

# SEEKR2: Versatile Multiscale Milestoning Utilizing the OpenMM 7.5 Molecular Dynamics Engine

*Lane W. Votapka, Andrew M. Stokely, Anupam A. Ojha, Rommie E. Amaro\**

University of California, San Diego, 9500 Gilman Dr. La Jolla, CA 92093

\*Email: [ramaro@ucsd.edu](mailto:ramaro@ucsd.edu)

**KEYWORDS** Markovian Milestoning with Voronoi Tessellations, Milestoning, OpenMM, NAMD, Browndye, Molecular Dynamics, Brownian Dynamics, Simulation Enabled Estimation of Kinetic Rates.

## **Abstract**

We present SEEKR2 (Simulation-Enabled Estimation of Kinetic Rates Version 2) – the latest iteration in the family of SEEKR programs for using multiscale simulation methods to computationally estimate the kinetics and thermodynamics of molecular processes, in particular, ligand-receptor binding. SEEKR2 generates equivalent, or improved, results compared to the earlier versions of SEEKR, but with significant increases in speed and capabilities. SEEKR2 has also been built with greater ease of usability with extensible features to enable future expansions of the method. Now, in addition to supporting simulations using NAMD, calculations may be run

with the fast and extensible OpenMM simulation engine. The Brownian dynamics portion of the calculation has also been upgraded to Browndye 2.

## **Introduction**

### ***Background***

The ability to computationally predict kinetics quantities, such as rate constants of reactions involving biomacromolecules, remains an active pursuit in computational and theoretical biophysics.<sup>1-9</sup> Many approaches rely on sampling possible reaction pathways using simulation methods such as molecular dynamics (MD)<sup>10-17</sup> and Brownian dynamics (BD),<sup>18-22</sup> however, the main challenge arises from the need to sample many MD simulation trajectories to obtain accurate predictions for important kinetics quantities, such as the  $k_{\text{off}}$ .<sup>23</sup> At present, the amount of brute force MD simulations required to obtain kinetics of ligand binding and unbinding remains intractable for most applications involving biologically relevant targets. Therefore, many clever approaches to avoid the cost of brute force MD simulations use a wide variety of schemes to expand the temporal and spatial reach available to the computational biophysics community to predict kinetics quantities. We, and others, have summarized these approaches elsewhere.<sup>24-36</sup>

SEEKR is one method we developed to utilize both MD and BD approaches such that we may exploit MD when explicit solvent and full molecular flexibility is required, but also exploit BD's speed when semi-rigid body molecules and implicit solvents will suffice.<sup>37-40</sup> SEEKR accomplishes this by partitioning the phase space of a system into smaller regions, then simulating trajectories within these regions using whichever is the most appropriate simulation approach, allowing each region to be simulated in parallel. The question of how to determine the best partitions of the MD and BD regions is still not completely understood. Minimally, the BD region should extend beyond the first, and probably second, solvation shell to minimize inaccuracies

caused by the implicit solvent. Within the solvation shells and the binding site itself, the explicit solvent and molecular flexibility of MD are likely required to obtain reasonable thermodynamic and kinetic quantities. In addition, by partitioning the phase space of molecular motion into smaller regions, one may ensure that events which are kinetically relevant but often rare, are adequately observed and characterized. The statistics obtained from short simulations in each of these smaller regions are then stitched together using milestoning theory.<sup>23,41-44</sup>

SEEKR performed well in predicting ligand-receptor kinetics,<sup>38,45</sup> and was mostly successful in rank-ordering the affinity and residence times of a series of ligands binding to a receptor.<sup>39,40</sup> The SEEKR approach was further augmented by utilizing a newer modification to milestoning theory, Markovian milestoning with Voronoi tessellations (MMVT).<sup>39</sup> MMVT-SEEKR performed comparably well to the classical milestoning approach used in the earlier versions, with some added benefits, including an increase in accuracy for some quantities and decreased computational cost. All previous versions of SEEKR, including the MMVT version, used the NAMD simulation software package for MD.<sup>46</sup>

Here, we present SEEKR2, which uses the OpenMM simulation software suite<sup>47</sup> for MD as an alternative to NAMD. OpenMM has enjoyed skyrocketing popularity among the scientific community because of its python interface, ease of extensibility, competitive performance on GPUs, and active development community. In addition, the design of OpenMM makes it relatively easy to develop independent plugins. Using SEEKR2, one may perform the MD portion of the SEEKR protocol in either OpenMM or NAMD, using either MMVT or the conventional milestoning method as developed by Elber and colleagues.<sup>23,41,42</sup> A number of auxiliary utility programs are also included to perform calculations using the typical SEEKR protocol.

## *Design and Implementation*

### **The SEEKR2 OpenMM Plugin**

The SEEKR2 OpenMM plugin design is based on OpenMM's own layered architecture; it contains: 1) a Python interface layer for easy interactions with a user, 2) CPU and GPU kernels which implement a number of integrators that implement the dynamics defined by MMVT or Elber milestone, and 3) a C++ API layer to connect the Python layer with the kernels.

A key aspect of the MMVT protocol is the definition of Voronoi cells. When a system reaches one of the boundaries of the cell, it collides against the boundary, and the identities and timescales of these collisions are logged for eventual analysis. SEEKR2 leverages the powerful custom mathematical expressions within the OpenMM package in order to define the locations of Voronoi cells and boundaries. By supplying a mathematical expression, the user can define the boundaries of a Voronoi cell, and when the system crosses it, the integrator logs the crossing information, the atomic positions and velocities are restored to the previous step, and velocities are reversed.

The mathematical expression for a given boundary defines a function, which can be any function of the system atomic positions. The functions for a cell (one function for each boundary) are defined by the user such that when the system is inside the cell, the boundary function is negative. But if the system ever crosses the boundary, the value of the boundary function is positive for that system configuration.

Such custom mathematical expressions make it straightforward to define Voronoi cells and boundaries in high-dimensional spaces; up to 31 possible boundaries may be defined per cell in the SEEKR2 OpenMM plugin. Examples are included along with the plugin that show how the collective variables defining anchor points can be distances, angles, or absolute positions between atoms or any groups of atoms and more.

An additional feature of the plugin allows users to optionally save OpenMM state objects whenever the boundary is crossed. These states can be used to analyze or visualize the locations of MMVT collisions, or as a starting set of atomic positions/velocities for simulations in adjacent Voronoi cells. Users may also optionally set the plugin to compute MMVT rate matrices, incubation time vectors, and other quantities needed in post-processing analysis, and then update them all to a file with each collision. There is negligible slowdown caused by this feature, and can be used to compute convergence and other quantities "on-the-fly".

### **The SEEKR2 Application Programming Interface (API)**

We include additional scripts and programs (called the API) for preparation, running, and analysis of the simulations for each cell. For preparation of the SEEKR2 calculation, we include a simple utility that prepares the file tree and files for a SEEKR2 calculation using concentric spherical milestones defined by the distance between the center-of-mass (COM) of atoms of a receptor binding site and the COM of a ligand molecule used in our first MMVT SEEKR paper.<sup>39</sup> For the convenient control of the SEEKR2 calculation, we include software that can be called with a single line that can run the simulations for any or all of the simulation regions. Lastly, we include an analysis package that will perform the milestoning calculations to obtain mean first passage times, rate constants, free energy profiles, and other quantities that are potentially desirable to obtain from an SEEKR2 calculation. Highly automated programs are also included to compute convergence

In addition to these utilities, a number of example calculation scripts are included, as well as comprehensive unit tests, documentation, and tutorials, all within a MolSSI cookiecutter, which we found at <https://github.com/MolSSI/cookiecutter-cms.git>.

## Results and Discussion

Benchmarking was performed to compare the performance of SEEKR2 against the original MMVT-SEEKR implementation in NAMD and a conventional OpenMM simulation of the same molecular system. Comparisons are listed in Table 1. For the trypsin-benzamidine system, the SEEKR2 OpenMM implementation performs almost 20 times faster than the MMVT-SEEKR code running on a GPU, which were re-run for the purposes of this study, and almost six times faster than the old MMVT-SEEKR running on a 68-core CPU node.<sup>39</sup>

Compared to a conventional OpenMM simulation (without SEEKR2), only a ~25% loss of speed was observed for these systems when the milestoning protocols were included in the SEEKR2 plugin.

To ensure that SEEKR correctly replicated the rate constants as predicted in the original MMVT-SEEKR implementation,<sup>39</sup> we repeated the simulations in a nearly identical process (details listed in the Materials and Methods) to obtain kinetic and thermodynamic quantities, which are reported in this section.

Several of our previous papers have run SEEKR calculations on the trypsin-benzamidine system. In Table 2, the rate constants for those calculations are listed alongside the values computed using the SEEKR2 program. SEEKR2 obtains a  $k_{\text{off}}$  and a  $k_{\text{on}}$  that are within an order of magnitude of the experimentally-measured quantities. Compared to previous versions of SEEKR, and compared to the experimental values, SEEKR2 obtains a  $k_{\text{off}}$  that is slightly too fast ( $990 \pm 70 \text{ s}^{-1}$  from SEEKR2 compared to the experimental  $600 \pm 300 \text{ s}^{-1}$ ). The  $k_{\text{on}}$  obtained by SEEKR2 is too fast relative to experiment ( $(3.40 \pm 0.05) \times 10^8 \text{ M}^{-1} \text{ s}^{-1}$  from SEEKR2 compared to the  $2.9 \times 10^7 \text{ M}^{-1} \text{ s}^{-1}$  from experiment), but nearly within an order of magnitude. Finally, the  $\Delta G_{\text{bind}}$  computed by SEEKR2 was off by a little more than 0.8 kcal/mol ( $-7.55 \pm 0.04 \text{ kcal/mol}$  from SEEKR2 compared

to  $-6.71 \pm 0.05$  kcal/mol from experiment. For the trypsin-benzamidine system, the total MD SMD simulation time was 20 ns, the total MD MMVT simulation time was 5  $\mu$ s, and two million BD trajectories were performed, in total. This is slightly more, though similar, to the 4.4  $\mu$ s of MD simulation time in our previous MMVT study.<sup>39</sup>

In addition to the trypsin-benzamidine system, previous SEEKR studies have focused on computing the kinetics of a so-called “host-guest” model system, composed of the  $\beta$ -cyclodextrin and a series of small organic molecules.<sup>40,48</sup> Using SEEKR2, as we did with MMVT SEEKR, we divided the space surrounding the  $\beta$ -cyclodextrin into twelve concentric spherical Voronoi cells (Figure 1) and recomputed  $k_{\text{off}}$  (Figure 2),  $k_{\text{on}}$  (Figure 3), and  $\Delta G_{\text{bind}}$  values (Figure 4) for the seven ligands mentioned in previous publications.<sup>39,40,48</sup> SEEKR2 now correctly ranks the compounds by  $k_{\text{off}}$ . SEEKR2 also represents a substantial improvement in the calculations of absolute  $k_{\text{on}}$ s for the host-guest system, although SEEKR2 does not correctly rank  $k_{\text{on}}$ s for all seven host-guest systems, which is difficult for any method since the host-guest  $k_{\text{on}}$ s differ by magnitudes that are relatively small compared to experimental margins of error. The  $\Delta G_{\text{bind}}$  values were computed with fairly similar accuracy to previous calculations, with the exception of aspirin, which showed an anomalous  $\Delta G_{\text{bind}}$  value. It is likely that the noise seen in the  $k_{\text{on}}$  and  $\Delta G_{\text{bind}}$  calculations are primarily caused by the concentric spherical milestone shapes used in this study, which may not adequately approximate isosurfaces of the committor function for the host-guest system, as would be required by exact milestone theory.<sup>23</sup> At this time, SEEKR2 only supports concentric spherical milestone shapes, but we plan to add support for many other milestone shapes in the near future, which will allow us to investigate whether other types of milestones will improve the accuracy of  $k_{\text{on}}$  calculations for the host-guest system. For each individual host-guest system, the cost of the SMD simulations to generate a starting structure was 110 ns, the MD MMVT ran for

700 ns each, as well as 110,000 BD trajectories per guest molecule. This is again slightly more simulation, though similar, to the ~560 ns total MD of our previous MMVT study.<sup>39</sup>

## **Conclusion**

In summary, SEEKR2 performs MMVT and conventional milestoning simulations with an extensible interface for milestone/Voronoi cell definitions, and performs the simulations using either a CPU platform, or, much more quickly and efficiently, using the GPU platform. At this time, only one-dimensional concentric spherical milestone shapes have been formally implemented, but more dimensions and other milestone shapes will be straightforward to implement. We've shown that SEEKR2, in general, performs better than earlier versions of SEEKR to recreate the kinetics and thermodynamics for two benchmark systems and contains many of the utilities that members of the biophysics community may find useful for their own milestoning calculations.

## **Materials and Methods**

The SEEKR2 calculations in this paper required both MD and BD simulations, which require different sorts of inputs. The analysis (calculation of kinetics and thermodynamics quantities, error analysis, and convergence analysis) were performed by SEEKR2. When possible, temperatures, salt concentrations, and protonation states were set to recreate experimental conditions as closely as possible. One exception to this was the MD simulations of the  $\beta$ -cyclodextrin “host” with the 1- or 2-naphylethanol “guest”, which had 0.5M  $\text{MnSO}_4$  dissolved in solution for the experiment.<sup>49</sup> Due to the inadequacy or lack of parameters for these divalent ions, we elected to use pure water in these MD simulations, as we did with the other host-guest systems, which is consistent with previous studies.<sup>48,50</sup> However, these salts were added to the BD simulations for the naphthylethanol molecules, since they were likely to significantly affect the rates of encounter. All

$\Delta G_{\text{bind}}$  values were computed using the formula  $\Delta G_{\text{bind}} = RT \ln(k_{\text{off}} / k_{\text{on}})$ , where  $R$  is the gas constant,  $T$  is the temperature, and  $\ln()$  is the natural logarithm function.

### ***Molecular dynamics simulations***

All MD were performed with OpenMM 7.5 using the SEEKR2 OpenMM Plugin. Simulations are initiated through the SEEKR2 API. MMVT simulation were performed according to the prescribed procedure.<sup>51</sup> All starting structures and parameters were reused from the previous SEEKR MMVT study, and the same collective variable definitions, site locations, and concentric spherical milestone shapes were used.<sup>39</sup>

### ***Brownian dynamics simulations***

All BD were performed using the Browndye 2 program.<sup>52</sup> As with OpenMM, all simulations were prepared and controlled through the SEEKR API

. Interior dielectrics were set to 4, while exterior dielectric constants were set to 78. All atomic positions, charges, and radii were reused from the previous SEEKR-MMVT study. The APBS program was used to compute electrostatic grids.<sup>53</sup> Desolvation forces and hydrodynamic interactions were enabled for all calculations, and other physical quantities (such as viscosity, solvent radius, etc.) were left at their defaults.

### ***Trypsin-Benzamidine system***

Simulations of the trypsin-benzamidine system were performed in an almost identical fashion to our previous study,<sup>39</sup> with all simulations performed at 298.15 K and, in the MD simulations, with rigid hydrogen-heavy atom bonds, and a nonbonded cutoff of 9 Å, and a timestep of 2 fs, using OpenMM. The OpenMM implementation allowed us to check for a collision every time step, instead of every 10 time steps of the previous implementation,<sup>39</sup> which likely improved calculation accuracy. We added some additional milestones to the trypsin-benzamidine system, such that

milestones were located at 1, 2, 3, 4, 6, 8, 10, 12, 14, 16, and 18 Å from the center of mass of the binding site. Starting structures within each Voronoi cell were generated from a steered molecular dynamics (SMD) simulation, where the system was started from a bound state configuration and pulled out to a site-ligand from a distance of 1 Å to a distance of 13 Å over the course of 20 ns of constant volume (NVT) MD using a moving harmonic restraint with a spring constant of 9000  $\text{kJ}\cdot\text{mol}^{-1}\cdot\text{nm}^{-2}$ . Upon examination, we determined that the original 14 Å milestones showed anomalous results in the BD, probably due to solvation shell effects and steric hindrances caused by the rigid body approach in BD. We added milestones beyond the original 14 Å to improve the calculation by moving the BD region beyond the solvation shells. Starting structures beyond the 14 Å milestone were extracted sequentially from the states generated at the moments of MMVT collisions against lower milestones. Using the generated starting structures for each Voronoi cell, a total of 500 ns of MMVT MD simulations per cell were performed using SEEK2. All collisions against the milestones were recorded for later analysis. For the BD simulations, the ligand was started at the b-surface and proceeded until it either escaped or satisfied the “reaction criteria” of touching the 18 Å radius milestone (b-surface stage). Then, among those that touched the 18 Å milestone, 1000 structures were extracted, and from each of these, 1000 independent BD simulations were run until the ligands either escaped or touched the 16 Å milestone (BD milestone stage). The purpose of the b-surface stage is to compute the rate of initial encounter with the outermost milestone, while the BD milestone stage computes transition probabilities used in the milestone model which are used, in combination with the initial encounter rate, to compute the  $k_{\text{on}}$ .

### ***Host-guest systems***

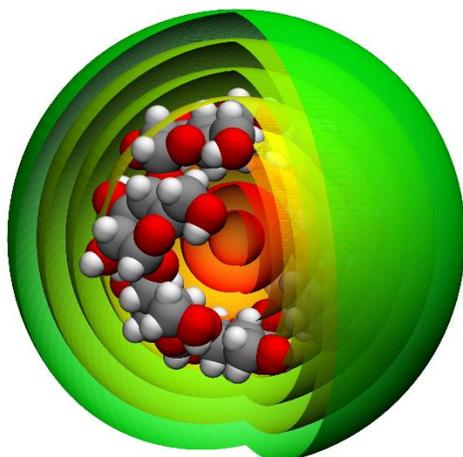
For the host-guest systems, all parameters and starting structures were identical to our previous SEEKR papers where we used the Q4MD force field.<sup>39,40</sup> However, in order to better recreate experimental conditions, a few adjustments to the host-guest system simulations are summarized in Table 3.

New SMD simulations were performed where the centers of masses (COMs) of “guest” ligands were restrained to 0.5 Å from the COM of the  $\beta$ -cyclodextrin “host” for 10 ns of constant pressure MD (NPT). Following this, the guest molecules were pulled by a moving harmonic restraint with a spring constant of 90000 kJ•mol<sup>-1</sup>•nm<sup>-2</sup> in an SMD simulation from the 0.5 Å starting location to a final COM-COM distance of 13.5 over the course of 100 ns of NVT MD. The purpose of the SMD simulations was to generate starting structures between each pair of milestones (near the center of each Voronoi cell). Using these starting structures for the MD simulations, MMVT simulations were run using SEEKR for 50 ns per Voronoi cell, which is equal to the maximum simulation length per anchor of the previous MMVT study we performed.<sup>39</sup> BD simulations of both the b-surface stage and the BD milestone stage were run to compute host-guest  $k_{onS}$ .

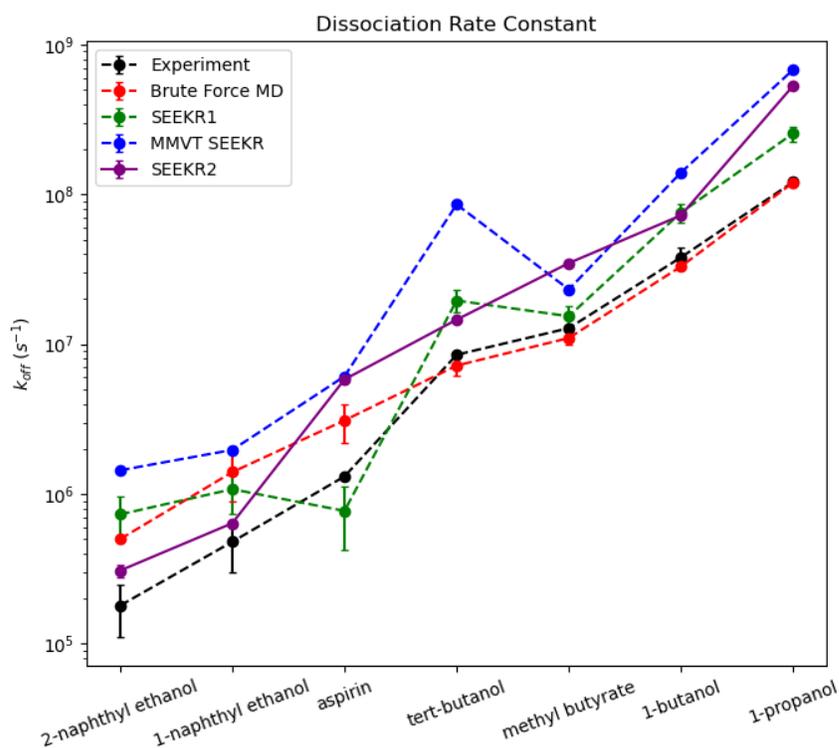
### **Data and Software Availability**

SEEKR2 can be found at <https://github.com/seekrcentral/seekr2.git>, where development is ongoing. Documentation for installation and usage, as well as tutorials, can also be found there. In order to run SEEKR2 with OpenMM, it is necessary to also install the SEEKR2 OpenMM plugin at [https://github.com/seekrcentral/seekr2\\_openmm\\_plugin.git](https://github.com/seekrcentral/seekr2_openmm_plugin.git). The data for this study can be obtained at [https://amarolab.ucsd.edu/files/seekr2\\_data.tgz](https://amarolab.ucsd.edu/files/seekr2_data.tgz).

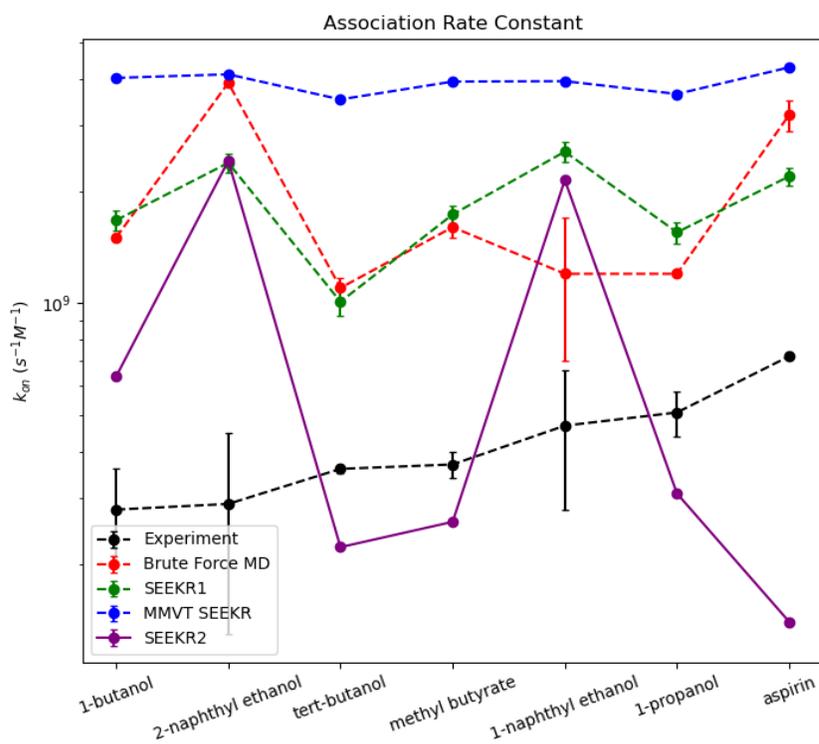
## **FIGURES**



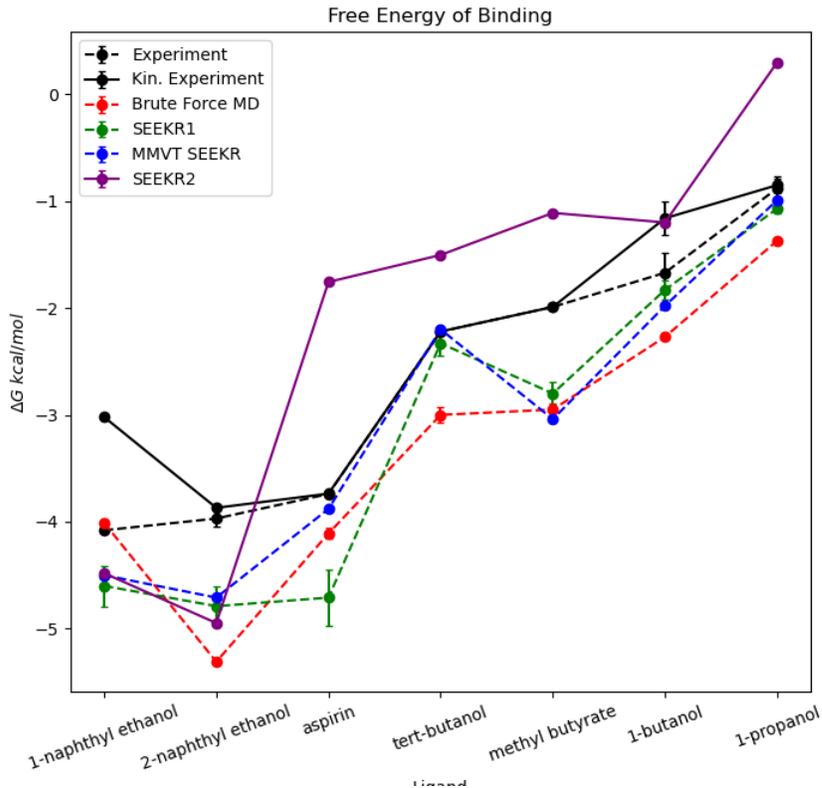
**Figure 1.** The space surrounding the  $\beta$ -cyclodextrin molecule is divided into twelve concentric spherical Voronoi tessellations (eight of the innermost ones are shown here), with boundaries that exist at 1 Å increments from 1 Å to 13 Å, and so on. OpenMM is used to run MMVT using MD within each of these cells, and trajectories collide against the boundaries between each cell, giving the transition times and statistics which are analyzed with milestoning theory.



**Figure 2.** The  $k_{\text{off}}$ s of each of the “guest” molecules to dissociate from the “host” molecule are listed. SEEKR2 performs comparably or better than brute force MD and other computational methods. SEEKR2 is also the only method (aside from brute force MD) that correctly ranks the “guest” compounds by  $k_{\text{off}}$  according to experiment. Error bars are present for the SEEKR2 data, but they are sometimes too small to see in this figure.



**Figure 3.** The  $k_{on}$ s of each “guest” compound as they bind to the “host” molecule are shown in order of increasing experimentally measured  $k_{on}$ . SEEKR2 performs the best of all methods for estimating absolute  $k_{on}$  values. No methods were able to correctly rank  $k_{on}$ s. This is likely due to the very small differences between experimentally measured  $k_{on}$ s. Error bars are present for the SEEKR2 data, but they are sometimes too small to see in this figure.



**Figure 4.** The  $\Delta G_{\text{bind}}$  of each “guest” compound when binding to the “host” molecule ranked from lowest to highest  $\Delta G_{\text{bind}}$ . Note that two experimental values may be listed – one where the  $\Delta G_{\text{bind}}$  was computed directly, and one where  $\Delta G_{\text{bind}}$  was computed from the experimentally-measured  $k_{\text{on}}$  and  $k_{\text{off}}$ . SEEKR2 correctly predicts compound ranking, with the exception of the first two compounds, which have very similar  $\Delta G_{\text{bind}}$ . Error bars are present for the SEEKR2 data, but they are sometimes too small to see in this figure.

#### TABLES

MD Engine	SEEKR version	Computing Resource	Performance (ns/day)
NAMD2.13	MMVT SEEKR	Expansive V100 GPU (1 CPU)	12

		Stampede CPU node (68 CPUs)	47
OpenMM7.5	SEEKR2	Expanse V100 GPU (1 CPU)	300
	Conventional	Expanse V100 GPU (1 CPU)	416

**Table 1.** Trypsin-benzamidine system performance (~23000 atoms).

Trypsin/ benzamidine	$k_{on}$ ( $M^{-1} s^{-1}$ )	$k_{off}$ ( $s^{-1}$ )	$\Delta G_{bind}$ (kcal/mol)
Experimental	$2.9 \times 10^7$	$600 \pm 300$	$-6.71 \pm 0.05$
SEEKR1 (2017) (ref)	$(2.1 \pm 0.3) \times 10^7$	$83 \pm 14$	$-7.4 \pm 0.1$
MMVT SEEKR (2020) (ref)	$(12.0 \pm 0.5) \times 10^7$	$174 \pm 9$	$-7.9 \pm 0.04$
SEEKR2 (2021)	$(3.40 \pm 0.02) \times 10^8$	$990 \pm 70$	$-7.55 \pm 0.04$

**Table 2.** Thermodynamics and Kinetics of binding results computed for the trypsin-benzamidine system in current and previous studies.

Guest Molecule	Experimental Study	Temperature (K)	Salt	pH
1-propanol	Fukahori et. al. 2004 <sup>54</sup>	298.15	pure water	7
1-butanol	Fukahori et. al. 2004 <sup>54</sup>	298.15	pure water	7
tert-butanol	Fukahori et. al. 2004 <sup>54</sup>	298.15	pure water	7
1-naphthyl-ethanol	Barros et. al. 1998 <sup>49</sup>	293.15	0.5 M MnSO <sub>4</sub>	7
2-naphthyl-ethanol	Barros et. al 1998 <sup>49</sup>	293.15	0.5 M MnSO <sub>4</sub>	7
methyl butyrate	Nishikawa et. al. 2002 <sup>55</sup>	298.15	pure water	7
aspirin (protonated)	Fukahori et. al. 2006 <sup>56</sup>	298.15	pure water	1.7

**Table 3.** Experimental conditions when measuring the kinetics of binding/unbinding for the host-guest system. We attempted to recreate these conditions as closely as possible in our simulations/calculations.

#### ASSOCIATED CONTENT

The Supporting Information is available free of charge.

SEEKR2 can be found at <https://github.com/seekrcentral/seekr2.git>. The SEEKR2 OpenMM plugin can be found at [https://github.com/seekrcentral/seekr2\\_openmm\\_plugin.git](https://github.com/seekrcentral/seekr2_openmm_plugin.git). The data for this study can be obtained at [https://amarolab.ucsd.edu/files/seekr2\\_data.tgz](https://amarolab.ucsd.edu/files/seekr2_data.tgz).

#### AUTHOR INFORMATION

##### Corresponding Author

Rommie E. Amaro – Department of Chemistry and Biochemistry, University of California San Diego, La Jolla, California 92093, United States; orcid.org/0000-0002-9275-9553; Email: 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. LWV and AMS coded the SEEKR2 API and OpenMM plugin. LWV ran the calculations, wrote the text of the manuscript, and made tables and some figures. AMS contributed to the writing, editing, and some figures. AAO tested the SEEKR2

code and assisted with bug fixes. REA is the principal investigator of the project, providing the resources, funding, environment, and guidance on the project, as well as some of the manuscript writing and editing.

## **Funding Sources**

This work was funded in part by the National Science Foundation through XSEDE supercomputing resources provided via TG-CHE060073 to R.E.A.

## **Notes**

The authors declare no competing financial interest.

## ACKNOWLEDGMENT

We thank J. Andrew McCammon and Gary Huber for insightful and helpful discussions. We also thank Benjamin Jagger for providing useful content, code, advice, and feedback on the manuscript. We also thank Jeffrey Wagner for assistance with the Github repository, and Ilker Deveci and Hilliary Frank for assistance with documentation.

## ABBREVIATIONS

SEKR, simulation enabled estimation of kinetic rates; MMVT Markovian milestoning with Voronoi Tessellations; MD molecular dynamics, BD Brownian dynamics; COM center of mass; NVT constant number of particles, volume, and temperature simulation; NPT constant number of particles, pressure, and temperature simulation; API application programming interface.

## 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.; Segala, E.; Amaral, M.; Guo, D.; Andres, D.; Georgi, V.; Stoddart, L. A.; Hill, S.; Cooke, R. M.; De Graaf, C.; Leurs, R.; Frech, M.; Wade, R. C.; de Lange, E. C. M.; IJzerman, A. P.; Müller-Fahrnow, A.; Ecker, G. F. 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. In *Thermodynamics and Kinetics of Drug Binding*; John Wiley & Sons, Ltd, 2015; pp 211–235. <https://doi.org/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) 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>.

(11) 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.; Moraes, M. A.; Dror, R. O.; Piana, S.; Shan, Y.; Towles, B.; Salmon, J. K.; Grossman, J. P.; Mackenzie, K. M.; Bank, J. A.; Young, C.; Deneroff, M. M.; Batson, B. 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>.

(12) 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.; Forte, A.; Gagliardo, J.; Gill, G.; Greskamp, B.; Ho, C. R.; Ierardi, D. J.; Iserovich, L.; Kuskin, J. S.; Larson, R. H.; Layman, T.; Lee, L. S.; Lerer, A. K.; Li, C.; Killebrew, D.; Mackenzie, K. M.; Mok, S. Y. H.; Moraes, M. A.; Mueller, R.; Nociolo, L. J.; Peticolas, J. L.; Quan, T.; Ramot, D.; Salmon, J. K.; Scarpazza, D. P.; Ben Schafer, U.; Siddique, N.; Snyder, C. W.; Spengler, J.; Tang, P. T. P.; Theobald, M.; Toma, H.; Towles, B.; Vitale, B.; Wang, S. C.; Young, C. 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>.

(13) 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>.

(14) 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>.

(15) 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>.

(16) 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>.

(17) 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>.

(18) 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>.

- (19) 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>.
- (20) 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>.
- (21) 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>.
- (22) 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>.
- (23) Bello-Rivas, J. M.; Elber, R. Exact Milestoning. *J. Chem. Phys.* **2015**, *142* (9), 94102. <https://doi.org/10.1063/1.4913399>.
- (24) Elber, R. A New Paradigm for Atomically Detailed Simulations of Kinetics in Biophysical Systems. *Q. Rev. Biophys.* **2017**, *50*, e8. [https://doi.org/DOI: 10.1017/S0033583517000063](https://doi.org/DOI:10.1017/S0033583517000063).
- (25) Doerr, S.; de Fabritiis, G. On-the-Fly Learning and Sampling of Ligand Binding by High-Throughput Molecular Simulations. *J. Chem. Theory Comput.* **2014**, *10* (5), 2064–2069. [https://doi.org/doi: 10.1021/ct400919u](https://doi.org/doi:10.1021/ct400919u).

(26) 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>.

(27) 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>.

(28) 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>.

(29) 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>.

(30) 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>.

(31) 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>.

(32) Gobbo, D.; Piretti, V.; Di Martino, R. M. C.; Tripathi, S. K.; Giabbai, B.; Storici, P.; Demitri, N.; Giroto, 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) 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>.

(34) Plattner, N.; Noé, F. Protein Conformational Plasticity and Complex Ligand-Binding Kinetics Explored by Atomistic Simulations and Markov Models. *Nat. Commun.* **2015**, *6* (May), 7653. <https://doi.org/10.1038/ncomms8653>.

(35) 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>.

(36) Mollica, L.; Theret, I.; Antoine, M.; Perron-Sierra, F.; Charton, Y.; Fourquez, J.-M.; Wierzbicki, M.; Boutin, J. A.; Ferry, G.; Decherchi, S.; Bottegoni, G.; Ducrot, P.; Cavalli, A. 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>.

(37) Votapka, L. W.; Lee, C. T.; Amaro, R. E. Two Relations to Estimate Membrane Permeability Using Milestoning. *J. Phys. Chem. B* **2016**, *120* (33), 8606–8616. <https://doi.org/10.1021/acs.jpcc.6b02814>.

(38) 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.jpcc.6b09388>.

(39) Jagger, B. R.; Ojha, A. A.; Amaro, R. E. Predicting Ligand Binding Kinetics Using a Markovian Milestoning with Voronoi Tessellations Multiscale Approach. *J. Chem. Theory Comput.* **2020**. <https://doi.org/10.1021/acs.jctc.0c00495>.

(40) 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.jpcclett.8b02047>.

(41) 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>.

(42) 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>.

(43) 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>.

(44) 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>.

(45) 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>.

(46) 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>.

(47) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L.-P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D.; Wiewiora, R. P.; Brooks, B. R.; Pande, V. S. OpenMM 7: Rapid Development of High Performance Algorithms for Molecular Dynamics. *PLOS Comput. Biol.* **2017**, *13* (7), e1005659. <https://doi.org/10.1371/journal.pcbi.1005659>.

(48) 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>.

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

(50) Fenley, A. T.; Henriksen, N. M.; Muddana, H. S.; Gilson, M. K. Bridging Calorimetry and Simulation through Precise Calculations of Cucurbituril–Guest Binding Enthalpies. *J. Chem. Theory Comput.* **2014**, *10* (9), 4069–4078. <https://doi.org/10.1021/ct5004109>.

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

(52) Huber, G.; McCammon, J. A. Brownian Dynamics Simulations of Biological Molecules. *Trends Chem.* **2019**, *1* (8), 727–738.

(53) Jurrus, E.; Engel, D.; Star, K.; Monson, K.; Brandi, J.; Felberg, L. E.; Brookes, D. H.; Wilson, L.; Chen, J.; Liles, K.; Chun, M.; Li, P.; Gohara, D. W.; Dolinsky, T.; Konecny, R.; Koes, D. R.; Nielsen, J. E.; Head-Gordon, T.; Geng, W.; Krasny, R.; Wei, G.-W.; Holst, M. J.; McCammon, J. A.; Baker, N. A. Improvements to the APBS Biomolecular Solvation Software Suite. *Protein Sci.* **2018**, *27* (1), 112–128. <https://doi.org/https://doi.org/10.1002/pro.3280>.

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

(55) 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+>.

(56) Fukahori, T.; Kondo, M.; Nishikawa, S. Dynamic Study of Interaction between  $\beta$ -Cyclodextrin and Aspirin by the Ultrasonic Relaxation Method. *J. Phys. Chem. B* **2006**, *110* (9), 4487–4491. <https://doi.org/10.1021/jp058205n>.