximate coordinates for ligand–receptor unbinding or conformational transitions using principal component analysis (PCA) and numerical analysis such as interpolation. The top few principal component (PC) modes extract major motions from data in trajectories, and the projection of the dataset into the PC space creates a PC plot for easily visualizing the molecular motions. In addition to providing approximated unbinding or conformational transition coordinates, we implemented a new partitioning approach to assign indexes along the approximated coordinates in the PC plot. The indexes can be used as milestones or microstates for further investigation. Users can easily select projected points on the PC plot to examine the molecular system of interest. Using the indexes as milestones, BKiT generates input files to run many short 100-ps classical MD (cMD) and uses milestoning theory to construct the free energy profile and estimate binding residence time.

We demonstrate the construction of the free energy landscape and residence time estimation for two systems: a chemical host–guest system, aspirin and β-cyclodextrin (β-CD), and a protein system that is also a promising cancer drug target, cyclin-dependent kinase 8 (CDK8) and cyclin C (CycC) and two pyrazolourea (PL) ligands: a known ligand, PL1, and our designed ligand, PL1-OH [8]. Cyclodextrins (CDs) are a class of cyclic oligosaccharide compounds and have various applications as chemical hosts or carriers[45,46]. CDK8/CycC plays a key role in regulating transcription activities[47–49]. As typical kinases, CDK8 has an ATP binding site and an activation loop with a DMG (Asp-Met-Gly) motif. The pyrazolourea ligands are considered type-II ligands because the compounds bind to CDK8–CycC with DMG-out loop conformations. Results produced with BKiT agree with existing findings and illustrate the molecular motions and conformations that lead to local free energy minima of barriers. We also demonstrate how to use the transient conformations

revealed in the free energy profile to modify compounds to achieve desired binding kinetic behavior.

Methods

The source code of BKiT is written in Python, and the current implementation is the Python version, available in GitHub. It uses popular libraries such as numpy, scikit-learn, scipy and pytraj[50], a python implementation of cpptraj[51]. The toolkit is of general applicability and can be used to examine changes of molecular conformations, but here we focus on the ligand–receptor binding landscape. The protocol and utilities are detailed in the following paragraphs, and the main approaches are summarized below (Fig. 1):

1. Obtaining a trajectory (or trajectories) describing the event under investigation, for example, the association/dissociation of ligand–protein complexes. Notably, BKiT does not perform sampling to obtain the trajectories, and usually an enhanced sampling technique is needed because the binding/unbinding processes can comprise several rare events.
2. Analyzing the given trajectories to generate an approximated binding/unbinding path in the 3D or 2D PC projection space (PC plot). The program can estimate the mean path and perform smoothing and interpolation of the path. Users can select frames of the trajectories in the PC plot to visualize the conformations.
3. Assigning unbinding coordinates in the PC plot by using disks with 3D projection or using lines with 2D projection in the PC plot. The disks or lines serve as coordinate indexes based on the approximated unbinding path in the PC plot. BKiT can also further optimize the unbinding indexes for assigning milestones or microstates for further kinetics and thermodynamics studies.
4. Generating input files using frames of the given trajectories or other user-selected conformations as an initial structure to run many cMD simulations.
5. Analyzing results from item 4 for transitions between the microstates (milestones) to compute the transition kernel, flux, free energy profile and residence time.

Trajectory generation via molecular simulations.

Solely studying endpoints such as the free and final bound complexes of both a ligand and a receptor is usually not sufficient to characterize the binding/unbinding free energy landscape and investigate binding kinetics. Atomistic detailed ligand unbinding or binding pathways need to be sampled first (Fig. 1(i)). Because ligand unbinding usually encounters large energy barriers during dissociation, researchers typically apply enhanced sampling methods such as accelerated MD or steered MD to generate the ligand unbinding trajectories. BKiT does not provide the sampling tools to sample ligand unbinding or long time-scale protein motions, and any given trajectory or trajectories with a reasonable physical pathway can be examined by the program.

Analyzing a dataset from given trajectories to generate a visually observable unbinding path.

BKiT processes a trajectory with an event of interest by performing PCA and using the eigenvectors of the first 2 or 3 PC modes to represent the approximated unbinding path (the solid lines in Fig. 1(ii)). BKiT allows various atom selections for PCA. For example, users can select heavy atoms of the ligand and receptor, or only Cα of proteins to perform PC analysis. PCA is a mathematical approach that can extract the major motions from the given trajectory[37,38]. The equation $PC_i(X(j)) = R^T(X(j) - \langle X \rangle)$ is used to project frames from the given trajectory onto the PC_i space, where i indicates a PC mode, ranging from 1 to 3 in the current package; R^T is the eigenvector for PC_1, PC_2 or PC_3; and $X(j)$ and $\langle X \rangle$ are the Cartesian coordinates of the selected atoms at frame j and the average over the trajectory, respectively. Using the PC plot reduces high-dimensional ligand–protein motions (3N-6 degrees of freedom where N is number of atoms) to a few dimensions for plotting the coordinates to the PC plot for further investigation. As illustrated in Fig. 1(ii), each frame of the given trajectory is shown as a dot in the 3D PC plot. As a result, the molecular motions along a trajectory can be visualized in the constructed PC plot, and users can select dots on the PC plot to observe the conformations of those frames.

Because a smooth approximated unbinding path is required for assigning indexes (i.e., milestones) for binding kinetics calculations, we applied interpolation to smooth the approximated unbinding path and ensure equal distance between the disk centers. BKiT provides two different approaches to generate the approximated unbinding coordinates, which can accurately represent molecular motions during the event of interest. The first method is based on a rolling average in the PC plot. This strategy assumes that the dots represent smooth molecular movements during ligand unbinding. However, during a simulation, a ligand may slightly unbind away from a bound conformation and move back multiple times. Therefore, we developed the second method, which uses a mean path as a reference and re-evaluates every point using local neighborhood average with the K-dimensional tree (k-d tree) algorithm for fast neighbor search. However, both approaches cannot guarantee a smooth unbinding path, as illustrated in Fig. S1, Fig. S2. Thus, an additional path-smoothening steps are applied as described in SI.

Notably, before PC analysis, a proper structural alignment must be performed to eliminate molecular translation/rotation during simulations by using a user-defined selection (i.e., Cα of proteins and heavy atoms of ligands) and a reference frame, preferably the first frame of the trajectory. This paper considers all 3 PC modes in the examples, tests, and constructed free energy profile.

Assigning unbinding indexes.

Once an approximated unbinding path is built, BKiT inserts disks (3D) or lines (2D) on the PC plot to represent unbinding coordinates. Here we focus on the 3D PC plot. Disks are inserted at the positions of the smoothed path, with each disk called an index. In most cases, the index starts from the bound states to ligand unbound conformations. Normal vectors, A_i,

to a disk i surface are gradually optimized over a few thousand iterations to avoid overlaps between the neighboring disks:

$$A_i^{t+1} = A_i^t(1-\alpha) + \frac{\alpha}{2}\left(A_{i+1}^t + A_{i-1}^t\right)$$

where superscript t denotes iteration steps, and α is a learning rate that controls the optimization speed of disk orientations Fig. S3. By default, α is set to a small value, 0.01, to ensure gradual convergence.

Depending on the molecular system, users can assign the radius of the disk and the space between two adjacent disks (indexes). The default for a ligand–protein system is set to both 10 eigenvalue units (eu) for the disk radius and their space apart, and users can assign different disk radii and intervals when needed. The disk radius and interval do not need to be the same throughout the unbinding path. For example, one may use an 8-eu interval to place disks when a ligand locates near the crystal structure bound form and a 10-eu interval for the rest. For smaller host–guest systems, smaller disks (i.e., 2 eu) are sufficient.

Guided by the PC plot, generating input files for selected frames of the given trajectories to perform MD simulations.

BKiT allows users to visualize and select dots on the PC plot (Fig. 1(iii and iv)), which are frames of the given trajectory, for further investigation. The program generates input files for running cMD using the Amber program. However, users can easily modify the open source package to run other simulations to accommodate their needs. Our example here is computing a binding free energy profile and kinetics; therefore, our examples used BKiT to generate numerous short cMD runs. Multiple replicas using different random number seeds can be assigned for each initial structure for each cMD run. For example, users can select explicit dots from the PC plot, dots near a disk/line, dots between indexes, or frames resaved from the given trajectories for cMD runs. Users can perform various post-analyses for the new simulation results or use our utility to process the data to construct a free energy profile and estimate binding residence time.

Analyzing short MD runs to construct a free energy profile and investigate dissociation kinetics.

BKiT provides utilities to analyze many short MD runs and apply the adaptive milestoning theory to construct the transition kernel. Frames of the cMD runs are projected with the same eigenvectors used to construct the PC plot. Because of the vast number of frames (~0.5 Tb data), this operation is usually the bottleneck of the post-analysis. Thus, users are encouraged to perform projections in chunks by using the provided script (MD2PCA.py), and the process can be completed within 5 min by using a multi-core CPU processor. Each frame of the cMD run is again represented as a dot on the PC plot, and an ID is assigned to each dot based on the frame number saved from a trajectory. Notably, BKiT can analyze many short MD trajectories to construct the transition kernel, and the initial structures for MD do not need to be exactly located on a milestone.

To describe the position of the newly generated dots (frames of MD runs) in the PC plot, a dot (frame) is counted as in an index space i if it lies between the disks (indexes) i and $i + 1$. If a dot is outside the space between any two disks, it does not belong to any index space. If the next frame moves from index space i to $i + 1$, it crosses index $i + 1$. Similarly, when the next frame moves from index space i to $i - 1$, it crosses index i. The first time that a frame hits an index, for example, index i, it is termed an initial point. A molecular system may cross the same index multiple times before hitting an adjacent index; then a transition is counted and the lifetime τ_i is recorded. BKiT tracks the movements of each short trajectory to record the transitions on each index (milestone). The number of the transitions in index i per unit time is defined as flux and denoted as q_i. The transition kernel is a matrix that contains probabilities of a transition between milestones i and j denoted as K_{ij}. It can be directly calculated by counting transitions between the indexes (milestones):

$$K_{ij} \simeq \frac{n_{ij}}{n_i}$$

In this, n_i is the number of trajectories initiated at milestone i, and counting these trajectories that end up at index j are denoted as n_{ij}.

Currently, BKiT offers two methods for calculating the stationary flux. The first method iteratively updates flux values by taking a weighted average between fluxes flowing through neighboring indexes until the flux in all indexes (milestones) converge to the stationary value[30,52].

$$q_i^{n+1} = \frac{q_{i-1}^n K_{i-1,i} + q_{i+1}^n K_{i+1,i}}{\sum_1^N q_i^n}$$

Convergence is achieved when an overall flux change between the consecutive iterations is less than the allowed error, $\sum_{i=1}^{N} \left| q_i t + 1 - q_i^t \right| \leqslant \varepsilon$. The second method solves the eigenvalues of the transition kernel. Eigenvectors corresponding to the highest eigenvalue, 1, represent the stationary flux. Results of these two methods are in a good agreement, as demonstrated in Fig. S4 and S5. By multiplying the stationary flux to the average lifetime τ_i of the transition, we get the population and calculate the free energy per unbinding index by using $F_i = k_B T \log(q_i \tau_i)$. The overall mean first passage time (or residence time) for a ligand crossing index f can be estimated by $MFPT_f = \sum_1^f q_i \tau_i / q_f$ [30]. Note that an absorbing boundary was applied here for the transition matrix when computing q.

System Setup.

β-CD and aspirin complex conformations were taken from a published trajectory, and our short cMD runs used the same MD setting as those in the paper[53]. Amber general force field (GAFF)[54] and q4MD-CD[55] force field were used for aspirin and β-CD, respectively. Short cMD runs for CDK8–CycC–PL complexes were the same as the existing work [8] which used Amber14SB[56] force field and GAFF for the protein and compounds,

respectively. All 100ps short cMD runs used TIP3P [57] water model and the Amber package [58] in 298K, and a frame was saved every 100 fs. In β-CD and aspirin system, all frames were aligned to a reference frame using the heavy atoms of β-CD, and PCA projection was performed by selecting all O1, C1 and C2 atoms from β-CD and all heavy atoms of aspirin. In CDK8–CycC–PL complexes, the frames were aligned to a reference frame using the Cα of CDK8, and PCA projection was performed by selecting Cα of CDK8–CycC and all heavy atoms of PL compounds.

Error analysis.

When constructing the transition matrix, one needs many short cMD runs. BKiT counts numerous transitions and computes the average lifetime τ_i and its standard deviation $\sigma_{\tau,i}$ of each transition. Based on error propagation rule[61], standard deviation of free energy at unbinding index i can be derived:

$$\sigma_{F,i} = k_B T \frac{\sigma_{\tau,i}}{\tau_i}$$

Furthermore, standard deviation of the mean first passage time MFPT that start at milestone b and terminate at milestone f is estimated using the form:

$$\sigma_{MFPT} = \frac{\sqrt{\sum_{i=b}^{f} (q_i \sigma_{\tau,i})^2}}{q_f}$$

Results and Discussion

First, we use one unbinding event obtained from a long timescale cMD trajectory, guest aspirin dissociated from β-CD (Fig. 2a), to describe and benchmark the computational strategy. In the second part of the Results section, we outline the results for an existing and a designed ligand unbinding to the CDK8–CycC complex (Fig. 2b).

β-CD and aspirin complex.

We used guest aspirin unbinding from β-CD as our first example; the experimental measurement of $\Delta G_{exp}(-3.7 \pm 0$ kcal/mol), $k_{exp\_on}\left(7.2 \pm 0.04 \times 10^8\ 1/\text{sM}\right)$ and $k_{exp\_off}\left(1.3 \pm 0.03 \times 10^6\ 1/\text{s}\right)$ and computation results for $\Delta G_{comp}(-4.11 \pm 0.05$ kcal/mol), $k_{comp\_on}\left(3.2 \pm 0.3 \times 10^9\ 1/\text{sM}\right)$ and $k_{comp\_off}\left(3.1 \pm 0.9 \times 10^6\ 1/\text{s}\right)$ are both available [53,59]. β-CD has 7 glucopyranose units (D-glucose), which results in a hydrophobic cavity. The host has asymmetrically wide and narrow rims with hydrophilic hydroxyl groups. We took one unbinding event from a 6-μs cMD trajectory from an existing publication, reporting a total of 17 binding/unbinding events [53]. In this event, aspirin dissociated from a narrow rim, and we selected a brief period of time, 0.43 ns, when aspirin was ready to dissociate from the bound conformation until it barely left β-CD (Fig. 2a). The first 3 PC modes covered 83.4% of the overall motion. BKiT built the approximated unbinding path with the k-d tree algorithm, as detailed in Fig. 1(ii). Using the interval 0.8 eu, a total of 62 2-eu radius disks

were inserted in the approximated unbinding path. Disk orientations were optimized over 8000 iterations to result in the final indexes (milestones) shown in Fig. 1(iii,iv). To set up multiple short cMD runs, we resaved the 0.43-ns trajectory every 0.1 ps, for a total of 430 initial structures for MD. We used 20 replicates with different random number seeds, and each run was set to 100 ps with a frame saved every 100 fs.

Using BKiT to analyze the short MD runs, we computed an aspirin-unbinding free energy landscape and residence time (Fig. 3). The free energy landscape reveals multiple local free energy minima and barriers and estimates the binding free energy of the bound complex of $\Delta G = \sim -4$ kcal/mol with residence time 56 ns. As compared with $\Delta G_{comp} = -4.11$ kcal/mol, the approximated binding free energy is reasonable, considering that aspirin bound to the narrow rim. Notably, the primary (wide rim) and secondary (narrow rim) cavities of β-CD are not identical, and existing studies showed that guests bind preferentially to the primary cavity of β-CD because of stronger attractions provided by this cavity. At the first local barrier shown in Figure 3A, aspirin is bound directly in the middle of cyclodextrin's narrow rim. The cavity of cyclodextrin began to slightly distort, resulting in more intermolecular attractions between the hydrophilic tips and aspirin. The increased intermolecular contacts brought down the energy of the system to the global minima (Fig. 3B). As the system approached the next barrier (Fig. 3C), the tips of cyclodextrin moved outward, similar to the first barrier. Aspirin continued drifting to one side of the cavity and started to associate with a single glucopyranose unit of cyclodextrin (Fig. 3D). As the local interactions were strengthened, aspirin slid toward the rim, thus resulting in another local free energy barrier (Fig. 3E). Aspirin then began to escape the tips of the rim and diffused outward toward the solvent (Fig. 3F).

Because the calculations rely on accurately counting the transitions between the unbinding indexes (milestones), we examined the simulation lengths of the short cMD runs and points for counting the initial-point distribution (IPD) to validate the suggested simulation length and our strategies for counting conformational transitions to build the transition kernel K_{ij}. Unlike exact milestoning theory, in which researchers need to perform restrained MD simulations to obtain initial points on a milestone i for short MD runs, we used frames from a given trajectory for short cMD. We did not terminate a cMD run when a molecular system moved to an adjacent milestone $i + 1$ *or* $i - 1$. Figure 4 shows that running 20 replicas of 80-ps and 100-ps short MD runs resulted in asymptotic potential of mean force (PMF), which suggests that running 80 ps with 20 replicas is sufficient to converge the free energy calculation. We also checked the IPD for 4 energy barriers, which usually have less sampling as compared with local energy minima. The population of initial points in a barrier index (milestone) satisfies statistical ensemble distribution (Fig. 5), which demonstrates that our "on-the-fly" strategy is an efficient strategy to provide IPD, with no need to run additional restrained MD on an index to generate the initial conformations. The population is converged after running 80-ps MD runs using 20 replicas, although the 100-ps runs provided more initial points on a milestone (numbers in parentheses in Fig. 5).

CDK8–CycC interactions with pyrazolourea ligands.

Our second model system is a real case study of pharmaceutical interest, protein kinase CDK8–CycC (Fig. 2c). We applied BKiT to a known tight binder pyrazolourea ligand PL1 ($\Delta G_{exp} = -10.7 \pm 0.07$ kcal/mol and residence time = 1944 min) and our designed ligand PL1-OH that was predicted to increase both binding affinity and residence time. We investigated published PL1 and PL1-OH unbinding trajectories from CDK8–CycC by using metadynamics [8]. We used the Cartesian coordinates of Cα and ligand heavy atoms for PC analysis. The first 3 PC modes covered 81.5% and 88.5% of the overall motions for PL1 and PL1-OH, respectively. BKiT built the approximated unbinding path with the k-d tree algorithm, as detailed in Fig. S1 and S2. For protein systems, we used a 10-eu radius disk to present each index along the approximated unbinding path in the PC plot. Using the 4-eu interval on the smoothed unbinding path, a total of 84 and 96 disks were inserted to present the unbinding indexes of PL1 and PL1-OH, respectively. Disk orientations were optimized over 4000 iterations to bring the final indexes (milestones) shown in Fig. S3. To start multiple short cMD runs, we took one in every 10 frames from the provided metadynamics trajectories to obtain 500 and 650 frames for the PL1–CDK8–CycC and PL1-OH–CDK8–CycC complexes as initial structures for cMD, respectively. Each initial structure was set to run 100-ps cMD with 20 replicas, and a frame was saved every 100 fs. Fig. S4 and S5 showed that using a total of 10,000 and 13,000 100-ps short cMD runs for both protein systems was sufficient to generate a converged ligand unbinding free energy landscape.

The unbinding free energy profile in Figure 6 suggested that additional intermolecular attractions may be achieved by adding a functional group in the alkane chain between two nitrogen atoms (Fig. 2b). Therefore, a hydroxyl group was added to PL1. Our prediction shown in Figure 7 suggested that PL1-OH and CDK8–CycC formed more stable bound conformations and prolonged the bound states, resulting in ~70 times longer residence time than with PL1. More specifically, a local free energy barrier (Fig. 7C) was the result of hydrogen bond breaking between E66 and PL1-OH. As PL1-OH continuously moved away from the binding pocket, the length of a hydrogen bond between E66 and PL1-OH increased from 1.88 Å at index #55 to 2.04 Å at index #56. However, E66 remained near the same places, leading to a locally rugged energy landscape before proceeding to the next free energy barrier in Figure 7D.

For both systems, we examined free energy convergence with different simulation lengths (Fig. S4 and S5), which showed good convergence after longer than 80-ps simulation length cMD. IPD of selected indexes were checked to ensure that the distribution satisfied statistical ensembles (Fig. S6 and S7). Although the absolute binding free energy of PL-1 is ~ −7 kcal/mol (Fig. 6), which is higher than the experimental value, the predicted ΔG shows ~ 3 kcal/mol stronger binding affinity for the designed compound PL1-OH. It is worth mentioning that the calculation focused on the transient bound CDK8–CycC–PL conformations and we did not perform sufficient sampling to capture surface diffusion of both PL compounds. However, these bound forms are the main determinants for relative binding free energy between compounds of interest. In addition, although we assigned many milestones, we used Cα and equal interval to assign indexes (milestones). As a result, some fluctuations contributed to unbinding free energy barriers are ignored, which results

in errors. The calculations may be further improved by considering side-chains rotations which are critical in intermolecular interactions and incorporating machine learning to capture ignored motions and energy barriers. The calculation provides another approach to investigate transient barriers and minima as a computer-aided guidance for drug design that can alter drug binding affinity and kinetics.

Conclusions

We present a new toolkit, BKiT, that offers robust utilities for post-analyzing simulation trajectories using a reduced dimensionality PC plot, guiding researchers to select frames of interest for further investigation, and for computing the free energy landscape and ligand binding residence time using milestoning theory. BKiT aligns a given trajectory or trajectories to a reference frame, performs PC analysis, and uses the eigenvectors of the first 2 or 3 PC modes to project frames of the trajectory to a PC plot. Researchers can examine molecular motions on the 2D or 3D PC plot and select frames in the python based BKiT interface for additional analysis. The approaches also include assigning an approximated unbinding path and creating ligand unbinding indexes in the PC plot. The indexes can serve as milestones or microstates for using theories that connect multiple MD runs to extract kinetic information. In the current setting, BKiT applies milestoning theory to construct a free energy landscape along the indexes and estimate binding residence time. BKiT assists in preparing MD simulations and analyzes many short MD runs to count transitions between indexes (milestones). We used one event of aspirin unbinding from β-CD to benchmark our approaches, which showed good agreement in computed ΔG and residence time with results from cMD runs. The computed unbinding free energy landscape also provides a more complete picture of host–guest binding barriers and stable local minima to deepen our understanding of molecular recognition. Using CDK8–CycC–ligand complexes as examples, we computed free energy landscape using 3D PC plots. Similar to use of 2D PC plot[8], the free energy profile again revealed key interactions to keep the ligand PL1 in a bound conformation and reiterated our previously suggested design strategies for creating PL1-OH. The analysis yielded good results that PL1-OH could enhance binding affinity and residence time. The python library exploited in the program is easy to use and has extensible features for future expansion of more utilities for post-analysis and calculations. With its efficiency, BKiT is well suited for studying protein–drug binding systems for structure-based drug design.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This study was supported by the US National Science Foundation (MCB-1932984), US National Institutes of Health (GM-109045) and San Diego Supercomputer Center.

Data and Software Availability

All software is open-source and is made available on GitHub, which can be accessed from our group website: http://chemcha-gpu0.ucr.edu/software/ and a method toolbox KBbox[60]: https://kbbox.h-its.org/toolbox/tools/data-analysis-tools/bkit/. The BKiT page of the site also contains a helpful user manual, installation tips, example input files, and bash scripts to streamline the system setup: http://chemcha-gpu0.ucr.edu/bkit/

References

1. Bruce NJ, Ganotra GK, Kokh DB, Sadiq SK, Wade RC. New approaches for computing ligand–receptor binding kinetics. Current Opinion in Structural Biology. p. 1–10 (2018).
2. Decherchi S, Cavalli A. Thermodynamics and Kinetics of Drug-Target Binding by Molecular Simulation. Chem Rev. 120:12788–12833 (2020). [PubMed: 33006893]
3. Mobley DL, Gilson MK. Predicting Binding Free Energies: Frontiers and Benchmarks [Internet]. Dill KA, editor. Annual Review of Biophysics, Vol 46. p. 531–558 (2017). Available from: <Go to ISI>://WOS:000402908700024
4. Abel R, Wang LL, Harder ED, Berne BJ, Friesner RA. Advancing Drug Discovery through Enhanced Free Energy Calculations. Accounts of Chemical Research. p. 1625–1632 (2017). [PubMed: 28677954]
5. Song LF, Lee TS, Zhu C, York DM, Merz KM. Using AMBER18 for Relative Free Energy Calculations. Journal of Chemical Information and Modeling. p. 3128–3135 (2019). [PubMed: 31244091]
6. Re SY, Oshima H, Kasahara K, Kamiya M, Sugita Y. Encounter complexes and hidden poses of kinase-inhibitor binding on the free-energy landscape. Proceedings of the National Academy of Sciences of the United States of America. p. 18404–18409 (2019). [PubMed: 31451651]
7. Ribeiro JML, Tsai ST, Pramanik D, Wang YH, Tiwary P. Kinetics of Ligand-Protein Dissociation from All-Atom Simulations: Are We There Yet? Biochemistry. p. 156–165 (2019).
8. Tang ZY, Chen SH, Chang CEA. Transient States and Barriers from Molecular Simulations and the Milestoning Theory: Kinetics in Ligand-Protein Recognition and Compound Design. Journal of Chemical Theory and Computation. p. 1882–1895 (2020). [PubMed: 32031801]
9. Tonge PJ. Drug-Target Kinetics in Drug Discovery. Acs Chemical Neuroscience. p. 29–39 (2018). [PubMed: 28640596]
10. Huang Y-MM. Multiscale computational study of ligand binding pathways: Case of p38 MAP kinase and its inhibitors. Biophys J. 120:3881–3892 (2021). PMCID: PMC8511166 [PubMed: 34453922]
11. Ganesan A, Coote ML, Barakat K. Molecular “time-machines” to unravel key biological events for drug design [Internet]. Wiley Interdisciplinary Reviews-Computational Molecular Science. (2017). Available from: <Go to ISI>://WOS:000403439500003
12. Pan AC, Xu HF, Palpant T, Shaw DE. Quantitative Characterization of the Binding and Unbinding of Millimolar Drug Fragments with Molecular Dynamics Simulations. Journal of Chemical Theory and Computation. p. 3372–3377 (2017). [PubMed: 28582625]
13. Lotz SD, Dickson A. Unbiased Molecular Dynamics of 11 min Timescale Drug Unbinding Reveals Transition State Stabilizing Interactions. J Am Chem Soc. 140:618–628 (2018). [PubMed: 29303257]
14. Dixon T, Uyar A, Ferguson-Miller S, Dickson A. Membrane-Mediated Ligand Unbinding of the PK-11195 Ligand from TSPO. Biophys J. 120:158–167 (2021). PMCID: PMC7820730 [PubMed: 33221248]
15. Scafuri N, Soler MA, Spitaleri A, Rocchia W. Enhanced Molecular Dynamics Method to Efficiently Increase the Discrimination Capability of Computational Protein-Protein Docking. Journal of Chemical Theory and Computation. p. 7271–7280 (2021). [PubMed: 34653335]
16. Miao YL, McCammon JA. Unconstrained enhanced sampling for free energy calculations of biomolecules: a review. Molecular Simulation. p. 1046–1055 (2016). [PubMed: 27453631]

17. Kokh DB, Amaral M, Bomke J, Gradler U, Musil D, Buchstaller HP, Dreyer MK, Frech M, Lowinski M, Vallee F, Bianciotto M, Rak A, Wade RC. Estimation of Drug-Target Residence Times by tau-Random Acceleration Molecular Dynamics Simulations. Journal of Chemical Theory and Computation. p. 3859–3869 (2018). [PubMed: 29768913]
18. Schuetz DA, Bernetti M, Bertazzo M, Musil D, Eggenweiler HM, Recanatini M, Masetti M, Ecker GF, Cavalli A. Predicting Residence Time and Drug Unbinding Pathway through Scaled Molecular Dynamics. Journal of Chemical Information and Modeling. p. 535–549 (2019). [PubMed: 30500211]
19. Barducci A, Bonomi M, Parrinello M. Metadynamics. Wiley Interdisciplinary Reviews-Computational Molecular Science. p. 826–843 (2011).
20. Doshi U, Hamelberg D. Towards fast, rigorous and efficient conformational sampling of biomolecules: Advances in accelerated molecular dynamics. Biochimica Et Biophysica Acta-General Subjects. p. 878–888 (2015).
21. Grazioli G, Andricioaei I. Advances in milestoning. I. Enhanced sampling via wind-assisted reweighted milestoning (WARM) [Internet]. Journal of Chemical Physics. (2018). Available from: <Go to ISI>://WOS:000444035800008
22. Zuckerman DM, Chong LT. Weighted Ensemble Simulation: Review of Methodology, Applications, and Software. Annual Review of Biophysics, Vol 46. p. 43–57 (2017).
23. Tran DP, Takemura K, Kuwata K, Kitao A. Protein-Ligand Dissociation Simulated by Parallel Cascade Selection Molecular Dynamics. Journal of Chemical Theory and Computation. p. 404–417 (2018). [PubMed: 29182324]
24. Wong CF. Steered molecular dynamics simulations for uncovering the molecular mechanisms of drug dissociation and for drug screening: A test on the focal adhesion kinase. Journal of Computational Chemistry. p. 1307–1318 (2018). [PubMed: 29498075]
25. Tang ZY, Roberts CC, Chang CEA. Understanding ligand-receptor non-covalent binding kinetics using molecular modeling. Frontiers in Bioscience-Landmark. p. 960–981 (2017).
26. Ahmad K, Rizzi A, Capelli R, Mandelli D, Lyu W, Carloni P. Enhanced-Sampling Simulations for the Estimation of Ligand Binding Kinetics: Current Status and Perspective. Front Mol Biosci. 9:899805 (2022).
27. Kastner J. Umbrella sampling. Wiley Interdisciplinary Reviews-Computational Molecular Science. p. 932–942 (2011).
28. Awasthi S, Kapil V, Nair NN. Sampling Free Energy Surfaces as Slices by Combining Umbrella Sampling and Metadynamics. Journal of Computational Chemistry. p. 1413–1424 (2016). [PubMed: 27059305]
29. You WL, Tang ZY, Chang CEA. Potential Mean Force from Umbrella Sampling Simulations: What Can We Learn and What Is Missed? Journal of Chemical Theory and Computation. p. 2433–2443 (2019). [PubMed: 30811931]
30. Elber R. A new paradigm for atomically detailed simulations of kinetics in biophysical systems [Internet]. Quarterly Reviews of Biophysics. (2017). Available from: <Go to ISI>://WOS:000402400900002
31. Maragliano L, Vanden-Eijnden E, Roux B. Free Energy and Kinetics of Conformational Transitions from Voronoi Tessellated Milestoning with Restraining Potentials. Journal of Chemical Theory and Computation. p. 2589–2594 (2009). [PubMed: 20354583]
32. Votapka LW, Stokely AM, Ojha AA, Amaro RE. SEEKR2: Versatile Multiscale Milestoning Utilizing the OpenMM Molecular Dynamics Engine. J Chem Inf Model. 62:3253–3262 (2022). [PubMed: 35759413]
33. Narayan B, Fathizadeh A, Templeton C, He P, Arasteh S, Elber R, Buchete N-V, Levy RM. The transition between active and inactive conformations of Abl kinase studied by rock climbing and Milestoning. Biochim Biophys Acta Gen Subj. 1864:129508 (2020). PMCID: PMC7012767
34. Chodera JD, Noe F. Markov state models of biomolecular conformational dynamics. Current Opinion in Structural Biology. p. 135–144 (2014). [PubMed: 24836551]
35. Berezhkovskii AM, Szabo A. Committors, first-passage times, fluxes, Markov states, milestones, and all that [Internet]. Journal of Chemical Physics. (2019). Available from: <Go to ISI>://WOS:000458109300007

36. Husic BE, Pande VS. Markov State Models: From an Art to a Science. Journal of the American Chemical Society. p. 2386–2396 (2018). [PubMed: 29323881]
37. David CC, Jacobs DJ. Principal Component Analysis: A Method for Determining the Essential Dynamics of Proteins. Protein Dynamics: Methods and Protocols. p. 193–226 (2014).
38. Ahmad M, Helms V, Kalinina OV, Lengauer T. Relative Principal Components Analysis: Application to Analyzing Biomolecular Conformational Changes. J Chem Theory Comput. 15:2166–2178 (2019). [PubMed: 30763093]
39. Balsera MA, Wriggers W, Oono Y, Schulten K. Principal component analysis and long time protein dynamics. Journal of Physical Chemistry. p. 2567–2572 (1996).
40. Allison JR. Computational methods for exploring protein conformations. Biochemical Society Transactions. 48:1707–1724 (2020). [PubMed: 32756904]
41. Noé F, Clementi C. Collective variables for the study of long-time kinetics from molecular trajectories: theory and methods. Current Opinion in Structural Biology. 43:141–147 (2017). [PubMed: 28327454]
42. Sidky H, Chen W, Ferguson AL. Machine learning for collective variable discovery and enhanced sampling in biomolecular simulation. Molecular Physics. 118:e1737742 (2020).
43. Tao P, Sodt AJ, Shao YH, Konig G, Brooks BR. Computing the Free Energy along a Reaction Coordinate Using Rigid Body Dynamics. Journal of Chemical Theory and Computation. p. 4198–4207 (2014). [PubMed: 25328492]
44. Bai F, Xu Y, Chen J, Liu Q, Gu J, Wang X, Ma J, Li H, Onuchic JN, Jiang H. Free energy landscape for the binding process of Huperzine A to acetylcholinesterase. Proceedings of the National Academy of Sciences of the United States of America. p. 4273–4278 (2013). [PubMed: 23440190]
45. Amiri S, Amiri S. Cyclodextrins : Properties and Industrial Applications. (2018).
46. Suarez D, Diaz N. Affinity Calculations of Cyclodextrin Host-Guest Complexes: Assessment of Strengths and Weaknesses of End-Point Free Energy Methods. Journal of Chemical Information and Modeling. p. 421–440 (2019). [PubMed: 30566348]
47. 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. p. 5073–5092 (2018). [PubMed: 29266937]
48. Knuesel MT, Meyer KD, Bernecky C, Taatjes DJ. The human CDK8 subcomplex is a molecular switch that controls Mediator coactivator function. Genes Dev. 23:439–451 (2009). PMCID: PMC2648653 [PubMed: 19240132]
49. Schneider EV, Bottcher J, Huber R, Maskos K, Neumann L. Structure-kinetic relationship study of CDK8/CycC specific compounds. Proceedings of the National Academy of Sciences of the United States of America. p. 8081–8086 (2013). [PubMed: 23630251]
50. Nguyen H, Case DA, Rose AS. NGLview-interactive molecular graphics for Jupyter notebooks. Bioinformatics. 34:1241–1242 (2018). PMCID: PMC6031024 [PubMed: 29236954]
51. Roe DR, Cheatham TE. Parallelization of CPPTRAJ Enables Large Scale Analysis of Molecular Dynamics Trajectory Data. Journal of Computational Chemistry. p. 2110–2117 (2018). [PubMed: 30368859]
52. Bello-Rivas JM, Elber R. Exact milestoning. The Journal of Chemical Physics. 142:094102 (2015). [PubMed: 25747056]
53. Tang ZY, Chang CEA. Binding Thermodynamics and Kinetics Calculations Using Chemical Host and Guest: A Comprehensive Picture of Molecular Recognition. Journal of Chemical Theory and Computation. p. 303–318 (2018).
54. Wang JM, Wolf RM, Caldwell JW, Kollman PA, Case DA. Development and testing of a general amber force field. Journal of Computational Chemistry. p. 1157–1174 (2004). [PubMed: 15116359]
55. Cezard C, Trivelli X, Aubry F, Djedaini-Pilard F, Dupradeau FY. Molecular dynamics studies of native and substituted cyclodextrins in different media: 1. Charge derivation and force field performances. Physical Chemistry Chemical Physics. p. 15103–15121 (2011).

56. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. J Chem Theory Comput. 11:3696–3713 (2015). [PubMed: 26574453]

57. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. The Journal of Chemical Physics. 79:926–935 (1983).

58. Case DA, Aktulga HM, Belfon K, Ben-Shalom I, Brozell SR, Cerutti DS, Cheatham III TE, Cruzeiro VWD, Darden TA, Duke RE. Amber 2021. University of California, San Francisco; (2021).

59. Fukahori T, Kondo M, Nishikawa S. Dynamic study of interaction between beta-cyclodextrin and aspirin by the ultrasonic relaxation method. Journal of Physical Chemistry B. p. 4487–4491 (2006). [PubMed: 16509753]

60. Bruce NJ, Ganotra GK, Richter S, Wade RC. KBbox: A Toolbox of Computational Methods for Studying the Kinetics of Molecular Binding. Journal of Chemical Information and Modeling. p. 3630–3634 (2019). [PubMed: 31381336]

61. Taylor JR An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements. 2nd ed.; University Science Books: Sausalito, CA, 1996.

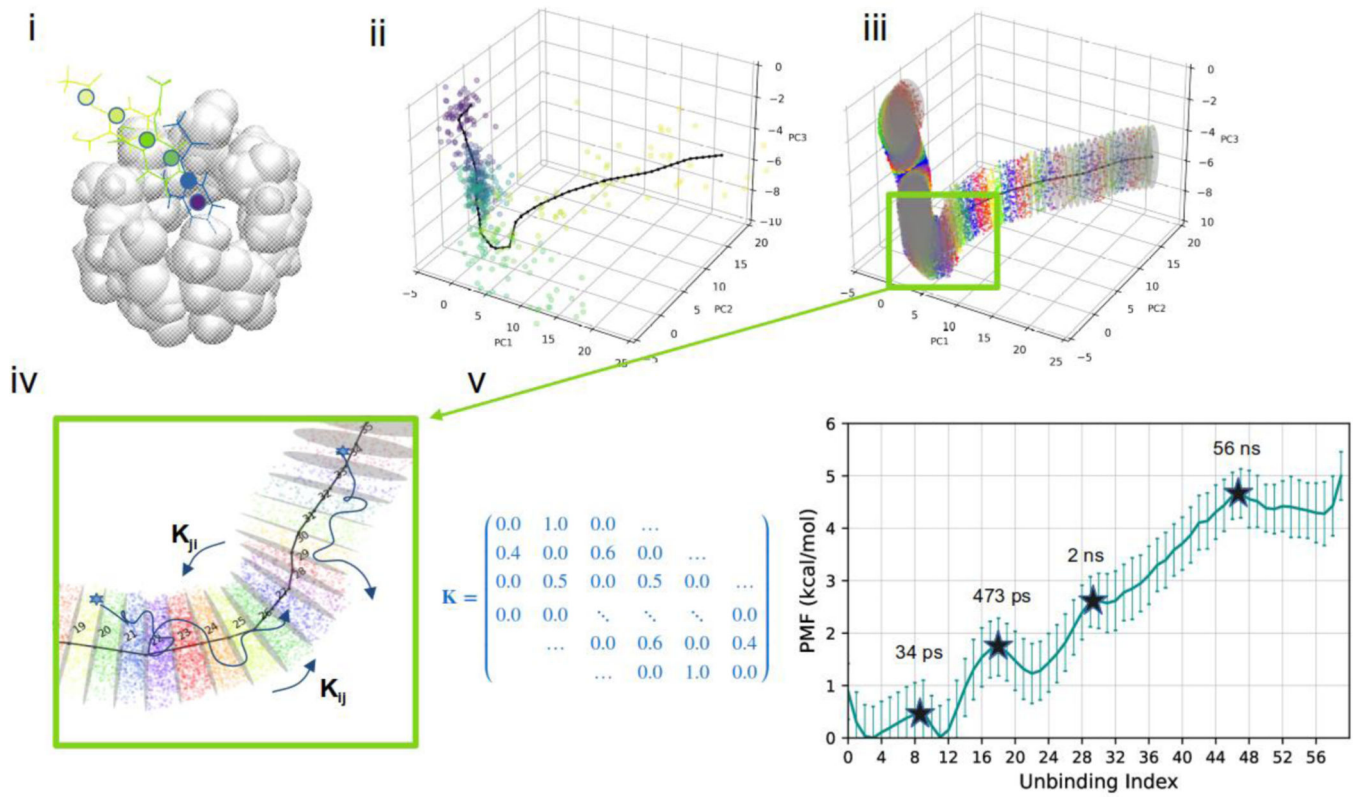


Fig. 1.

Graphical summary of the utilities offered in BKiT using the β-cyclodextrin (β-CD) and aspirin complex as an example. i) A trajectory of aspirin dissociation from β-CD. The thick line is a cartoon representation of aspirin unbinding with three distinct positions: bound in the middle of the cavity (purple), near the rim (light green) and diffusing to solvent (yellow). ii) All frames of the dissociation trajectory are projected into a 3D principal component (PC) plot, where each dot presents a frame. Continuous change in color from violet to yellow represents unbinding. A smoothed unbinding path is shown as a black line. iii) Optimized disks (indexes) are placed along the smoothed path shown in ii). iv) Projections of two 100-ps molecular dynamics (MD) trajectories are shown in blue curves, with start points marked as stars. Different indexes are labeled using a different color, and BKiT counts the transitions between two adjacent indexes. v) Illustration of a transition kernel and free energy plot (potential mean force [PMF]) where the calculated mean first passage times for major barriers are reported.

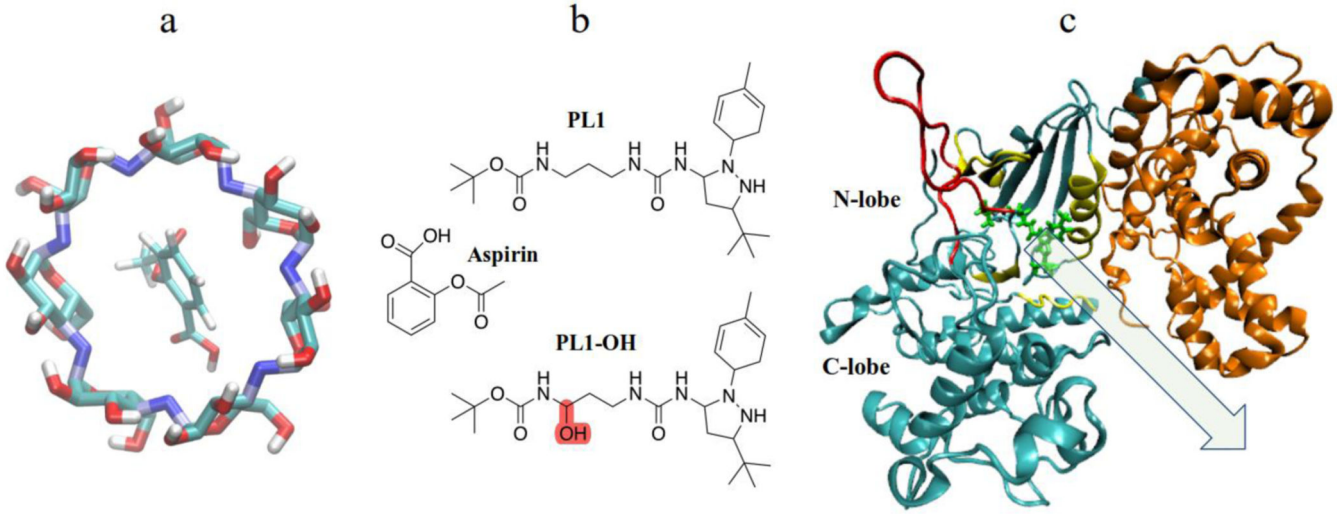


Fig. 2.

Molecular structures used in this study. a) β-CD and aspirin bound complex. b) Chemical structure of ligands pyrazolourea 1 (PL1), PL1-OH and aspirin. c) A complex structure of the proteins cyclin-dependent kinase 8 (CDK8; cyan), cyclin C (CycC; orange) and PL1 (green). Regions with significant motions during ligand unbinding are presented with different colors. Yellow: αC helices, β1, β2, and β8 sheets, and residues 146–148. Red: activation loop. Unbinding path is shown with a transparent arrow.

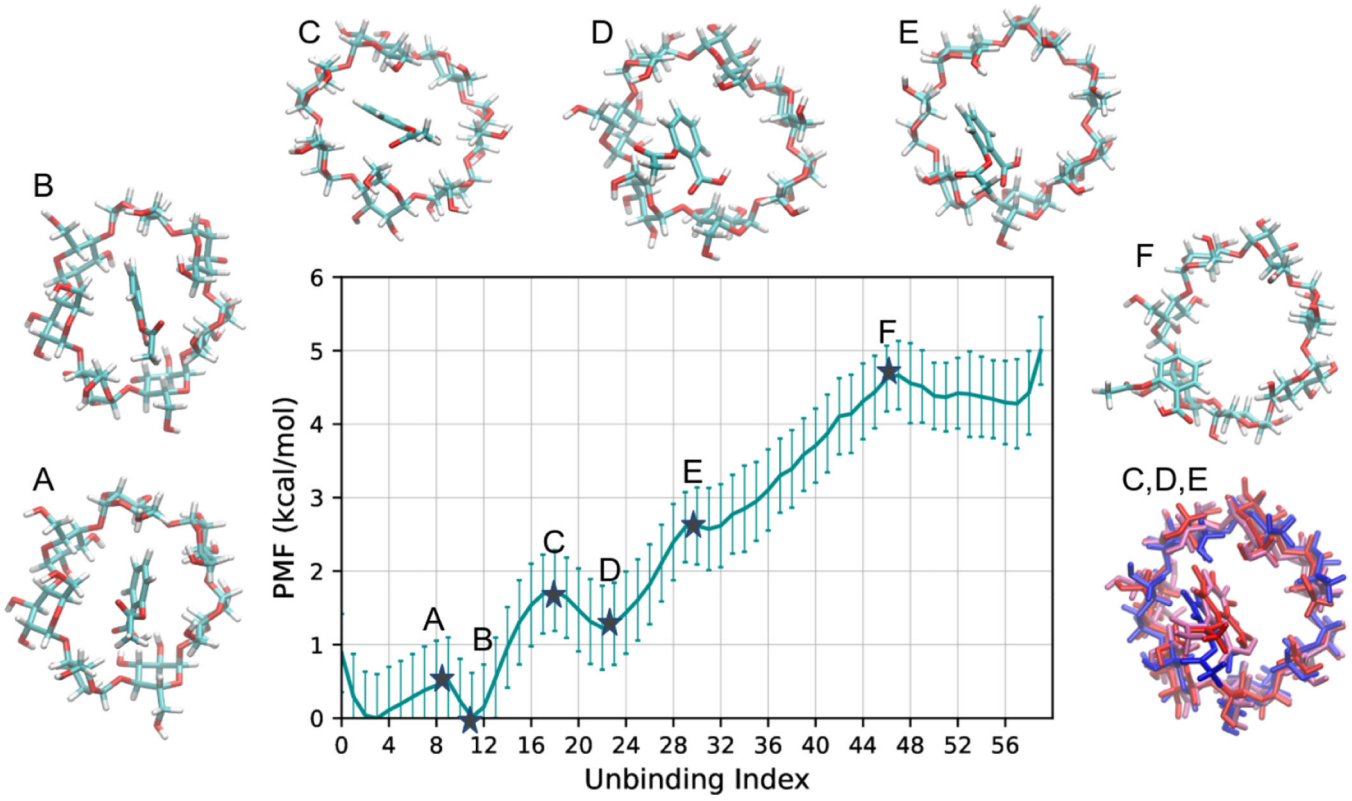


Fig. 3.

Free energy profile (PMF) of β-CD and aspirin with molecular structures. Structures depicted from left to right (A-F) show dissociation and conformational changes associated with the labeled local free energy minimum or barrier marked with stars for clarity. Structures for microstates C, D and E are superimposed for better visual comparison between different conformations.

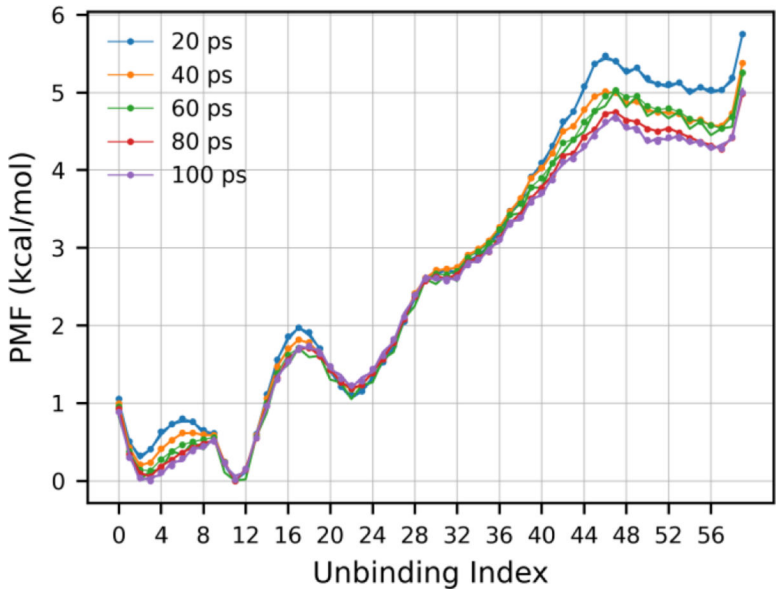


Fig. 4.

PMF profile of unbinding β-CD and aspirin complex using a 100-ps MD run with 20 replicates of each initial structure. Free energy results are calculated using a series of different simulation lengths (20, 40, 60, 80, 100 ps) for the β-CD and aspirin dissociation. Circles and lines represent iterative and eigenvalue methods, respectively, to compute the stationary flux. Local barriers increase when MD runs are shorter than 60 ns because of insufficient transition counts.

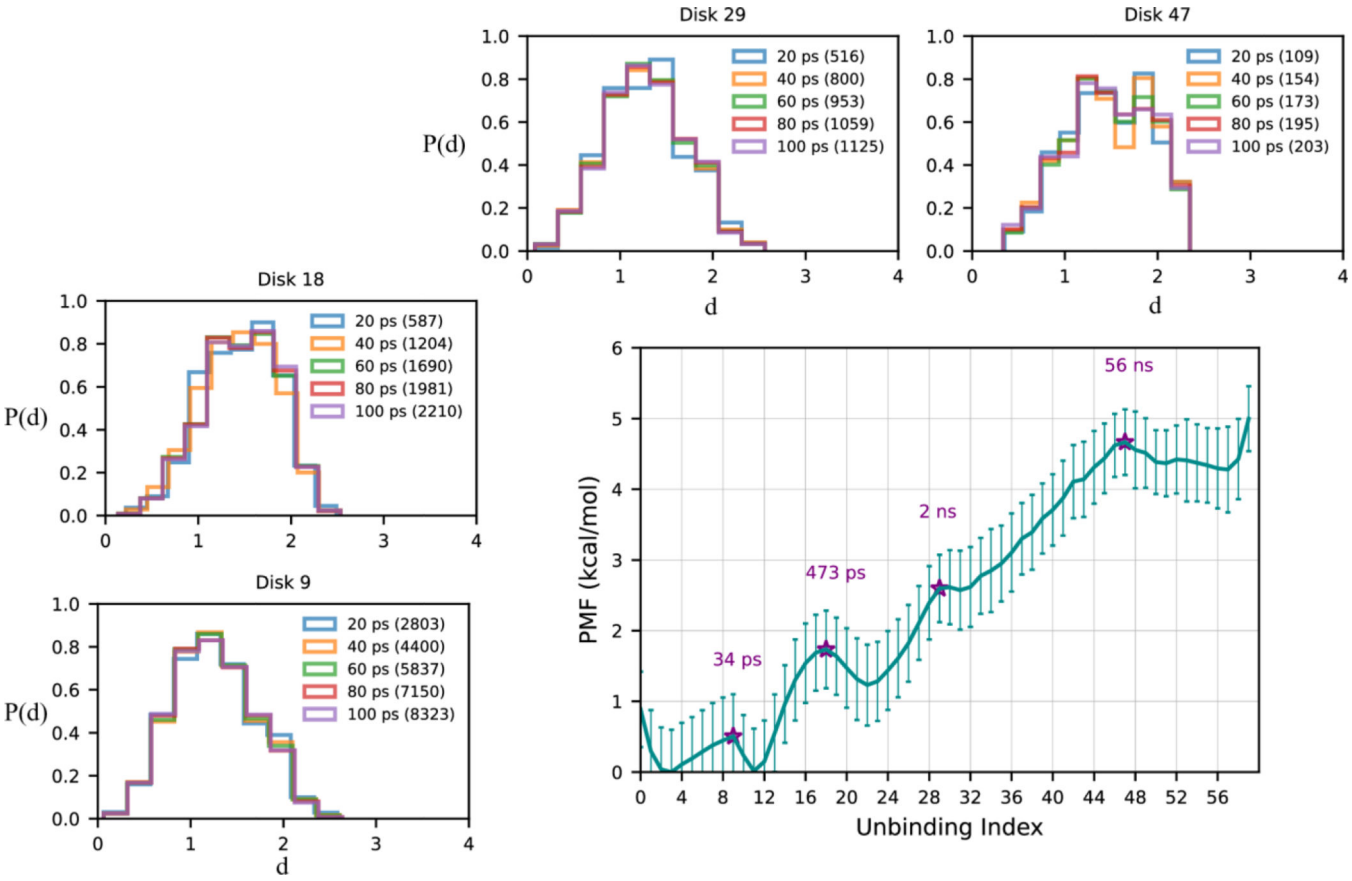


Fig. 5.

Examination of initial point distributions (IPD) for β-CD and aspirin complex. IPDs of each index marked with a star were plotted using the Euclidean distances between the center of the disk (index) and all points that hit the disk. Distributions were reported from 20-, 40-, 60-, 80- and 100-ps MD runs, and convergence was achieved after running longer than 80-ps MD simulations. The total number of initial points on the disks is shown in parentheses. The overall first passage time for aspirin passing major barriers along the PMF profile is reported.

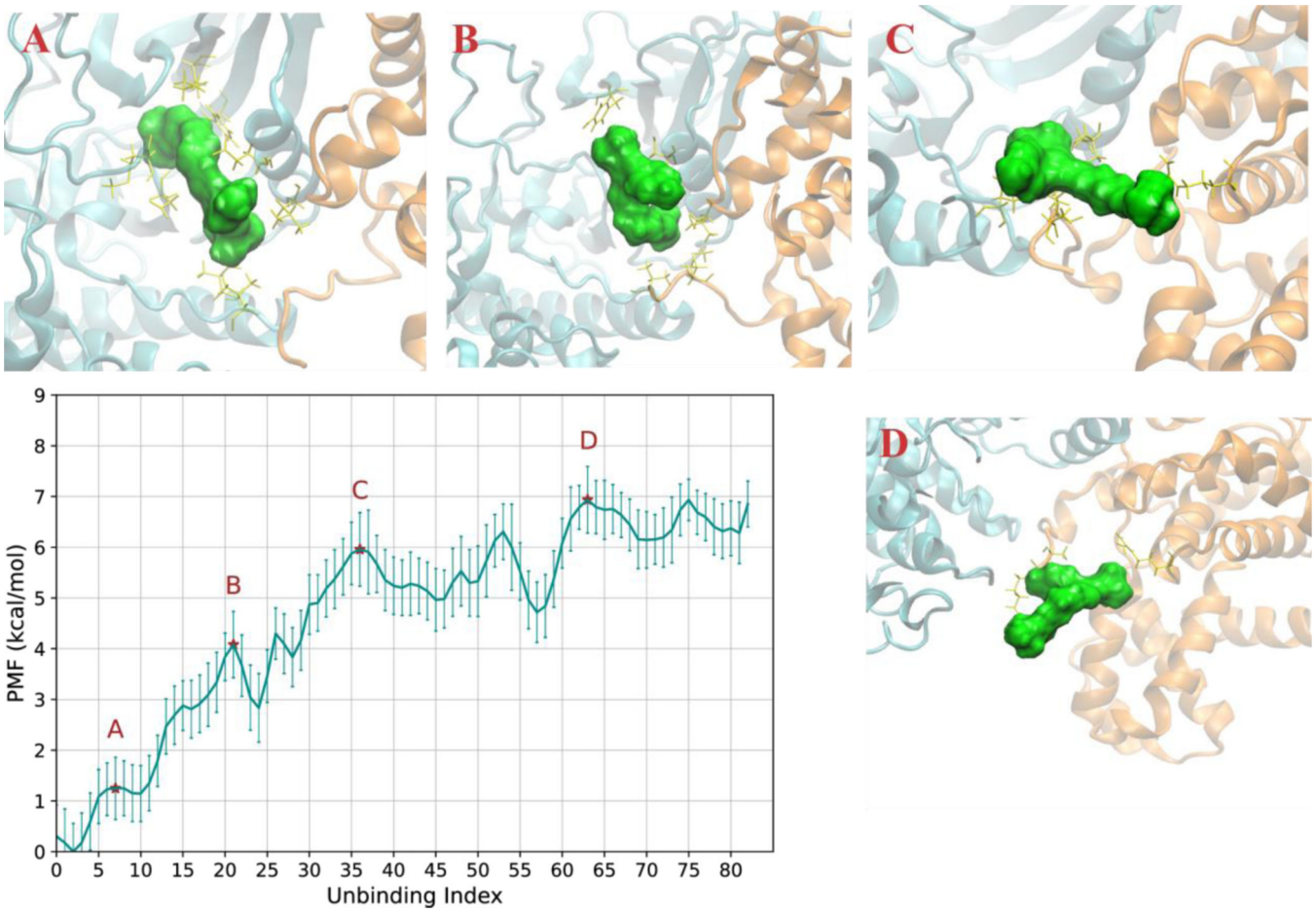


Fig. 6.

Free energy profile (PMF) for PL1 unbinding from CDK8–CycC. Local barriers are labeled with reported overall first passage time of PL1 at barrier A = 1 ns, B = 138 ns, C = 4 μs and D = 13 μs. Snapshots for each labeled energy barrier illustrate key residue interactions with PL1.

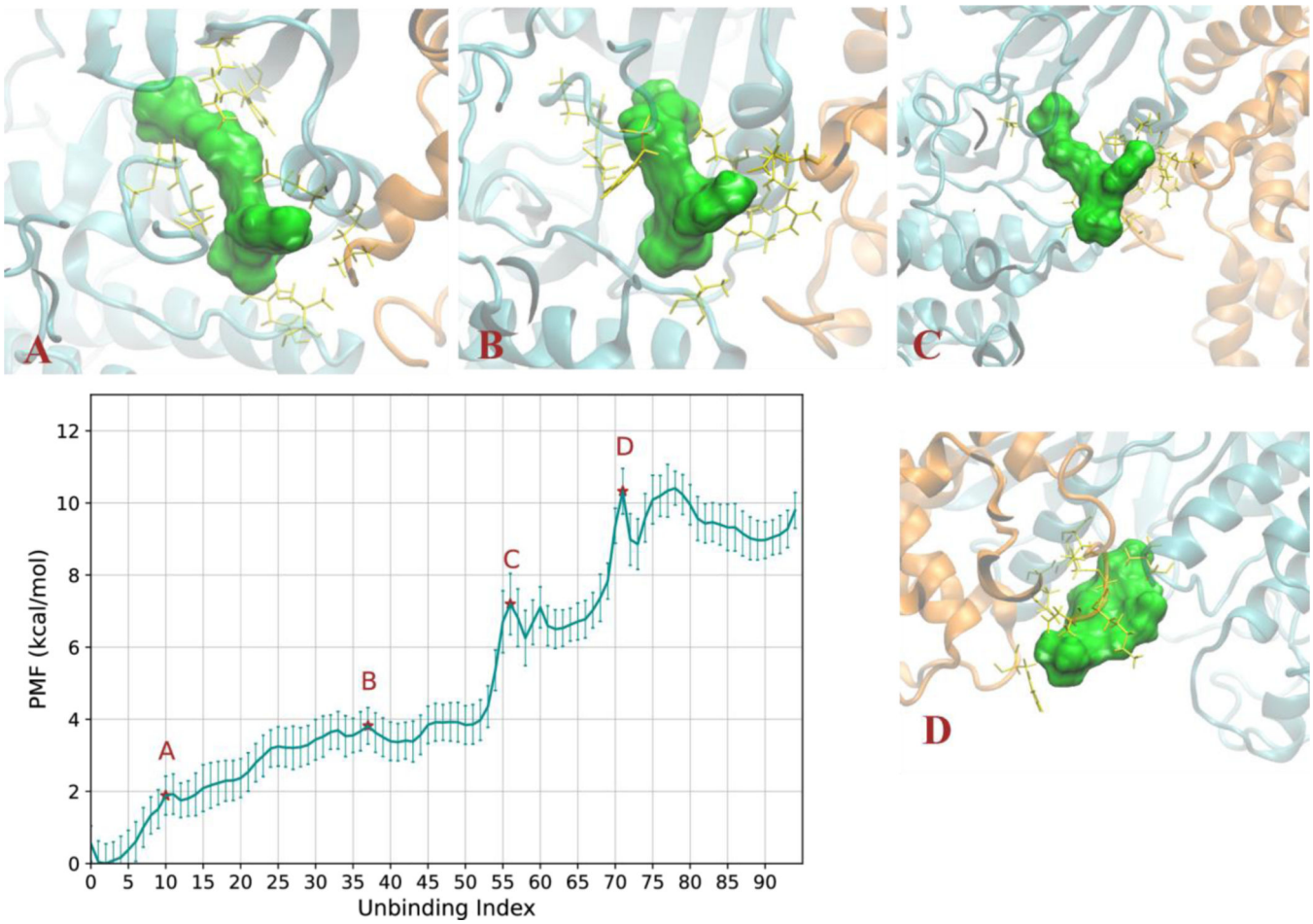


Fig. 7.

Free energy profile (PMF) for PL1-OH unbinding from CDK8–CycC. Local barriers are labeled with reported overall first passage time of PL1-OH at barrier A = 9 ns, B = 118 ns, C = 38 μs and D = 1 ms. Snapshots for each labeled energy barrier illustrate key residue interactions with PL1-OH.