Predicting Ligand Binding Kinetics Using a Markovian Milestoning with Voronoi Tessellations Multiscale Approach

Benjamin R. Jagger, Anupam A. Ojha, Rommie E. Amaro

Department of Chemistry and Biochemistry, University of California San Diego, 9500 Gilman Drive, La Jolla, California 92093, United States

ABSTRACT

Accurate and efficient computational predictions of ligand binding kinetics can be useful to inform drug discovery campaigns, particularly in the screening and lead optimization phases. Simulation Enabled Estimation of Kinetic Rates, SEEKR, is a multiscale molecular dynamics, Brownian dynamics, and milestoning simulation approach for calculating receptor-ligand association and dissociation rates. Here we present the implementation of a Markovian milestoning with Voronoi tessellations approach that significantly reduces the simulation cost of calculations as well as further improving their parallelizability. The new approach is applied to a host-guest system to assess its effectiveness for rank-ordering compounds by kinetic rates and to the model protein system, trypsin, with the noncovalent inhibitor benzamidine. For both applications, we

demonstrate that the new approach requires up to a factor of 10 less simulation time to achieve results with comparable or increased accuracy.

1. INTRODUCTION

Historically drug discovery campaigns have focused on equilibrium metrics, such and binding affinity, to inform screening and lead optimization of prospective compounds. However, kinetic parameters of binding, such and the on rate (k_{on}) and the off rate (k_{off}), are receiving increased attention as effective predictors of a compound's *in vivo* efficacy.[1,2] Of particular interest is the residence time ($1/k_{off}$) of compounds, which accounts for the effects of protein conformational flexibility on binding, unbinding, and rebinding as well as other factors.[3–6] Furthermore, the Kinetics for Drug Discovery Consortium database reports that only 0.4% of compounds uploaded with experimentally measured kinetics have diffusion-controlled association rate constants, suggesting k_{on} may also be an informative parameter to aid in lead optimization and the prediction of efficacy.[2] Ligand binding kinetics (k_{on} and k_{off}) are determined by a combination of effects such as: protein conformational flexibility, ligand induced receptor conformational changes, binding site water rearrangements, and drug rebinding, all of which influence the potency as well as selectivity of prospective compounds. Multiple compounds can have the same equilibrium binding affinity, yet corresponding values of k_{on} and k_{off} can vary by orders of magnitude. The additional level of detail afforded by knowledge of both the association and dissociation rate can therefore be critical for rationalizing why some compounds have efficacy, while others do not, aiding the lead optimization effort and reducing the high attrition rates currently associated with lack of *in vivo* efficacy .

Computational binding kinetics predictions have the potential to reduce the time and cost associated with experimental synthesis, assay development, and testing of many candidate compounds.[7,8] In particular, molecular simulation approaches are attractive for the structural, dynamical, and mechanistic insights they can provide of the drug binding/unbinding pathways in addition to predicting rate constants.[9] Brownian dynamics (BD) simulations are routinely used to efficiently estimate protein-ligand association rates and identify binding pathways.[10–14] Atomistic molecular dynamics (MD) simulations can also be used to study ligand binding and unbinding, however the increased model complexity necessitated by MD makes it limited by sampling. Hardware and software improvements such as exascale computing, the Anton super computer, increasingly powerful graphical processing units (GPUs), and volunteer distributed computing have made the study of binding kinetics with brute-force type approaches possible.[15–22] Generally, these approaches are limited to a small number of compounds and observe only a few association events and no dissociation events. Brute force MD simulations can access timescales on the order of milliseconds; however drug molecules often have residence times on the order of seconds, minutes, or even longer. As such, sampling remains the foremost limitation for these approaches. Furthermore, for simulation-based techniques to be useful in a drug discovery campaign, they must be able to provide predictions for 10s-100s of compounds in a reasonable timeframe. To overcome these challenges, many MD-based approaches have been developed that utilize biasing forces or other statistical mechanical techniques to access the timescales needed to predict binding and unbinding kinetics.[9,23] These include methods such as Markov State Models (MSMs),[22,24–28] metadynamics,[29–34] milestoning,[35–40] and others.[41–46] Additionally, multiscale methods exist that integrate MD with other approaches such as quantum

mechanics or BD, or continuum approaches to better predict kinetic parameters by improving either accuracy or scalability.[47–53]

One such multiscale approach is the MD/BD/milestoning methodology "Simulation Enabled Estimation of Kinetic Rates" (SEEKR) which we develop and have shown to be effective for the calculation of both k_{on} and k_{off} as well as the rank ordering of compounds by their rates.[54–56] Milestoning theory facilitates the division of simulation space into smaller regions called milestones that can be simulated independently and in parallel.[57–60] SEEKR uses atomistic, fully flexible MD simulations for milestones close to the binding site where these interactions are critical for describing the binding/unbinding process. Rigid body BD simulations are used in regions far from the binding site to dramatically reduce the computational cost, while still providing a sufficient description of the binding process, which is primarily diffusive in these regions. SEEKR is a freely available software package that automates the preparation, simulation and analysis of these binding kinetics calculations using the existing software NAMD[61] for MD simulations and Browndye[62] for BD simulations. While the effectiveness of SEEKR was previously demonstrated for predicting kinetic rates as well as rank ordering compounds, these calculations required a significant computational cost that would make the screening of many compounds challenging. It was therefore necessary to develop improvements to this methodology to reduce the amount of MD simulation required as well as improve the parallelizability of calculations.

Here we present a new implementation of SEEKR which utilizes the theory of Markovian Milestoning with Voronoi Tesselations (MMVT).[63,64] This new approach overcomes the primary sampling bottleneck associated with our previous implementation; obtaining an equilibrium distribution on each milestone. Instead, trajectories are confined to a Voronoi cell with the use of

a reflective boundary condition. Figure 1 shows a general schematic of an MMVT SEEKR model which combined MD and BD simulations.
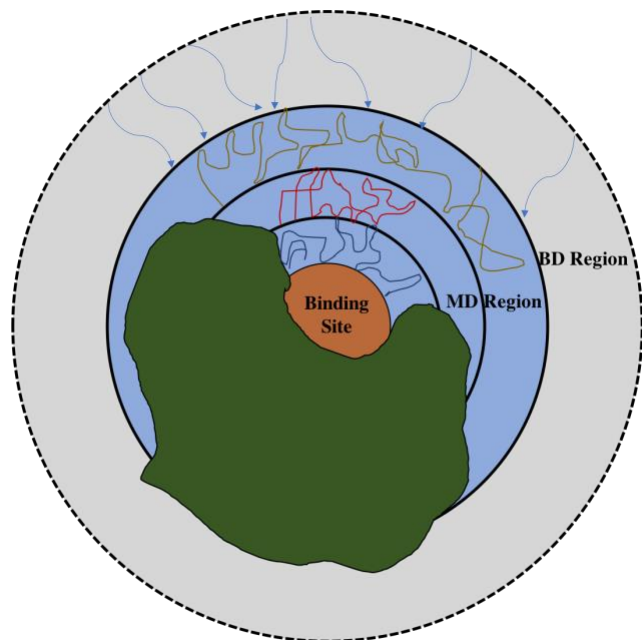

Figure 1. Cartoon depiction of a MMVT SEEKR rate calculation using spherical milestones representing radial distances from the binding site (black circles). The blue shaded regions are treated with MD simulations, while the grey shaded region employs computationally less expensive BD simulations. MD trajectories (colored lines) are confined to a particular cell with the use of a reflective boundary condition when a milestone is touched. Many BD trajectories (blue arrows) efficiently simulate the association of the ligand from large distances. Milestoning theory enables the statistics from many independent cells and both simulation modalities to be combined for the calculation of binding and unbinding rates .

We test this new implementation on a model host-guest system: β-cyclodextrin with seven small molecule ligands, as well as the model protein system: trypsin with the noncovalent inhibitor benzamidine. The accuracy and efficiency of the MMVT SEEKR results are directly compared

to experimentally measured kinetics, the previous SEEKR implementation, and other simulation approaches for each system. MMVT SEEKR produces results that are in agreement with experimental measurements and comparable to the previous SEEKR implementation for both model systems, while benefiting from up to a 10-fold reduction in simulation cost. Finally, we discuss convergence estimates for the sampling of each milestone as a way to further reduce the simulation cost by adaptively terminating or extending individual simulations.

2. METHODS

2.1 MMVT SEEKR package

The MMVT SEEKR package is a series of python scripts (python 3.7 or later), freely available on Github, that automates the preparation, running, and analysis of all simulations necessary for ligand binding kinetics calculations. MMVT SEEKR utilizes user-defined inputs of structures and model parameters to generate all files necessary for a SEEKR calculation. Files are organized into a filetree with branches for each independent milestone. MMVT SEEKR uses the freely available softwares NAMD[61] for MD simulations and Browndye[62] for BD simulations and generates all necessary input files for running these simulations in the appropriate portions of the filetree. MMVT SEEKR uses the Colvars collective variable module of NAMD to define and monitor milestones during MD simulation, with any collective variable defined in the module able to be used for milestoning in MMVT SEEKR.[65] The appropriate colvar input files are created by the SEEKR preparation scripts. The SEEKR package also includes an analysis module containing functions to extract results from the simulation outputs, calculate rates, assess simulation convergence, perform error analysis, and easily plot relevant quantities. The module

is designed to be imported into a Jupyter notebook, and a sample notebook and tutorial are included in the distribution.

2.2 Markovian Milestoning with Voronoi Tessellations: theory and implementation

Our previous implementation of SEEKR employed a traditional milestoning procedure where short trajectories were initiated on each milestone and run only until they touched another milestone.[54,55,57,58] The primary challenge of this approach is that one must know the correct probability distribution from which to reinitialize new trajectories on each milestone, called the first hitting point distribution (FHPD). The FHPD is obtained by first running a long, harmonically restrained trajectory to sample the equilibrium distribution on each milestone. Position/velocity configurations from this equilibrium distribution are then used to launch new trajectories which are propagated backward in time by reversing the velocity of the system. Only equilibrium configurations which touch another milestone before touching the milestone on which they started are included as part of the FHPD. This procedure is computationally expensive, in particular, long-timescale (microsecond) trajectories must be run for each milestone in order to adequately sample all configurations of the equilibrium distribution. This creates a computational bottleneck for the SEEKR method as this portion of the method has limited parallelizability. The MMVT procedure proposed by Vanden-Eijnden and Venturoli overcomes this barrier by eliminating the requirement of initializing all trajectories from a configuration in the FHPD.[63] Instead, milestones are defined as the edges of a Voronoi tessellation and trajectories are confined to a Voronoi cell with the use of a reflective boundary condition. Here we will briefly describe the key aspects of this theory necessary for our implementation and refer the reader to the original paper for a deeper theoretical description and the paper of Maragliano et. al. for an implementation employing restraining potentials.[63,64]

The central assumption of MMVT is that the evolution of the system through time can be described as a continuous-time Markov-jump process between milestone states with the rate matrix, **Q,** having off-diagonal elements q_{ij} for $i{\neq}j$ and diagonal elements $q_{ii} = -\Sigma_{j\neq i} q_{ij}$ where i and j correspond to the starting and ending milestone indices. From a maximum likelihood estimation of **Q**, the off-diagonal elements ($i{\neq}j$) are defined as

$$q_{ij} = \begin{cases} \dfrac{N_{ij}}{R_i} & if \ R_i \neq 0 \\ 0 & if \ R_i = 0 \end{cases} \tag{1}$$

where N_{ij} is the number of transitions between milestone i and milestone j and R_i is the total time spent having last touched milestone i. The quantities N_{ij} and R_i can be estimated from independent simulations confined to Voronoi cells as described below.
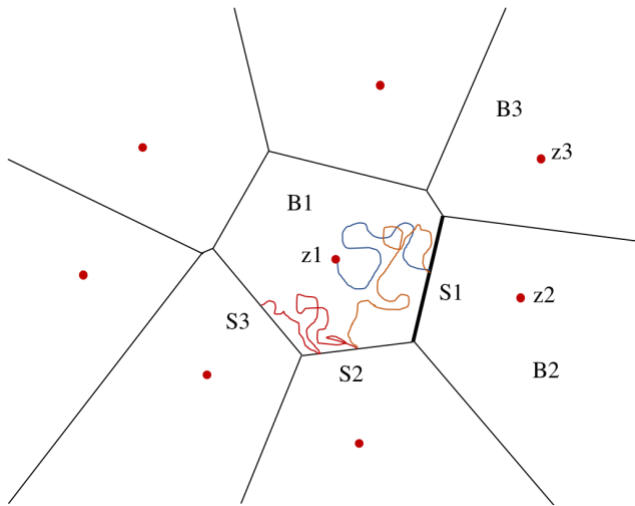

Figure 2. Sample Voronoi tessellation from the red generating points, z. The edges of the cells define the milestones. Milestone S1 (thick line) represents the shared boundary between cells B1 and B2. The colored lines represent a hypothetical trajectory confined to cell B1 using reflective boundary conditions. Changes in color correspond to successful transitions between milestones

for a single, continuous trajectory. The yellow portion is a transition from milestone S1 to S2 and the red portion from S2 to S3. The same simulation procedure is conducted independently in each Voronoi cell.

The definition of the Voronoi cells can be generalized from Cartesian space to collective variable space (i.e. bond distances, angles, etc.) with the collective variables denoted as $\theta(x) = (\theta_1(x),\ldots,\theta_M(x))$. A set of generating points, $z_\alpha \in \mathbb{R}_M$, with $\alpha = 1,2,\ldots,\Lambda$, define a unique partition of configuration space, Ω, into Voronoi cells (figure 2). The cell B_α from generating point z_α is the region

$$B_\alpha = \{x \in \Omega: \|\theta(x) - z_\alpha\| < \|\theta(x) - z_\beta\| \text{ for all } \beta \neq \alpha\} \tag{2}$$

The milestones are therefore defined as the common boundary, or edges, of adjacent cells. Independent simulations can then be carried out in each of the cells, propagated by the appropriate dynamical integrator, with the addition of a collision rule at the cell boundaries to keeps the trajectory confined to the appropriate cell. This collision rule is defined as

$$x_\alpha(t + \Delta t) = \begin{cases} x_\alpha^* & if \ x_\alpha^* \in B_\alpha \\ x_\alpha(t) & otherwise \end{cases} \tag{3}$$

and

$$v_\alpha(t + \Delta t) = \begin{cases} v_\alpha^* & if \ x_\alpha^* \in B_\alpha \\ -v_\alpha(t) & otherwise \end{cases} \tag{4}$$

These boundary conditions, in essence, result in the velocity of the system being reversed whenever a trajectory collides with a boundary in order to keep the system inside the appropriate

cell. The underlying justification for this rule is that, from time reversibility, every trajectory leaving the cell has a statistically indistinguishable trajectory entering the cell at the same point, but with its velocity reversed. Therefore, the correct Boltzmann-Gibbs distribution is maintained within the cell, as long as some thermal bath ensures that the trajectory does not perfectly retrack itself when its velocity is reversed upon collision with a boundary. Importantly, this procedure eliminates the need to determine the FHPD and equilibrium distribution on each milestone.

From the simulations confined to a Voronoi cell, B_α, with edges (or milestone) indices i and j, one can obtain the quantities N_{ij}^α and R_i^α, where N_{ij}^α is the number of times a trajectory collides with a milestone after having last touched a different milestone and R_i^α is the total time the simulation spends having last touched milestone i. These two quantities can be related to the quantities N_{ij} and R_i needed for the determination of $\mathbf{Q}$ by weighting the cell specific values by the equilibrium probability of that cell:

$$N_{ij} = T \sum_{\alpha=1}^{\Lambda} \pi_\alpha \frac{N_{ij}^\alpha}{T_\alpha} \tag{5}$$

$$R_i = T \sum_{\alpha=1}^{\Lambda} \pi_\alpha \frac{R_i^\alpha}{T_\alpha} \tag{6}$$

Here T_α is the total simulation time in cell α and T is the reciprocal sum of time spent in all cells, which ensures dimensional consistency. The equilibrium probability, π_α, can then be computed by solving the system of equations defined by 7 and 8.[66]

$$\sum_{\beta=1,\beta\neq\alpha}^{\Lambda} \pi_\beta k_{\beta,\alpha} = \sum_{\beta=1,\beta\neq\alpha}^{\Lambda} \pi_\alpha k_{\alpha,\beta} \ , \tag{7}$$

$$\sum_{\alpha=1}^{\Lambda} \pi_\alpha = 1 \tag{8}$$

Where we assume that the flux in and out of each cell is zero at steady state for the unrestrained system. The quantity $k_{\alpha,\beta}$ is defined as:

$$k_{\alpha,\beta} = \frac{N_{\alpha,\beta}}{T_\alpha} \tag{9}$$

Where $N_{\alpha,\beta}$ is the total number of collisions with the common boundary of cells B_α and Bβ. It is important to note that the key quantities to determine the rate matrix **Q** are N_{ij}^α and R_i^α, which can be obtained independently for each Voronoi cell. This independence facilitates the embarrassingly parallel nature of the MMVT SEEKR simulations. Furthermore, the convergence of these key quantities can be monitored as an estimate of the convergence of sampling for a particular MMVT cell, which will be discussed in further detail in section 2.4. The off rate can then be approximated as the reciprocal of the mean first passage time (MFPT) from the bound state to the outermost milestone using the standard expression for the MFPT in a continuous-time Markov chain

$$\hat{Q}T^N = -\mathbf{1} \tag{10}$$

$\hat{Q}$ is the N-1 by N-1 matrix obtained by deleting the last row and column of Q and **-1** is the unit vector in ℝN-1. TN is the a vector with entries T_i^N that are the MFPTs from milestone i to

milestone N. It has been shown that the MFPTs computed from MMVT are exact if optimal milestones are used, as is the case for traditional milestoning simulations.[58,63]

For SEEKR, the MMVT algorithm is implemented directly in the NAMD configuration file using the existing TCL interface. The colvars module is used to monitor the milestones defined by existing collective variables (with no biasing force used).[65] The existing NAMD TCL commands "checkpoint" and "revert" as well as the "rescalevels" command are used to facilitate the reflective boundary conditions needed for the MMVT algorithm when the monitored collective variable crosses a predefined milestone boundary. To improve calculation efficiency, boundary crossings can be checked after a user-defined number of steps, rather than every step of the simulation. All transition events are output in the simulation output file for post-processing with the SEEKR analysis package.

2.3 Incorporating Brownian dynamics simulations to calculate k_{on}

BD simulations are extremely useful for efficiently simulating the portion of the association process where the ligand is far from the binding site and therefore the atomistic detail of MD is not required to obtain an adequate description of the process. Instead, solvent is approximated by a dielectric and solute molecules (receptor and ligand) are treated as rigid or semi-rigid bodies with dynamics propagated according to the general equation of Brownian motion. The Northrup Allison McCammon (NAM) method can be used to estimate k_{on} from BD simulations[11,14] using the equation

$$k_{on} = k_b \beta \tag{11}$$

where k_b is the rate of diffusion to a spherical surface of radius b (b surface) from the receptor calculated by

$$k_b = 4\pi \left[\int_b^\infty \frac{e^{\left(\frac{U(r)}{k_b T}\right)}}{r^2 D(r)} dr \right]^{-1} \tag{12}$$

Where U(r) is the potential energy between the receptor and ligand at distance r, $k_b T$ is the Boltzmann constant times temperature, and D(r) is the diffusion coefficient. β from eq. 11 is the probability that a ligand on the b surface will continue on to react, rather than escaping to an infinite distance. In practice, k_b is calculated automatically by the Browndye software used by SEEKR for the BD simulations.[62] Traditionally the value of β is calculated from many BD simulations, however in the SEEKR implementation, we calculate this probability from a combination of MD and BD simulations. BD simulations are first conducted from the b surface to the outermost milestone. Successful trajectories from this simulation are a FHPD on the outermost milestone. Subsequent BD simulations are then carried out from each point in this FHPD until they successfully touch the second outermost milestone or escape to infinity. The rate matrix, Q, constructed from the MMVT portion of the model can then be converted to a transition probability matrix, K, and modified to include the probability of binding/escape determined from the BD simulations. β is then calculated as the stationary flux, q_{stat}, of the bound state milestone by solving the equation

$$q_{stat} (\boldsymbol{I} - \boldsymbol{K}) = 0 \tag{13}$$

Where **I** is the identity matrix. The values of k_b and β can then be used to calculate k_{on} with eqn. 11. While the transition probabilities are obtained differently for the MMVT implementation, the calculation of k_{on} described here is the same as in the original implementation of SEEKR.

2.4 Error analysis simulation convergence estimates

The statistical error associated with the calculation of k_{on} and k_{off} was estimated using a Markov chain Monte Carlo (MCMC) procedure based on the procedure detailed by Noé in 2008 that was modified to sample the rate matrix rather than the transition kernel.[67] Each nonzero entry of the rate matrix, q_{ij} is sampled by pulling a new value from the appropriate gamma distribution with parameters N_{ij} and $1/R_i$ which is accepted or rejected based on a Metropolis criteria. The standard deviation of the rate constants calculated from many iterations of the MCMC procedure is used as an estimate of the statistical error of the calculation. The convergence of the MCMC calculated rate constants is monitored to ensure the rate matrix has been sufficiently sampled. Finally, it should be noted that the average calculated MCMC rate constant and the maximum likelihood estimate described in section 2.2 should converge with sufficient sampling, which can also be monitored as a measure of convergence of the simulations.

The convergence of sampling in each Voronoi cell is essential for determining the amount of simulation needed to accurately calculate the rate constants of intertest. As described in section 2.2, the two key quantities necessary to construct the rate matrix, Q, are $N_{ij}^{\alpha}/T_{\alpha}$ and R_i^{α}/T_{α}, which are independently obtained from simulations in each Voronoi cell. As a result, these two quantities can be monitored as a function of simulation time in each cell to estimate the convergence of sampling from the simulations. The SEEKR analysis package described in section 2.1 contains functions to extract and plot these quantities, providing the user with
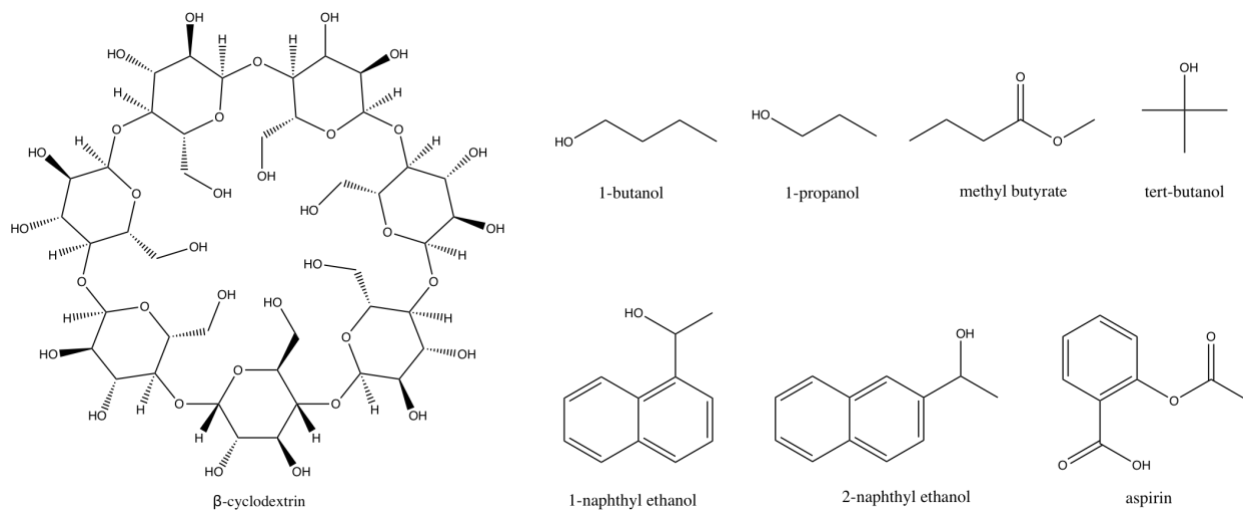
qualitative, visual estimates of the convergence of sampling. A quantitative metric, however, is desirable, as it can then be utilized to provide a more rigorous and reproduceable metric for convergence that is transferrable between systems. In the SEEKR package, we have implemented a sliding window root mean square deviation (RMSD) function to provide such an estimate. The quantities $N_{ij}^{\alpha}/T_{\alpha}$ and R_i^{α}/T_{α} are calculated for user-defined strided portions of the data. A window of user-defined length is then moved through the data, grouping it into samples. For each sample, the RMSD from the average value is calculated. A user defined cutoff is specified as a percentage of the magnitude of the quantity (i.e 5% of the magnitude of the value). If the RMSD of each quantity remains below this cutoff for a user-defined length of simulation (i.e. 100ns) then the sampling in that cell is considered converged. Leaving these parameters to be specified by the user allows the convergence estimate to be adaptable to the particular application being considered. This allows the user to balance strictness of convergence, required level of accuracy, and amount of simulation time invested based on the particular question being answered. For example, a rank-ordering application may not require the same strictness of convergence as trying to estimate the true value of the rate constant within experimental error. By estimating the convergence of each Voronoi cell independently, sampling can be adaptively terminated or extended on a cell by cell basis. Allowing more simulation time to be devoted to difficult to sample areas, while eliminating excess simulation in easier to sample regions. We note that a metric such as this could also be used in the future to monitor convergence "on-the-fly" during simulations, rather than after a portion of simulation is run.

3. Results and Discussion

3.1 Host-guest molecule rank ordering

We assessed the effectiveness of the our new MMVT SEEKR implementation for rank-ordering compounds by their binding/unbinding rates. The model host-guest system β-cyclodextrin with seven different guest molecules was studied (chart 1), as this was the same system studied with the original SEEKR implementation. Therefore, it was possible to directly compare the accuracy and efficiency of the new MMVT approach to the previous implementation[56] as well as to brute-force MD simulation calculated rates[39] and experimentally measured kinetics.[68–72]

Chart 1. Structures of β-cyclodextrin and the seven ligands tested

β-cyclodextrin   1-butanol   1-propanol   methyl butyrate   tert-butanol   1-naphthyl ethanol   2-naphthyl ethanol   aspirin

System and simulation details can be found in section S1 of the Supporting Information. A one dimensional pseudo-Voronoi tessellation was generated using anchor points that resulted in milestones being placed between the center of mass (COM) of the host and the COM of the guest from 1.5 Å to 13.5 Å in 1.5 Å increments. Many short (~20-50 ns) MMVT MD simulations were carried out for each Voronoi cell for a combined total of ~560 ns of simulation per ligand (exact

simulation lengths are presented in table S1). Additionally, BD simulations were performed for the 13.5 Å milestone as described in section 2.3 with additional details in section S1. Both k_{on} and k_{off} as well as the binding free energy were calculated for each ligand. The convergence estimates described in section 2.4 were also used to determine the minimum simulation necessary to produce a converged result for each cell and the rates were recalculated using only that portion of the data. A sliding window of 30 samples was used pulled from the data with a stride of 1 ns after skipping the first 10 ns. Cells were considered converged when values remained less than 5% of the average value for 20 windows (20 ns). Rate constants calculated with this minimum RMSD cutoff produced results consistent with the data from the full simulations, while benefitting from an additional ~20% reduction in simulation time. The values of the rate constants and binding free energies calculated using both methods as well as the brute force MD simulations and experiment are presented in tables S2-S4. Figure 3 shows the calculated values for a) k_{on} and b) k_{off} and c) ΔG ordered by increasing magnitude of the experimentally measured value. The values calculated with MMVT SEEKR are in good agreement with values calculated from the previous SEEKR implementation. The values of k_{on} remain approximately one order of magnitude faster than experiment. As in the previous implementation, rank-ordering by k_{on} was not possible due to the limited variation in the experimental and computed values which are all near the diffusion limit.[73] The rank-ordering of ligands by k_{off} was improved with MMVT SEEKR; incorrectly ordering only two ligands, rather than three. The k_{off} values calculated with MMVT SEEKR were consistently faster than the experimental rates, which was also observed in the original SEEKR implementation. In addition, the binding free energy of each ligand can be determined because k_{on} and k_{off} are known. Calculated binding free energies are also in good agreement with experiment (Figure 3c), with

only the value for methyl butyrate differing from the experimental value by more than 1 kcal/mol. As the on rates for these compounds are similar, the binding free energy is primarily dominated by the values of k_{off}. As such the rank-ordering is also similar to that for k_{off}; incorrectly ordering methyl butyrate and tert-butanol, but also misordering 2-naphthylethanol as a result of a faster k_{on}. The MMVT SEEKR method was able to produce comparable predictions of k_{on} and k_{off} and binding free energy to the original SEEKR implementation, with improved rank-ordering of the ligands by k_{off} and comparable rank-ordering by free energy. Furthermore, MMVT SEEKR benefits from a roughly 10-fold reduction in the amount of simulation used to achieve this result. The minimum simulation estimates, which produce nearly identical results to the full simulation data, save an additional 20% (~100ns) on the total MMVT simulation cost.
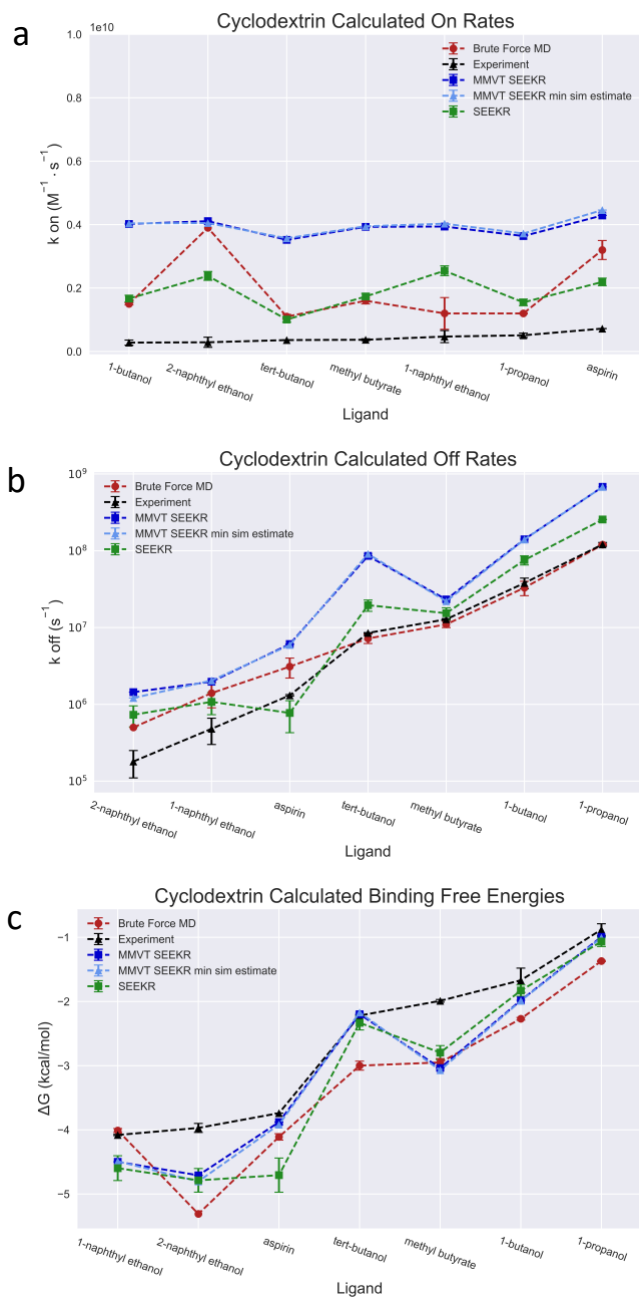
Figure 3. Comparison of experimentally measured values[68–72] (black), brute force MD[39] (red) original SEEKR implementation (green) MMVT SEEKR[56] (dark blue) and MMVT SEEKR minimum simulation estimates for a) k_{on} b) k_{off} and c) binding free energy. Compounds are ordered by increasing experimentally measured values.

3.2 Trypsin-benzamidine application

We also tested the MMVT SEEKR method on the well-studied model system trypsin with the noncovalent inhibitor, benzamidine. This system has been used as a benchmark by many simulation-based approaches, including the original SEEKR implementation.[22,24,31,42,44,55] The increased complexity resulting from protein dynamics as well as a k_{off} value multiple orders of magnitude slower than those tested in the host-guest systems serve as a test for the efficiency and accuracy of the new MMVT SEEKR implementation. Detailed system and simulation details can be found in section S2 of the supporting information. The same collective variable representing the distance of the ligand from the binding site was used as in the original SEEKR implementation.[55] Voronoi Cells were generated from anchor points that resulted in milestones with distances of 1.0, 2.0, 3.0, 4.0, 6.0, 8.0, 10.0, 12.0, and 13.0 Å (figure 4). We note that the MMVT algorithm samples the regions between milestones differently than the previous implementation, and therefore the spacing of milestones is not identical to the previous implementation to account for sampling challenges associated with large energy barriers and to ensure that the Markov assumption remains valid. Minimum simulation estimates of k_{on} and k_{off} and binding free energy were obtained using a stride of 2 ns after skipping the first 20 ns, an RMSD sample window of 200 ns and a cutoff of 5% for at least 100 ns.
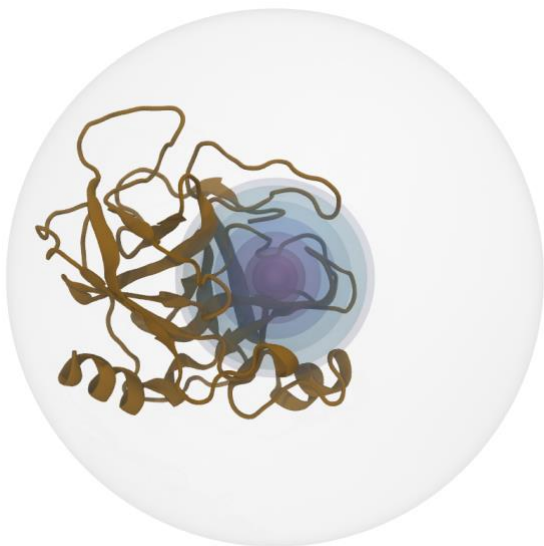
Figure 4. Structure of trypsin (cartoon) with milestones drawn as colored spheres. The outermost (grey) sphere represents the BD simulations from the "b surface" described in section 2.3. Many MD milestones are placed close to the binding site, while the BD region covers a much larger portion of the system.

MMVT SEEKR effectively reproduces the experimentally measured on and off rates and binding free energy (table 1).[74] Using the full 4.4 µs of simulation data, the MMVT SEEKR result more closely reproduces the experimental residence time than the original SEEKR result, which required 19 µs; a ~4-fold reduction in simulation time. The association rate calculated with MMVT SEEKR is faster than both experiment and the previous SEEKR implementation. However, we tested the sensitivity of the BD results to ionic strength, and identified the same trend observed in our previous work (table S5). Benzamidine association is highly sensitive to the ionic strength used in the BD simulations, with lower ionic strengths resulting in slower association rates. This suggests that the benzamidine molecule experiences a repulsive force as it approaches the binding site. Furthermore, the MMVT minimum simulation estimate again

produces a comparable result to the original SEEKR implementation, requiring only 2.9 μs of simulation. This minimum simulation estimate saves an additional 35% of simulation from the full MMVT data, a ~7 fold reduction in simulation from the original implementation. The enhancement in sampling provided by the MMVT SEEKR approach is evident, as it predicts residence times that are over 1000 times longer than the simulation time invested. Statistically robust estimates of such residence times on the order of a millisecond would likely pose a significant challenge and expense for a brute force simulation approach, highlighting the value of the SEEKR approach for both its enhancement in sampling as well as parallelism.

Table 1. Trypsin-benzamidine calculated rates and binding free energies, simulation time, and experimentally measured values

Method	k_{off} (s-1)	Residence Time (μs)	k_{on} (M-1 s-1)	ΔG (kcal/mol)	Simulation Time (μs)
Experiment[74]	600 ± 300	1700	2.9×10^7	-6.7 ± 0.05	
SEEKR[55]	83 ± 14	12000	$(2.1 \pm 0.3) \times 10^7$	-7.4 ± 0.1	19
MMVT SEEKR	174 ± 9	5750	$(1.2 \pm 0.05) \times 10^8$	-7.9 ± 0.04	4.4
MMVT SEEKR minimum simulation estimate	62 ± 6	16000	$(1.7 \pm 0.1) \times 10^8$	-8.8 ± 0.07	2.9

4. Conclusion

We have presented a new MMVT algorithm implemented in the SEEKR package for calculating receptor-ligand binding and unbinding rate constants as well as binding free energies. The results of the two applications we have described here demonstrate that MMVT SEEKR is effective for both rank-ordering compounds by their kinetics as well as reproducing the magnitude of experimentally measured kinetics. MMVT SEEKR benefits from a significant reduction in simulation cost compared to the previous SEEKR implementation by eliminating the need to determine equilibrium distributions and FHPDs for each milestone. We have also described a method for estimating the convergence of sampling for each Voronoi cell and adaptively extending or terminating simulations accordingly. This convergence estimate was shown to further reduce the simulation cost of MMVT SEEKR calculations while retaining accuracy. The MMVT algorithm can also be used to construct and simulate models with multiple dimensions of milestones, unlike the one dimensional models used in this study. Additional milestones could be useful for improving sampling of other slow degrees of freedom that may exist in a more complicated system. This may be particularly important when studying larger drug molecules with longer residence times and more complicated binding/unbinding mechanisms. The improvements to efficiency, as well as the embarrassingly parallel nature of milestoning simulations, make MMVT SEEKR well suited for use in future prospective studies for larger systems of pharmaceutical relevance.

ASSOCIATED CONTENT

Supporting Information.

The following files are available free of charge:

MMVT SEEKR code, documentation, and tutorial are available at

https://github.com/brjagger/mmvt_seekr

Additional system/simulation details, numerical values of calculated rates and binding free

energies of cyclodextrin systems (PDF)

Jupyter notebook analysis code/output for each system (PDFs)

AUTHOR INFORMATION

Corresponding Author

*Rommie E Amaro, ramaro@ucsd.edu

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval

to the final version of the manuscript.

ABBREVIATIONS

SEEKR, simulation enabled estimation of kinetic rates; MMVT, Markovian milestoning with Voronoi tessellations; MD, molecular dynamics; BD, Brownian dynamics; MFPT, mean first passage time; COM, center of mass

REFERENCES

(1)   Swinney, D. C. Opinion: Biochemical Mechanisms of Drug Action: What Does It Take for Success? *Nat. Rev. Drug Discov.* **2004**, *3* (9), 801–808. https://doi.org/10.1038/nrd1500.

(2)   Schuetz, D. A.; de Witte, W. E. A.; Wong, Y. C.; Knasmueller, B.; Richter, L.; Kokh, D. B.; Sadiq, S. K.; Bosma, R.; Nederpelt, I.; Heitman, L. H.; et al. Kinetics for Drug Discovery: An Industry-Driven Effort to Target Drug Residence Time. *Drug Discov. Today* **2017**, *22* (6), 896–911. https://doi.org/10.1016/j.drudis.2017.02.002.

(3)   Copeland, R. A.; Pompliano, D. L.; Meek, T. D. Drug–Target Residence Time and Its Implications for Lead Optimization. *Nat. Rev. Drug Discov.* **2006**, *5* (9), 730–739. https://doi.org/10.1038/nrd2082.

(4)   Tummino, P. J.; Copeland, R. A. Residence Time of Receptor-Ligand Complexes and Its Effect on Biological Function. *Biochemistry* **2008**, *47* (20), 5481–5492. https://doi.org/10.1021/bi8002023.

(5)   Copeland, R. A. The Drug-Target Residence Time Model: A 10-Year Retrospective. *Nat. Rev. Drug Discov.* **2016**, *15* (2), 87–95. https://doi.org/10.1038/nrd.2015.18.

(6)   Lu, H.; Tonge, P. J. Drug–Target Residence Time: Critical Information for Lead Optimization. *Curr. Opin. Chem. Biol.* **2010**, *14* (4), 467–474. https://doi.org/10.1016/j.cbpa.2010.06.176.

(7)   Romanowska, J.; Kokh, D. B.; Fuller, J. C.; Wade, R. C. Computational Approaches for Studying Drug Binding Kinetics; John Wiley & Sons, Ltd, 2015; pp 211–235. https://doi.org/10.1002/9783527673025.ch11.

(8)   De Vivo, M.; Masetti, M.; Bottegoni, G.; Cavalli, A. Role of Molecular Dynamics and Related Methods in Drug Discovery. *J. Med. Chem.* **2016**, *59* (9), 4035–4061. https://doi.org/10.1021/acs.jmedchem.5b01684.

(9)   Bruce, N. J.; Ganotra, G. K.; Kokh, D. B.; Sadiq, S. K.; Wade, R. C. New Approaches for Computing Ligand–Receptor Binding Kinetics. *Curr. Opin. Struct. Biol.* **2018**, *49*, 1–10. https://doi.org/10.1016/j.sbi.2017.10.001.

(10)   Ermak, D. L.; McCammon, J. A. Brownian Dynamics with Hydrodynamic Interactions. *J. Chem. Phys.* **1978**, *69* (4), 1352–1360. https://doi.org/10.1063/1.436761.

(11)   Northrup, S. H.; Allison, S. A.; McCammon, J. A. Brownian Dynamics Simulation of Diffusion-Influenced Bimolecular Reactions. *J. Chem. Phys.* **1984**, *80* (4), 1517–1524.

https://doi.org/10.1063/1.446900.

(12)   McCammon, J. A.; Northrup, S. H.; Allison, S. A. Diffusional Dynamics of Ligand-Receptor Association. *J. Phys. Chem.* **1986**, *90* (17), 3901–3905. https://doi.org/10.1021/j100408a015.

(13)   Northrup, S. H.; Erickson, H. P. Kinetics of Protein-Protein Association Explained by Brownian Dynamics Computer Simulation. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89* (8), 3338–3342. https://doi.org/10.1073/pnas.89.8.3338.

(14)   Luty, B. A.; McCammon, J. A.; Zhou, H. X. Diffusive Reaction Rates from Brownian Dynamics Simulations: Replacing the Outer Cutoff Surface by an Analytical Treatment. *J. Chem. Phys.* **1992**, *97* (8), 5682–5686. https://doi.org/10.1063/1.463777.

(15)   Lee, C. T.; Amaro, R. E. Exascale Computing: A New Dawn for Computational Biology. *Comput. Sci. Eng.* **2018**, *20* (5), 18–25. https://doi.org/10.1109/MCSE.2018.05329812.

(16)   Shaw, D. E.; Bowers, K. J.; Chow, E.; Eastwood, M. P.; Ierardi, D. J.; Klepeis, J. L.; Kuskin, J. S.; Larson, R. H.; Lindorff-Larsen, K.; Maragakis, P.; et al. Millisecond-Scale Molecular Dynamics Simulations on Anton. In *Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis - SC '09*; ACM Press: New York, New York, USA, 2009; p 1. https://doi.org/10.1145/1654059.1654126.

(17)   Shaw, D. E.; Grossman, J. P.; Bank, J. A.; Batson, B.; Butts, J. A.; Chao, J. C.; Deneroff, M. M.; Dror, R. O.; Even, A.; Fenton, C. H.; et al. Anton 2: Raising the Bar for Performance and Programmability in a Special-Purpose Molecular Dynamics Supercomputer. In *International Conference for High Performance Computing, Networking, Storage and*

Analysis, SC; IEEE, 2014; Vol. 2015-Janua, pp 41–53. https://doi.org/10.1109/SC.2014.9.

(18) Shan, Y.; Kim, E. T.; Eastwood, M. P.; Dror, R. O.; Seeliger, M. A.; Shaw, D. E. How Does a Drug Molecule Find Its Target Binding Site? *J. Am. Chem. Soc.* **2011**, *133* (24), 9181–9183. https://doi.org/10.1021/ja202726y.

(19) Dror, R. O.; Pan, A. C.; Arlow, D. H.; Borhani, D. W.; Maragakis, P.; Shan, Y.; Xu, H.; Shaw, D. E. Pathway and Mechanism of Drug Binding to G-Protein-Coupled Receptors. *Proc. Natl. Acad. Sci.* **2011**, *108* (32), 13118–13123. https://doi.org/10.1073/pnas.1104614108.

(20) Pan, A. C.; Borhani, D. W.; Dror, R. O.; Shaw, D. E. Molecular Determinants of Drug-Receptor Binding Kinetics. *Drug Discov. Today* **2013**, *18* (13–14), 667–673. https://doi.org/10.1016/j.drudis.2013.02.007.

(21) Huang, D.; Caflisch, A. The Free Energy Landscape of Small Molecule Unbinding. *PLoS Comput. Biol.* **2011**, *7* (2), e1002002. https://doi.org/10.1371/journal.pcbi.1002002.

(22) Buch, I.; Giorgino, T.; De Fabritiis, G. Complete Reconstruction of an Enzyme-Inhibitor Binding Process by Molecular Dynamics Simulations. *Proc. Natl. Acad. Sci.* **2011**, *108* (25), 10184–10189. https://doi.org/10.1073/pnas.1103547108.

(23) Elber, R. A New Paradigm for Atomically Detailed Simulations of Kinetics in Biophysical Systems. *Q. Rev. Biophys.* **2017**, *50* (May), e8. https://doi.org/10.1017/S0033583517000063.

(24) Plattner, N.; Noé, F. Protein Conformational Plasticity and Complex Ligand-Binding Kinetics Explored by Atomistic Simulations and Markov Models. *Nat. Commun.* **2015**, *6*

(1), 7653. https://doi.org/10.1038/ncomms8653.

(25)   Wu, H.; Paul, F.; Wehmeyer, C.; Noé, F. Multiensemble Markov Models of Molecular Thermodynamics and Kinetics. *Proc. Natl. Acad. Sci.* **2016**, *113* (23), E3221–E3230. https://doi.org/10.1073/pnas.1525092113.

(26)   Bowman, G. R.; Pande, V. S.; Noé, F. *An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation*; Bowman, G. R., Pande, V. S., Noé, F., Eds.; Advances in Experimental Medicine and Biology; Springer Netherlands: Dordrecht, 2014; Vol. 797. https://doi.org/10.1007/978-94-007-7606-7.

(27)   Prinz, J. H.; Wu, H.; Sarich, M.; Keller, B.; Senne, M.; Held, M.; Chodera, J. D.; Schtte, C.; Noé, F. Markov Models of Molecular Kinetics: Generation and Validation. *J. Chem. Phys.* **2011**, *134* (17), 174105. https://doi.org/10.1063/1.3565032.

(28)   Doerr, S.; de Fabritiis, G. On-the-Fly Learning and Sampling of Ligand Binding by High-Throughput Molecular Simulations. *J. Chem. Theory ...* **2014**, *10* (5), 2064–2069. https://doi.org/doi: 10.1021/ct400919u.

(29)   Mollica, L.; Decherchi, S.; Zia, S. R.; Gaspari, R.; Cavalli, A.; Rocchia, W. Kinetics of Protein-Ligand Unbinding via Smoothed Potential Molecular Dynamics Simulations. *Sci. Rep.* **2015**, *5* (1), 11539. https://doi.org/10.1038/srep11539.

(30)   Mollica, L.; Theret, I.; Antoine, M.; Perron-Sierra, F.; Charton, Y.; Fourquez, J.-M.; Wierzbicki, M.; Boutin, J. A.; Ferry, G.; Decherchi, S.; et al. Molecular Dynamics Simulations and Kinetic Measurements to Estimate and Predict Protein–Ligand Residence Times. *J. Med. Chem.* **2016**, *59* (15), 7167–7176.

https://doi.org/10.1021/acs.jmedchem.6b00632.

(31)   Tiwary, P.; Limongelli, V.; Salvalaglio, M.; Parrinello, M. Kinetics of Protein–Ligand Unbinding: Predicting Pathways, Rates, and Rate-Limiting Steps. *Proc. Natl. Acad. Sci.* **2015**, *112* (5), 201424461. https://doi.org/10.1073/pnas.1424461112.

(32)   Gobbo, D.; Piretti, V.; Di Martino, R. M. C.; Tripathi, S. K.; Giabbai, B.; Storici, P.; Demitri, N.; Girotto, S.; Decherchi, S.; Cavalli, A. Investigating Drug–Target Residence Time in Kinases through Enhanced Sampling Simulations. *J. Chem. Theory Comput.* **2019**, *15* (8), 4646–4659. https://doi.org/10.1021/acs.jctc.9b00104.

(33)   Bernetti, M.; Masetti, M.; Recanatini, M.; Amaro, R. E.; Cavalli, A. An Integrated Markov State Model and Path Metadynamics Approach To Characterize Drug Binding Processes. *J.   Chem.   Theory   Comput.* **2019**, *15*   (10),   5689–5702. https://doi.org/10.1021/acs.jctc.9b00450.

(34)   Casasnovas, R.; Limongelli, V.; Tiwary, P.; Carloni, P.; Parrinello, M. Unbinding Kinetics of a P38 MAP Kinase Type II Inhibitor from Metadynamics Simulations. *J. Am. Chem. Soc.* **2017**, *139* (13), 4780–4788. https://doi.org/10.1021/jacs.6b12950.

(35)   Bucci, A.; Yu, T.-Q.; Vanden-Eijnden, E.; Abrams, C. F. Kinetics of O 2 Entry and Exit in Monomeric Sarcosine Oxidase via Markovian Milestoning Molecular Dynamics. *J. Chem. Theory Comput.* **2016**, *12* (6), 2964–2972. https://doi.org/10.1021/acs.jctc.6b00071.

(36)   Yu, T.-Q.; Lapelosa, M.; Vanden-Eijnden, E.; Abrams, C. F. Full Kinetics of CO Entry, Internal Diffusion, and Exit in Myoglobin from Transition-Path Theory Simulations. *J. Am. Chem. Soc.* **2015**, *137* (8), 3041–3050. https://doi.org/10.1021/ja512484q.

(37)   Ma, P.; Cardenas, A. E.; Chaudhari, M. I.; Elber, R.; Rempe, S. B. The Impact of
        Protonation on Early Translocation of Anthrax Lethal Factor: Kinetics from Molecular
        Dynamics Simulations and Milestoning Theory. *J. Am. Chem. Soc.* **2017**, jacs.7b07419.
        https://doi.org/10.1021/jacs.7b07419.

(38)   Kirmizialtin, S.; Nguyen, V.; Johnson, K. A.; Elber, R. How Conformational Dynamics of
        DNA Polymerase Select Correct Substrates: Experiments and Simulations. *Structure* **2012**,
        20 (4), 618–627. https://doi.org/10.1016/j.str.2012.02.018.

(39)   Tang, Z.; Chang, C. A. Binding Thermodynamics and Kinetics Calculations Using
        Chemical Host and Guest: A Comprehensive Picture of Molecular Recognition. *J. Chem.
        Theory Comput.* **2018**, *14* (1), 303–318. https://doi.org/10.1021/acs.jctc.7b00899.

(40)   Tang, Z.; Chen, S.-H.; Chang, C. A. Transient States and Barriers from Molecular
        Simulations and the Milestoning Theory: Kinetics in Ligand–Protein Recognition and
        Compound   Design.   *J.   Chem.   Theory   Comput.*   **2020**,   acs.jctc.9b01153.
        https://doi.org/10.1021/acs.jctc.9b01153.

(41)   Kokh, D. B.; Amaral, M.; Bomke, J.; Grädler, U.; Musil, D.; Buchstaller, H.-P.; Dreyer, M.
        K.; Frech, M.; Lowinski, M.; Vallee, F.; et al. Estimation of Drug-Target Residence Times
        by τ-Random Acceleration Molecular Dynamics Simulations. *J. Chem. Theory Comput.*
        2018, *14* (7), 3859–3869. https://doi.org/10.1021/acs.jctc.8b00230.

(42)   Teo, I.; Mayne, C. G.; Schulten, K.; Lelièvre, T. Adaptive Multilevel Splitting Method for
        Molecular Dynamics Calculation of Benzamidine-Trypsin Dissociation Time. *J. Chem.
        Theory Comput.* **2016**, acs.jctc.6b00277. https://doi.org/10.1021/acs.jctc.6b00277.

(43) Dickson, A.; Lotz, S. D. Ligand Release Pathways Obtained with WExplore: Residence Times and Mechanisms. *J. Phys. Chem. B* **2016**, *120* (24), 5377–5385. https://doi.org/10.1021/acs.jpcb.6b04012.

(44) Dickson, A.; Lotz, S. D. Multiple Ligand Unbinding Pathways and Ligand-Induced Destabilization Revealed by WExplore. *Biophys. J.* **2017**, *112* (4), 620–629. https://doi.org/10.1016/j.bpj.2017.01.006.

(45) Lotz, S. D.; Dickson, A. Unbiased Molecular Dynamics of 11 Min Timescale Drug Unbinding Reveals Transition State Stabilizing Interactions. *J. Am. Chem. Soc.* **2018**, jacs.7b08572. https://doi.org/10.1021/jacs.7b08572.

(46) Miao, Y.; Huang, Y. M.; Walker, R. C.; McCammon, J. A.; Chang, C. A. Ligand Binding Pathways and Conformational Transitions of the HIV Protease. *Biochemistry* **2018**, *57* (9), 1533–1541. https://doi.org/10.1021/acs.biochem.7b01248.

(47) Lonsdale, R.; Fort, R. M.; Rydberg, P.; Harvey, J. N.; Mulholland, A. J. Quantum Mechanics/Molecular Mechanics Modeling of Drug Metabolism: Mexiletine N-Hydroxylation by Cytochrome P450 1A2. *Chem. Res. Toxicol.* **2016**, *29* (6), 963–971. https://doi.org/10.1021/acs.chemrestox.5b00514.

(48) Haldar, S.; Comitani, F.; Saladino, G.; Woods, C.; Van Der Kamp, M. W.; Mulholland, A. J.; Gervasio, F. L.; Van Der Kamp # §, M. W.; †department, ‡ *. A Multiscale Simulation Approach to Modelling Drug-Protein Binding Kinetics A Multiscale Simulation Approach to Modelling Drug-Protein Bind-Ing Kinetics. *Just Accept. Manuscr.* • **2018**. https://doi.org/10.1021/acs.jctc.8b00687.

(49)   Zeller, F.; Luitz, M. P.; Bomblies, R.; Zacharias, M. Multiscale Simulation of Receptor–
        Drug Association Kinetics: Application to Neuraminidase Inhibitors. *J. Chem. Theory
        Comput.* **2017**, *13* (10), 5097–5105. https://doi.org/10.1021/acs.jctc.7b00631.

(50)   Aboelkassem, Y.; McCabe, K. J.; Huber, G. A.; Regnier, M.; McCammon, J. A.;
        McCulloch, A. D. A Stochastic Multiscale Model of Cardiac Thin Filament Activation
        Using Brownian-Langevin Dynamics. *Biophys. J.* **2019**, *117* (12), 2255–2272.
        https://doi.org/10.1016/j.bpj.2019.08.003.

(51)   Chang, C.-E. A.; Trylska, J.; Tozzini, V.; Andrew McCammon, J. Binding Pathways of
        Ligands to HIV-1 Protease: Coarse-Grained and Atomistic Simulations. *Chem. Biol. Drug
        Des.* **2007**, *69* (1), 5–13. https://doi.org/10.1111/j.1747-0285.2007.00464.x.

(52)   Jagger, B. R.; Kochanek, S. E.; Haldar, S.; Amaro, R. E.; Mulholland, A. J. Multiscale
        Simulation Approaches to Modeling Drug–Protein Binding. *Curr. Opin. Struct. Biol.* **2020**,
        61, 213–221. https://doi.org/10.1016/j.sbi.2020.01.014.

(53)   Amaro, R. E.; Mulholland, A. J. Bridging Biological and Chemical Complexity in the
        Search for Cures: Multiscale Methods in Drug Design. *Nat Rev Chem* **2018**, *2* (4), 0148.
        https://doi.org/10.1038/s41570-018-0148.

(54)   Votapka, L. W.; Amaro, R. E. Multiscale Estimation of Binding Kinetics Using Brownian
        Dynamics, Molecular Dynamics and Milestoning. *PLOS Comput. Biol.* **2015**, *11* (10),
        e1004381. https://doi.org/10.1371/journal.pcbi.1004381.

(55)   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. *J. Phys. Chem. B* **2017**, *121* (15), 3597–3606. https://doi.org/10.1021/acs.jpcb.6b09388.

(56) Jagger, B. R.; Lee, C. T.; Amaro, R. E. Quantitative Ranking of Ligand Binding Kinetics with a Multiscale Milestoning Simulation Approach. *J. Phys. Chem. Lett.* **2018**, *9* (17), 4941–4948. https://doi.org/10.1021/acs.jpclett.8b02047.

(57) Faradjian, A. K.; Elber, R. Computing Time Scales from Reaction Coordinates by Milestoning. *J. Chem. Phys.* **2004**, *120* (23), 10880–10889. https://doi.org/10.1063/1.1738640.

(58) Vanden-Eijnden, E.; Venturoli, M.; Ciccotti, G.; Elber, R. On the Assumptions Underlying Milestoning. *J. Chem. Phys.* **2008**, *129* (17), 174102. https://doi.org/10.1063/1.2996509.

(59) West, A. M. A.; Elber, R.; Shalloway, D. Extending Molecular Dynamics Time Scales with Milestoning: Example of Complex Kinetics in a Solvated Peptide. *J. Chem. Phys.* **2007**, *126* (14), 145104. https://doi.org/10.1063/1.2716389.

(60) Bello-Rivas, J. M.; Elber, R. Simulations of Thermodynamics and Kinetics on Rough Energy Landscapes with Milestoning. *J. Comput. Chem.* **2016**, *37* (6), 602–613. https://doi.org/10.1002/jcc.24039.

(61) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.* **2005**, *26* (16), 1781–1802. https://doi.org/10.1002/jcc.20289.

(62) Huber, G. A.; McCammon, J. A. Browndye: A Software Package for Brownian Dynamics. *Comput. Phys. Commun.* **2010**, *181* (11), 1896–1905.

https://doi.org/10.1016/j.cpc.2010.07.022.

(63)   Vanden-Eijnden, E.; Venturoli, M. Markovian Milestoning with Voronoi Tessellations. *J. Chem. Phys.* **2009**, *130* (19), 194101. https://doi.org/10.1063/1.3129843.

(64)   Maragliano, L.; Vanden-Eijnden, E.; Roux, B. Free Energy and Kinetics of Conformational Transitions from Voronoi Tessellated Milestoning with Restraining Potentials. *J. Chem. Theory Comput.* **2009**, *5* (10), 2589–2594. https://doi.org/10.1021/ct900279z.

(65)   Fiorin, G.; Klein, M. L.; Hénin, J. Using Collective Variables to Drive Molecular Dynamics Simulations. *Mol. Phys.* **2013**, *111* (22–23), 3345–3362. https://doi.org/10.1080/00268976.2013.813594.

(66)   Vanden-Eijnden, E.; Venturoli, M. Revisiting the Finite Temperature String Method for the Calculation of Reaction Tubes and Free Energies. *J. Chem. Phys.* **2009**, *130* (19). https://doi.org/10.1063/1.3130083.

(67)   Noé, F. Probability Distributions of Molecular Observables Computed from Markov Models. *J. Chem. Phys.* **2008**, *128* (24), 244103. https://doi.org/10.1063/1.2916718.

(68)   Nishikawa, S.; Fukahori, T.; Ishikawa, K. Ultrasonic Relaxations in Aqueous Solutions of Propionic Acid in the Presence and Absence of Beta-Cyclodextrin. *J. Phys. Chem. A* **2002**, *106* (12), 3029–3033. https://doi.org/10.1021/jp012065+.

(69)   Fukahori, T.; Nishikawa, S.; Yamaguchi, K. Kinetics on Isomeric Alcohols Recognition by α- and β-Cyclodextrins Using Ultrasonic Relaxation Method. *Bull. Chem. Soc. Jpn.* **2004**, *77* (12), 2193–2198. https://doi.org/10.1246/bcsj.77.2193.

(70)   Fukahori, T.; Kondo, M.; Nishikawa, S. Dynamic Study of Interaction between β-

Cyclodextrin and Aspirin by the Ultrasonic Relaxation Method. *J. Phys. Chem. B* **2006**, *110* (9), 4487–4491. https://doi.org/10.1021/jp058205n.

(71)   Rekharsky, M. V; Inoue, Y. Complexation Thermodynamics of Cyclodextrins. *Chem. Rev.* **1998**, *98* (5), 1875–1918. https://doi.org/10.1021/cr970015o.

(72)   Barros, T. C.; Stefaniak, K.; Holzwarth, J. F.; Bohne, C. Complexation of Naphthylethanols with β-Cyclodextrin. *J. Phys. Chem. A* **1998**, *102* (28), 5639–5651. https://doi.org/10.1021/jp9803844.

(73)   Berg, O. G.; von Hippel, P. H. Diffusion-Controlled Macromolecular Interactions. *Annual review of biophysics and biophysical chemistry*. June 1985, pp 131–160. https://doi.org/10.1146/annurev.biophys.14.1.131.

(74)   Guillain, F.; Thusius, D. The Use of Proflavin as an Indicator in Temperature-Jump Studies of the Binding of a Competitive Inhibitor to Trypsin. *J. Am. Chem. Soc.* **1970**, *92* (9), 5534–5536. https://doi.org/10.1021/ja00721a051.