

WEIGHTED ENSEMBLE ORIENTED METHOD DEVELOPMENT FOR INCREASED
EFFICIENCY OF DRUG DESIGN

By

Nicole Marie Roussey

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Biochemistry and Molecular Biology – Doctor of Philosophy

2022

ABSTRACT

In drug design, there are many biologically relevant values of interest such as (un)binding rates and free energies. Recent works show that values such as the unbinding rate, k_{off} , may be equally as important to the drug design process as the free energy, ΔG . This is due the relationship between k_{off} and the residence time (RT) for a given drug, as RT can be a predictor of drug efficacy. In order to determine k_{off} , information about the transition path ensemble is required, but the instability and short lifespan of this state makes it difficult to gain atomic level insights with experimental methods. Many computational methods oriented around Molecular dynamics (MD) have been developed to predict these values. However, events of interest, such as binding and unbinding events necessary to compute rates and energies, are rare and often occur on prohibitively long timescales ranging from milliseconds to hours. This is problematic as most straight-forward MD simulations are computationally limited to the microsecond timescale. Enhanced sampling algorithms such as the path sampling method Weighted Ensemble (WE) permit the simulation of these pathways in less computational time, through the resampling process of merging away redundant trajectories in an ensemble, and cloning trajectories of interest. In this thesis, a new resampling algorithm that is based on the Jarzynski Equality (JE) for use with the WE method is developed that allows the use of short, efficient, nonequilibrium simulations to predict (un)binding free energies with a Lennard-Jones pair. This method is found to be generally inefficient with larger, more protein-ligand like test systems called Statistical Assessment of Modeling of Proteins and Ligands (SAMPL) systems (host-guest pairs). In order to better this method, solvent-based features of these systems are studied in detail, and it is determined that high-probability events of interest have very little guest-ion interaction. This is of great interest due to the prevalence of SAMPL systems in method development. It is also found that there is a physical point in the unbinding path for these systems that represents a "point of no return", or a commitment to unbinding point. With this in mind, a new resampling algorithm based off of the preexisting Resampling of Ensembles by Variation Optimization (REVO)

algorithm is developed. This prevents cloning operations from occurring on trajectories that have surpassed the commitment to unbinding point. This allows the probabilities of these trajectories to remain high when the target state is reached and leaves room open in the ensemble for the exploration of other pathways. This new resampler called "cutoff-REVO" produces far fewer, but much higher probability unbinding events than its predecessor and more accurately and consistently predicts both ΔG and k_{off} for four systems of interest. Overall, this work has provided insights into systems of interest and new means of obtaining information essential to the drug design process.

Copyright by
NICOLE MARIE ROUSSEY
2022

ACKNOWLEDGEMENTS

There are many people and institutions I would like to acknowledge as having been invaluable contributors to both my degree as well as my pursuit of it. Firstly, I would like to thank my advisor, Dr. Alex Dickson. Your kindness, understanding, and excellent mentorship have never gone unnoticed or unappreciated. Thank you for always pushing me to learn and grow as a scientist and thank you for taking a leap of faith and allowing someone with no computational experience to work for you - it has really meant so much to me. Secondly, I would like to thank all of the Dickson lab members, both past and current; working with all of you has been an incredible experience for which I am extremely grateful.

I would also like to thank the members of my committee for their insight into my research, as well as the Department of Biochemistry and Molecular Biology for the opportunity to further my education, expand my scientific horizons, and foremost for the opportunity to pursue this degree. Thank you to Oakland University and the wonderful research team at the Eye Research Institute for giving me my first research experience. Thank you to Dr. Andrew Goldberg for opening your lab to sophomore with no research experience and providing the mentorship that has set the foundation for the scientist I am today.

I would like to thank my friends for all of the fun and all of the support provided over the years. To Gina, your quick wit and crude humor have provided many highly necessary laughs. Here's to you, Dr. Shoemaker. To Nazanin and Tom, for always being there to help each other and for getting through grad school together. Thank you to my family for always being just one (inevitably long, loud) phone call away. Thank you to my father, Brian, and Linda, for your never ending support and encouragement of my education and helping me become the person I am today. Thank you to my mother, Jill, and Mike, for always being there for me. To my brother Jon and his family, and in an homage to Jon's dissertation, thank you for always being willing to share in the family gossip and being able to share in a unique sense of humor. To my nephews Jeremy, Cameron, Milo and Nolan, thank you for reminding me to never lose my sense of imagination. To my grandfather, Al, for always

having a crisp five dollar bill ready to give to his grand-kids for good grades, even when they were in grad school. You never failed to make me smile. Although they'll never read this, thank you to all the wonderful pets I have had in my life over the years, Maggie, Sadie, Simba, Max and Luna. Your wagging tails and exasperated meows have always brought a comforting sense of joy when it was needed most.

Lastly, I would like to thank my husband Timothé. Your love and support has helped me get through grad school more than you'll ever know. Je t'aime énormément et de tout mon cœur, et franchement, je pense que sans toi, ça aurait été très difficile d'être diplômée.

TABLE OF CONTENTS

CHAPTER 1 INTRODUCTION	1
1.1 The Significance of Ligand (Un)binding Kinetics in Drug Design . .	1
1.2 In-Silico Methods for the Determination of Kinetics and Free Energies	3
1.3 The Jarzynski Equality (The Jarzynski Nonequilibrium Work Theorem)	16
1.4 SAMPL Systems	18
1.5 Outline of Work	19
CHAPTER 2 ENHANCED JARZYNSKI FREE ENERGY CALCULATIONS USING WEIGHTED ENSEMBLE	23
2.1 Introduction	23
2.2 Methods	27
2.3 Results	35
2.4 Discussion	44
CHAPTER 3 LOCAL ION DENSITIES CAN INFLUENCE TRANSITION PATHS OF MOLECULAR BINDING	48
3.1 Introduction	48
3.2 Materials and Methods	50
3.3 Results	53
3.4 Discussion	59
CHAPTER 4 QUALITY OVER QUANTITY: SAMPLING HIGH PROBABILITY RARE EVENTS WITH THE WEIGHTED ENSEMBLE ALGORITHM	61
4.1 Introduction	61
4.2 Methods	65
4.3 Results	72
4.4 Discussion	79
4.5 Conclusions	82
4.6 Acknowledgements	82
CHAPTER 5 SUMMARY, IMPACTS, & OUTLOOK	83
BIBLIOGRAPHY	88
APPENDIX A: ENHANCED JARZYNSKI FREE ENERGY CALCULATIONS USING WEIGHTED ENSEMBLE	104
APPENDIX B: LOCAL ION DENSITIES CAN INFLUENCE TRANSITION PATHS OF MOLECULAR BINDING	107

APPENDIX C: QUALITY OVER QUANTITY: SAMPLING HIGH
PROBABILITY RARE EVENTS WITH THE WEIGHTED
ENSEMBLE ALGORITHM 112

CHAPTER 1

INTRODUCTION

1.1 The Significance of Ligand (Un)binding Kinetics in Drug Design

In drug design, ligand efficacy is often predicted by measuring thermodynamic values for a molecule of interest. The significance of one of these values, free energy (ΔG), for biological systems of interest has long since been established. Free energy is the key thermodynamic quantity that determines the reversibility of a process and is used to describe many biological functions of interest, including but not limited to protein-protein interactions and conformational changes, permeability, and ligand (un)binding[1]. The determination of binding affinity for a molecule-pair has been used in the design and understanding of drugs in many cases [2, 3, 4] through means often done under equilibrium conditions.

While these methods may be efficient, the body lacks a true equilibrium as it works to maintain homeostasis. This can be seen in the case of a drug being administered: while some of the drug is still being absorbed and distributed throughout the body, the body is also working to metabolize and excrete the drug. To compensate for this, it has been suggested that the “residence time” (RT) which is equal to $\frac{1}{k_{off}}$ (where k_{off} is the off rate), or the amount of time it takes the ligand to unbind, should be a key feature considered during the optimization process, a methodology referred to as “kinetics oriented drug design”[5, 6, 7, 8, 9]. This is because of how RT often correlates linearly with drug efficacy (Figure 1.1A). An example of such a relationship can be seen in Lu 2009[10, 11] where the relationship between the RT of an antibiotic (specifically a FabI enoyl-reductase inhibitor) and drug efficacy or survival of the model organism used is observed to be nearly linear. It is important to note that in some cases while this relationship is linear, the relationship with binding affinity (and potentially efficacy) may not be. An example of these non-linear affinity and linear RT relationships can be seen in Costa, 2016[12] with the Translocator Protein (TSPO) and various ligands. If a drug has high affinity but a very low RT, it is possible for efficacy to be low, as the drug will frequently unbind making it regularly susceptible to being metabolized

or excreted.

For a drug to be effective, a minimum concentration that varies on a drug-by-drug basis must be reached and maintained. This value is the “minimum effective concentration”, whereas the “maximum effective concentration (minimum toxic concentration)” is the highest concentration possible prior to the onset of toxic effects from the drug[13] (Figure 1.1B). The issue of toxicity from a drug can arise when the drug induces off target effects through the interaction with molecules other than the intended target, or when a drug is administered orally or through injection and absorption speeds cannot be easily controlled[13]. Between the minimum and maximum effective concentrations exists the “therapeutic window”, where a drug can work effectively within the body[14].

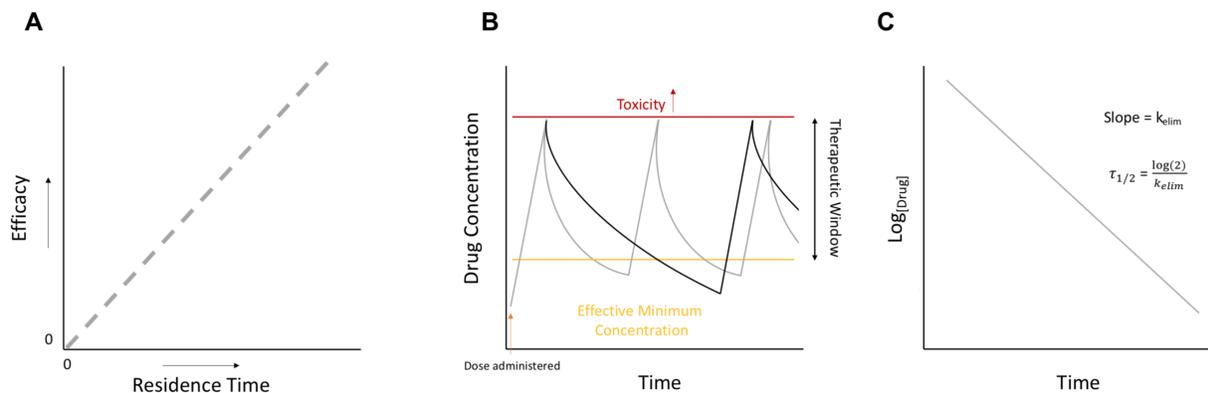


Figure 1.1: Essential Drug Characteristics in Drug design A) The relationship between the RT and efficacy of a drug. B) The amount of drug in the blood changing with time for a short RT drug (gray) and a long RT drug (black). The minimum concentration required for an effect is shown as a yellow line, and the maximum, nontoxic concentration is shown in red. C) The log of drug concentration over time following peak concentration. The slope of C is the elimination rate of the drug.

It is of significant interest to both increase the time the drug concentration remains within the therapeutic window as well as to increase the overall size of that window. One of the key means of increasing the length of time a drug is present at a concentration within this window is by increasing the drug’s half-life ($\tau_{1/2}$) within the body. The half-life represents the time it takes for half of the drug to be metabolized or excreted from the body. The longer the drug’s half-life, the lower the frequency with which it will need to be administered. The half-life of a drug within the body can be calculated by first determining the elimination rate, k_{elim}

(Figure 1.1 C). This is done by finding the concentration of the drug in the blood at various time points following the maximum concentration being met and determining the slope of this plot. $\tau_{1/2}$ can then be found as $\frac{\log(2)}{k_{elim}}$ (or $\frac{\ln(2)}{k_{off}}$)[15]. The half-life can be increased by optimizing the ligand in such a way that its k_{off} is altered. This can be done through ligand modifications that increase affinity for a receptor to such an extent that pharmacokinetic values change[16, 17].

These drug modifications should aim to lower the k_{off} and subsequently increase the drug's RT. This helps make the drug less vulnerable to elimination as it will spend more time bound to the target. If the RT can be increased to a time that is longer than the drug's half-life, it is possible to maintain drug action within the body after much of the drug has been expelled[6], reducing the need for regular administration of the drug.

Due to the significance of RT in ligand behavior, there is growing interest in understanding and optimizing ligand kinetics in drug design. This method adds another option to the existing tools for drug designers to use as new tools can permit breakthroughs in druggability [18]which may be particularly beneficial as only $\sim 10\%$ of the human genome has thus far been found to be druggable[19].

1.2 In-Silico Methods for the Determination of Kinetics and Free Energies

There are many methods available experimentally to determine values of interest for both drug design and numerous other biological motions of interest, such as fluorescence assays[20, 21, 22] and radio-ligand binding[23, 24, 25]. Despite the advances made in biochemistry techniques for the determination of rates and free energies, we are limited by the loss of knowledge of the mechanism or pathway taken in the unbinding process at the atomistic level.

Currently, experimental characterization of points of interest in the unbinding pathway such as the transition state ensemble is not readily doable due to the incredibly short lifespan of this state. Other than protein-ligand structural information, experimental methods also

do not provide information on the various other factors that can influence the unbinding process, such as fluctuations in solvent microstates. The fluctuations would include factors such as ion densities around the protein or the ligand, varying densities of waters within the binding pocket, amongst other factors. These fluctuations are known to have an impact on the (un)binding process [26] for small molecules, and in the future may become a factor for consideration in the drug design process. The importance of structural information in the drug design process has necessitated the development of various computational methods to simulate the (un)binding process. In this thesis, a method called “molecular dynamics” is used. This method permits the simulation of numerous biomolecules including proteins, DNA, RNA, lipids, carbohydrates, etc., at atomic resolution, thus allowing us to observe (un)binding pathways and many other processes of interest.

Molecular Dynamics

MD is a highly powerful computational algorithm that is used for the simulation of atomic movement and interactions at the atomistic level[27, 28]. The original MD algorithm was first developed in 1957[29] with the use of hard-sphere representations of atoms. Since its development, major advances have happened in both hardware and software available for running MD simulations that have permitted the study of far more complex and biologically relevant systems. This method was selected for use here, as it now permits the study and generation of ensembles of long timescale events such as ligand (un)binding and unstable states of interest (such as transition states) at atomic resolution in feasible timelines[28].

In recent years, MD has been used for the study of numerous biological motions and actions of interest, including but not limited to, the opening and closing of the SARS-CoV-2 spike protein[30], unbinding of ligands from proteins[9, 31, 32], DNA-ion interactions[33], rRNA-ion interaction[34], and large, charged systems such as water-model dependent diffusion through biological membranes[35, 36].

A brief explanation of a general MD simulation is as follows. Initially, the positions and velocities are first set for all atoms in the system, often set to match the temperature of the

simulation. Using this information, the force on each atom is then determined via:

$$\mathbf{F}(t) = -\nabla U(\mathbf{\Gamma}(t)), \tag{1.1}$$

in which U is the potential energy at time t for all atomic positions in set $\mathbf{\Gamma}$ (a trajectory at time t).

Often, U is reported in $\frac{kcal}{mol}$, the force is in $kcal/mol\text{\AA}$ (or nm), the respective positions are presented in either nm or \AA with this unit often depending on the MD engine used for the simulation, and the time is frequently reported in femtoseconds ($10^{-15}s$).

Force Fields

The specifics of the potential energy function used above are dependent on the choice of another essential algorithm called a “force field” that is used. Numerous force field options exist, with some of the most commonly used ones including the AMBER[37, 38] force fields and the CHARMM[39, 40] force fields, with the CHARMM36m force field being used for the work done in this thesis. Generally speaking, force field calculations of forces on an atom include terms that include electrostatic and Lennard-Jones energies (non-bonded interactions) as well as bonds, angles, dihedral angles and Urey-Bradley terms (bonded interactions). Following the determination of the force on each atom, Newton’s equations of motion can then be used to determine the corresponding change in positions and velocities (\mathbf{v} , in \AA or nm per femtosecond) for each atom in space for the respective time step. This process is repeated N times per cycle of dynamics, where N is the desired number of dynamic steps per cycle.

Integration

The goal of force fields is often to maintain a constant energy in the simulation as an NVE or NPE ensemble (N = constant number of atoms, E = constant energy, and V and P = either constant volume or pressure). However, it is common practice for simulations of biological systems to be done at a constant temperature T (as an NPT or NVT ensemble) as in done in this work. This is done to better compare with experiments at constant temperature. In order to maintain a constant temperate in a simulation, it is necessary to use a “thermostat”.

In the work done in this thesis, the Langevin Integrator that uses Langevin dynamics is utilized[41].

In Langevin dynamics, the force on each atom is calculated with a modified form of the equation shown above, where force is solved as:

$$\mathbf{F}(t) = -\nabla U(\mathbf{\Gamma}(t)) - \gamma m \mathbf{v}(t) + \sqrt{2m\gamma k_b T} \mathbf{R}(t), \quad (1.2)$$

in which T is the temperature of the system in Kelvin, the Boltzmann constant is represented as k_b in $\frac{kcal}{molK}$, γ , a value most often set to 1, is a friction coefficient for the system in ps^{-1} , and the $\mathbf{R}(t)$ variable is a 0 centered Gaussian distribution used for the addition of randomness to the system. In this equation, the first term is the standard force term. The added second and third term (whose sum maintains temperature) are as follows: the second term removes energy from the system as friction, and the third term adds randomness to the system as represented by thermal fluctuations[42] through the addition of energy back into the simulation.

Overall, the MD method is an incredibly powerful tool that can be utilized to learn about a vast array of biological activities and systems of interest. Despite this, MD does come with one major limitation. The timescales possible to simulate at atomistic levels are quite short due to these simulation being performed with very small time steps; whereas biological events such as ligand (un)binding often occur on the millisecond to minute timescale. That being said, significant advancements in both hardware and software have greatly increased both what systems can be simulated and the sampling range possible. Hardware advancements range from increased graphics processing unit (GPU) power (allowing for the simulation of up to hundreds of nanoseconds per day [32, 9]) to the development of supercomputers dedicated to MD simulations (such as D.E. Shaw's Anton series of computers[43]) capable of even greater simulation power. Despite these incredible improvements in computing power since the conception of the original MD algorithm, there are still limitations on timescales that can be simulated. Consequently, there is a need for the development of software that

can greatly increase simulation timescales and sufficiently utilize modern hardware.

Enhanced Sampling and Other Simulation Techniques

While the MD method is able to simulate (un)binding events and many other biological events of interest, there is still a limitation presented by the timescales of biological motions of interest. Many natural phenomena of interest occur on timescales (seconds to minutes) longer than currently possible to simulation with standard brute force dynamics. To overcome this limitation, many “enhanced sampling” and alchemical algorithms have been developed to either simulate these long time scale events or to more efficiently determine values of interest such as ΔG .

There are many types of algorithms and simulation techniques that have been developed, including alchemical methods[1], altered potential energy methods[44, 45], and trajectory parallelization methods[46]. These methods present us with a means of gaining information about long timescale processes through the use of short-timescale simulations that can feasibly be run on modern computers.

Alchemical Methods

Taking a brief turn away from focusing on rates, there are numerous simulation types that take advantage of unphysical events to quickly obtain free energies of transitions often without the simulation of complete paths. In these methods, it is common for the (un)binding process to not be directly simulated. Hence the name ”alchemical“, often a ligand is transformed into either another molecule or a non-interacting entity through an arbitrary intermediate [47]. These methods are beneficial in that they are often computationally inexpensive when compared to other methods[47] that explicitly simulate reactive paths. There are numerous simulation techniques that take advantage of these methods, including but not limited to free energy perturbation (FEP) and nonequilibrium work (NEW) methods [1].

The concept of FEP calculations, comes from the identity $e^{-\beta\Delta A} = \langle e^{-\beta\Delta U} \rangle_0^{eq}$ [48] where

$\Delta U \equiv U(\mathbf{\Gamma}, \lambda_1) - U(\mathbf{\Gamma}, \lambda_0)$ [1] where the angled brackets represent an equilibrium ensemble average and $\beta = 1/k_B T$ (where T is the temperature the process is performed at, and k_B is the Boltzmann constant). λ_n in this case represents the state of the system where λ_0 is used to signify an initial state, and λ_1 a final state. λ is a malleable variable and can represent, for example, a system beginning with side chain A at λ_0 and ending with side chain B at λ_1 .

FEP determines the difference in free energy between two states, A and B through mutating the system from one state to another through unphysical intermediates. This can be done to determine relative binding free energies by morphing one ligand into another or absolute binding free energies by eliminating ligands from a binding pocket[1]. While this is a powerful tool, FEP calculations are generally limited to mutations between ligands with the same charge, and only converge well for systems with small differences[1].

NEW methods rely on driving systems away from equilibrium. The identity for this method is the JE, $\langle e^{-\beta W} \rangle = e^{-\beta \Delta F}$, which is explained further in Section 1.3. In this method, systems begin in some state λ_0 and end in λ_1 , and the work (W) applied is the work required to move the system from one value of λ to the next. This can be measured as the work required to update forces on the system. Bidirectional forms of both of these methods exist and can prove highly beneficial for determining accurate and converged free energies. For NEW, the bidirectional starting point is the Crooks Fluctuation Theorem[1, 49].

While these methods are widely used for varying systems[1, 50, 51, 52], some sources of error for alchemical methods include residual charges from the vanishing atoms in single topology methods and the overlap of groups in a system in double topology models[1]. Alchemical methods are also limited in that they take advantage of ΔG being a state function, in that only the initial and final states matter. Therefore, these methods only determine the free energy differences between the points of interest for a transition, and subsequently do not produce free energy profiles for the unbinding/transition pathway of interest nor do they permit the determination of rates. These are not limitations for many of the other methods discussed here.

Altered Potential Energy Methods

There are many types of altered potential energy techniques that have been developed, including but not limited to metadynamics[45, 53], umbrella sampling[44], and accelerated MD methods such as temperature accelerated MD [54, 55]. While all of these methods may work differently, they all accelerate the simulation of long time scale events through the addition of biasing forces applied along one or more collective variable (CV)s that describe the transition of interest. Biasing along these CVs helps the system cross high energy barriers faster. In each method the biasing force is added in a different way. Despite these differences, in all of these methods, the addition of a bias to the system changes the potential energy U to:

$$U_{sys} = U + U_{bias}, \tag{1.3}$$

where U_{sys} is the new potential energy and U_{bias} is the energy added by the biasing forces.

In a general implementation of the metadynamics method, one trajectory is run at a time over a CV that describes the free energy landscape. To explain how the bias is added to this system, an analogy of adding sand to a landscape until valleys are filled and peaks are crossed is used. As the low energy basins, or valleys fill, a system can more easily cross the corresponding peak. The simulation equivalent to this analogy is as follows: as the system reaches a new location along the CV, a Gaussian, or addition of energy, located at that CV is added to the energy of the system. As the simulation moves forward, these energy additions accumulate in low energy basins until the system has been “elevated” enough to cross the barrier. In contrast to the single simulation run in metadynamics, umbrella sampling runs multiple simulations at points along the CV called windows. In each of this independent windows, a harmonic force is added to the system, to even out the free energy landscape at that point along the CV and lock the system in to that given value of the CV. The data from each of these windows can be combined through various methods including Weighted Histogram Analysis Method (WHAM)[56, 57].

Despite the benefits of these methods, there are various drawbacks to each of them. One

of the main drawbacks is the need to have a CV for the unbinding process (or other process of interest) for the system which requires prior knowledge of the reaction coordinate that the system follows, or user intuition to develop a potentially correct CV[58]. CVs can also miss out on important aspects of the unbinding process, leading to incorrect or unphysical events to occur within the system and subsequently result in incorrect free energy calculations. One reason this can happen is because relaxation of systems along “orthogonal degrees of freedom” or degrees of freedom that are independent of the CV used can be very slow. An example of this issue can be seen when simple distance based CVs are used for pulling charged molecules (such as peptides) through a hydrophobic core of a membrane. This results in poor convergence of the free-energy with the most significant hysteresis happening as the peptide enters and exists the core - a crucial step associated with a very large free energy barrier[59].

The bias must also be removed from the system for the determination of rate constants if they are of interest, and the assumption must be made that this bias did not impact system dynamics at critical points such as the transition state[53]. To overcome the complex issue of determining a proper CV, numerous methods have been developed to predict a CV for a system or reaction. This includes and is not limited to AMINO [60], DiffNets [61], Deep-TICA [62], SGOOP [63], and VAMPnets [64].

Trajectory Parallelization Methods

The above simulation methods all take advantage of potentially unphysical events through the addition or removal of various energies from the systems. This has implications for the validity of the dynamics that are derived from these simulations, the accuracy of the poses and pathways sampled, and what values can be determined. As stated above, the altered potential energy methods also require prior knowledge of the reaction of interest which can be very difficult to accurately predict. However, a group of methods exists that does not make modifications to the total energy of the system.

In these methods, no bias is added to the system and the trajectories are continuous which results in the generation of complete pathways and permits the determination of rate

constants. Trajectories are run at once in parallel, and then combined through means such as Markov state modeling, run with mid simulation checkpoints, or analyzed via trajectory and cycle specific probabilities[46]. These are methods such as milestoning[65, 66, 67] and the weighted ensemble method[68, 69, 32, 70] among others including Markov state modeling[71], transition path sampling[72], and forward flux sampling[73, 46]. These methods use multiple parallel trajectories to sample as much of the free energy landscape as possible.

The milestoning method connects two end points of interest that represent predefined basins for a reaction. Between these basins, there are numerous bins that connect the basins, and to define these bins and basins prior knowledge of the reaction coordinate is required. This method aims to determine the flux moving from one bin to another and using this flux, bin-to-bin rates can be determined. These individual rates can then be combined to find the rate constant for the transition between basins. A similar method called WE, is used in this thesis. In the WE method[68, 70, 69, 32] which runs semi-independent trajectories in parallel, trajectories are periodically “resampled”, a process where the number of interesting trajectories is increased via “cloning” and their probability drops with the goal of escaping a local energy minimum, and redundant trajectories are “merged” away. Newly developed versions of this method have been created that do not require the use of a CV. This method has been heavily used and further developed in our lab[74, 31, 9, 75, 32, 76, 26] and will be described in greater detail in the section below.

Weighted Ensemble

In the previous section, various types of enhanced sampling algorithms designed to increase the breadth of sampling from MD simulations were explored. Here, we will go into greater detail about the enhanced sampling algorithm utilized for all simulations in this thesis, the WE algorithm. WE is a rare-event oriented method that works to focus sampling on low-probability states that are relevant to some system-specific process of interest. In the original development of this method (created in 1996 by Huber and Kim [68]), Brownian motion between a product and reactant basin was simulated through the division of

conformational space into bins. Modern use of this method is more often interested in low-probability events such as ligand (un)binding processes, with various methodologies having been developed including bin-less options. While ligand unbinding is a common use case for the WE method [9, 74, 31, 32, 77] it is not the only one. The WE method has also been used at atomistic and coarse-grained levels for many other biologically relevant motions of interest [70] including but not limited to ion permeation[78], large-scale protein conformational changes[79, 80], protein-protein binding [81], and protein folding [82].

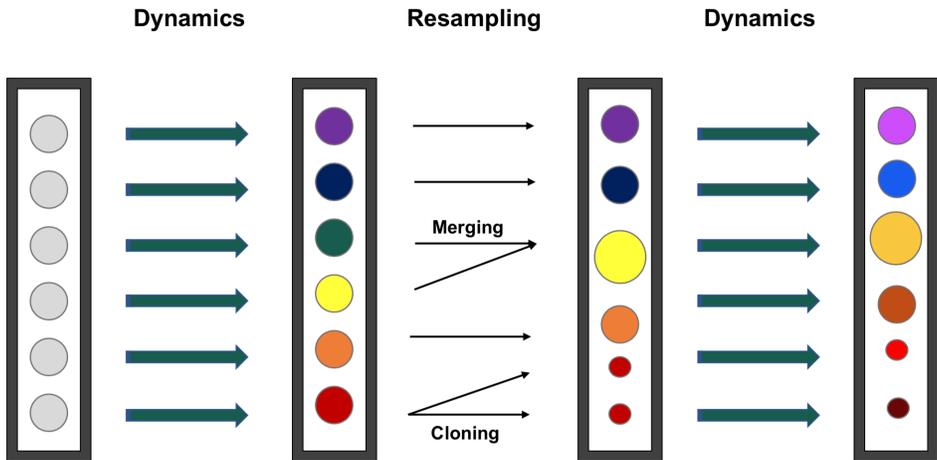


Figure 1.2: The WE Algorithm. Each circle represents a trajectory in the ensemble. The colors represent conformations of the system and the circle size represents the trajectory weight. All trajectories start with the same conformation and weight. In the dynamics step, the trajectories are moved forward in time by the molecular dynamics algorithm for a predetermined number of steps. In the resampling step, merging and cloning operations can be performed on members of the trajectory set. The dynamics-resampling procedure repeats until the end of the simulation.

In a general overview of the method, simulations run with the WE algorithm have two distinct steps: a cycle of MD, followed by a “resampling” step (Figure 1.2). If a trajectory is selected to be resampled in this step, it can undergo one of two operations. One operation is merging, where trajectories i and j that are similar based on a system-specific feature of interest have their weights w combined, where $w_i + w_j = w_{ij}$, and one of their two conformations is kept. The other operation, cloning, is where a trajectory k that is underrepresented based on a system-specific feature of interest has its conformation duplicated, and its weight

evenly distributed amongst its clones.

As previously mentioned, the standard format of the WE algorithm is to break down conformational space into “bins” along a predetermined reaction coordinate for resampling purposes. In this format, resampling is done to maintain a constant number of trajectories in each bin, where cloning can be done to increase the number of trajectories in a bin, and merging to decrease. While effective, this method is hindered by applications to high-dimensional spaces (such as ligand (un)binding processes and other biologically relevant motions of interest) that require a large number of bins/trajectories, as this becomes computationally less feasible [69]. Many modifications have been made to this method including the creation of bins in hierarchical Voronoi polyhedra[69] and bin-less resampling algorithms [32] to circumvent these issues.

Resampling of Ensembles by Variation Optimization

The WE simulation method is able to generate reactive trajectories for (un)binding paths and other long timescale processes without the addition of external biasing forces. However, the traditional WE algorithm divides the landscape into bins to guide the resampling process. Many modifications have been made to this method, including the WExplore method [69], that breaks down space into exponentially increasing numbers of Voronoi polyhedra. However, even these modified methods often have high dimensionality resulting in a high computational cost. To combat these common issues, a bin-less WE algorithm was developed: Resampling of Ensembles by Variation Optimization (REVO) [32].

In the REVO algorithm, the merging and cloning process is guided by the “trajectory variation” function, where the variation (V) is defined as:

$$V = \sum_i V_i = \sum_i \sum_j \left(\frac{d_{ij}}{d_\star} \right)^\alpha \phi_i \phi_j, \tag{1.4}$$

where, V_i is the trajectory variation contributed by walker i . d_{ij} is the distance, calculated by the distance metric, between walker i and walker j . d_\star is a “characteristic distance”, which is equal to the mean of the distance metric after one cycle of dynamics, and is used to make the

variation function unitless. This does not impact resampling behavior and is implemented for ease of comparison between different distance metrics for the same system. The ϕ_i term, is a “novelty” term that describes the significance of individual walkers in the ensemble. The term balances the “exploration” (distance) term with the “exploitation” (novelty) in the REVO algorithm. The novelty term ϕ is often defined according to the walker weight, w_i , represented by the function:

$$\phi_i = \log(w_i) - \log\left(\frac{p_{min}}{100}\right), \quad (1.5)$$

where p_{min} is the predefined minimum weight allowed. The p_{min} value is often set to 1×10^{-12} .

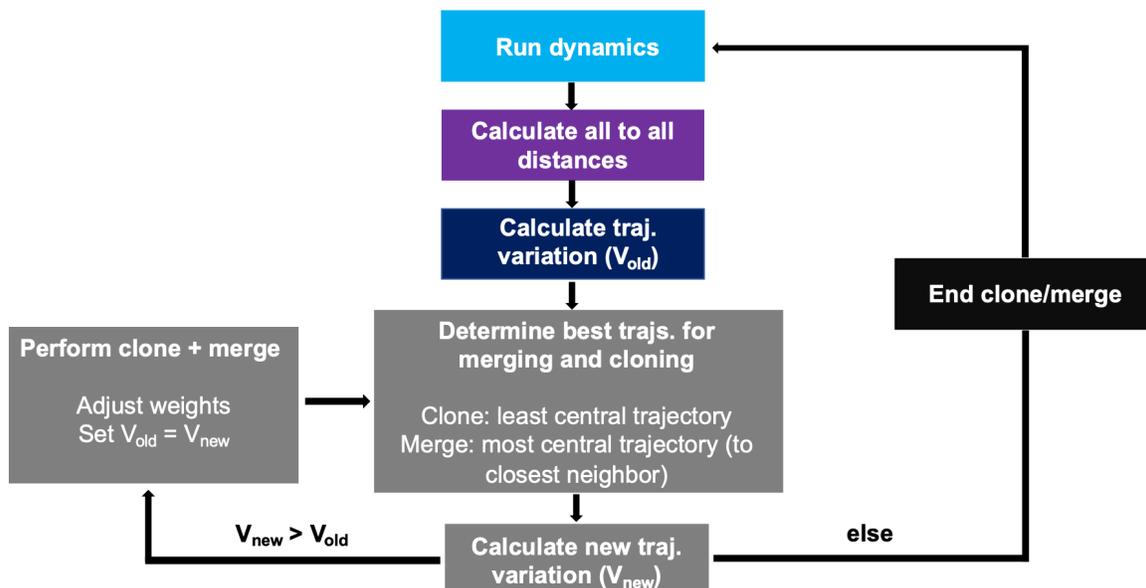


Figure 1.3: The REVO Algorithm. First, the dynamics are run and the all-to-all distances are calculated for the ensemble (cyan, purple boxes). The variation is calculated (blue). The the best options for cloning (least central trajectory) and merging (most central trajectory to closest neighbor) are determined (gray). The new trajectory variation is calculated as if this step has been performed and if the new variation is larger than the old variation, the step is performed; however, if the new variation is smaller than the old variation, resampling is ended (black) and dynamics are run again.

After a cycle of dynamics has been run and the distances calculated, the REVO algorithm, Figure 1.3, guides merging and cloning by first calculating V . Then, walkers are proposed for cloning and merging operations. To clone some walker i , it must have the highest V_i and a cloned-weight greater than p_{min} . To merge some walkers j and k , their distance d_{jk} must

be within a predefined cutoff value called the “merge distance”. The sum of their weights, w_{jk} , needs to be below the predefined maximum weight, p_{max} (often set to 0.1). Once walker selection is complete, V is recalculated as though the merging and cloning operation was performed. If the value of V increases, the merging and cloning operations are performed. This resampling process repeats until V has been maximized. Once V is maximized (it does not increase after a proposed merging and cloning step), resampling is ended and a new cycle of MD is run.

Ensemble Splitting for the Determination of Rates

Depending on the simulation method used to learn about the binding and unbinding kinetics of a system, it may be possible to determine the ΔG of (un)binding for the system directly from the binding and unbinding paths generated. If the WE method (Section 1.2) is utilized, a process called ensemble splitting[74, 83, 84, 85, 86] can be done to efficiently determine ΔG , k_{on} , k_{off} , the RT, and the binding affinity or K_d for the system. This method can technically be utilized with straight forward trajectories; however, this would likely be very inefficient.

In this method, we divide the ensemble of trajectories into “binding” and “unbinding” ensembles where the “unbinding ensemble” consists of trajectories that have most recently been in the bound basin, and the “binding ensemble” consists of the trajectories that have most recently been in the unbound basin. These basins do not overlap and can be described based on a system-appropriate definition of bound and unbound. In this thesis, applicable simulations fall into either only the unbinding or the binding ensemble and the corresponding simulations are conducted separately in either the "binding" or "unbinding" ensemble. When a trajectory exits its ensemble by entering the other basin, its probability is saved and in using these threshold crossing probabilities, we can determine transition rates between the basins. If rates are calculated for both the binding and unbinding processes, we can subsequently determine the ΔG of unbinding as well. The transition rates are calculated directly from the trajectory flux that exits one ensemble and enters the other: k_{off} (4.4), k_{on} (4.5) and ΔG

(4.6) can be determined as:

$$k_{off} = \frac{\overline{\phi}_u}{T} = \frac{\sum_{i \in U} w_i}{T} = \frac{1}{RT}, \quad (1.6)$$

$$k_{on} = \frac{\overline{\phi}_b}{C} = \frac{\sum_{i \in R} w_i}{CT}, \quad (1.7)$$

$$\Delta G = kT \ln \frac{k_{off}}{C_0 k_{on}}. \quad (1.8)$$

Here, $\overline{\phi}_b$ and $\overline{\phi}_u$ are the binding and unbinding “flux”. These fluxes are equivalent to the k_{on} and k_{off} rates for the system. T is the elapsed time for all corresponding transitions from one ensemble to another, the summed values are the total probabilities for the corresponding transitions, and C is the standard concentration of 1 mol/L.

1.3 The Jarzynski Equality (The Jarzynski Nonequilibrium Work Theorem)

The JE, or the Jarzynski Nonequilibrium Work Theorem is another means of calculating the free energy for a process of interest. To first understand the JE, one must first understand the exploitation of the relationship between ΔF and the work (W) applied to a system by some nonequilibrium force. This applied force could be used for a process that, for example, drives a system from some state A to some state B .

In a nonequilibrium process, the ensemble average of work necessary for said process, $\langle W \rangle$, must be greater than or equal to the difference in free energy, ΔF , between the initial and final states of the process[87] as referred to as the Maximum Work Theorem (MWT):

$$\langle W \rangle \geq \Delta F. \quad (1.9)$$

For microscopic trajectories, even small differences between the starting and subsequent states can be significant, and can necessitate drastic differences in the necessary amount of work required to perform some process. While the majority of iterations of this process will result in an amount of work done that is greater than ΔF between the initial and final states of interest, the work required *can* be less than the ΔF between those two states (which appears to violate the second law of thermodynamics[88, 87]).

Repeating an applied-work process on a system will result in the generation of a statistical distribution of the work values ($p(W)$) obtained for that process. Once generated, a statistical distributions of W can be assessed through an "equality transformation" of the MWT [87] referred to as the "Jarzynski Equality"[88] which is as follows:

$$\langle e^{-\beta W} \rangle = e^{-\beta \Delta F}, \quad (1.10)$$

where, $\beta = 1/k_B T$, T is the temperature the process is performed at, and k_B is the Boltzmann constant. The JE thus provides a means of determining *equilibrium properties*, such as unbinding free energies, from continuous, potentially short, nonequilibrium trajectories. The JE also does not require that the final state of the system is equilibrated.

However, the JE has one major hindrance. In $p(W)$ most iterations of a process will result in the generation of a "typical" work value near the peak of the $p(W)$. However, $\langle e^{-\beta W} \rangle$ is dominated by values of W near the peak of $i(W) = p(W)e^{-\beta W}$ [89] (Fig. 2.1), where $i(W)$ as the "importance" value of some specific value of work. The importance is a measurement of the contribution for some work value to $\langle e^{-\beta W} \rangle$ and consequently to ΔF in Eq. 1.10.

The JE has been utilized in many applications, including the determination of binding free energies for simple host-guest systems used in the SAMPL6 Challenge[90] as well as in combination with Steered Molecular Dynamics (SMD) for complex biological systems such as TSPO to construct PMF profiles and find binding free energies for different ligands[91].

In this thesis, the JE is used with the WE method in two ways to determine unbinding free energies for a Lennard-Jones pair (a simple two-atom test system). The first method developed is a combination of the JE and WE in the form of a new resampler. We refer to this method as "Importance Resampling", where a resampling framework that focuses resampling on trajectories with a high value of $i(W)$ is developed with the aim of increasing the number of trajectories in an ensemble with a high importance. The other method developed is history-REVO Resampling, a variation on the REVO resampler that considers

the work done for a whole trajectory for resampling purposes that resamples with the aim of diversifying the work distribution for an ensemble of trajectories.

1.4 SAMPL Systems

Briefly, the SAMPL systems used in this thesis are small molecule pairs that come from the SAMPL Challenges[92, 93]. These systems are used to test and develop new computational methods used for the prediction of properties that are relevant to drug discovery (of particular interest, the unbinding free energy, ΔG). Relevant work with these systems can be submitted to the SAMPL Challenges officially, such as the SAMPL6[90, 94] Challenge, or used in general method development.

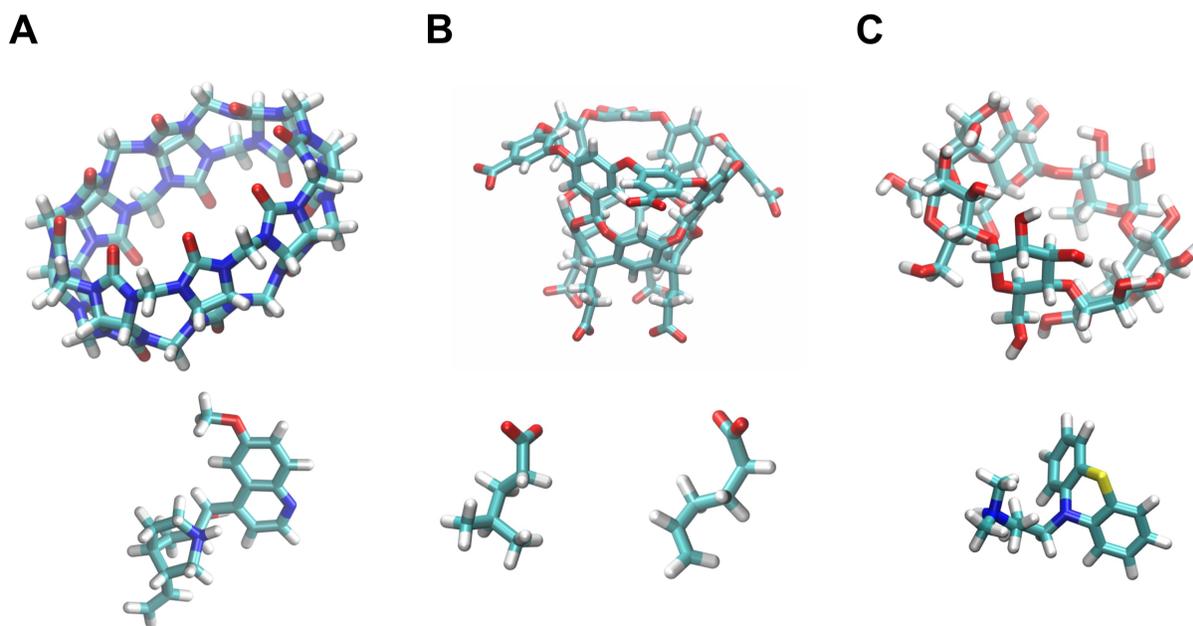


Figure 1.4: Host-guest systems. Hosts are shown on top, and guests are shown on bottom. A) The CB8 host and its ligand, G3. B) The OA host and its guests, G6 (left) and G3 (right). C) The β -CD host and its corresponding ligand, PMZ. All panels use the default “atom name” coloring scheme from VMD [95]: white=Hydrogen, cyan=Carbon, blue=Nitrogen, red=Oxygen, and yellow=Sulfur.

The systems used in this thesis include the OA-G6, OA-G3, and CB8-G3, systems from the SAMPL6[96] Challenge and the β -CD-PMZ system from the SAMPL9[97] Challenge. OA (octa-acid), CB8 (cucurbit[8]uril) and β -CD (β -cyclodextrin) are “host” molecules with binding pockets to which their corresponding “guests”, G6 (4-methyl pentanoic acid) and

G3 (5-hexenoic acid), G3 (quinine), and PMZ (promazine hydrochloride), respectively, bind (Figure 1.4). The OA molecule has a -8 charge and four-fold symmetry along the vertical axis. Its guests G6 and G3 are both small molecules with an explicit -1 charge. The CB8 host is a neutral molecule (as is its quinine guest) with two-fold symmetry along the horizontal axis, and eight-fold symmetry along the vertical axis. β -CD is a neutral host molecule with seven-fold symmetry along its vertical axis and no net charge. Its guest, PMZ has an explicit charge of $+1$.

1.5 Outline of Work

The overall goal of this thesis is to develop new methods for use with the weighted ensemble enhanced sampling method to increase the accuracy and efficiency of simulations for the purpose of efficient, kinetics oriented drug design. More specifically, the goals are:

- Combine the JE with WE to develop methods that will allow us to determine unbinding free energies from short, low computational-cost, nonequilibrium simulations.
- Learn important solvent-based features of host-guest test system (un)binding pathways to guide (un)binding simulations.
- Use features learned about host-guest unbinding with a new variation of the REVO method to decrease variation between simulations of the same system and increase confidence in values determined, such as ΔG .

We first combine the JE with the WE method - more specifically the REVO method described in Chapter 2. In this work we utilize a Lennard-Jones Pair, a two atom test system used to test and validate the methods developed. This system is ideal as it is smaller than most other common test systems such as host-guest pairs or other common biological test systems, and provides an analytical solution for the binding free energy. Two methods are developed here, with those methods being:

- Importance Resampling: A resampler that focuses resampling on trajectories with a high value of Jarzynski-importance.
- history-REVO Resampling: A variation on the resampler that considers the work done for a whole trajectory for resampling purposes.

The convergence of free energies, work curves, and importance curves calculated by the history-REVO method are compared to straight-forward simulations. The methods developed are also compared to Diffusion Monte Carlo (DIFFMC) method, an established path sampling method that uses unperturbed dynamics and resamples based on a “potential” determined from the applied work. From this work we found that the Importance Resampler method had only sporadic success due to the ever-changing value of work for a trajectory, but found promising results with the the broad work distributions generated by the history-REVO resampler.

Following the completion of this project, a large amount of unpublished work was done using the history-REVO method and the host-guest system OA-G6. This host-guest system consists of two small molecules that have non-covalent interactions to form a complex. The larger molecule containing the binding pocket is referred to as the host, and the smaller molecule that (un)binds is referred to as the guest. In this work, it was found that the history-REVO method was inefficient with larger systems, likely due to insufficient descriptors for the unbinding of the system. More specifically, we found that very large variations existed between simulations of the same length and setup, with free energies varying up to $+/- 15$ kcal/mol from the known unbinding free energy for the system.

Due to these issues, we proceeded to learn about the important solvent-based features of unbinding for the host-guest system, as shown in Chapter 3 for use in our simulation to increase the accuracy of the free energy calculations performed. Specifically, we were interested in the solvent-based differences between the high and low probability unbinding events. High weight events are of particular interest due to the fact that they strongly

influence ensemble properties, and the generation of many high weight unbinding events can lead to more accuracy and consistency between data-sets. For our analysis we took advantage of large WE datasets that had previously been generated for our host-guest systems of interest OA-G6 and OA-G3 (described in more detail in Chapter 3). Using this data, a time and weight, or probability based analysis is done on physical features of interest, including but not limited to:

- The number of waters in the binding site.
- The number of ions around the negatively charged regions of the host.
- The number of ions around the guest molecule.

It is found that there is a significant difference in A) the guest-ion interactions and B) the local ion densities between the high and low weight unbinding events. Specifically, high weight unbinding events average less than 0.5 ions present in the region of space that guest enters when it leaves the binding pocket. This results in the high weight events quickly hitting the boundary condition and the trajectory ending with a higher weight than had it been repeatedly cloned. Importantly, we also find that there is a "commitment to unbinding" point for all reactive trajectories in this data-set, which is found to be a center of mass (COM)-to-COM distance of 0.7 nm. We refer to this point as t_0 .

Using the concept of this t_0 point, a new resampler based off of the REVO resampler is developed that we call 'cutoff-REVO', which is explored in Chapter 4. In this work, t_0 , or more broadly a commitment to event-of-interest point is added to the REVO resampler as a cutoff point for cloning operations. This is done because once a trajectory commits to an event of interest, cloning it only reduces its weight and takes up more room in the ensemble that could have otherwise been used to explore other paths. In this work specifically, once a system-specific COM-to-COM distance is reached for a trajectory, it is no longer eligible to undergo a cloning operation during the resampling step of a WE simulation. For four SAMPL systems (OA-G6, OA-G3, CB8-G3, and β -CD-PMZ), we compare:

- The accuracy of the unbinding free energy to the computational and experimental reference values when available.
- The convergence of the k_{off} rate.
- The differences in the distribution of the unbinding-event weights.
- Differences in resampling patterns between traditional REVO and cutoff-REVO.

From this work we find that the cutoff-REVO resampler produced significantly fewer unbinding events for all systems, but the events produced have much higher weights. This promotes faster convergence of the k_{off} value for the systems and resulted in more accurate free energies of unbinding. In the final Chapter, 5, we revisit the highest level goals of this thesis and assess the progress, pitfalls, innovations, and potential steps for improvements to the methods developed.

CHAPTER 2

ENHANCED JARZYNSKI FREE ENERGY CALCULATIONS USING WEIGHTED ENSEMBLE

This is work reproduced from “Nicole M. Roussey and Alex Dickson, “Enhanced Jarzynski free energy calculations using weighted ensemble”, *The Journal of Chemical Physics*. 153, 134116 (2020) <https://doi.org/10.1063/5.0020600>”[76] with the permission of AIP Publishing. The work is presented here as published except that the supplemental figures are included in the appendix.

2.1 Introduction

The importance of free energy calculations from atomic simulations has been well established for many biological and chemical applications. The free energy of a transition in a biological system is the key thermodynamic quantity governing the transition’s reversibility and the work required to perform the transition in an unfavorable direction. In this way, free energy forms the theoretical underpinnings of a range of biological processes including ligand (un)binding, protein folding, solubility, and biologically significant conformational changes in protein structures[1]. Free energy calculations also provide a means of estimating other relevant values for biological processes such as permeability coefficients and rate constants [1]. Due to the significance of this value, new, computationally efficient methods of calculating free energies are always of interest.

Increases in computational power, new developments in enhanced sampling, and advances in statistical mechanics have led to the development of a wide variety of methods to calculate free energies (see reviews [1, 98, 99]). Despite decades of systematic improvement, there are complications and drawbacks associated with each method. Relative binding free energies can now be routinely calculated by FEP, with robust computational tools available (e.g. FEP+, Schrodinger [100]), however these are limited to sets of ligands with high similarity [101]. There are also a set of methods to calculate the absolute binding free energy for a protein-ligand or protein-protein complex. This can either be done through physical sam-

pling of the unbinding pathway (e.g. umbrella sampling [102, 103], Markov state modeling [104, 105], Metadynamics [106, 107], Milestoning [108, 109] and weighted ensemble strategies [68, 110]) or through alchemical transformations[111] (e.g. using double decoupling [112] with thermodynamic integration [113] or Hamiltonian replica exchange [114]). and varying combinations of these free energy estimators and methods. The two main obstacles of physical sampling strategies are i) convergence of the free energy estimate, and ii) projection of the binding process onto one or more order parameters. In alchemical transformations, single topology models are hindered by residual charges from the vanishing atoms that can create large forces in the system and double topology models can have overlap between groups leading to instability in the system [1]. Alchemical transformations also only predict the free energy difference of the endpoints and cannot produce complete free energy profiles. Even for small systems such as host-guest pairs used in recent SAMPL challenges[94], computational free energies obtained with these methods are often not consistent with each other, and can deviate by several kcal/mol from experimentally determined values [94, 90].

An alternative method for calculating the free energy of a process at equilibrium (ΔF) is to exploit a relationship between ΔF and the work applied to a system by a nonequilibrium force (W). The maximum work theorem states that in a nonequilibrium process, the ensemble average of work required, $\langle W \rangle$, is greater than the difference in free energy, ΔF , between the initial and final states of the process[87]:

$$\langle W \rangle \geq \Delta F. \tag{2.1}$$

Differences between individual trajectories are significant in microscopic systems and can result in very different values of work being observed for the same process. For most realizations of a process, the work done is greater than ΔF ; however, the work can be smaller than the free energy difference (seemingly violating the second law of thermodynamics)[88, 87]. Performing many realizations of a process produces a statistical distribution of work values for that process. These statistical distributions of W can be accounted for in an "equality

transformation" of the maximum work theorem [87] referred to as the Jarzynski Equality[88]:

$$\langle e^{-\beta W} \rangle = e^{-\beta \Delta F}, \quad (2.2)$$

where, $\beta = 1/k_B T$, where k_B is the Boltzmann constant and T is the temperature at which the transformation process is performed. Remarkably, this equation generates equilibrium properties from short, continuous, nonequilibrium trajectories that have little sensitivity to the quality of the reaction coordinate. Another appealing aspect of the Jarzynski Equality is that the equilibration of the final state of the system is not required, as it would not change the measured nonequilibrium work.

The Jarzynski Equality has been utilized in a broad range of applications. Procacci and Guarnieri [115] determined water-octanol partition coefficients for small molecules using the solvation free energies in both solvents. Binding free energies have been found using Jarzynski-based unidirectional estimators for small host-guest systems, such as those used in SAMPL6[90]. Similarly, SMD has been used with more complex systems, such as the TSPO, in tandem with the Jarzynski Equality to reconstruct PMF profiles and determine ΔF_{off} for different ligands[91]. Xiong et al [116] applied the Jarzynski Equality in a number of contexts: calculating free energies with QM-MM to study conversion reactions by the enzyme chorismate mutase; mechanical unfolding of the Ace-Alanine₈-NMe biopolymer; and creating free energy profiles for ligand diffusion in globin proteins.

Unfortunately, the trajectories that dominate the expectation value in Eq. 2.2 are those with low (or negative) values of W , which occur rarely and can be difficult to observe in simulation. These dominant trajectories in a forward process occur with a probability roughly approximated as

$$P_f \sim e^{-\beta \langle W_d^r \rangle} \quad (2.3)$$

where W_d^r is the dissipative work of the reverse process, as estimated by Jarzynski[89]. This value is equal to $\frac{1}{N_c^f}$, where N_c^f the number of realizations of a processes necessary for convergence of free energy. This suggests that the number of realizations required for a

process increases relative to the exponential of the average work and therefore the system size [117, 89, 118]. In pulling experiments done to determine helix propensities for different amino acids, the Ala₁₂ peptide was found to require $\sim 10^5$ realizations for convergence of free energy for the forward process (and thus a $\sim 10^{-5}$ probability of a dominant realization occurring) and $\sim 10^7$ realizations for convergence of free energy for the reverse process [119].

Efforts have been made previously to apply path sampling algorithms to preferentially generate trajectories with lower values of work. One difficulty is that of the algorithms that have been developed to sample low probability events, only a small subset of these are applicable to nonequilibrium systems. Transition Path Sampling (TPS) methods [72, 120] have been applied to generate ensembles of trajectories with lower work values that contribute most significantly to the ensemble average [120, 121, 122]. One challenge of this approach is that TPS is only able to select low-work trajectories after they have been generated and does not aim to accelerate their production. Also it is difficult to sample new low-work trajectories as there are typically substantial free energy barriers that separate different low-work pathways. Other methods compatible with nonequilibrium systems – Nonequilibrium Umbrella Sampling (NEUS) [123, 83] and DIFFMC [124] – have also been applied to preferentially sample low-work trajectories for model systems [125, 126]. The WE method [68, 127] – which utilizes a set of weighted, unbiased trajectories that can be cloned and merged – is another path sampling algorithm that is applicable to nonequilibrium systems and has been used to increase the chances of observing rare biomolecular conformations and to simulate long-timescale processes such as protein folding [128], large conformational transitions [129], and ligand (un)binding [130, 9, 131].

Here we investigate the capability of the WE method to enhance the sampling of rare *trajectories*, specifically those with atypical values of work. Doing so requires a customized WE algorithm that considers the whole trajectory history during resampling. We adapt our recently developed algorithm called REVO (“Resampling Ensembles by Variation Optimization”) [32] for this purpose. We also implement a new resampling algorithm that

explicitly considers the importance of individual trajectories and specifically focuses sampling on trajectories with the most significant work values. This new algorithm includes a tunable parameter we call "amplification" that governs the depth to which the method will dig into the work distribution tail. We apply these methods to unbinding trajectories of a two-particle Lennard-Jones system with well depths ranging from 2 to 20 kcal/mol. The efficiency of these different resampling strategies is then analyzed for systems over a wide range of binding free energies. We also compare these results to the DIFFMC approach, where the applied work is used as the selection potential [132]. Finally, we conclude with a discussion of this Enhanced Jarzynski method and history based resampling, including future calculations of protein-ligand binding free energies.

2.2 Methods

Generalized outline of Weighted ensemble sampling

The framework for a generalized WE algorithm includes two main steps: the propagation of a group of trajectories (or "walkers") forward in time by MD, and resampling, which performs merging and cloning operations on the ensemble of walkers. Resampling operations function with the goal of cloning walkers with desirable features and merging together less-desirable walkers based on some feature of interest.

In a WE simulation, all walkers have a statistical weight or probability. When a walker is cloned, two independent walkers are created that have the conformation of the original cloned walker and half of its weight. In a merging operation, two walkers, A and B, are combined to create walker C with a weight of $w_C = w_A + w_B$, and C takes on the conformation of either A or B with a probability proportional to their weights. A resampling function takes in an ensemble of walkers and returns a new ensemble with conformations drawn from the original input ensemble. In general, the returned ensemble can have the same or different number of walkers, but the sum of walker weights, typically equal to one, is unchanged. In this work we maintain a constant number of walkers over the course of all simulations.

History-Dependent REVO Resampling

To perform work-based resampling simulations, the Resampling of Ensembles by Variation Optimization, or REVO [32] resampler is used. This method is briefly described below, with specific details regarding this application in Section 2.2. REVO governs cloning and merging through the maximization of an objective function called the "trajectory variation" (V), which is a scaled sum of all-to-all pairwise distances between the walkers:

$$V = \sum_i V_i = \sum_i \sum_j \left(\frac{d_{ij}}{d_0} \right)^\alpha \phi_i \phi_j \quad (2.4)$$

where d_{ij} is the distance between two walkers i and j , α is an exponent that modifies the distances' influence on the overall value of V , and d_0 is a "characteristic distance" that does not affect merging and cloning and serves to make V unitless. The distance metric is defined to capture the system-specific event of interest and is described below. ϕ is a non-negative function that measures the relative importance of each walker and can again be designed in a system-specific fashion based on walker attributes, such as conformation, history, or weight. Here ϕ is defined as:

$$\phi_i = \log(w_i) - \log\left(\frac{p_{min}}{100}\right) \quad (2.5)$$

where p_{min} is the minimum statistical weight that a walker can hold, and w_i is the current weight of walker i . Setting a value of p_{min} is useful to avoid spending simulation time on trajectories that will not contribute meaningfully to the observables you wish to calculate. Similarly, a maximum value of the walker weight p_{max} is also useful to prevent the agglomeration of all of the weight into a single walker. Both p_{min} and p_{max} are enforced by simply preventing the resampling algorithm from suggesting at-risk walkers for cloning and merging, respectively. The overall goal of the REVO resampler is the optimization of V . To do this, walkers with a high V_i (Eq. 3.1) are selected for cloning, and walkers with a low V_i are selected for merging.

While REVO is designed with high-dimensional systems in mind, here we found it sufficient to compute distance in a one-dimensional space, where d_{ij} is simply the difference

in the sum of the time-elapsed work (as explained in Section 2.2) between a pair of walkers i and j . Notably, this distance metric relies on the entire history of a trajectory and not just the instantaneous conformation of the system. These history-dependent quantities, which we refer to as “activities” can be utilized with the `wepy` simulation package [133] and the `wepy-activity` plugin [134]. The value of the activity, in this case, the work, for each walker is calculated after every cycle of dynamics. This value is then added to a running sum of work for that walker. The value of the running sum is then passed to a distance metric to do resampling.

Importance Resampling

Multiple realizations of a process will result in the generation of a statistical distribution of work values ($p(W)$). While most realizations of a process result in a "typical" work value near the peak of $p(W)$, $\langle e^{-\beta W} \rangle$ is dominated by realizations near the peak of $i(W) = p(W)e^{-\beta W}$ [89] (Fig. 2.1). We refer to $i(W)$ as the "importance" of a specific work value, as it measures the extent to which that work value contributes to the ensemble average $\langle e^{-\beta W} \rangle$ and subsequently to ΔF using Eq. 2.2.

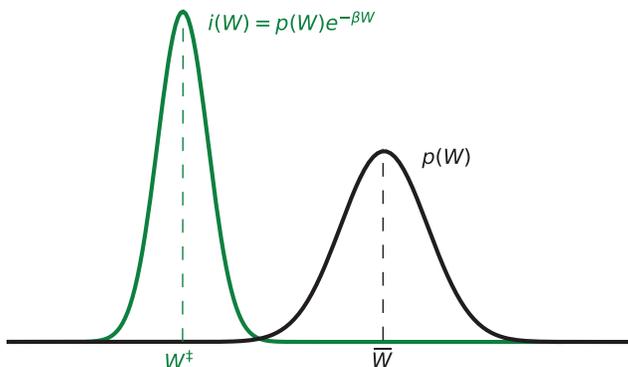


Figure 2.1: Probability and importance curves. Most values of work observed are near the peak of $p(W)$, or around the mean of the work distribution, \bar{W} . Values of work that dominate $\langle e^{-\beta W} \rangle$ occur near the peak of $i(W)$, around W^\ddagger .

With this in mind we introduce a new trajectory ensemble resampling algorithm, "Impor-

tance Resampling", for the WE framework that works with the WEPY simulation package. The overall goal of this process is to generate an ensemble with multiple trajectories that have weights at or around the peak of the importance curve. The Importance resampler uses a new algorithm to select walkers for merging and cloning, which is comprised of two selection steps, one based on trajectory importance, and another based on weight. Note that resampling based on selection (rather than cloning and merging) in a weighted ensemble context was described previously in Ref. [135, 136]. For importance-based selection, N_I trajectories are randomly selected with a probability proportional to their relative importance of that trajectory:

$$I_j = \frac{w_j e^{-\mu\beta W_j}}{\sum_k w_k e^{-\mu\beta W_k}}, \quad (2.6)$$

where I_j is the relative importance of trajectory j , w_i is the weight of trajectory i , and W_i is the total work applied to the system so far in trajectory i . Eq. 2.6 also includes an additional factor called the "amplification" (μ). This can be set to any positive number to tune the strength with which we want to amplify the importance of low-work walkers. We examine different values of μ below. If a walker is selected once, it retains its weight for the next cycle. If a walker is selected more than once, its weight is divided evenly among the clones.

The remainder of the slots ($N_W = N - N_I$) are reserved for walkers that are selected based on walker weight. All walkers that were not chosen by importance are eligible for weight-based selection. Once selected, each walker takes on the probability:

$$p_W = \frac{1}{N_W} \left(1 - \sum_{i \in \mathbb{I}} w_i \right) \quad (2.7)$$

where \mathbb{I} is the set of walkers that were selected based on their importance. Note that weight-based selection is necessary for the importance-based trajectories to maintain the proper relative walker weights. While, intuitively, low N_W values would lead to better sampling of the high-importance walkers, here we aim to maintain a number of weight-based walkers (e.g. $N_W = 5$) to ensure some diversity in the weight-based ensemble. A schematic of the Importance resampler is shown in Fig. 2.2 for one round of resampling.

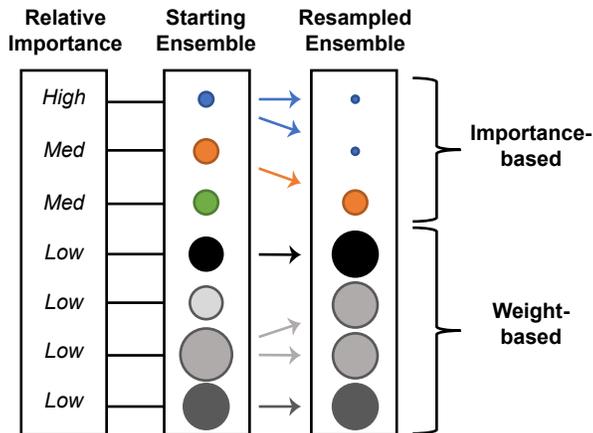


Figure 2.2: Importance resampler schematic. Each walker is represented by a different colored circle. The walker weight is represented by circle size and different colors indicate different conformations of the system. In resampling, importance based walkers are selected randomly with a probability proportional to their relative importance. The weight of walkers selected by their importance is equal to the weight of the parent walker divided by the number of times that walker was selected. Weight based walkers are selected randomly with a probability proportional to their weight. Each walker selected by weight is given an equal weight, given by Eq. 2.7. In this schematic, $N_I = 3$ and $N_W = 4$.

Lennard-Jones Pair Test System and Simulation Setup

We use a Lennard-Jones pair test system modified from the `LennardJonesPair` module of `OpenMMTools` [137] to analyze the performance of both the REVO resampler and the Importance resampler. We chose this system due to its simplicity and the fact that its difficulty can be easily tuned by changing ϵ , the depth of the inter-particle interaction energy well.

$$V_{LJ} = 4\epsilon \left(\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right), \quad (2.8)$$

The default values of particle mass and σ – 39.9 daltons and 3.35 Å respectively – were used for all simulations with a box size set to 40 Å with the `NonBondedForce` set to `CutOffPeriodic`. ϵ was varied from 2 to 20 kcal/mol. For this system, the target free energy as a function of the inter-particle separation, r , can be analytically solved as:

$$F_r = -2kT \ln(r) + V_{LJ}. \quad (2.9)$$

Using this system, nonequilibrium pulling simulations were performed without resampling (referred to as “straightforward MD”), with the REVO resampler, and with the Importance resampler. All simulations were run with OpenMM[138] version 7.4.0 and OpenMMTools version 0.18.3. The system is run at a constant pressure of 1 atm and temperature of 300 K using Langevin dynamics with a friction coefficient of 1 ps^{-1} and an integration step-size of 2 fs. For REVO simulations, a minimum and maximum walker weight (p_{\min} and p_{\max}) of 10^{-100} and 0.5, respectively, were used with a maximum merging distance of 2.5 kJ/mol and $\alpha = 4$. For Importance resampler simulations, 5 of 50 slots were designated for weight-based walkers ($N_W = 5$ and $N_I = 45$).

To perform the pulling simulations, a harmonic CustomBondForce from OpenMM is applied to the system. The force applied is:

$$U(\vec{x}; r_0) = \frac{k}{2}(r(\vec{x}) - r_0)^2, \quad (2.10)$$

where k is a spring-constant with a constant value of 2000 kJ/mol/nm², r_0 is the target interatomic distance for that cycle with a starting value of 0.32 nm, and $r(\vec{x})$ is the interatomic distance calculated from a state vector \vec{x} . After every cycle of dynamics, the value of r_0 is updated for all trajectories by a predefined value, $\Delta r_0 = (r_0^f - r_0^i)/n_c$, where r_0^f is the final separation distance (2.0 nm) and n_c is the number of cycles, set to either 500 (examined first in Results section below), or 53 (examined later).

Starting positions for each well depth are generated from straightforward simulations run at r_0^i . A set of 1000 starting positions was generated using the final positions of 1000 independent, 20 ns trajectories. These positions were saved and are used to randomly initiate pair starting positions for the nonequilibrium simulations.

The Work Equation and Free Energy Surfaces

The nonequilibrium trajectories use a set of biasing potentials (Eq. 2.10), to restrict the system to progressively increasing values of r_0 . The potential energy related to this biasing force at cycle t is denoted: $U(\vec{x}_t; r_0^t)$, and depends on both the current state of the system

\vec{x}_t (in this case used to calculate r , the observed interatomic distance) and the location of the biasing force r_0 at cycle t , which we denote as r_0^t .

Both the REVO resampler and the Importance resampler require that work is calculated after every cycle by an activity metric and added to a running sum. For the REVO resampler this total activity is then used for determining distances for the resampling process. For the Importance resampler, the total activity is passed to resampler to determine importance values for each walker. The work observed along a trajectory (W_t) up to and including cycle t is calculated as follows:

$$W_t = \sum_{c=0}^t U(\vec{x}_c, r_0^{c+1}) - U(\vec{x}_c, r_0^c) \quad (2.11)$$

where the work from cycle c in the summand is equal to the resultant energy change from instantaneously changing the control parameter r_0 [49].

Using work values and walker weights, free energy surfaces can be generated for pulling simulations using a nonequilibrium adapted weighted histogram method [139] [140]. Here we adapt Eq. 8 from Hummer and Szabo[139] to incorporate averages over weighted trajectory sets as:

$$F_0(r) = -kT \ln \frac{\sum_t \frac{\langle \delta(r-r(\vec{x}_t)) e^{-\beta W_t} \rangle}{\langle e^{-\beta W_t} \rangle}}{\sum_t \frac{e^{-\beta U(r; r_0^t)}}{\langle e^{-\beta W_t} \rangle}}. \quad (2.12)$$

where r denotes a specific value of the interparticle separation. Here, $\langle \dots \rangle$ indicates an ensemble average for a specific timepoint, t , which for weighted ensemble simulations is calculated as:

$$\langle f(\vec{x}_t) \rangle = \frac{1}{N_{\text{runs}}} \sum_{\text{runs}} \sum_{i=1}^N w_i f(\vec{x}_t^i) \quad (2.13)$$

where N is the number of walkers, and \vec{x}_t^i is the state vector of walker i at cycle t . The free energy (F_0) in Eq. 2.12 is calculated for each value of r and is used here to generate a free energy profile.

As this is a simple system, the target free energy profile for the Lennard-Jones pair can be solved for analytically using Eq. 2.9. To analyze convergence of binding free energies,

we define ΔF as the difference in free energy between the interatomic distances of 0.38 nm and 1.5 nm. We then track $\Delta\Delta F$ as a function of simulation time, which is the difference between the target ΔF determined analytically and the ΔF determined from simulation.

Diffusion Monte Carlo

An existing path sampling method, called Diffusion Monte Carlo, is similar to the Importance Resampler in a number of ways. Both use an ensemble of trajectories (of size N_{traj}) that evolve according to unperturbed dynamics and are periodically resampled. Diffusion Monte Carlo, a quantum Monte Carlo method, uses a "selection potential" for each trajectory in an ensemble that can be determined from the applied work after each cycle of dynamics. Briefly, dynamics are run, the work increment for each trajectory is calculated, and a weighting factor for each trajectory, is determined as $w_i = e^{-\beta W_{\text{inc}}}$, where W_{inc} is the incremental work for that trajectory. Note that these weights wholly determined by this work increment and are unrelated to the weights resulting from resampling in the weighted ensemble algorithm.

Following determination of the weight, a resampling process is done by determining branching numbers for each trajectory according to their weighting factors, and choosing the resulting trajectories for the next cycle using a stochastic process, where the average number of times a trajectory is chosen is equal to $w_i / \sum_j w_j$.

For a given cycle of dynamics (say, c), the average weight is determined and saved as a variable, $z_c = \sum_i w_i / N_{\text{traj}}$. The z values calculated for an ensemble can be utilized for analysis with Eq. 2.12 through the substitution of the running product of these terms ($Z_c = \prod_{i=0}^c z_i$) for the existing $e^{-\beta W}$ term for each cycle. A full explanation of the algorithm implemented in this work can be found in Algorithm 6.3 from Rousset et al[132].

2.3 Results

Reconstruction of free energy curves

We first ran a set of 200 independent simulations with 50 walkers each for both straightforward dynamics (with no resampling) and the REVO resampler. Separate sets were run for $\epsilon = 2, 5, 10$ and 20 kcal/mol (or 8.37, 20.9, 41.8 and 83.7 kJ/mol). All simulations are 500 cycles in length with 100 steps per cycle and were set up as described in Section 2.2. Figure 2.3 shows free energy predictions as well as target free energy curves computed analytically (see Eq. 2.9), for each value of ϵ . Both straightforward and REVO simulations accurately recreate the target free energy surfaces for all values of ϵ . Free energies are calculated from simulation using the Hummer and Szabo equation above (Eq. 2.12).

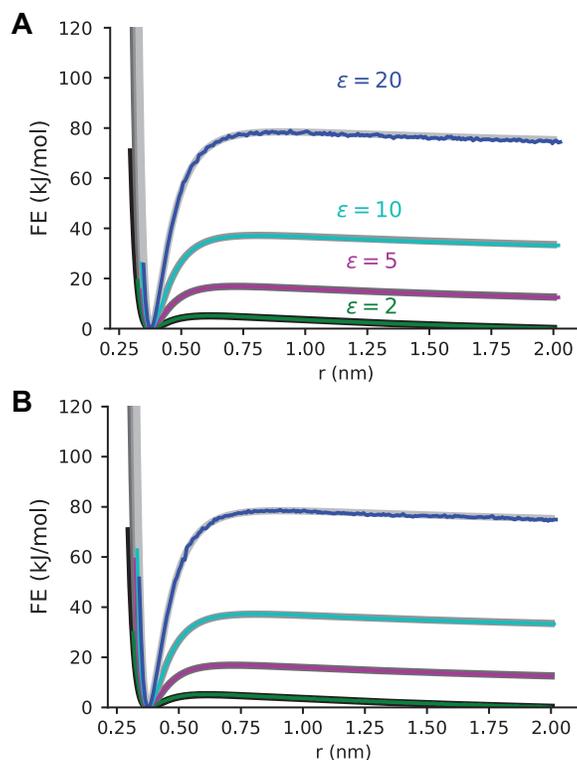


Figure 2.3: Calculated free energy curves. Reconstructions of free energy surfaces are shown as a function of inter-particle separation (r) for (A) straightforward resampling and (B) REVO. Four different well-depths, $\epsilon = 2, 5, 10$ and 20 kcal/mol are shown. In each panel the target free energy for each ϵ is shown as a thick line.

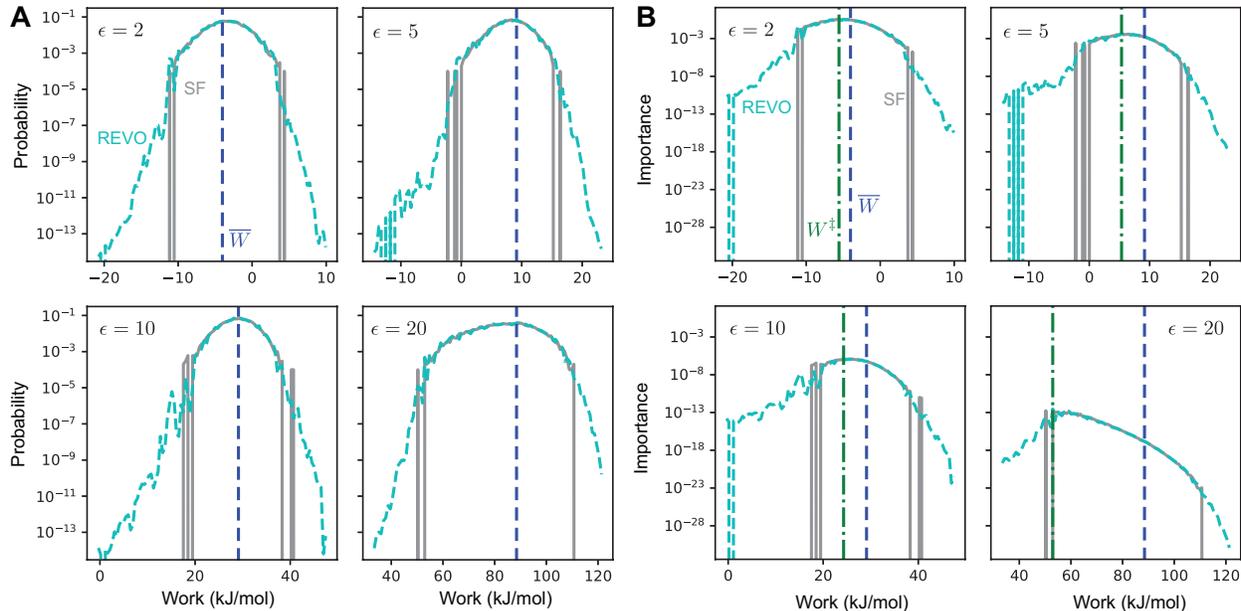


Figure 2.4: Probability distributions of applied work and trajectory importance. (A) The corresponding work distributions for 2.3 are shown for $\epsilon = 2, 5, 10,$ and 20 kcal/mol. Probability peaks are marked by dashed blue lines and are found at $-4.02, 9.14, 29.12,$ and 88.48 kJ/mol, respectively. (B) The probability distribution for the importance ($i(W) = p(W)e^{-\beta W}$) are shown for $\epsilon = 2, 5, 10,$ and 20 kcal/mol. Importance peaks are marked by dashed green lines and are found at $-5.6, 5.3, 24,$ and 53 kJ/mol, respectively.

Fig. 2.4A shows probability distributions for the cumulative work applied to the system during the dissociation process. We observe that REVO more extensively samples the tails of the work distributions for all four values of ϵ . These also show how the most probable values of work depend on the depth of the free energy well. For $\epsilon = 20$, large values of applied work are necessary to perform dissociation, while for $\epsilon = 2$ many trajectories can be observed with low, or even negative applied work. Importantly, we also observe that the probability distributions of applied work are not Gaussian, especially at higher values of ϵ . This indicates that direct sampling of low-work trajectories is important to obtaining accurate free energies with the JE.

As noted above, the importance of a given work value can be calculated as $i(W) = p(W)e^{-\beta W}$ [89]. Importance distributions are averaged over the entire trajectory set and shown in Fig. 2.4B. As expected, peaks in $i(W)$ (dashed green lines) do not correspond

to the peaks in $p(W)$ (solid blue lines). They are shifted to lower work values by a gap that increases with ϵ . For all values of ϵ , both methods are able to successfully sample the peak of the importance distributions; however, for $\epsilon = 20$, this peak lies at the limit of the distribution sampled by straightforward dynamics.

The REVO method is thus able to sample both the high and low tails of the work distribution more effectively than straightforward dynamics. To visualize how the ensemble of trajectories in REVO evolves in time, we show a “resampling tree”, which is a directed acyclic graph where the nodes represent walker states at each cycle and the edges show how walkers are cloned during REVO resampling (Fig. 2.5). Note that the bottom of the tree shows the initial state of the walkers and the time axis is in the upward direction. The highest weight walkers have a final value of work between 80-100 kJ/mol, which corresponds to the peak of the $\epsilon = 20$ probability curve in Fig. 2.4A. Low weight walkers have values of work that represent the high and low tails of the work distribution. This tree shows how REVO identifies and amplifies low-work walkers early on in the simulation, leading to better sampling of the final work distribution and more efficient calculation of binding free energies.

Convergence of free energy predictions

To analyze the relative efficiency of history-based REVO, we study the convergence of the free energy profile towards the target profile as a function of simulation time. Rather than considering the whole free energy curve, we monitor the free energy difference, between two points, one for the bound state ($r = 0.38$ nm) and one for the unbound state ($r = 1.5$ nm). We denote this as ΔF and refer to it as the unbinding free energy. $\Delta\Delta F$ is the difference between the unbinding free energy determined from simulation and the unbinding free energy for the analytically determined solution given by Eq. 2.9. Figure 2.6 shows expectation values of the root mean squared $\Delta\Delta F$ and its uncertainty for both straightforward dynamics and REVO simulations.

For smaller numbers of trajectories straightforward dynamics generally outperforms REVO, although this difference is not statistically significant for $\epsilon = 20$. The final root mean square

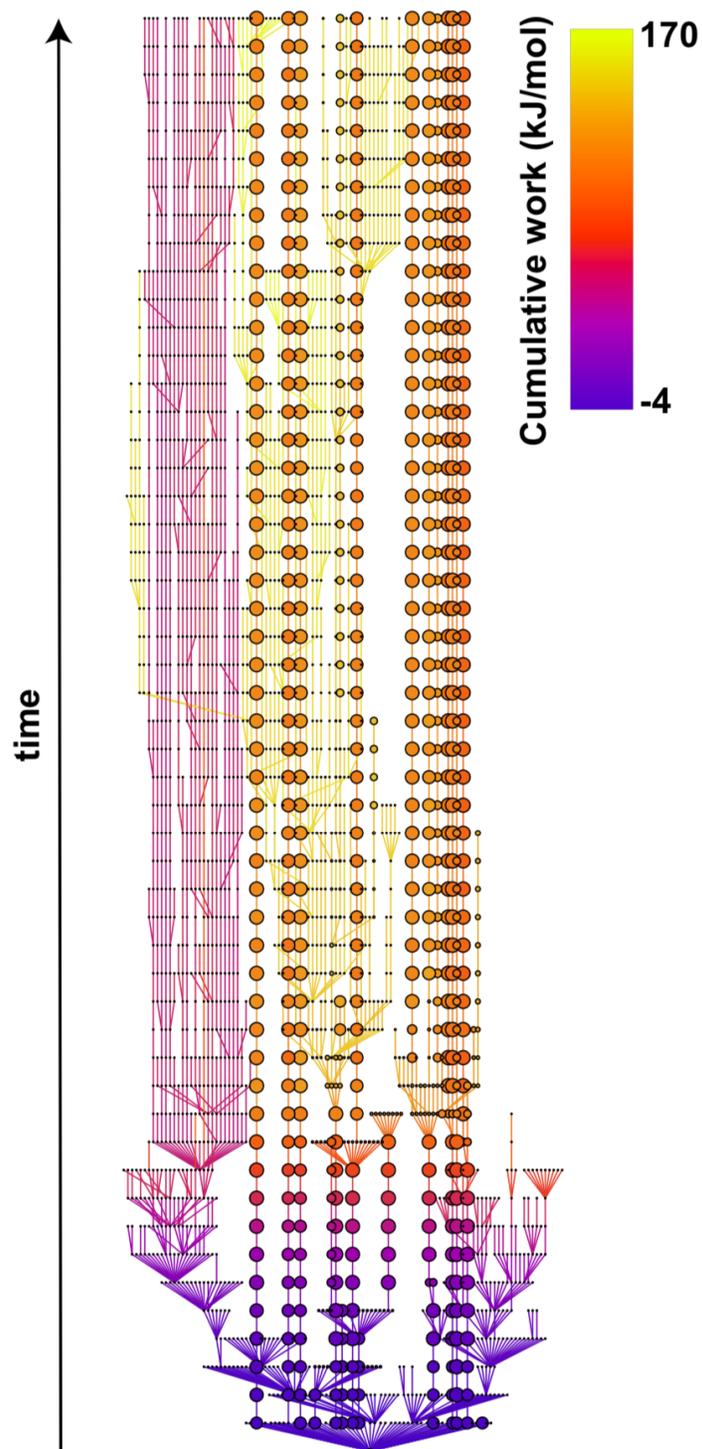


Figure 2.5: A REVO resampling tree. A resampling tree for a single REVO simulation for $\epsilon = 20$. Node color and size correspond to work and walker weight, respectively. For clarity, we show here a 53 cycle simulation, which generally samples a wider distribution of work values than 500 cycle simulations shown in Fig. 2.4A. This figure was made with Gephi[141].

error (RMSE)s for 8000 trajectories are roughly equivalent for $\epsilon = 2$ and 5. For a well depth of 10 kcal/mol, REVO performs significantly better than straightforward dynamics after 400 trajectories. For $\epsilon = 20$, higher error is observed for both methods throughout, although REVO outperforms SF at long times, obtaining a final RMSE (0.49 kJ/mol) that is roughly half of the final value obtained with straightforward dynamics (0.78 kJ/mol).

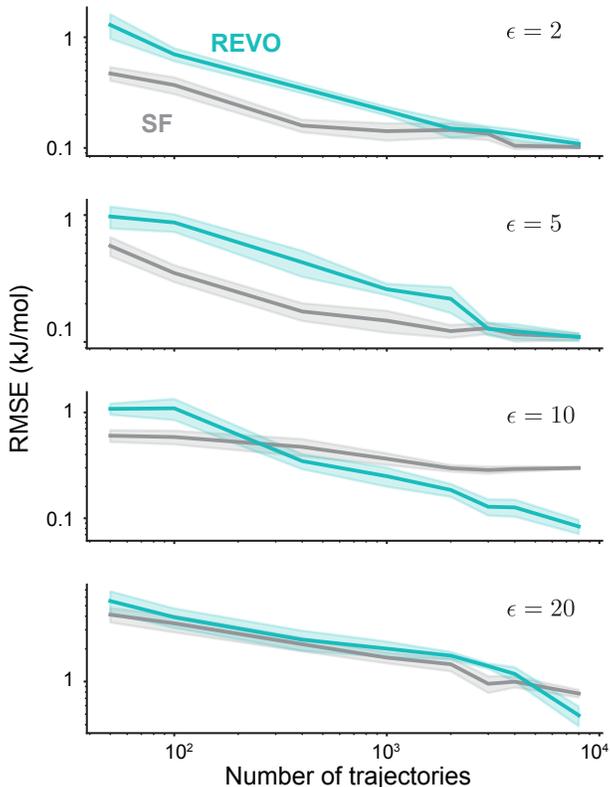


Figure 2.6: Error vs Time. The convergence of the binding free energy for $\epsilon =$ (A) 2, (B) 5, (C) 10, and (D) 20 kcal/mol for straightforward (SF) and REVO as a function of the total number of walkers used to calculate each free energy value. For REVO simulations an ensemble of 50 was used, and multiple simulations are combined to reach the number of walkers shown. The shaded areas show the uncertainty calculated at each point using standard error of the mean over 15 trajectory sets.

Importance resampler results

Although REVO is able to broadly enhance sampling of the work distribution, we investigate whether a method that specifically amplifies sampling of walkers at the peak of the importance distribution would more efficiently calculate ΔF . The Importance resampler is

described in Section 2.2 and uses a parameter, μ , to govern how deeply it samples low-work trajectories. Five sets of 200 independent, 50 walker simulations were run for multiple values of μ , for each ϵ analyzed above. Datasets were generated for simulations with both 53 cycles and 500 cycles and setup as described in Section 2.2. We found that values of $\mu < 1$ perform best, with the RMSE gradually increasing with μ for simulations with 53 cycles. Errors were consistently within a small range for increasing values of μ for 500 cycle simulations (Fig. A.1). However, these errors were in general larger than those observed for both SF and REVO for similar numbers of trajectories.

There is variability across different sets of trajectories obtained with the same value of μ as shown in Fig. A.2 for $\mu = 0.1$. This is true even for very large trajectory sets. We find that this is due to differences in the sampling of the work distribution as shown for two different, 10000-trajectory sets in Fig. 2.7. In Fig. 2.7A, the free energy surfaces for two sets of data for $\epsilon = 20$ with $\mu = 0.4$ are compared, with one set outperforming the other with a lower overall RMSE. The work distributions for these datasets (Fig. 2.7B) show differences in work values below ~ 65 kJ/mol, and it is seen in 2.7C, that these low-work values have very high importance (consistent with results in Fig. 2.4B). The dataset with higher resolution in the low-work tail (blue) has lower RMSE in the free energy surface.

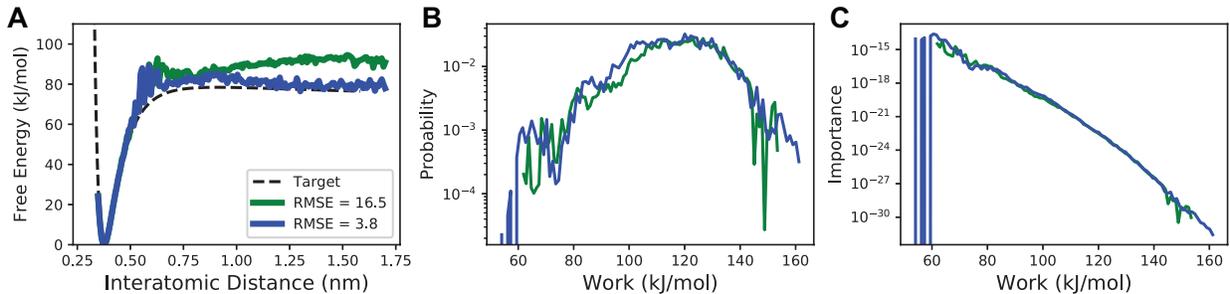


Figure 2.7: Comparison of two importance sampling trajectory sets with $\mu = 0.4$. The (A) free energy surfaces, (B) work distributions, and (C) importance curves for two sets of 53 cycle simulations.

For 500 cycle simulations, Importance resampling performed the worst among the three resampling strategies examined here, for all values of ϵ . However, we wanted to investigate

whether it could be useful for shorter simulations, in systems where generating sufficiently long trajectories is computationally demanding. Figures A.2 and A.3 analyze the efficiency of the convergence of the error of free energy profile ($\Delta\Delta F$) in comparison to straightforward dynamics and REVO for 53 cycle simulations. Broadly, Importance resampling performs with comparable efficiency to REVO and straightforward using 53 cycle simulations, although it performs poorly for $\epsilon = 20$ kcal/mol. Individual values of μ were found to outperform both SF and simulations for each value of ϵ ; however, the performance was inconsistent between different independent sets of 200 simulations with the same value of μ as shown in Fig. A.2 for $\mu = 0.1$.

To show the differences in work distributions sampled by all methods as well as the effect of μ in the Importance resampler, work distributions and corresponding importance curves for 53 and 500 cycle simulations are shown in Fig. 2.8. Each curve shown is an average over 1000 trajectories each. In the 53 cycle simulations the Importance resampler can more thoroughly sample the tails of the work distribution, with a depth that can be tuned by μ . This results in a bimodal distribution of work values with one group of low-work, low-probability trajectories, and another group of average-work, high-probability trajectories. As μ increases, the separation between the two halves of the bimodal distribution increases with peaks moving further into both the low and high work values. However, for 500 cycle simulations, the Importance resampler only samples a limited range of work values, similar to that of straightforward dynamics. For longer simulations, the peak of the importance curve is not reached for all values of μ .

To further visualize μ 's effects on resampling, Figure 2.9 shows resampling trees for $\mu = 0.5$ and 5.0. For low values of μ , nearly all walkers have high final values of work. In this case, the majority of these end walkers originate from a single early cloning event of a high work walker roughly 1/3 of the way through the simulation. At this point the weight of the low-work walkers decreased enough that the high-weight, high-work walker had a larger $i(W)$ value. This does not accomplish the original goals of importance-based resampling,

and results in final distributions that are similar to those obtained with straightforward resampling.

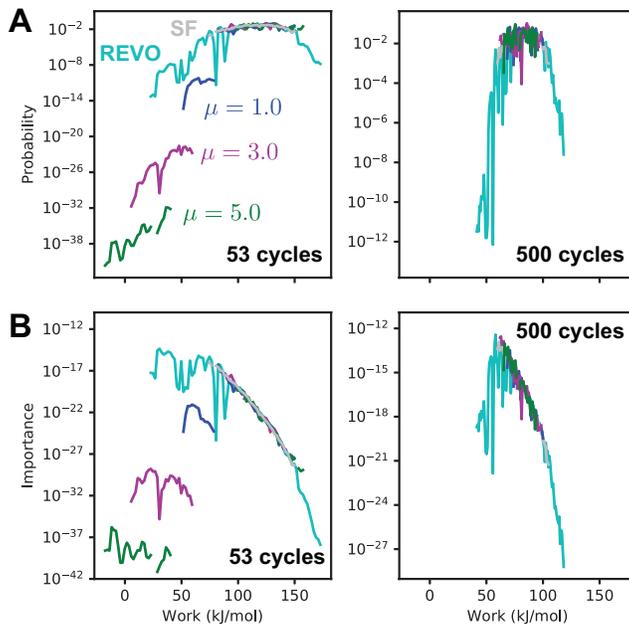


Figure 2.8: Work and importance curves for 53 and 500 cycle simulations. The work curves for (A) 53 and (B) 500 cycle simulations is shown for REVO, SF and Importance resampling with three different amplification factors (μ). Corresponding importance curves shown in (C) and (D).

For high values of μ there is a stark separation of the weight-based and importance-based walkers beginning early on in the simulation. The weight-based walkers maintain high values of work throughout the remainder of the simulation whereas the value of work for the importance-based walkers increases and then decreases over time. For high μ , many cloning events also occur en masse from single walkers in some cycles. High μ , 500 cycle simulations have very limited sampling of the work distribution, unlike 53 cycle simulations. Although resampling patterns for these simulations initially resemble their 53 cycle counterparts, eventually many cloning events from the high-weight, high-work walkers overwhelm the low-work walkers. This results in final work distributions that resemble straightforward data (Fig. 2.8B) and again is not able to accomplish the goals of importance-based resampling.

Comparison with Diffusion Monte Carlo

To analyze the relative efficiency of REVO in comparison to Diffusion Monte Carlo, we study convergence towards the target free energy profile, as above in Fig. 2.6. For both methods 800 simulations with 50 walkers each are run with $\epsilon = 20$ kcal/mol. All simulations are 500 cycles in length with 100 steps per cycle. REVO simulations were run as described in Sec. 2.2, and Diffusion Monte Carlo simulations were run in a similar fashion, but using the Diffusion Monte Carlo resampling algorithm and free energy calculation strategy described in Sec. 2.2. Resampling was done every cycle, which corresponds to a relative entropy cutoff of 0.

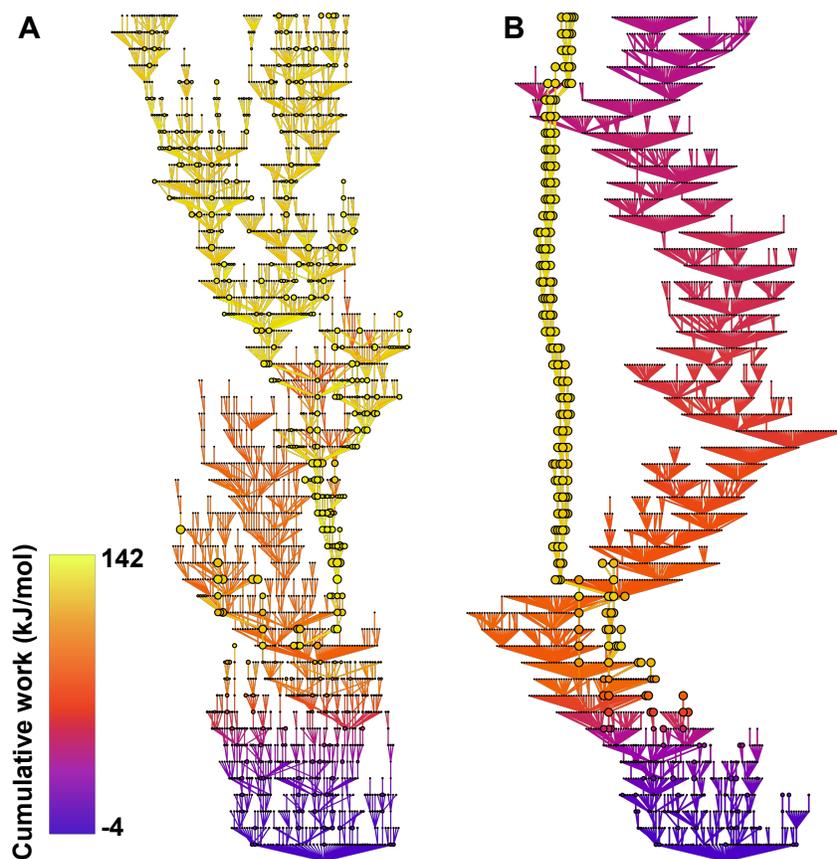


Figure 2.9: Resampling trees for a low and high value of μ . Resampling trees for $\mu =$ (A) 0.5 and (B) 5 for $\epsilon = 20$ simulations with 53 cycles. Node color and size correspond to work and weight, respectively.

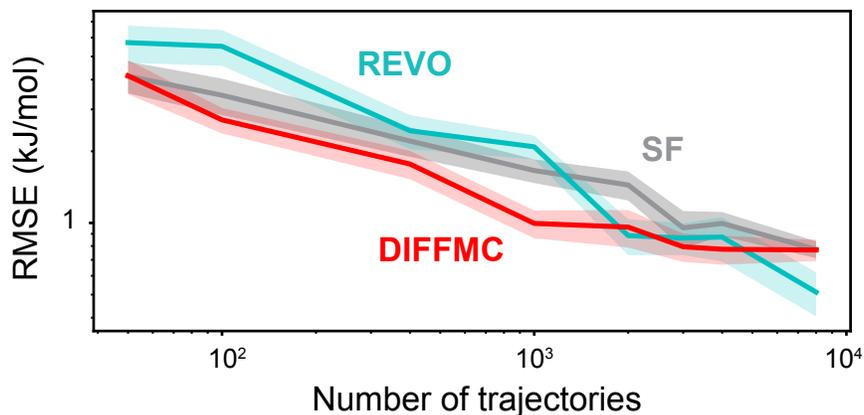


Figure 2.10: Error vs Time for REVO and Diffusion Monte Carlo. The convergence of the binding free energy for $\epsilon = 20$ kcal/mol for straightforward (SF), REVO, and DIFFMC as a function of the total number of trajectories used to calculate each free energy value. For REVO and DIFFMC simulations an ensemble size of 50 was used, and multiple simulations are combined to reach the number of trajectories shown. The shaded areas show the uncertainty calculated at each point using the standard error of the mean over 15 independent trajectory sets.

DIFFMC out-performs both straightforward dynamics as well as REVO for low numbers of trajectories. While Diffusion Monte Carlo performs well for smaller datasets, the rate of RMSE decrease appears to slow for larger trajectory sets, whereas REVO and straightforward dynamics continue to gradually decrease. After this point, Diffusion Monte Carlo and straightforward dynamics obtain similar final values of 0.77 ± 0.08 and 0.78 ± 0.06 kJ/mol, respectively with REVO obtaining a final value of 0.51 ± 0.10 kJ/mol.

2.4 Discussion

The Jarzynski Equality generally requires a very large number of realizations of a process to converge the free energy due to the low probability of observing high-importance work values. Due to the potentially large computational cost, this is a major limiting factor in the use of the nonequilibrium work theorem for calculating free energies. The results above demonstrate that thorough sampling of the tails of work distributions can increase the accuracy of free energy calculations, potentially unlocking the promise of the Jarzynski Equality. This can be achieved through the use of a promising combination of methods – weighted ensemble

rare-event algorithms with short duration nonequilibrium simulations – that efficiently generates a work distribution with low-probability tails. A history-dependent REVOresampler generally outperformed direct straightforward sampling of unbinding trajectories especially for dissociation events with large free energy barriers.

We also examined a novel Importance-based resampler that showed sporadic success using short trajectories, but generally failed to work as designed. The efficiency of the Lennard-Jones pair allowed us to run many simulations at increasing amplification factors for the Importance resampler. While the inclusion of the amplification factor allowed for further exploration into the tails of the work distribution, it presented a puzzling result. As amplifications increased beyond ~ 1 , the accuracy of free energy calculations began to decrease for shorter simulations (53 cycles in duration). As is shown in Fig. 2.9A, although the importance resampler is able to sample trajectories with extremely low work values, in practice it underestimates the probability of these trajectories. This occurs at higher amplifications because the resampler picks one or two high-importance trajectories each round to be cloned repeatedly, although these often have a change in work that results in a lower importance in the next cycle. In contrast, the REVO algorithm only seeks to diversify the work values at each cycle, picking not only the highest importance walkers but a more diverse group. This results in a more thorough sampling of the whole work distribution, and indicates that the REVO resampler will be more efficient to work with and develop further moving forward.

REVO is also compared to an existing Diffusion Monte Carlo algorithm and was found to show smaller errors in free energy for large ensembles of trajectories for the $\epsilon = 20$ kcal/mol system. In these calculations, for comparative purposes we used the same parameters between the REVO, Importance Resampler and Diffusion Monte Carlo algorithms, whenever possible. In theory, a more exhaustive search of parameter space, such as the use of a nonzero entropy trigger may improve the results obtainable by Diffusion Monte Carlo. Larger trajectory sets may also be beneficial for improving results, as it has been shown that

the (intrinsic) RMSE of Diffusion Monte Carlo goes to zero as the number of trajectories goes to infinity[132]. It could be surprising that REVO, which enhances sampling of the entire work distribution, including higher-work trajectories, would outperform an algorithm that explicitly focuses only on those trajectories that are most important for the calculation of the work distribution. However, like the Importance Resampler, Diffusion Monte Carlo is making predictions of the overall importance of a trajectory based on incomplete work histories. These results imply that maintaining a set of trajectories with diverse work values can result in better sampling of low work trajectories at the end of the simulation. We expect this behavior to be even more important when moving to protein-ligand systems with more orthogonal degrees of freedom.

We note that there is an interesting connection between the amplification factor used in the Importance resampler and large deviation theory, in which paths are re-weighted according to a factor e^{-sA} , where A is an activity measured along a path and s is the strength of an applied field [142]. This has been used previously to study dynamical phase transitions, where the $\langle A \rangle$ or its derivative shows a discontinuity as a function of s , that give insight into phenomena such as glass transitions [143, 144], synchronization to external fields [84], and kinetic trapping during protein folding [145]. Methods such as Transition Path Sampling can apply the e^{-sA} field directly in their path acceptance step and sample paths within a given s -ensemble. In contrast, in the Importance resampler only a subset of the weighted ensembles are chosen according to $I(W)$. This said, the possibility of a dynamical phase transition of $\langle \beta W \rangle$ in response to μ is an intriguing direction for future work.

To further improve upon the promising results of the REVO resampler for work calculations, new distance metrics can be created that focus on increasing the cloning of the low work values exclusively. Based on these results we also anticipate that traditional weighted ensemble binning of the work distribution would also perform well. However, unlike REVO, suitable bins would need to be defined along the work axis, which would be different for each

system and would be difficult to predict beforehand. In addition, sampling efficiency and accuracy could be improved by performing simulations in both the forward and reverse direction [146]. The potential for REVO to accelerate free energy calculations within a bi-directional framework will be studied in future work.

Here we use the notion of a trajectory “activity”, which is a history-dependent observable associated with each trajectory. This is implemented with the WEPY software package and can be written to be system specific and work with either a distance metric (e.g. in the REVO algorithm) or directly with a specific resampler (e.g. in the Importance resampler). The utilization of activities in weighted ensemble resampling could find use in systems looking at mobility. For instance, glass transitions could be studied using an activity measuring mean squared displacement, and ion transport could be studied using an activity measuring ion flux through a pore. Another example of a trajectory activity is the classical action, which could be used to calculate probabilities of trajectories[147].

In this work we examine a very simple two-atom system with umbrella potentials defined by the inter-particle separation. For larger, more complex systems (protein-peptide, protein-ligand), there can be many possible (un)binding pathways [148, 149]. While general reaction coordinates, such as the inter-particle separation, can be used to describe the unbinding process, the efficiency of generating low-work trajectories will likely depend on the proper choice of reaction coordinate. Fortunately, the combinations of methods presented here offer many opportunities for modification. For instance, the functional form of the pulling force could be changed, such as making it a double-well potential that gradually destabilizes the bound state. The REVO resampling algorithm could also take into account distances between unbinding trajectories to promote a diversity of transition paths as well as work values. Alternatively, instead of restraint potentials, small, randomly oriented forces can be applied to an unbinding ligand, such as in the Random Accelerated Molecular Dynamics (RAMD) method [150]. This could allow for efficient free energy calculations while requiring no previous knowledge of the unbinding pathway.

CHAPTER 3

LOCAL ION DENSITIES CAN INFLUENCE TRANSITION PATHS OF MOLECULAR BINDING

This work is reproduced from “Nicole M. Roussey and Alex Dickson, Local Ion Densities can Influence Transition Paths of Molecular Binding, *Frontiers in Molecular Biosciences*, 9, (2022) <https://doi.org/10.3389/fmolb.2022.858316>”[26]. The work is presented here as published except that the supplemental figures are included in the appendix.

3.1 Introduction

Atomistic simulations are a broadly used method to better understand the microscopic interactions that govern ligand binding and unbinding and to calculate critical values such as transition rates and free energies. Both rates and free energies can in principle be computed with straightforward molecular simulations, starting in either the bound or unbound state. However, the cost required to simulate binding transition paths is typically prohibitive due to high energetic barriers separating the bound and unbound states. To overcome these barriers, a variety of enhanced sampling techniques can be employed, which commonly require the use of a predefined reaction coordinate: a single collective variable that describes the progress of the (un)binding reaction.

The use of proper reaction coordinates can lead to improvements in the convergence of free energies for enhanced sampling methods [63] and is necessary for accurate path-based free energy calculations in biological systems [151]. Many methods have been developed to seek out optimal reaction coordinates including but not limited to VAMPnets [64], DiffNets [61], Deep-TICA [62], SGOOP [63], and AMINO [60]. All of the above methods construct a reaction coordinate from a set of candidate features that are either predefined or require user intuition of the (un)binding process.

Significant effort has been dedicated to understanding the role of water in the ligand (un)binding process, including binding pocket solvation effects and bulk and single molecule effects [152, 153, 154, 155]. Water molecule density has been included in reaction coordi-

nates through the utilization of Deep-LDA [156]. This method successfully found a complex reorganization of the water structure in unbinding for use as a reaction coordinate and has been able to produce accurate binding free energies [155]. The role of ions along molecular binding pathways is much less understood. Ion distributions surrounding molecules such as double stranded DNA [33] and RNA [34] have been studied and it has been found that ion affinity for molecules such as cyclodextrins and DNA is dependent on the force field used [35] as well as the water model employed [33]. A difference in unbinding rates has been found between implicit and explicit ions in simulation, with implicit ion representations overestimating unbinding rates across a broad range of ion concentrations [157]. However, it appears that little is known about the effects of changes in ion densities along ligand (un)binding pathways.

Recent studies have demonstrated that adaptations of the WE method [68, 69, 32] can efficiently generate ligand binding and unbinding pathways that can then be used to determine rates and binding free energies [74, 9, 158]. Specifically, an extensive analysis was conducted on a series of host-guest systems containing small, organic guest molecules (“G3” and “G6”) interacting with “octa-acid” hosts (“OA”) (Fig. 3.1), which were originally part of the SAMPL6 (Statistical Assessment of the Modeling of Proteins and Ligands) SAMPLing challenge [94, 90]. The REVO variant of the weighted ensemble method allowed for efficient generation of large numbers of binding and unbinding events without employing biasing forces that could perturb the (un)binding mechanism. This is notable as mean first passage times of unbinding ranged up to hundreds of seconds for these systems. It accomplishes this by running an ensemble of trajectories and periodically “resampling” this ensemble to shift computational emphasis toward unique trajectories that are moving towards a target state, and adjusting the probabilities of the trajectories accordingly. As a result, each unbinding pathway has an associated statistical weight (ranging from 10^{-12} to 10^{-6}) that governs how strongly it contributes to the calculation of observables, including the unbinding rate constant, k_{off} .

During these resampling steps, only the geometric relationship between the host and guest molecules was used; the positions of the water molecules and ions were neglected. Here, a time- and probability-dependent analysis of solvent based features including water and ions is presented for unbinding trajectories from the OA-G3 and OA-G6 SAMPL systems. We explore the significant differences in guest-ion interactions between high- and low-probability unbinding events, also referred to as “exit points”, as well as differences in spatial arrangements of ions during unbinding. In these simulations, we have found that the generation of the most probable reactive paths requires fluctuations toward low ion densities within certain regions of the simulation box, particularly in the space immediately above the binding pocket. Differences in these ion densities along transition paths are associated with up to 10^6 -fold differences in unbinding probabilities, which motivates the future inclusion of ion densities in (un)binding progress variables.

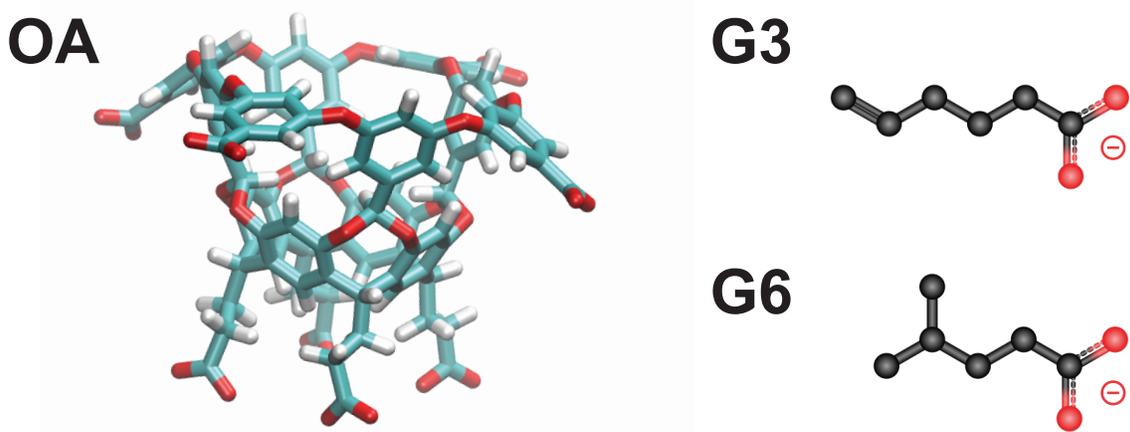


Figure 3.1: The OA-G3 and OA-G6 systems. The OA host molecule (left). The G3 (top right) and G6 (bottom right) guest molecules.

3.2 Materials and Methods

Weighted Ensemble Sampling

The simulations analyzed here were previously generated [74, 158] with a variant of the WE [68] method called "REVO" [32] utilizing the Wepy [75] software package. A generalized

framework for WE is as follows (Fig. 3.2). WE uses an ensemble of trajectories that are evolved forward in time in a parallel fashion. Each trajectory carries with it a statistical weight (w) that governs the extent to which it contributes to ensemble averages. Generally, WE simulations include two main steps: 1) An MD simulation step that moves trajectories forward in time by a predetermined time interval, and 2) a resampling step that include cloning and merging operations. Resampling is designed to both use cloning to increase the number of trajectories that have a desirable value for a feature of interest, and to decrease redundancy by merging trajectories that are similar based on the feature of interest. Together, this process aims to diversify the trajectories within the ensemble with the goal of increasing the probability of sampling rare or long-timescale events of interest for a given system. When cloning a trajectory, two new independent trajectories with the same conformation are created with half the probability, or weight (w) of the original. Merging two trajectories A and B leads to the creation of trajectory C with weight $w_c = w_a + w_b$. Trajectory C inherits either conformation A or B with a probability proportional to w_a or w_b , respectively.

A central feature of a WE simulation is the resampling function (also referred to as a “resampler”) that determines which trajectories are selected for cloning and which are selected for merging. The resampler takes in an initial set of trajectories and returns a new set, which is the outcome of a series of merging and cloning steps following the rules described above. These new trajectories thus have conformations that are a subset of the initial conformation set and the sum of trajectory weights is unchanged (typically equal to 1).

In order to determine transition rates, these WE simulations were run in a nonequilibrium ensemble, where trajectories are created in the bound state and terminated in the unbound state. The unbound state was defined using a boundary condition boundary condition (BC) that is satisfied when the minimum host-guest distance is greater than 1.0 nm, following previous work [148]. When the BC is reached, the trajectory contributes to the reactive flux calculation according to its weight at the time of crossing, which we refer to as its “exit point probability”. The exit point probability can be anything between the minimum

and maximum values set when the simulation was run. An exit point or unbinding event being considered “high-weight” or “low-weight” is relative, with this being dependent on the weights of all exit points within the dataset. The weights of trajectories vary because they are changed during the resampling steps that are done between rounds of dynamics in the weighted ensemble algorithm.

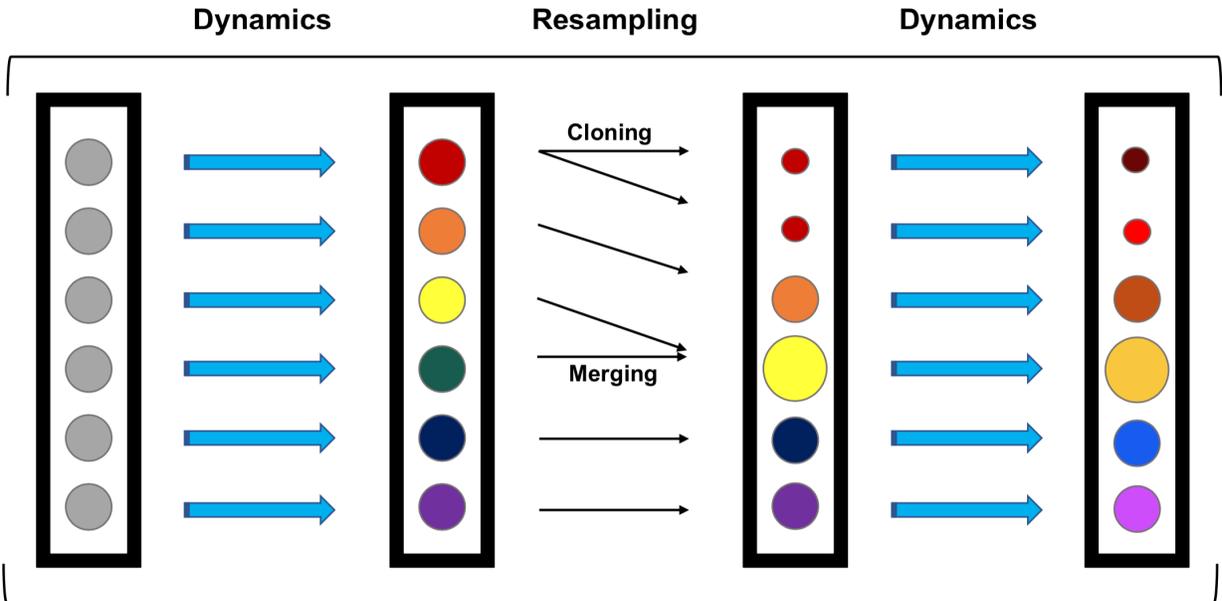


Figure 3.2: General WE Framework. Every circle represents a trajectory in the ensemble. Colors represent conformations and circle size represents probability, with all trajectories beginning with the same conformation and probability. Trajectories are run for a predetermined number of steps (dynamics), followed by a resampling step containing merging and cloning procedures. This cycle repeats until the end of the simulation.

Resampling of Ensembles by Variation Optimization (REVO)

REVO [32] is a resampling algorithm for use with Wepy that works by maximizing a function called the **trajectory variation** (V). V is a scaled sum of the all to all pairwise distances between trajectories in the ensemble (Eq. 3.1), where d_{ij} is the distance between trajectory i and trajectory j and V_i is the variation for trajectory i .

$$V = \sum_i V_i = \sum_i \sum_j \left(\frac{d_{ij}}{d_\star} \right)^\alpha \phi_i \phi_j \tag{3.1}$$

The measurement of distance between two trajectories can be arbitrarily defined in the REVO method. In this case it was defined as the root mean squared deviation of the ligand after aligning the host molecules. As the host molecules have four-fold symmetry, four separate distances were calculated after aligning the hosts in the four symmetrically-equivalent positions, upon which the smallest such distance was used for d_{ij} . ϕ_i is a non-negative function referred to as a “novelty” that signifies the importance of individual trajectories. In this work it was solely a function of walker weight. d_* , the “characteristic distance” is the average distance after one cycle of dynamics, and is only used to make the variation function unitless. The α parameter balances the value of the distance and novelty terms and was set equal to 4. Other methodological details pertinent to data generation are available in Ref. [74] and Ref. [158]. The overall goal of REVO is to optimize the value of V by cloning trajectories with a high value of V_i and merging trajectories with a low value of V_i . See Ref. [32] for more details of the REVO method.

Dataset information

The weighted ensemble data used for this analysis comes from papers published in 2020 [158] (the primary OA-G6 data set) and 2018 [74] (OA-G3 data set and a secondary OA-G6 data set). Briefly, the primary OA-G6 data set contains 10 simulations with 48 trajectories each and 1500 cycles per trajectory that begin in the initial OA-G6-0 pose provided in the SAMPL6 SAMPLing challenge [90]. The 2018 data sets contain five simulations each with 48 trajectories and 2000 cycles per trajectory, each beginning at one of the five initial poses for the corresponding system. Reactive paths begin in the bound state and end in the unbound state when a BC is hit. The BC is defined as a 1.0 nm minimum distance between the host and guest molecules.

3.3 Results

We find that each reactive path can be split into two phases: i) initial departure from the bound state, and ii) full separation of the host and guest. There are often many cycles

between the guest physically leaving the binding pocket of the host and the BC being hit. It was determined that in all of the reactive paths generated, a COM-to-COM distance of 0.7 nm indicated an irreversible transition between these two parts (Fig. B.1). This can be seen as a physical “commitment to unbinding” point after which rebinding does not occur, where the guest has just been released from the partially solvated binding pocket (Fig. 3.3A). The cycle corresponding to this point is found for all reactive paths and used for analysis; we refer to this point as t_0 .

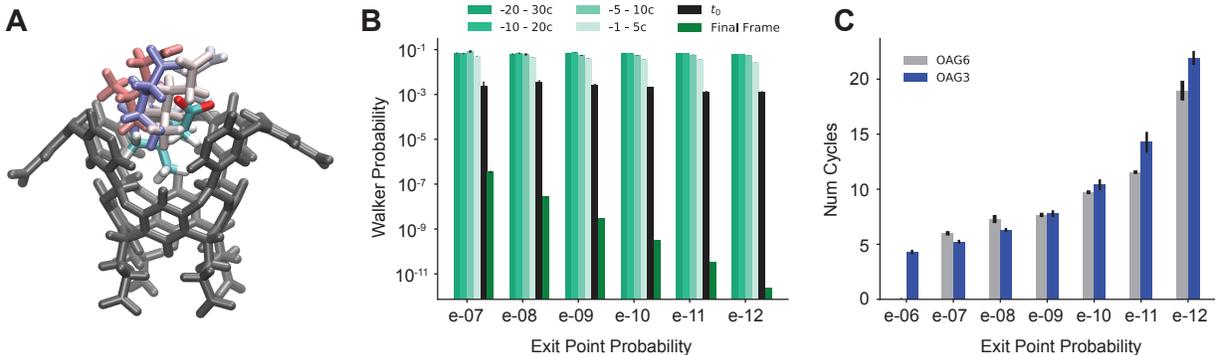


Figure 3.3: Analysis of t_0 poses. A) The OA host molecule with the G6 ligand in the starting pose (multi-color) and example t_0 poses (pastels). Some atoms from the host have been removed in A for clarity. B) Average probabilities from -30 cycles to the final unbinding event organized by unbinding probability for the 2020 OA-G6 data set. C) The average number of cycles between t_0 and the unbinding event for OA-G3 (blue) and the 2020 OA-G6 data set (gray) organized by unbinding probability.

When the BC is hit for the reactive paths, the unbinding probabilities varied between 10^{-12} and 10^{-6} for OA-G3 and between 10^{-12} and 10^{-7} for OA-G6. The low probability exit points are highly abundant for both OA-G6 and OA-G3, whereas the high probability exit points occur with a very low frequency for both systems.

Overall, the number of exit points *increases* as the probability of the exit points *decreases* (Table 3.1). Before the t_0 point, the probabilities of the reactive paths are roughly the same, with a value of 10^{-3} with only the probabilities following t_0 varying based on exit point probability (Fig. 3.3B). The number of cycles between t_0 and the unbinding event also correlates to the exit point probability, with high probability exit points having ~ 5 cycles between the two points, and ~ 20 cycles for low probability exit points (Fig. 3.3C). There is

Table 3.1: The number of observed unbinding events grouped by exit point probability. The OA-G6 row corresponds to the OA-G6 2020 dataset.

Number of Unbinding Events							
System	10^{-6}	10^{-7}	10^{-8}	10^{-9}	10^{-10}	10^{-11}	10^{-12}
OA-G6	0	7	17	112	359	652	2220
OA-G3	10	18	88	195	483	1116	4103

a steady increase in the average number of cycles between t_0 and the unbinding event as the probability of the trajectories decreases.

These differences prompt the question: are there differences in physical features associated with this large variation in exit point probability? To answer this question, a set of physical features was chosen and the values of those features were calculated for every cycle of every reactive path generated. The features in question include the number of waters in the binding site of the host, the number of ions around the upper negative charges of the host molecule, the number of ions around the guest, and the number of waters around the guest molecule (Fig. 3.4A). To calculate these features, a continuous logistic function was used: $f(r) = 1 - \frac{1}{1+(e^{-S(r-r_0)})}$, where r is the minimum atomic distance between the two entities. We use two different sets of values for the interaction radius (r_0) and steepness (S) parameters: $r_0 = 3 \text{ \AA}$, $S = 17$ or $r_0 = 5 \text{ \AA}$ and $S = 12$) (Fig. 3.4A). The sum of $f(r)$ across all ions (or waters) is a continuous count of the number of molecules of that species surrounding the host (or guest) for that cycle.

It was found that some features were consistent or had only a slight variation across all exit point probabilities, such as the number of binding site waters and the number of waters surrounding the guest molecule (Fig. B.2). However, some features were found to show trends that differentiated the high- and low-probability exit points. These features included the total number of positive ions surrounding the upper negative charges of the host (Fig. 3.4B and C) and the number of positive ions surrounding the guest molecule (Fig. 3.4D and E). In both OA-G6 and OA-G3 there is a general trend of the number of ions surrounding the upper negative charges of the host increasing as the exit point probability

increases, although this is not observed for $1e-7$ exit points in the OA-G6 dataset. There is also a clear trend of increasing guest- Na^+ interaction as the exit point probability decreases including before, at, and after the t_0 point. Similar trends were observed for features on the 3 Å scale (Fig. B.3).

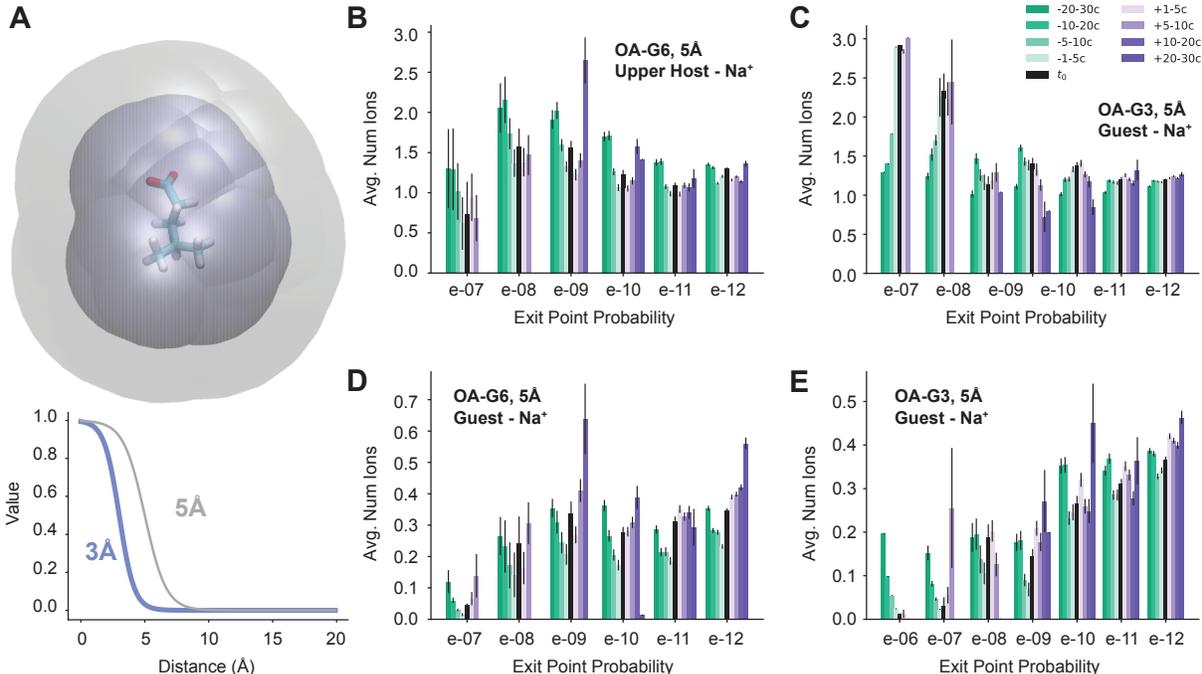


Figure 3.4: Feature Analysis. A) A visualization of the region of space considered for the guest-ion features using the G6 ligand. The maximum distance for the 5 Å scale is in gray and the 3 Å scale is in blue (top). The two logistic functions used to determine the molecule counts (bottom). B-E) Molecule counts for Na^+ ions with results organized by both time and exit point probability. The legend in C applies to all four plots. The average total ion count (5 Å scale) around the upper negative charges of the host for B) OA-G6 and C) OA-G3. The average total ion count (5 Å scale) around the guest for D) OA-G6 and E) OA-G3. OA-G6 results correspond to the 2020 data set.

As we find that the interaction between the guest and Na^+ ions correlates with the probability of the unbinding trajectories, we now examine Na^+ ion densities in the region of space directly above the host. Specifically, we examine a cylindrical region of space beginning immediately above the host and ending at the top of the box (Fig. 3.5A). We find that this region is critical to determine the outcome of dissociation trajectories that have reached t_0 . The autocorrelation of ion density in this region ($C(\tau)$) is surprisingly long-lived; it follows

a single exponential decay with a timescale of 77 ns (Fig. 3.5B).

Fig. 3.5C shows the average number of ions in the cylinder for cycles $[t_0 - 3, t_0 + 3]$ for each reactive trajectory. An average of 1.47 ions was found in upper cylinder space when averaged over all available data (including reactive and non-reactive trajectories). The cylinder ion densities of *reactive trajectories* were found to be significantly lower than the bulk average regardless of exit point probability. A striking relationship was observed between this ion density and the exit point probability that was consistent across all data sets with highly weighted exit point probabilities (Fig. 3.5C), where highly-weighted exit points had a significantly lower average number of ions in the cylinder. Overall, highly weighted exit points had less ions above the host, and subsequently near the guest at t_0 , with this number gradually increasing as the exit point probability decreased (Fig. B.4).

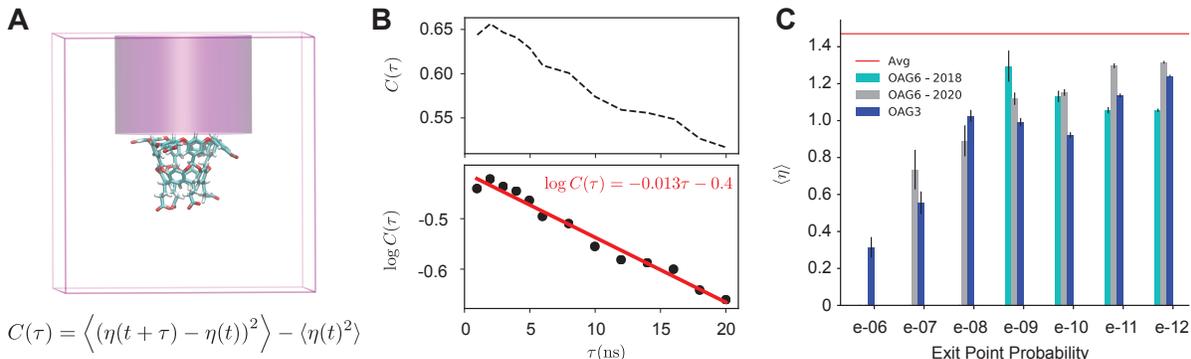


Figure 3.5: Ion Density Analysis. A) A diagram showing the simulation box and the cylindrical space above the host where the number of ions ($\eta(t)$) is determined. The equation for calculating the autocorrelation of this quantity ($C(\tau)$) is shown. B) An autocorrelation plot of the cylindrical ion density ($\eta(t)$) is calculated using all reactive and non-reactive trajectory data. C) The average number of ions in the cylinder space above the host for OA-G3 (dark blue), OA-G6 (2020, gray), and OA-G6 (2018, cyan). The average from 4500 random simulation cycles is shown in red.

To explain these findings, we first analyzed the electrostatic forces on the guest molecule for all OA-G6 reactive trajectories at the t_0 point for one ensemble of the OA-G6 2020 data set. This was done by first removing all forces from the system other than the nonbonded (electrostatic) forces. Then the force on the ligand was determined at key points along the unbinding trajectories and average forces were determined for each exit point probability

group. Results are shown for the initial bound cycles (cycles 0-6) and for the t_0 -surrounding cycles used for the cylinder-ion analysis (Fig. B.5). We find that the net electrostatic force is pushing the guest outward from the host, and that the magnitude of this force is about 20 kJ/mol/Å higher in the initial pose (80 kJ/mol/Å) than it is at t_0 (60 kJ/mol/Å). No significant difference is found for exit points of different weights for both the overall magnitude of the electrostatic force or the z-axis contribution to the force (B.5 A and B). We found no significant difference at t_0 across all exit point probabilities despite the difference in cylinder ion-occupancy.

An alternative explanation is that differences in occupancy change the likelihood of ion interaction after the t_0 point. This is consistent with our observations in Fig. 3.4D and E and would increase the number of cycles required to hit the BC (Fig. 3.3C) as well as the extent of their exploration of the simulation box. Exit point locations were determined for both the highest and lowest probability exit points for both OA-G6 (10^{-7} and 10^{-12}) and OA-G3 (10^{-6} and 10^{-12}). For both systems, it was found that for high probability exit points, most guest molecules reach the BC directly above the host, whereas the low probability exit points hit the BC at a wide distribution of points surrounding the host molecule (Fig. 3.6).

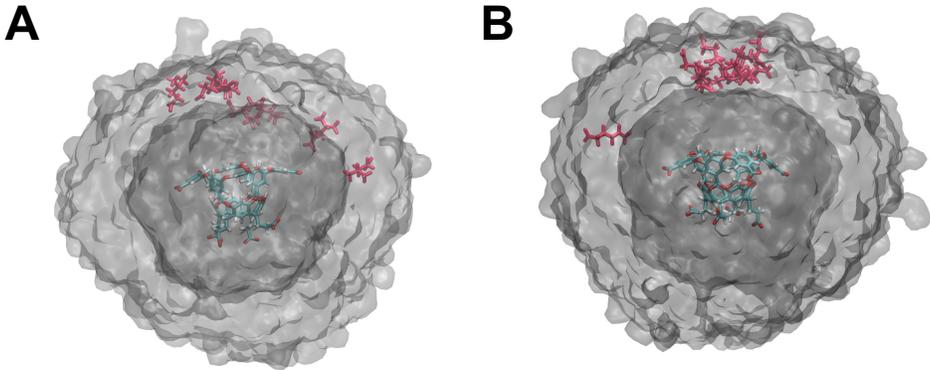


Figure 3.6: Exit Point Analysis. A) Unbinding event locations for exit points with probabilities 10^{-7} (red) and 10^{-12} (gray VolMap) for OA-G6. B) Unbinding event locations for exit points with probabilities 10^{-6} (red) and 10^{-12} (gray VolMap) for OA-G3. The surfaces show a density contour (Isoval) of 0.0001 in both panels.

3.4 Discussion

In summary, we find that location-dependent ion densities play a significant role in the unbinding process for the OA-G6 and OA-G3 systems. These systems are widely used for both the testing and development of force fields and numerous computational methods [90, 94, 74, 159, 160, 161] necessitating a thorough understanding of the mechanics of their unbinding. It is likely that ion densities play such a prominent role due to the charged nature of these systems (-8 for the host and -1 for the guest). Similar effects might also be observed in biological systems with even more significant charge densities such as calsequestrin [162], a protein necessary for muscle relaxation/contraction, with a net charge of -64, as well as systems with nucleic acids, which have a charge of -1 per nucleotide.

Further exploration and utilization of the effects of ion densities on ligand (un)binding could be done via various methods. Constraints on spatial densities of ions could be included in simulations to further examine the relationship between ion densities and unbinding rates or free energies. One possible strategy would be to conduct 2D Umbrella Sampling [163, 164] simulations that include a direct descriptor of (un)binding, such as the host-guest center-of-mass distance, and the ion density added as a second collective variable. Ion densities (and other features of interest) could also be utilized for resampling purposes for weighted ensemble simulations for the determination of distances between trajectories. This could encourage cloning operations of trajectories with ions in desirable locations, potentially allowing for more efficient generation of high probability unbinding events.

In weighted ensemble sampling, the equilibrium probability of a state is obtained by summing over the weights of all trajectories that have visited that state. This is similarly true for reactive paths: the overall probability of a path is determined by a weighted sum of trajectories. The analysis above breaks down a reactive trajectory set by weight, but it is important to note that relationship between the weight of a trajectory and the probability of the corresponding reaction path is not one-to-one. While high-weight trajectories in general sample from high-probability regions of space, it is possible that a low-weight trajectory could

visit a high-probability reaction path. For this reason, we should consider the low-probability trajectories (e.g. $p = 10^{-12}$) as a heterogeneous group that could contain observations of high-probability reaction paths. However, the high-weight trajectories (by definition) correspond only to high-probability paths.

Overall, these results suggest that greater attention may be required for ligand-ion interactions across various simulation methods, including those that require a predefined reaction coordinate. We find that there are many microscopic trajectories that contribute to the unbinding path ensemble, some of which are much more likely than others. Methods that only sample unlikely reactive paths could have difficulty computing accurate measurements of transition rates and free energies. In addition, incorrect transition states (including inaccuracies in solvent degrees of freedom) can lead to incorrect hypotheses about the molecular interactions that govern kinetics. This work underscores the importance of proper consideration of ion densities along unbinding pathways, especially for charged systems.

CHAPTER 4

QUALITY OVER QUANTITY: SAMPLING HIGH PROBABILITY RARE EVENTS WITH THE WEIGHTED ENSEMBLE ALGORITHM

This work has been published as a pre-print and is currently under review. The work is presented here as submitted except that the supplemental figures are included in the appendix.

4.1 Introduction

The prediction of ligand (un)binding free energies and rates is important to make systematic improvements in potency during the pre-clinical stages of drug design [165]. Over the years, numerous computational methods have been developed to predict both relative and absolute binding free energies (and more recently, their binding and unbinding rates, k_{on} and k_{off} , as well) [48, 166, 88, 114, 167, 168, 127, 169, 170]. These methods make different assumptions that affect accuracy [1, 171] and methodological errors are compounded by more general sources of error such as approximations in force-fields or the choice of water model. The performance of free energy predictors can be rigorously examined by comparison with experimental quantities [172, 173]. However, reported retrospective predictions are often overly optimistic. A more authentic way of assessing these tools is through participation in blind challenges such as the Statistical Assessment of Modeling of Proteins and Ligands (SAMPL): a series of challenges for the prediction of quantities such as unbinding free energies for small molecule (host-guest) pairs [90, 94]. In this challenge, the computational predictions are made without knowledge of the experimental results. Methods for the determination of free energies used in the SAMPL challenges vary from path sampling algorithms such as weighted ensemble [90, 74] to alchemical perturbation techniques such as double decoupling methods with thermodynamic integration or Hamiltonian replica exchange [90].

While there have been significant improvements in the methods used to determine free energies, all methods come with limitations and drawbacks. Even for small systems including the host-guest pairs provided by the SAMPL challenge, computationally calculated values of

ΔG for processes such as ligand (un)binding are often inconsistent, and can vary up to several kcal/mol from experimental reference values [90, 94]. Efforts to determine “computational reference values” using extensive sampling for a given set of conditions (e.g. force field, binding pose, temperature) are important to isolate methodological sources of error for competing approaches. In the SAMPL6 challenge, the results of the OpenMM/HREX method – double decoupling with Hamiltonian Replica Exchange – were used as the computational reference value and were primarily used for the determination of relative efficiencies of different methods. It was found that even for simulations with seemingly converged predictions for ΔG and nearly identical simulation parameters, results varied up to ~ 1.0 kcal/mol, and while sources of methodological error were determined, the exact source of these discrepancies was unknown.

For alchemical methods, a major bottleneck is achieving adequate sampling of different binding modes [174]. In addition, include the residual charges of vanishing atoms in single topology methods or overlap of groups in the system in double topology models can also be significant sources of error [1]. Alchemical methods also have the major drawback of only determining free energy differences between the points of interest, and subsequently do not provide free energy profiles for the unbinding/transition pathway of interest. Physical sampling methods, such as metadynamics [106, 107], umbrella sampling [103, 44], and weighted ensemble (WE) [68, 69, 32, 175, 76], can determine these free energy profiles, but are limited by the determination of collective variables and convergence issues. In particular, the WE method allows for the generation of continuous trajectories between points of interest without employing any biasing forces. It does this by iteratively “resampling” a collection of otherwise independent trajectories, referred to as “walkers”, each of which carrying a statistical weight. The resampling process clones walkers in underrepresented regions of interest – dividing their weight evenly among the clones – and merges walkers in overrepresented regions – combining their weights in order to conserve probability. The use of the “REVO” resampler (Resampling of Ensembles by Variation Optimization) [32] builds

upon these benefits as it resamples trajectories based upon their distances to one another, instead of constructing a set of bins. This makes it easier to sample high-dimensional spaces of collective variables, while avoiding exponential increases in simulation time associated with regularly-spaced bins.

Weighted ensemble methods are particularly useful for determining binding and unbinding rates. This is done by measuring the transition flux into either the bound or unbound state as the sum of walker weights per unit time. These binding and unbinding rates are interesting in their own right, as they are necessary to model drug behavior in typical nonequilibrium conditions. In addition, the unbinding rate (or equivalently, its inverse, the residence time) is the crucial quantity in some systems to determine drug efficacy *in vivo* [6, 12, 74]. Although much progress has been made in the calculation of residence times for pharmaceutically-relevant ligands, as of now there are no blind challenges for the prediction of binding or unbinding kinetics. While researchers can compare to experimental data, as discussed above sampling errors are hard to isolate, and connecting computational and experimental models (for instance, in the definition of the bound and unbound states) is not trivial. For this reason our approach is to indirectly validate the transition rate calculations by using them to calculate binding affinities in the SAMPL challenges. Specifically, we calculate the binding free energy (ΔG_b) as a function of transition rates using: $\Delta G_b = -kT \ln k_{off}/(C_0 k_{on})$.

Although weighted ensemble has been successful in simulating a wide variety of long-timescale events ranging from seconds, such as the opening of the SARS-CoV-2 spike protein [30] to multi-minute long events such as ligand unbinding [9, 31], a common problem is the high variation of walker weights between independent replicates, which causes large uncertainties in calculated observables such as free energies and rate constants [74]. REVO itself was developed as a way to lower the variation between independent replicates, however previous applications of REVO to the same systems used in this work yielded rate calculations with large differences from one run to another [74, 158]. This is caused by

the rare occurrence of high probability walkers crossing into the unbound state in unbinding simulations, which introduces large jumps in transition fluxes and unbinding rates. Interestingly, our previous work sought to identify microscopic determinants of these differences in probability and found that there was a correlation between guest-ion interactions and the unbinding weight of a given trajectory for several SAMPL systems [26]. Specifically, rare high-probability trajectories that contributed most significantly to the unbinding rate had a lower amount of guest-ion interaction than the far more numerous low-probability trajectories. By tracking the center of mass (COM)-to-COM distance (d_{COM}), we also found that the high- and low-probability trajectories could be distinguished even relatively early on in the unbinding pathway ($d_{COM} = 7\text{\AA}$) based on their guest-ion interaction, prior to there being a discernible difference in the trajectory weight. This d_{COM} distance also corresponded to a “commitment to unbinding” point, where reactive trajectories left the the binding pocket and no re-crossings were observed. This observation implies that the final weight of an unbinding trajectory was, to some extent, already determined early on. If so, it suggests that cloning trajectories with $d_{COM} > 7\text{\AA}$ is wasteful, as it is known that these trajectories are going to unbind and cloning them takes up unnecessary space in the ensemble that could be used to observe other pathways and events. Therefore, we believed that it may be beneficial to prevent cloning of trajectories that had already reached $d_{COM} = 7\text{\AA}$, as this leaves more spots available in the ensemble for other events to occur and will result in the generation of a high weight unbinding events.

Here, we determine unbinding free energies for four systems through binding and unbinding simulations with a slightly modified version of the REVO algorithm. Three of these systems come from the SAMPL6 challenge [90, 94]: “OA-G6”, “OA-G3”, and “CB8-G3”. For these systems, we had prior knowledge of ΔG , however the binding free energy for a fourth system, β -CD-PMZ, was predicted prospectively as part of the SAMPL9 challenge. To determine these free energies, a modification of REVO [32] was used for unbinding simulations. In this method, which we refer to as cutoff-REVO, cloning is prevented for trajectories that

have met a predefined physical requirement that, when reached, means the system is likely to commit to some activity of interest (here, unbinding). In the case of this work, that requirement is a system-specific d_{COM} distance between the host and guest molecules. This cloning cutoff value is determined from previously run standard REVO simulation data. The use of a cloning cutoff produced ensembles with fewer unbinding events, but with much higher weights. Using these simulations, k_{off} , k_{on} , and ΔG are calculated for all systems for both standard and cutoff-REVO. The ΔG values are compared to computational (SAMPL6 systems) and experimental (all systems) reference values. Differences in resampling patterns between standard REVO and cutoff-REVO are also analyzed and visualized using a network algorithm. A comprehensive tutorial for running standard REVO unbinding simulations, cutoff-REVO unbinding simulations, and rebinding simulations (along with all relevant code) is provided along with this paper [176].

4.2 Methods

Host-Guest Systems

The SAMPL systems used here are small molecule pairs that come from the SAMPL6 and SAMPL9 challenges. These are artificial systems whose primary purpose is to test the accuracy and efficiency of new computational methods for the prediction of binding free energy. The systems used here include the OA-G6, OA-G3, and CB8-G3, systems from SAMPL6 and β -CD-PMZ from SAMPL9 (Figure 4.1 and Figure C.1. These are all host-guest systems, where the host molecules are referred to as: OA (octa-acid), CB8 (cucurbit[8]uril) and β -CD (β -cyclodextrin). The SAMPL6 guest molecules are named according to their index in the challenge: OA-G6 refers to 4-methyl pentanoic acid, OA-G3 refers to 5-hexenoic acid, and CB8-G3 is quinine. PMZ refers to promazine hydrochloride.

The OA host molecule has a four-fold symmetry along the vertical axis and a -8 charge. Its guest molecules, G6 and G3 are both small molecules with an explicit charge of -1 . The CB8 host has eight-fold symmetry along the vertical axis and two-fold symmetry along the

horizontal axis. CB8, as well as its quinine guest are both neutral molecules. The β -CD host has no net charge and seven-fold symmetry along its vertical axis. Its guest, PMZ, has a +1 explicit charge.

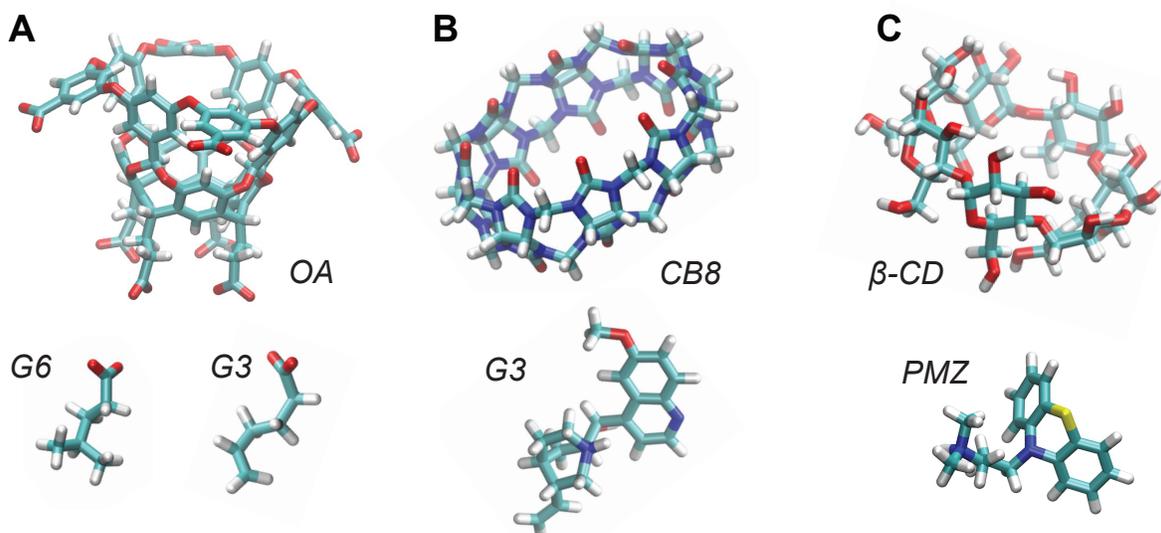


Figure 4.1: Host-guest systems. Hosts are shown on top, and guests are shown on bottom. A) The OA host and its guests, G6 (left) and G3 (right). B) The CB8 host and its ligand, G3. C) The β -CD host and its ligand, PMZ. All panels use the default “atom name” coloring scheme from VMD [95]: cyan=Carbon, white=Hydrogen, red=Oxygen, blue=Nitrogen and yellow=Sulfur.

The Weighted Ensemble method

The WE [68] method is an unbiased path sampling method that allows for the efficient generation of (un)binding paths. A generalized framework for this two-step method is as follows: in the first step an ensemble of trajectories, also called “walkers”, are independently propagated forward in time by molecular dynamics (MD), and in the second step these are “resampled”, using merging and cloning operations. The purpose of resampling is to clone walkers that are desirable and to merge walkers that are less desirable; this was originally guided by the populations of a set of bins that spanned the conformation space, but can be thought of more generally, as we discuss below. The goal of cloning, generally, is to help increase the chance of escaping local energy minima and observe a process of interest. The goal of merging is to reduce redundant trajectories in the ensemble and decrease the

computational cost of the simulation.

In a WE simulation, all of the walkers have a probability, or a statistical weight. When a walker undergoes a cloning operation, two new independent trajectories are created with the same conformation as the original walker and half of its weight. When two walkers A and B undergo a merging operation they are combined to create walker C with weight $w_c = w_A + w_B$. Walker C takes on either the conformation of walker A or B , with a probability proportional to their weights. Generally, a resampling algorithm can be thought of as a function that takes in the ensemble of trajectories, guides the merging and cloning process, and returns a new ensemble where the conformations are taken from the input ensemble. In this new ensemble, while the number of returned walkers can vary, the sum of the walker weights (most often $\sum_{i=1}^n w_i = 1$), remains unchanged.

REVO

The resampling algorithm used to perform both the binding and unbinding simulations in this work is the REVO (Resampling Ensembles by Variation Optimization) resampler [32]. After a cycle of dynamics has been run, REVO guides the merging and cloning operations for an ensemble by maximizing the “trajectory variation”, V (Eq. 4.1), which is a scaled sum of the all-to-all distances between the walkers in the ensemble.

$$V = \sum_i V_i = \sum_i \sum_j \left(\frac{d_{ij}}{d_\star} \right)^\alpha \phi_i \phi_j, \quad (4.1)$$

where V_i is the “trajectory variation” for walker i , d_{ij} is a distance calculated between walkers i and j , and ϕ_i is a “novelty” term that describes the significance of each individual walker in the ensemble. The distance metric d_{ij} is different for unbinding and rebinding REVO simulations, each of which is described below. The variable d_\star is a “characteristic distance”, which is used to make the variation function unit-less, and is equal to the mean of the distance metric after one cycle of MD. Note that this value does not have an impact on resampling behavior and is used for ease of comparison between varying distance metrics for the same system. α is a parameter that balances the “exploitation” (novelty) with the

“exploration” (distance) in the REVO algorithm. The novelty term, ϕ , is usually defined according to the walker weight, w_i , as follows:

$$\phi_i = \log(w_i) - \log\left(\frac{p_{min}}{100}\right), \quad (4.2)$$

where p_{min} is the predefined minimum walker weight allowed in the simulation. This should be set according to the anticipated probability of the events of interest. Here we set p_{min} to a value of 10^{-12} to be consistent with previous work [74, 158].

After a cycle of dynamics has been run and the distances have been calculated, REVO guides merging and cloning operations by first calculating the value of V . Walkers are then proposed for merging and cloning as follows. The walker i is proposed for cloning which has the highest V_i and a weight greater than $2p_{min}$. A walker j , with the lowest value of V_j and a weight less than p_{max} (here, set to 0.1) is chosen for merging. The walker k that is closest to j with a weight such that $w_j + w_k < p_{max}$ is identified as its merging partner. In order to proceed, their distance d_{jk} must be less than or equal to a predetermined value referred to as the “merge distance”. Once these three walkers have been selected for resampling, V is recalculated as though the merging and cloning operations have been performed. If V increased, the proposed resampling operations are carried out and the process of proposing merging and cloning steps repeats until V no longer increases, or suitable walkers cannot be found. At that point we terminate the resampling process and run the next cycle of dynamics.

The cutoff-REVO algorithm

Here we examine a modified version of the REVO resampler, referred to as cutoff-REVO. Overall, this algorithm works the same as the standard REVO algorithm described above, but with an additional criterion used to determine if a trajectory in the ensemble should be eligible for cloning. In this work, the eligibility function was a d_{COM} distance between the host and guest (with the value specifically-determined for each system). If a trajectory had a d_{COM} distance equal to or greater than the cutoff value, it was not eligible for cloning

in that cycle. This cutoff value was determined from d_{COM} plots of unbinding events from standard REVO simulations (described further in Section 4.3).

Although the eligibility function was based on COM in this work, in general, it is highly flexible and can be customized on a system by system basis. For instance, this criterion can be based on physical features such as RMSD, number of interactions or contacts, solvation of binding pockets, or other non-structural features such as energies or work. We implemented this eligibility function in the `wepy` software package [75]. Similar to other `wepy` objects, such as distance metrics or boundary conditions, the eligibility function is implemented in a modular fashion: users can implement and use their own customized eligibility function without changing the source code for the REVO resampler.

Simulation details

All of the simulations here were run with `wepy` [75] and OpenMM [138]. The following parameters were used for all simulations: 3000 cycles, 10000 dynamics steps per cycle with a 2 fs timestep, and a temperature of 300 K. Langevin integration was used for all simulations. The β -CD-PMZ system was built with CHARMM-GUI [177, 178, 179], and the SAMPL6 systems were built with the Gromacs input files provided by the challenge organizers [90]. The β -CD-PMZ system was run at constant pressure of 1 bar, and OA-G6, OA-G3, and CB8-G3 were run at constant volume.

For each SAMPL6 system, five potential binding poses were provided by the challenge organizers. In previous work we found that the choice of initial binding pose is inconsequential as all poses quickly interconvert with one another [74]. Here we chose pose 0 as the initial positions for all walkers in the REVO simulations for the OA-G6, OA-G3 and CB8-G3 systems. For β -CD-PMZ, we generated our own starting poses for the unbinding simulations. Three copies of the β -CD-PMZ system were built with the multicomponent assembler in CHARMM-GUI [177, 178, 179]. Following the 1 ns initial OpenMM simulation, a bound conformation was produced (Figure 4.2). The final frame of the bound run of the three initial simulations was used to initialize unbinding simulations.

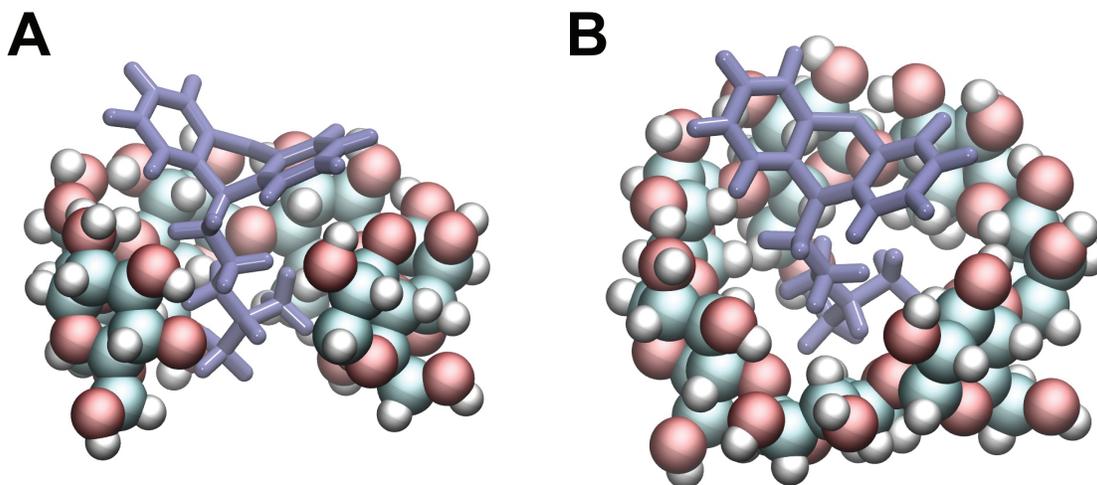


Figure 4.2: The β -CD-PMZ starting pose. A and B) Two different views of the binding pose obtained where the dimethylamino tail of PMZ is in the binding pocket of β -CD. Some atoms have been removed from β -CD in A for clarity. Both panels use the default “atom name” coloring scheme from VMD [95] for the host molecule: cyan=Carbon, white=Hydrogen, and red=Oxygen.

For the SAMPL6 systems, five “standard” REVO unbinding simulations were run (without the eligibility criterion). The distance metric between walkers, d_{ij} , is calculated as the RMSD of the guest atoms after alignment of the hosts. Note that in contrast to previous work [74, 158], the distance calculations did not take into account the symmetries of the host (4-fold for OA and 8-fold for CB8), for simplicity. The distance metric was implemented using the `ReceptorDistance` class in `wepy`. We use the standard unbinding boundary condition `UnbindingDistance` where warping occurs when the closest host-guest atomic distance is greater than 1 nm. In addition, five cutoff-REVO simulations were run, with a cutoff determined from the reactive trajectories in the standard REVO simulations (more information is available in Figure C.2) and the same distance metric and unbinding boundary conditions as the standard REVO simulations. For the β -CD-PMZ system we ran 5 standard REVO simulations and 5 cutoff-REVO simulations, using the same approach.

In the rebinding simulations, starting poses were taken from the unbinding events of the standard REVO simulations. For simplicity, we used a different distance metric in the

rebinding simulations for the β -CD-PMZ system, defined as follows:

$$d_{ij} = \frac{1}{d_{COM,i}} - \frac{1}{d_{COM,j}}. \tag{4.3}$$

This is similar to the `RebindingDistance` metric used in Dixon et al [74], but is based on the host-to-guest d_{COM} instead of the RMSD to a native structure. The difference of the inverse is used in order to emphasize difference between smaller values of d_{COM} . In addition we use a binding boundary condition that is also based on the d_{COM} . For this boundary condition, binding events were triggered when $d_{COM} < 0.4$ nm. This value was determined from Figure C.2, which shows a subset of the COM distances for the whole bound to unbound transitions for all four systems. For OA-G6, OA-G3, and CB8-G3, we use rebinding data that was previously generated in Ref. [74].

Figure C.2 was also used to determine the cutoff value for cloning eligibility. Coincidentally, the determined “commitment to unbinding” points for all four systems was 0.7 nm. While it appears that the cutoff COM distance chosen could have been reduced for the OA-G6 and OA-G3 systems, it was found in previous work (Ref. [26]) that for these systems, rebinding can often occur below a 0.7 nm d_{COM} distance. The tutorial repository associated with this article [176] contains everything needed to reproduce the results here, including: code for the custom distance metric and boundary condition, scripts to run the binding and unbinding simulations, scripts to plot the d_{COM} for reactive trajectories, and scripts to extract the rebinding simulation starting poses.

Calculation of rates and free energies

With the WE method, rates and subsequently free energies can be calculated through a method called “ensemble splitting” [74, 83, 84, 85, 86]. In this method, the equilibrium ensemble is split into “binding” and “unbinding” ensembles. The binding ensemble is composed of the trajectories that have most recently been in the unbound basin, and the unbinding ensemble is composed of the trajectories that have most recently been in the bound basin. This technique can be used with any set of non-overlapping basins. Here, the unbound

basin is defined as a set of structures where the minimum host-guest interatomic distance is greater than 1 nm, and the bound basin is either defined as a set of structures where the d_{COM} distance is below 0.4 nm (for β -CD-PMZ) or when the ligand RMSD to the native structure is < 0.1 nm (OA-G6, OA-G3, and CB8-G3).

Here, simulations are conducted entirely in either the binding or unbinding ensemble. When a walker leaves its ensemble by entering the other basin, its weight is saved, and the structure, which is recorded as an “exit point” then undergoes a process called warping. When a trajectory warps, it is set back to its starting pose but its weight remains unchanged. In this work, for the unbinding ensemble, the starting structure is the initial bound pose, and for the binding ensemble, the starting structure is one of the initial unbound poses. Rates are calculated via the trajectory flux that leaves one ensemble for the other: k_{off} (4.4), k_{on} (4.5) and ΔG (4.6) can be determined as:

$$k_{off} = \overline{\phi_u} = \frac{\sum_{i \in U} w_i}{T}, \quad (4.4)$$

$$k_{on} = \overline{\phi_b} = \frac{\sum_{i \in R} w_i}{CT}, \quad (4.5)$$

$$\Delta G = kT \ln \frac{k_{off}}{C_0 k_{on}}, \quad (4.6)$$

where, $\overline{\phi_b}$ and $\overline{\phi_u}$ are the binding and unbinding flux, T is the elapsed time, the sum is the sum of weights for the corresponding exit points, and C is the concentration of the guest molecule. The code for this analysis is provided in the associated tutorial.

4.3 Results

Resampling trees show differences in merging and cloning behavior

For each host-guest system used, five unbinding simulations were run for both the standard REVO resampler and the cutoff-REVO resampler. In this Section, we focus on resampling differences in the β -CD-PMZ system, as these are representative of the other systems as well. We find that these resamplers have significant differences in their cloning and merging behavior, which we visualize using “resampling trees” (Figure 4.3). These are directed acyclic

graphs where each cycle of each walker is represented by a node. The nodes for each cycle are evenly spaced along the vertical direction, with the earliest timepoints at the bottom of the resampling tree. Connections between nodes show the continuation of walkers over time: multiple connections between cycles show cloning events, and the absence of connections to the next cycle indicates that the walker was merged into one of the other walkers. The sizes of each node represent the weights of the walkers at a given timepoint, and in Figure 4.3 the nodes are colored according to their d_{COM} value.

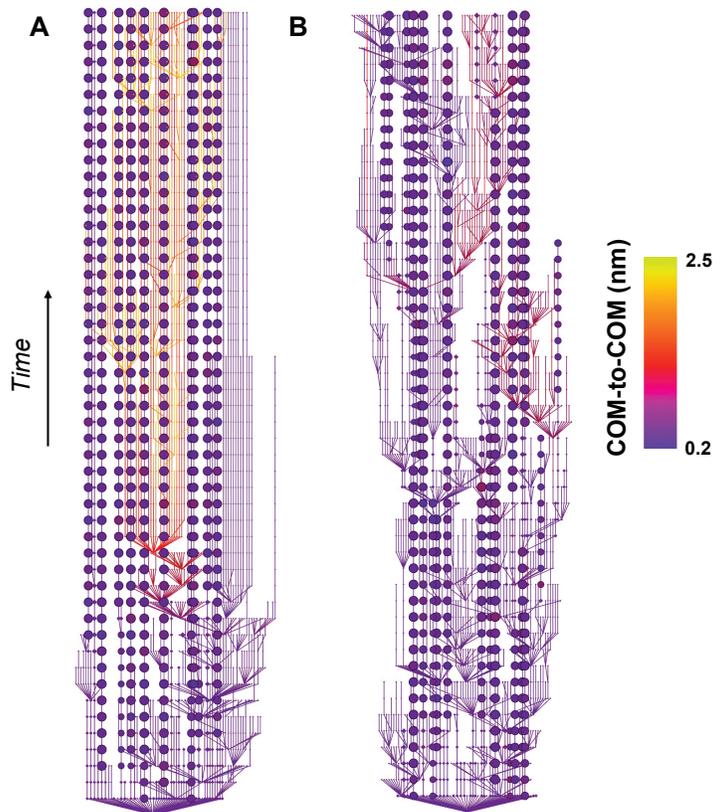


Figure 4.3: Resampling trees for A) REVO and B) cutoff-REVO. Each tree shows the cloning and merging behavior during their respective simulation, where each node represents a walker at a given point in time. The trees move forward in time from bottom to top. Nodes and edges are colored according to d_{COM} distance. Data from β -CD-PMZ. For clarity, only the first 50 cycles are shown from a single run. Similar behavior was observed in other runs.

We find that standard produces fewer cloning events per cycle ($\langle n_{wc} \rangle < 1$), with slightly fewer replicates ($\langle R \rangle$) produced at each of these events, whereas cutoff-REVO has more cloning events per cycle ($\langle n_{wc} \rangle = 1.9$) with slightly more replicates produced at each event,

on average (Table 4.1). As the simulation progresses the number of cloning events for standard REVO significantly decreases. This is in contrast to cutoff-REVO, where $\langle n_{wc} \rangle$ stays consistent over the entire simulation (Figure 4.4). After the first walker leaves the binding site in standard REVO, we find that the majority of cloning events focus on high d_{COM} walkers. This is consistent with REVO’s goal of maximizing the variation in the ensemble, but this limits our ability to clone walkers that are closer to the binding site and generate new, independent unbinding events. In contrast, for cutoff-REVO we see more consistent cloning of trajectories that are earlier in the unbinding process. This is due to the lack of mass cloning events that fragment the more unbound trajectories. Preventing this fragmentation results in more slots being available for early-unbinding cloning events, as the trajectories of interest that are beyond the cutoff are not taking up these spaces. Preventing the fragmentation of unbound walkers results in the generation of fewer, but higher-weighted unbinding events.

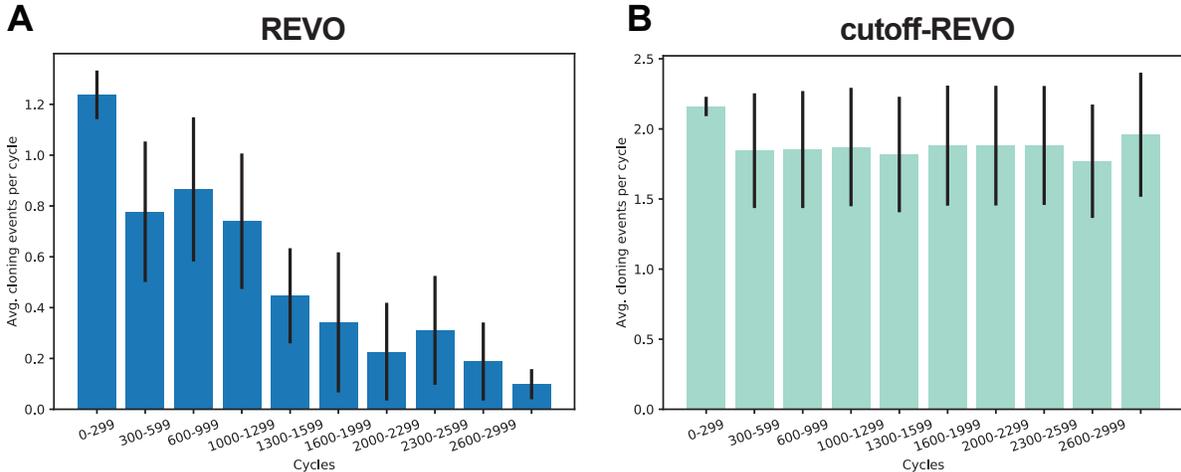


Figure 4.4: $\langle n_{wc} \rangle$ over time for β -CD-PMZ using A) REVO and B) cutoff-REVO. The average number of cloning events per cycle ($\langle n_{wc} \rangle$) broken down by simulation time (300 cycles per bin). Values are averaged over 5 runs and error are the standard error of the mean.

The average d_{COM} distance for a cloning event for standard REVO across the five β -CD-PMZ runs was found to be $1.1 \text{ nm} \pm 0.008$, whereas it was $0.35 \pm 0.0008 \text{ nm}$ for cutoff-REVO. As expected, the maximum d_{COM} of a cloning event was also much larger for

Table 4.1: The average number of walkers cloned per cycle ($\langle n_{wc} \rangle$) and the average number of replications made per cloning event ($\langle R \rangle$) is shown. In addition, we show the average ($\langle d_{COM} \rangle_c$), the minimum (Min. d_{COM}) and the maximum (Max. d_{COM}) value of d_{COM} in the set of cloned walkers. Values are shown separately for REVO and cutoff-REVO. These numbers are averaged across five runs and the errors are computed as the standard error of the mean across the set of runs.

Method	$\langle n_{wc} \rangle$	$\langle R \rangle$	$\langle d_{COM} \rangle_c$ (nm)	Min. d_{COM} (nm)	Max. d_{COM} (nm)
REVO	0.52 ± 0.14	2.71 ± 0.07	1.1 ± 0.008	0.14 ± 0.006	2.3 ± 0.013
cutoff-REVO	1.90 ± 0.38	3.25 ± 0.03	0.35 ± 0.0008	0.10 ± 0.005	0.76 ± 0.004

REVO (2.3 ± 0.013 nm) than for cutoff-REVO (0.76 ± 0.004 nm). By shifting the emphasis of the cloning events to lower d_{COM} walkers, cutoff-REVO also shifts emphasis towards higher weighted walkers. The relationship between the d_{COM} and walker weights are shown in Figure C.3. Figure C.3A shows that walkers with a weight $< 10^{-7}$ are mostly in the unbound state, with an average $d_{COM} > 1.0$ nm. As cutoff-REVO avoids the cloning of these walkers, the emphasis is dramatically shifted from low- to high-weight walkers (Figures C.3C and C.3D).

Unbinding events, free energy, and rate analysis

All of these simulations generate numerous unbinding events, but with stark differences in both the probabilities of the unbinding events and their number as seen in Figure C.4 (results for OA-G3 and CB8-G3 shown in Figure C.4). For all four systems, standard REVO unbinding simulations produce numerous unbinding events, although most of these had extremely low probabilities. The highest weighted unbinding events are particularly important as they dominate the calculation of unbinding rates in Eq. 4.4. For β -CD-PMZ, the highest weight event obtained was 10^{-2} for both standard REVO and cutoff-REVO. We observe differences in the highest weights for other systems: such as 10^{-7} and 10^{-3} (OA-G6), 10^{-6} and 10^{-3} (OA-G3), and 10^{-8} and 10^{-7} (CB8-G3), where the higher weights are always obtained by the cutoff-REVO algorithm. Interestingly, the number of unbinding events for cutoff-REVO did not increase as the exit point weight decreased; instead the unbinding event weights were roughly normally distributed. The number of unbinding events for all systems also

significantly decreased for all four systems as summarized in Table 4.2.

Table 4.2: The highest exit point weight obtained and the total number of unbinding events observed for all four systems for both standard REVO and cutoff-REVO over five simulations.

System	REVO		cutoff-REVO	
	Max. Wt.	Total	Max. Wt.	Total
β -CD-PMZ	10^{-2}	5303	10^{-2}	137
OA-G6	10^{-7}	16597	10^{-3}	69
OA-G3	10^{-6}	22901	10^{-3}	224
CB8-G3	10^{-8}	1246	10^{-7}	93

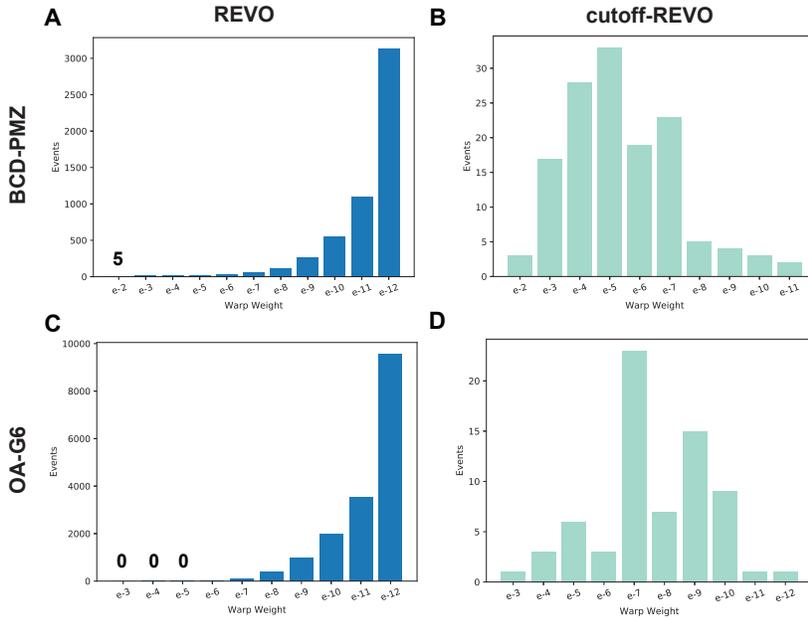


Figure 4.5: Unbinding walker weights for both methods. Walker weights at the time of unbinding for standard REVO (left) and cutoff-REVO (right) for the β -CD-PMZ system (A and B) and the OA-G6 system (C and D). The standard REVO algorithm showed a much larger quantity of unbinding events. The total number of unbinding events in each panel is as follows: A) 5303, B) 137, C) 16597, and D) 69.

The binding and unbinding rates for all systems were determined through the use of the “ensemble splitting” method as described in Section 4.2. The binding rate is found by dividing the rebinding trajectory flux by the guest concentration, where $C = \frac{1}{N_a V}$ and V is the box volume (Eq. 4.5). Rolling estimates of the k_{off} , k_{on} , and ΔG are shown for all systems in Figure C.5 for standard results and Figure C.6 for cutoff-REVO. We observe

differences between the standard REVO and cutoff-REVO results in both the accuracy of the final ΔG obtained and the consistency between runs as seen for k_{off} values in Figure 4.6 and Table 4.3. For the SAMPL6 systems, these values are compared to computational reference values determined through OpenMM-Hamiltonian Replica Exchange [94, 90] simulations that were run by the SAMPL organizers as a means to compare the efficiency of methods. No computational reference value is available for the β -CD-PMZ system. Experimental reference values are also available for all four systems. For the SAMPL6 systems, this value was determined with isothermal titration calorimetry [94, 180]. The same method was used to determine the free energy of unbinding for β -CD-PMZ, with more information on the method used at the SAMPL9 GitHub page [181].

Table 4.3: The off-rates from REVO and cutoff-REVO simulations. All k_{off} values are in s^{-1} . The mean first passage times are in ms. The uncertainties are calculated as the standard error of the mean of the log of k_{off} . They can be interpreted as the breadth of the uncertainty in orders of magnitude.

System	k_{off} (REVO)	REVO Error	k_{off} (c-REVO)	c-REVO Error
OA-G6	3.3×10^3	0.89	2.5×10^4	1.48
OA-G3	2.1×10^3	0.85	1.5×10^4	0.64
CB8-G3	1.77	1.54	2.38	0.39
β -CD-PMZ	5.0×10^5	2.26	5.5×10^5	1.15

For cutoff-REVO, all systems have a final value of ΔG that is very close to either the computational (OA-G3 and CB8-G3) or experimental (β -CD-PMZ and OA-G6) reference value. Standard REVO obtained a very similar final ΔG to cutoff-REVO for the β -CD-PMZ system, however standard REVO has a large uncertainty, with k_{off} estimates varying between runs by up to six orders of magnitude. The final ΔG s obtained for CB8-G3 for both methods were very similar; however, cutoff-REVO produced highly consistent runs for this system with minimal error in ΔG (Figure 4.7), whereas standard REVO had an uncertainty of 1.01 kcal/mol. The final free energies for REVO and cutoff-REVO as well as all available reference values are shown in Table 4.4, with plots available in Figures C.5 and C.6. The errors shown in Table 4.4 are the standard error for the final values determined for ΔG across all five runs for each data set. Coincidentally, both OA-G6 and β -CD-PMZ have exactly the

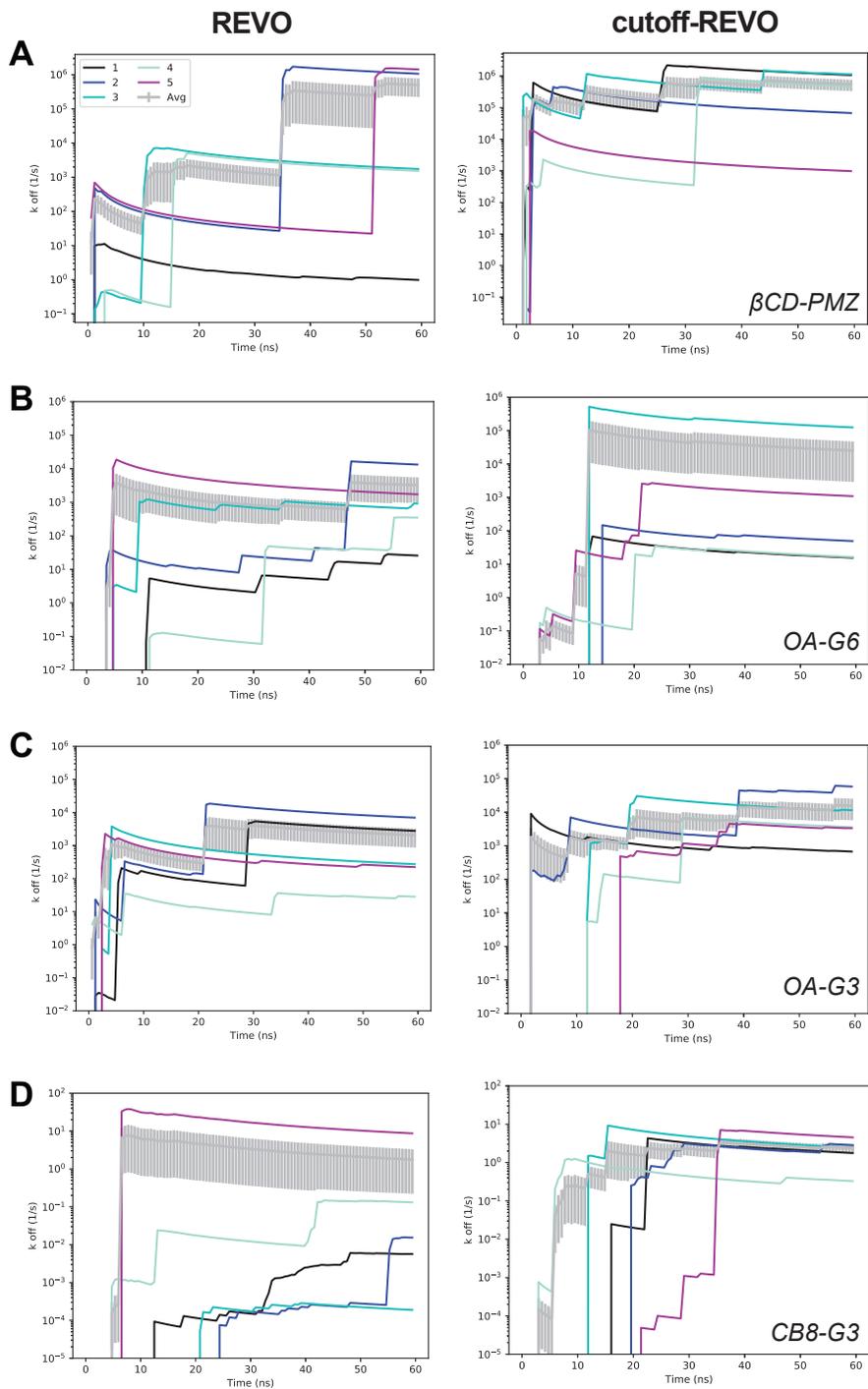


Figure 4.6: Comparison of unbinding rates for REVO and cutoff-REVO for all systems. Individual runs are shown as thin lines. Average unbinding rates are shown as thick grey lines and uncertainties, computed using the standard error of the mean across the 5 runs, are shown as the shaded grey regions. A) β -CD-PMZ. B) OA-G6. C) OA-G3. D) CB8-G3. The horizontal axis represents trajectory length and not aggregate simulation time.

same experimental binding free energy. With this in mind, both REVO and cutoff-REVO provide the best rank ordering possible for these systems.

Table 4.4: Binding free energies calculated for all systems, as well as experimental and computational reference values, where available. Computational reference values were calculated using extensive Hamiltonian Replica Exchange simulations (Ref. [90]). All ΔG values are in kcal/mol.

System	ΔG (REVO)	ΔG (c-REVO)	ΔG (comp. ref.)	ΔG (exp.)
OA-G6	-6.55 ± 0.53	-5.21 ± 0.81	-7.18 ± 0.06	-4.97 ± 0.02
OA-G3	-7.69 ± 0.55	-6.49 ± 0.41	-6.71 ± 0.05	-5.18 ± 0.02
CB8-G3	-11.35 ± 1.01	-11.17 ± 0.13	-10.80 ± 0.20	-6.45 ± 0.06
β -CD-PMZ	-5.05 ± 1.51	-5.00 ± 0.66	N/A	-4.97 ± 0.02
RMSE (comp.)	0.74	1.2	-	-
RMSE (exp.)	2.9	2.5	-	-

The root mean squared error for REVO and cutoff-REVO are also shown in Table 4.4 as compared to both the computational reference and experimental benchmarks. On average, the REVO calculations agree to the computational reference to within 0.74 kcal/mol, which is slightly lower than the cutoff-REVO RMSE of 1.2 kcal/mol. The difference between these RMSE values is comparable to the uncertainties of the free energy estimates. The lower computational reference RMSE for REVO is primarily due to better performance on the OA-G6 system, which is higher for cutoff-REVO (-5.21 kcal/mol). Conversely, the RMSE measured to experimental benchmarks is slightly lower for cutoff-REVO (2.5 kcal/mol) compared to standard REVO (2.9 kcal/mol). However, both of these values are driven higher by large over-estimates of the free energy for CB8-G3. As this is consistent with the computational reference, this suggests that force-field inaccuracies or some other feature of the computational model (e.g. protonation states, oligomerization states) is inconsistent with the experimental conditions. If this system is removed the RMSE drops to 1.7 kcal/mol for standard REVO and 0.77 kcal/mol for cutoff-REVO.

4.4 Discussion

The predicted free energies of unbinding from the cutoff-REVO algorithm improved upon those predicted by REVO in both their consistency across runs and their agreement with

benchmark values. For OA-G3 and CB8-G3 the agreement with the computational reference values was improved, significantly in the case of OA-G3. For CB8-G3, both REVO and cutoff-REVO showed good agreement with computational reference values, although the consistency across runs was significantly improved for cutoff-REVO. For the OA-G6 system, although agreement with the computational reference decreased for cutoff-REVO, the ΔG moved closer to the experimental value. We are only cautiously optimistic about the agreement with experiment in this case, as we also observe the largest uncertainties in ΔG for this system. In general, we view agreement with computational reference values as more meaningful, as they separate out inaccuracies of the forcefield.

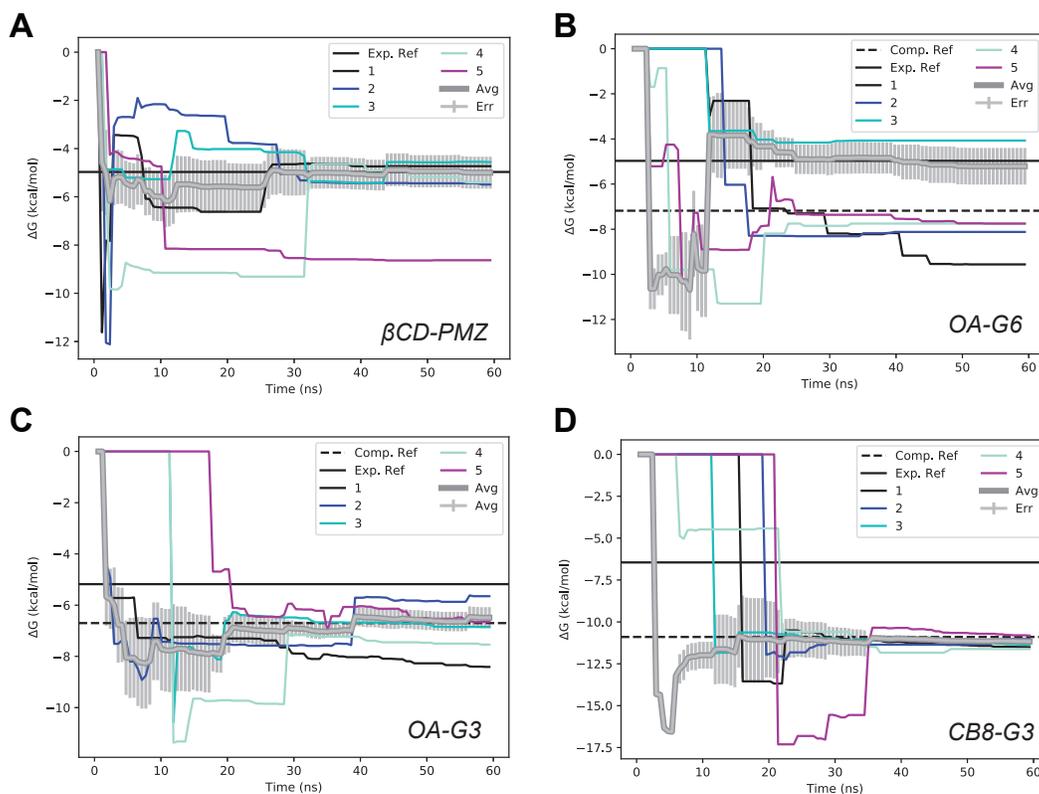


Figure 4.7: Prediction of cutoff-REVO unbinding free energy for all systems. Individual runs are shown as thin lines. Average free energies are shown as thick grey lines and uncertainties, computed using the standard error of the mean across the 5 runs, are shown as shaded grey regions. A) β -CD-PMZ, B) OA-G6, C) OA-G3, and D) CB8-G3 systems. The horizontal axis represents trajectory length and not aggregate simulation time.

We also prospectively predicted the binding free energy for the β -CD-PMZ system and

observe excellent agreement with the experimental quantity. This was achieved in a “blinded” fashion, where we did not have knowledge of the quantity at the time of prediction. Remarkably, the results in Figure 4.7A show that 4 of the 5 runs converged closely to the experimental quantity by the second half of the simulation, and the ensemble estimates of both REVO and cutoff-REVO agreed with the experimental value to within 0.1 kcal/mol. While we are encouraged by this agreement, we are again cautiously optimistic, as it is possible this resulted from fortuitous cancellation of error. However, a more complete assessment of REVO and cutoff-REVO requires prediction across a larger set of ligands, such as the complete ligand set in the SAMPL9 challenge, but this was not feasible in this work due to the computational costs involved.

Another factor to consider for the β -CD-PMZ system is the bound starting structure used to initialize the unbinding simulations. This was unknown beforehand, and was generated using short, straight-forward simulations, initialized in the unbound state. While β -CD has a ring-like structure similar to CB8, it is not C2-symmetric around the horizontal axis, allowing for distinct bound poses in the top and bottom of the ring. In addition, the PMZ ligand has many feasible modes of interaction with β -CD in addition to that shown in Figure 4.2, with the dimethylamino group inserted into the ring. In previous work on SAMPL6 systems, we found that five different starting poses were “kinetically indistinguishable”: they quickly interconverted with one another on a timescale that was much faster than the unbinding process [74]. In contrast, our previous work on the PK-11195 ligand dissociating from the TSPO membrane protein showed six poses that were not all able to interconvert before unbinding, and thus had pose-specific unbinding rates [31]. At this point, it is unclear which scenario applies to β -CD-PMZ, although we plan to address this in more detail in future work.

cutoff-REVO is slightly more computational expensive than standard REVO, due to the necessity of running a standard simulation to generate the data necessary to determine both a cutoff feature and its correspondingly value (if this information is not already known).

The feature used for cloning eligibility is highly flexible and can be whatever physical aspect of the system is relevant to the process that the user wants to observe. This could be a given number of contacts, RMSD to a reference structure, d_{COM} distances, energies, work, specific residue interactions etc., applicable a wide range of biological activities of interest such as ligand (un)binding, protein-protein interactions, or conformational changes, amongst others. The value of the cutoff used can be seen as a knob that can be turned, changing the eligibility function’s influence on the resampling patterns and subsequently the boundary crossing weights generated in the simulation. Overall, this method provides improvements in the prediction of free energies using complete transition paths.

4.5 Conclusions

Restricting the cloning behavior in the REVO sampling method resulted in dramatic changes in the resampling behavior. As a result, the new “cutoff-REVO” algorithm heavily emphasized cloning of walkers at the early stages of unbinding. Although this affected the number of unbinding events observed by the algorithm, the total weight of the unbound trajectories was higher for all four systems examined here. As such, cutoff-REVO was able to generate unbinding events with weights up to 10^4 times higher than standard REVO for the same system, indicating that these unbinding pathways occur with higher probability. As cutoff-REVO consistently outperformed standard REVO, we determine that for unbinding events, quality matters more than overall quantity.

4.6 Acknowledgements

This work was supported by R01GM130794 from the National Institutes of Health and DMS 1761320 from the National Science Foundation. Note that this work has not yet been peer-reviewed.

CHAPTER 5

SUMMARY, IMPACTS, & OUTLOOK

In section 1.5, the goals of this thesis were outlined. We can now go back and discuss these goals with a focus on what was accomplished, what could have potentially been improved, and the overall impacts of this collection of work.

The first goal of this thesis was to combine the JE with WE to determine unbinding free energies from short, nonequilibrium simulations. In this work two different methods were developed, one which focused resampling on trajectories with a high value of importance (Importance Resampling), and another method (history-REVO) that takes a trajectory's entire history of work into account for resampling purposes.

In Chapter 2 we tested these two different methods against straight forward simulations as well as a brief comparison to DIFFMC. We found that the Importance Resampler method was extremely sporadic. This is the result of one major pitfall of this method: regardless of how 'important' a trajectory is in one cycle, this does not necessarily mean that that trajectory would be equally/more 'important' in the following cycle(s). To better this method, we implemented an 'amplification factor' to assist exploration of the tails of the work distributions, but it was found that as this value increased beyond 1, the unbinding energies determined decreased in accuracy. This occurs at higher amplifications because the Importance Resampler selects several high-importance trajectories at each resampling step to be cloned repeatedly, and these trajectories would often have a change in work occurring in the next dynamics step that results in a lowered importance in the next cycle, the same issue previously found.

However, we found that the history-REVO method was faster to converge to the correct unbinding free energy than the Importance Resampler and straight-forward dynamics. This method initially appeared promising as a potential future method for the determination of unbinding free energies for systems of interest. It is worth noting that the history-REVO method is innovative in that the 'distance' between trajectories was not directly dependent

on root mean square deviation (RMSD)s between positions (or wholly position dependent in general) and considered an ‘activity‘ based distance. Also, this method is noteworthy as it considers the entire history of the trajectory in the resampling process represented by a running sum of work. These advancements open the door to be able to resample trajectories based off numerous, non-atom-index oriented features of a system. These features could include things such as agreement with desired quantities, ion transport, etc. among many other options relevant to biochemical systems.

While this method is a promising development, certain changes could increase both its impactfulness and overall success, such as obtaining a more complete picture of the (un)binding process. This would allow us to have a more system-specific and descriptive λ parameter to potentially increase the accuracy of free energy calculations. While the COM-to-COM parameter used can describe unbinding, it cannot distinguish between high- and low-probability pathways and consequently may generate low-probability pathways that require higher applied work. As we found with the work done on the OA-G6 system, a COM-to-COM only λ parameter was insufficient to describe the unbinding of the system and more detail could have been a mitigating factor for variability between simulations.

In Chapter 3, we aimed to learn important solvent-based features of the host-guest unbinding pathway. In this work we studied several differences between high- and low-weight unbinding events for the host-guest systems OA-G6 and OA-G3. In particular, we found differences between the guest-ion interactions, with high-weight unbinding events having fewer of these interactions. We also found that there are differences in ion densities in the simulation box with high-weight unbinding events having fewer ions in regions of space relevant to unbinding. Also discovered was what we called t_0 , or a commitment to unbinding point for both of these systems fairly early in the unbinding pathway. This is of significance due to the number of papers published on the (un)binding of these systems and their use in the development of new computational methods. In addition, as these weight-based differences are related to the charged nature of the host and guest, these results could give

insight into interactions involving highly charged biologically relevant molecules of interest such as DNA and RNA. This work is also of current relevance as learning more accurate descriptors of transitions and unbinding paths is an area of ongoing work. Many methods are currently being developed to define collective variables that represent these paths that can be used in resampling, such as VAMPnets [64], DiffNets [61], Deep-TICA [62], AMINO [60], and SGOOP [63] and these methods should include solvent based features. This work helped to show the importance of the role of solvent and ions along (un)binding pathways.

On the notion of t_0 and commitment to unbinding points, we reached the third goal of this thesis as discussed in Chapter 4, which was to use features learned about host-guest unbinding with a new variation of the REVO method to decrease variation between simulations of the same system and increase accuracy of ΔG calculations. In this work, a new modification to REVO was developed called cutoff-REVO. In this method, when the COM-to-COM distance between a molecule pair of interest reaches a certain value, that trajectory becomes ineligible for cloning operations during the resampling step in simulation. This was implemented because it was found in previous work that some systems commit to an action of interest (here, unbinding) early in the pathway. The prevention of cloning following this point allows trajectories to reach their desired state at a much higher weight and leave room open in the ensemble for the exploration of other pathways.

In this work, we found that REVO produces thousands of unbinding events for our systems of interest, whereas cutoff-REVO produces often one hundred or less for the same amount of sampling time. The key difference was that the events produced by REVO were of significantly lower weight, resulting in less weight reaching the unbound state despite the difference in the number of events. Subsequently, cutoff-REVO produced more accurate and/or lower error ΔG predictions for all four systems tested, and had better convergence of k_{off} , a critical value in the drug design process.

Two of the systems tested, OA-G6 and β -CD-PMZ obtained ΔG values that were very close to the experimental unbinding free energies. While this is an exciting result, it must

be seen with cautious optimism as these simulations (OA-G6) were run with the same parameters as the computational reference values, and the result may be due to a cancellation of errors.

To build upon this work, more time could be dedicated to determining the correct starting pose for the β -CD-PMZ system. Due to the non-symmetrical nature of both the host and guest in this system, there are numerous other poses possible that could be the correct starting pose. Analysis of rebinding simulations could be one means of determining the accuracy of our starting pose. Despite these potential pitfalls, this new cutoff-REVO method was more accurate and consistent than its predecessor REVO, and was only slightly more computationally expensive. The determination of both rates and free energies at the same time is challenging. This work is one of the first to try this in a blind challenge, and presents a starting point for developing best practices for a complex goal. Due to the highly flexible nature of what the cutoff can be, this algorithm can also be used to implement REVO in a way that considers things such as location of the walkers on the energy landscape, e.g. committor probability, and even the estimated Mean First Passage Time to the target, the latter of which has found to be an extremely effective binning strategy for WE[182].

Overall, the goal of the work done in this thesis was to learn about systems of interest and to develop new methods for the determination of values of interest such as ΔG and k_{off} . While some of the initial work with the JE was fairly unsuccessful when applied to more realistic systems, this work led to the development of other efficient and successful methods. To better the JE method, much was learned about highly-utilized test systems, for which knowledge of the (un)binding pathways is very important. In the study of this topic, the commitment-to-unbinding point was discovered. Using this information, a new resampling algorithm was developed that permitted accurate and consistent ΔG and k_{off} predictions. k_{off} predictions are of particular interest due to the value's relationship with RT and subsequently drug efficacy. The work done in this thesis provides insights into systems of interest and provides new, efficient means of obtaining information critical to the drug

design process.

BIBLIOGRAPHY

- [1] Andrew Pohorille, Christopher Jarzynski, and Christophe Chipot. Good practices in free-energy calculations. *Journal of Physical Chemistry B*, 114(32):10235–10253, 2010.
- [2] T. Tanaji Talele, A. Santosh Khedkar, and C. Alan Rigby. Successful applications of computer aided drug discovery: Moving drugs from concept to the clinic. *Current Topics in Medicinal Chemistry*, 10:127–141, 2010.
- [3] Johan Åqvist. Calculation of absolute binding free energies for charged ligands and effects of long-range electrostatic interactions. *Journal of computational chemistry*, 17(14):1587–1597, 1996.
- [4] Hao Li, Shuheng Dong, and Lili Duan. Difference in the binding mechanisms of abt-263/43b with bcl-xl/bcl-2: computational perspective on the accurate binding free energy analysis. *Journal of Molecular Modeling*, 27(11), 2021.
- [5] Robert A. Copeland, David L. Pompliano, and Thomas D. Meek. Drug–target residence time and its implications for lead optimization. *Nature Reviews Drug Discovery*, 5:730–739, 2006.
- [6] Robert .A Copeland. The drug–target residence time model: a 10-year retrospective. *Nature Reviews Drug Discovery*, 15:87–95, 2015.
- [7] Robert A. Copeland. The dynamics of drug-target interactions: drug-target residence time and its impact on efficacy and safety. *Expert Opinion in Drug Discovery*, 5(4):305–310, 2010.
- [8] Hao Lu and Peter J Tonge. Drug-target residence time: critical information for lead optimization. *Current Opinion in Chemical Biology*, 14(4):467–474, 2010.
- [9] Samuel D Lotz and Alex Dickson. Unbiased molecular dynamics of 11 min timescale drug unbinding reveals transition state stabilizing interactions. *Journal of the American Chemical Society*, 140(2):618–628, 2018.
- [10] Lu H. Slow-onset inhibition of the fabi enoyl reductase from francisella tularensis: residence time and in vivo activity. *ACS Chemical Biology*, 4:221–231, 2009.
- [11] Robert A. Copeland. *Evaluation of Enzyme Inhibitors in Drug Discovery: A Guide for Medicinal Chemists and Pharmacologists*, volume 2022. 2013.
- [12] Barbra Costa, Elenora Da Pozzo, Chiara Giacomelli, Elisabetta Barresi, Sabrina Taliani, Federico Da Settimo, and Claudia Martini. Tspo ligand residence time: a new parameter to predict compound neurosteroidogenic efficacy. *Scientific Reports*, 6, 2016.

- [13] H.Qu, S.Bhattacharyya, and P.Ducheyne. 4.34 sol–gel processed oxide controlled release materials. *Comprehensive Biomaterials II*, 4:617–643, 2017.
- [14] Augusto Caraceni, Cinzia Martini, Ernesto Zecca, and ElenaFagnoni. Chapter 27 - cancer pain management and palliative care. *Handbook of Clinical Neurology*, 104:391–415, 2012.
- [15] Javier Corzo. Time, the forgotten dimension of ligand binding teaching. *Biochemistry and Molecular Biology Education*, 34(6):413–416, 2006.
- [16] P. Dua, E. Hawkins, and P. H. van der Graff. A tutorial on target-mediated drug disposition (tmdd) models. *CPT: Pharmacometrics Systems Pharmacology*, 4:324–337, 2015.
- [17] Donald E. Mager. Target-mediated drug disposition and dynamics. *Biochemical Pharmacology*, 72(1):1–10, 2006.
- [18] T. Liu and R. Bl Altman. Identifying druggable targets by protein microenvironments matching: Application to transcription factors. *CPT: Pharmacometrics Systems Pharmacology*, 3(1):1–9, 2014.
- [19] Ahmet Bakan, Neysa Nevins, Ami S. Lakdawala, and Ivet Bahar. Druggability assessment of allosteric proteins by dynamics simulations in the presence of probe molecules. *Journal of Chemical Theory and Computation*, 8(7):2435–2447, 2012.
- [20] David A. Sykes, Steven J. Charlton, and Tahsin F. Kellici. *Single Step Determination of Unlabeled Compound Kinetics Using a Competition Association Binding Method Employing Time-Resolved FRET*, pages 177–194. Springer New York, New York, NY, 2018.
- [21] Ana M. Rossi and Colin W. Taylor. Analysis of protein-ligand interactions by fluorescence polarization. *Nature Protocols*, 6(3):365–387, 2011.
- [22] Biao Liu, Neelaabh Shankar, and Douglas H. Turner. Fluorescence competition assay measurements of free energy changes for rna pseudoknots. *Biochemistry*, 49(3):623–634, 2010.
- [23] Colleen A. Flanagan. GPCR-radioligand binding assays. *Methods in Cell Biology*, 132:191–215, 2016.
- [24] Sai Kiran Sharma, Serge K. Lyashchenko, Hijin A. Park, Nagavarakishore Pillarsetty, Yorann Roux, Jiong Wu, Sophie Poty, Kathryn M. Tully, John T. Poirier, and Jason S. Lewis. A rapid bead-based radioligand binding assay for the determination of target-binding fraction and quality control of radiopharmaceuticals. *Nuclear Medicine and Biology*, 71:32–38, 2019.

- [25] Edward C. Hulme and Mike A. Trevethick. Ligand binding assays at equilibrium: validation and interpretation. *British Journal of Pharmacology*, 161(6):1219–1237, 2010.
- [26] Nicole M Roussey and Alex Dickson. Local ion densities can influence transition paths of molecular binding. *Frontiers in Molecular Biosciences*, 9, 2022.
- [27] Sachin Potadia, Ashima Bagaria, and Deepak Chopra. Pmolecular dynamics simulation of proteins: A brief overview. *The Journal of Physical Chemistry and Biophysics*, 4(6), 2014.
- [28] Jacob D. Durrant and J. Andrew McCammon. Molecular dynamics simulations and drug discovery. *BMC Biology*, 9:1–9, 2011.
- [29] B.J. Alder and T.E Wainwright. Phase transition for a hard sphere system. *Journal of Chemical Physics*, 27:1208–1209, 1957.
- [30] Terra Sztain, Surl-Hee Ahn, Anthony T. Bogetti, Lorenzo Casalino, Jory A. Goldsmith, Evan Seitz, Ryan S. McCool, Fiona L. Kearns, Francisco Acosta-Reyes, Suvrajit Maji, Ghocheh Mashayekhi, J. Andrew McCammon, Abbas Ourmazd, Joachim Frank, Jason S. McLellan, Lillian T. Chong, and Rommie E. Amaro. A glycan gate controls opening of the sars-cov-2 spike protein. *Nature Chemistry*, 13:963–968, 2021.
- [31] Tom Dixon, Arzu Uyar, Shelagh Ferguson-Miller, and Alex Dickson. Membrane-mediated ligand unbinding of the pk-11195 ligand from tspo. *Biophysical Journal*, 120(1):158–167, 2021.
- [32] Nazanin Donyapour, Nicole M. Roussey, and Alex Dickson. Revo: Resampling of ensembles by variation optimization. *Journal of Chemical Physics*, 150(24):244112, 2019.
- [33] Egor S. Kolesnikov, Ivan Yu. Gushchin, Petr A. Zhilyaev, and Alexey V. Onufriev. Similarities and differences between na⁺ and k⁺ distributions around dna obtained with three popular water models. *Journal of Chemical Theory and Computation*, 17(11):7246–7259, 2021.
- [34] Pascal Auffinger, Lukasz Bielecki, and Eric Westhof. Symmetric k⁺ and mg²⁺ ion-binding sites in the 5 s rrna loop e inferred from molecular dynamics simulations. *Journal of Molecular Biology*, 335(2):555–571, 2004.
- [35] Mate Erdos, Michalis Frangou, Thijs J. H. Vlught, and Othonas A. Moulto. Diffusivity of α -, β -, γ -cyclodextrin and the inclusion complex of β -cyclodextrin: Ibuprofen in aqueous solutions; a molecular dynamics simulation study. *Fluid Phase Equilibria*, 528, 2021.

- [36] Alper T. Celebi, Seyed Hossein Jamali, André Bardow, Thijs J. H. Vlugt, and Othonas A. Moulton. Finite-size effects of diffusion coefficients computed from molecular dynamics: a review of what we have learned so far. *Molecular Simulation*, 47(10-11):831–845, 2021.
- [37] Romelia Salomon-Ferrer, David A. Case, and Ross C. Walker. An overview of the amber biomolecular simulation package. *WIREs Computational Molecular Science*, 3:198–210, 2013.
- [38] D.A. Case, III T.E. Cheatham, H. Gohlke T. Darden, R. Luo, Jr. K.M. Merz, A. Onufriev, C. Simmerling, B. Wang, and R. Woods. The amber biomolecular simulation programs. *The Journal of Computational Chemistry*, 26:1668–1688, 2005.
- [39] K. Vanommeslaeghe and A. D. MacKerell. Automation of the charmm general force field (cgenff) i: Bond perception and atom typing. *Journal of Chemical Information and Modeling*, 52(12):3144–3154, 2012.
- [40] K. Vanommeslaeghe, E. Prabhu Raman, and A. D. MacKerell. Automation of the charmm general force field (cgenff) ii: Assignment of bonded parameters and partial atomic charges. *Journal of Chemical Information and Modeling*, 52(12):3155–3168, 2012.
- [41] Nawaf Bou-Rabee. Time integrators for molecular dynamics. *Entropy*, 16:138–162, 2014.
- [42] Robert D. Skeel and Jesus A. Izaguirre. An impulse integrator for langevin dynamics. *Molecular Physics*, 100(24):3885–3891, 2002.
- [43] David E. Shaw, J.P. Grossman, Joseph A. Bank, Brannon Batson, J. Adam Butts, Jack C. Chao, Martin M. Deneroff, Ron O. Dror, Amos Even, Christopher H. Fenton, Anthony Forte, Joseph Gagliardo, Gennette Gill, Brian Greskamp, C. Richard Ho, Douglas J. Ierardi, Lev Iserovich, Jeffrey S. Kuskin, Richard H. Larson, Timothy Layman, Li-Siang Lee, Adam K. Lerer, Chester Li, Daniel Killebrew, Kenneth M. Mackenzie, Shark Yeuk-Hai Mok, Mark A. Moraes, Rolf Mueller, Lawrence J. Nociolo, Jon L. Peticolas, Terry Quan, Daniel Ramot, John K. Salmon, Daniele P. Scarpazza, U. Ben Schafer, Naseer Siddique, Christopher W. Snyder, Jochen Spengler, Ping Tak Peter Tang, Michael Theobald, Horia Toma, Brian Towles, Benjamin Vitale, Stanley C. Wang, and Cliff Young. Anton 2: Raising the bar for performance and programmability in a special purpose molecular dynamics supercomputer. In *SC '14: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, pages 41–53, 2014.
- [44] Johannes Kästner. Umbrella sampling. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 1(6):932–942, 2011.

- [45] Alessandro Laio and Michele Parrinello. Escaping free-energy minima. *Proceedings of the National Academy of Sciences*, 99(20):12562–12566, 2002.
- [46] Alex Dickson, Pratyush Tiwary, and Harish Vashisth. Kinetics of ligand binding through advanced computational approaches: A review. *Current Topics in Medicinal Chemistry*, 17:2626–2641, 2017.
- [47] John D. Chodera, David L. Mobley, Michael R. Shirts, Richard W. Dixon, Kim Branson, and Vijay S. Pande. Alchemical free energy methods for drug discovery: Progress and challenges. *Current Opinions in Structural Biology*, 21(2):150–160, 2011.
- [48] Robert W. Zwanzig. High-temperature equation of state by a perturbation method. i. nonpolar gases. *The Journal of Chemical Physics*, 22, 1954.
- [49] Gavin E Crooks. Nonequilibrium measurements of free energy differences for microscopically reversible markovian systems. *Journal of Statistical Physics*, 90(5-6):1481–1487, 1998.
- [50] John D. Chodera, David L. Mobley, Michael R. Shirts, Richard W. Dixon, Kim Branson, , and Vijay S. Pande. Alchemical free energy methods for drug discovery: Progress and challenges. *Current Opinions in Structural Biology*, 21(2):150–160, 2012.
- [51] Lin Frank Song and Kenneth M. Merz Jr. Evolution of alchemical free energy methods in drug discovery. *Journal of Chemical Information and Modeling*, 60(11):5308–5318, 2020.
- [52] Dharmeshkumar Patel, Jagdish Suresh Patel, and F. Marty Ytreberg. Implementing and assessing an alchemical method for calculating protein–protein binding free energy. *Journal of Chemical Theory and Computation*, 17(4):2457–2464, 2021.
- [53] Pratyush Tiwary and Michele Parrinello. From metadynamics to dynamics. *Physics Review Letters*, 111(23), 2013.
- [54] Richard J. Zamora, Blas P. Uberuaga, Danny Perez, and Arthur F Voter. The modern temperature-accelerated dynamics approach. *Annual Review of Chemical and Biomolecular Engineering*, 7(7):87–110, 2016.
- [55] Cameron F. Abrams and Eric Vanden-Eijnden. Large-scale conformational sampling of proteins using temperature-accelerated molecular dynamics. *Proceedings of the National Academy of Sciences*, 107(11):4961–4966, 2010.
- [56] Shankar Kumar, John M. Rosenberg, Djamal Bouzida, Robert H. Swendsen, and Peter A. Kollman. The weighted histogram analysis method for free-energy calculations on biomolecules. i. the method. *Journal of Computational Chemistry*, 13(8):1011–1021, 1992.

- [57] http://membrane.urmc.rochester.edu/wordpress/?page_id=126.
- [58] Giovanni Bussi and Alessandro Laoi. Using metadynamics to explore complex free-energy landscapes. *Nature Review Physics*, 2:200–212, 2020.
- [59] Radim Brozek Ivo Kabelka and Robert Vácha. Selecting collective variables and free-energy methods for peptide translocation across membranes. *Journal of Chemical Information and Modeling*, 61(2):819–830, 2021.
- [60] Pavan Ravindra, Zachary Smith, and Pratyush Tiwary. Automatic mutual information noise omission (amino): generating order parameters for molecular systems. *Molecular Systems Design and Engineering*, 5(1):339–348, 2020.
- [61] Michael D. Ward, Maxwell I. Zimmerman, Artur Miller, Moses Chung, S. J. Swamidass, and Gregory R. Bowman. Deep learning the structural determinants of protein biochemical properties by comparing structural ensembles with diffnets. *Nature Communications*, 12(3023), 2021.
- [62] Luigi Bonati, GiovanniMaria Piccini, and Michele Parrinello. Deep learning the slow modes for rare events sampling. *Proceedings of the National Academy of Sciences*, 118(44), 2021.
- [63] Pratyush Tiwary and B. J. Berne. Spectral gap optimization of order parameters for sampling complex molecular systems. *Proceedings of the National Academy of Sciences*, 113(11):2839–2844, 2016.
- [64] Andreas Mardt, Luca Pasquali, Hao Wu, and Frank Noe. Vampnets for deep learning of molecular kinetics. *Nature Communications*, 9(9), 2018.
- [65] Ron Elber. Long-timescale simulation methods. *Current Opinion in Structural Biology*, 15(2):151–156, 2005.
- [66] Peter Majek and Ron Elber. Milestoning without a reaction coordinate. *Journal of Chemical Theory and Computation*, 6(6):1805–1817, 2011.
- [67] Alexander T. Hawk, Sai Sriharsha M. Konda, and Dmitrii E. Marakov. Computation of transit times using the milestoning method with applications to polymer translocation. *Journal of Chemical Physics*, 139(6), 2013.
- [68] Gary A Huber and Sangtae Kim. Weighted-ensemble Brownian dynamics simulations for protein association reactions. *Biophysical Journal*, 70:97–110, 1996.
- [69] Alex Dickson and Charles L. Brooks. Wexplore: Hierarchical exploration of high-dimensional spaces using the weighted ensemble algorithm. *Journal of Physical Chemistry B*, 118(13):3532–3542, 2014.

- [70] Daniel M. Zuckerman and Lillian T. Chong. Weighted ensemble simulation: Review of methodology, applications, and software. *Annual Review of Biophysics*, 46:43–57, 2017.
- [71] J. D. Chodera and F. Noeé. Markov state models of biomolecular conformational dynamics. *Current Opinions in Structural Biology*, 25:135–144, 2014.
- [72] Peter G. Bolhuis, David Chandler, Christoph Dellago, and Phillip L. Geissler. Transition path sampling: Throwing ropes over rough mountain passes, in the dark. *Annual Review of Physical Chemistry*, 53:291–318, 2002.
- [73] R. J. Allen, C. Valeriani, and P. R. ten Wolde. Forward flux sampling for rare event simulations. *Journal of Physics: Condensed Matter*, 21(46), 2009.
- [74] Tom Dixon, Samuel D. Lotz, and Alex Dickson. Predicting ligand binding affinity using on- and off-rates for the sampl6 sampling challenge. *Journal of Computer-Aided Molecular Design*, 32:1001–1012, 2018.
- [75] Samuel D. Lotz and Alex Dickson. Wepy: A flexible software framework for simulating rare events with weighted ensemble resampling. *ACS Omega*, 5(49):31608–31623, 2020.
- [76] Nicole M. Roussey and Alex Dickson. Enhanced jarzynski free energy calculations using weighted ensemble. *Journal of Chemical Physics*, 153, 2020.
- [77] Matthew C. Zwier, Joshua L. Adelman, Joseph W. Kaus, Adam J. Pratt, Kim F. Wong, Nicholas B. Rego, Ernesto Suarez, Steveb Lettieri, David W. Wang, Michael Grabe, Daniel M. Zuckerman, and Lillian T. Chong. Westpa: An interoperable, highly scalable software package for weighted ensemble simulation and analysis. *Journal of Chemical Theory and Computation*, 11(2):800–809, 2015.
- [78] Joshua L. Adelman and Michael Grabe. Simulating current-voltage relationships for a narrow ion channel using the weighted ensemble method. *Journal of Chemical Theory and Computation*, 11:1907–1918, 2015.
- [79] Bin W. Zhang, David Jasnow, and Daniel M. Zuckerman. Efficient and verified simulation of a path ensemble for conformational change in a united-residue model of calmodulin. *Proceedings of the National Academy of Sciences*, 104(46):18043–18048, 2007.
- [80] Joshua L. Adelman, Amy L. Dale, Matthew C. Zwier, Divesh Bhatt, Lillian T. Chong, Daniel M. Zuckerman, and Michael Grabe. Simulations of the alternating access mechanism of the sodium symporter mhp1. *Biophysical Journal*, 101(10):2399–2407, 2011.
- [81] Ali S. Saglam and Lillian T. Chong. Highly efficient computation of the basal k_{on} using direct simulation of protein–protein association with flexible molecular models.

- Journal of Physical Chemistry B*, 120(1):117–122, 2016.
- [82] Badi’ Abdul-Wahid, Haoyun Feng, Dinesh Rajan, Ronan Costaouec, Eric Darve, Douglas Thain, and Jesús A. Izaguirre. AWE-WQ: Fast-Forwarding Molecular Dynamics Using the Accelerated Weighted Ensemble. *Journal of Chemical Information and Modeling*, 54(10):3033–3043, 2014.
- [83] Alex Dickson, Aryeh Warmflash, and Aaron R Dinner. Nonequilibrium umbrella sampling in spaces of many order parameters. *Journal of Chemical Physics*, 130(7):074104, 2009.
- [84] Alex Dickson, Mark Maienschein-Cline, Allison Tovo-Dwyer, Jeff R. Hammond, and Aaron R. Dinner. Flow-dependent unfolding and refolding of an rna by nonequilibrium umbrella sampling. *Journal of Chemical Theory and Computation*, 7(9):2710–2720, 2011.
- [85] Eric Vanden-Eijnden and Maddalena Venturoli. Exact rate calculations by trajectory parallelization and tilting. *Journal of Chemical Physics*, 131:044120, 2009.
- [86] Ernesto Suárez, Steven Lettieri, Matthew C. Zwier, Carsen A. Stringer, Sundar Raman Subramanian, Lillian T. Chong, and Daniel M. Zuckerman. Simultaneous computation of dynamical and equilibrium information using a weighted ensemble of trajectories. *Journal of Chemical Theory and Computation*, 10(7):2658–2667, 2014.
- [87] Christoph Dellago and Gerhard Hummer. Computing equilibrium free energies using non-equilibrium molecular dynamics. *Entropy*, 16:41–61, 2014.
- [88] Christopher Jarzynski. Nonequilibrium equality for free energy differences. *Physical Review Letters*, 78(14):2690, 1997.
- [89] Christopher Jarzynski. Rare events and the convergence of exponentially averaged work values. *Physical Review E*, 73(4), Apr 2006.
- [90] Andrea Rizzi, Travis Jense, David R. Slochow, Matteo Aldeghi, Vytautas Gapsys, Dimitris Ntekoumes, Stefano Bosisio, Michail Papadourkis, Neil M. Henriksen, Bert L. de Groot, Zoe Cournia, Alex Dickson, Julien Michel, Michael K. Gilson, Michael R. Shirts, David L. Mobley, and John D. Chodera. The sampl6 sampling challenge: assessing the reliability and efficiency of binding free energy calculations. *Journal of Computer Aided Molecular Design*, 34:601–633, 2020.
- [91] Agostino Bruno, Elisabetta Barresi, Nicola Simola, Eleonora Da Pozzo, Barbara Costa, Ettore Novellino, Federico Da Settimo, Claudia Martini, Sabrina Taliani, and Sandro Cosconati. Unbinding of translocator protein 18 kda (tspo) ligands: From in vitro residence time to in vivo efficacy via in silico simulations. *ACS Chemical Neuroscience*, 10:3805–3814, 2019.

- [92] Sampl challenges. <https://www.samplchallenges.org/>.
- [93] Sampl github. <https://github.com/samplchallenges>.
- [94] Andrea Rizzi, Steven Murkli, John N McNeill, Wei yao, Matthew Sullivan, Michael K Gilson, Michael W Chiu, Lyle Isaacs, Bruce C Gibb, David L Mobley, and John D Chodera. Overview of the sampl6 host-guest binding affinity prediction challenge. *Journal of Computer Aided Molecular Design*, 32(10):937–963, 2018.
- [95] William Humphrey, Andrew Dalke, and Klaus Schulten. VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics*, 14:33–38, 1996.
- [96] Sampl6 github. <https://github.com/samplchallenges/SAMPL6>.
- [97] Sampl9 github. <https://github.com/samplchallenges/SAMPL9>.
- [98] Clara D. Christ, Alan E. Mark, and Wilfred F. Van Gunsteren. Basic ingredients of free energy calculations: A review. *Journal of Computational Chemistry*, 31(8):1569–1582, 2010.
- [99] Zoe Cournia, Bryce Allen, and Woody Sherman. *Journal of Chemical Information and Modeling*, 57(12):2911–2937, 2017.
- [100] Lingle Wang, Jennifer Chambers, and Robert Abel. *Protein–Ligand Binding Free Energy Calculations with FEP+*, volume 2022, pages 201–232. 2019.
- [101] William L. Jorgensen and Laura L. Thomas. Perspective on free-energy perturbation calculations for chemical equilibria. *Journal of Chemical Theory and Computation*, 4(6):869–876, 2008.
- [102] J M Torrie and J P Valleau. Non-physical sampling distributions in Monte-Carlo free-energy estimation- umbrella sampling. *Journal of Computational Physics*, 23:187–199, 1977.
- [103] Comparison of the umbrella sampling and the double decoupling method in binding free energy predictions for SAMPL6 octa-acid host–guest challenges. *Journal of Computer-Aided Molecular Design*, 32(10):1075–1086, 2018.
- [104] Nina Singhal, Christopher D Snow, and Vijay S Pande. Using path sampling to build better Markovian state models: predicting the folding rate and mechanism of a tryptophan zipper beta hairpin. *Journal of Chemical Physics*, 121(1):415–25, 2004.
- [105] Shuo Gu, Daniel Adriano Silva, Luming Meng, Alexander Yue, and Xuhui Huang. Quantitatively Characterizing the Ligand Binding Mechanisms of Choline Binding Protein Using Markov State Model Analysis. *PLoS Computational Biology*,

- 10(8):e1003767, 2014.
- [106] Alessandro Laio and Michele Parrinello. Escaping free-energy minima. *Proceedings of the National Academy of Sciences*, 99(20):12562–12566, 2002.
- [107] Pratyush Tiwary, Jagannath Mondal, and B J Berne. How and when does an anticancer drug leave its binding site ? *Science Advances*, 3(5):e1700014, 2017.
- [108] A K Faradjian and R Elber. Computing time scales from reaction coordinates by milestoning. *Journal of Chemical Physics*, 120:10880, 2004.
- [109] Lane W. Votapka, Benjamin R. Jagger, Alexandra L. Heyneman, and Rommie E. Amaro. SEEKR: Simulation Enabled Estimation of Kinetic Rates, A Computational Tool to Estimate Molecular Kinetics and Its Application to Trypsin-Benzamidine Binding. *Journal of Physical Chemistry B*, 121(15):3597–3606, 2017.
- [110] Alex Dickson. Mapping the Ligand Binding Landscape. *Biophysical Journal*, 115(9):1707–1719, 2018.
- [111] Matteo Aldeghi, Joseph P. Bluck, and Philip C. Biggin. *Absolute Alchemical Free Energy Calculations for Ligand Binding: A Beginner’s Guide*, pages 199–232. Springer New York, New York, NY, 2018.
- [112] Michael K. Gilson, James A. Given, Bruce L. Bush, and J. Andrew McCammon. The statistical-thermodynamic basis for computation of binding affinities: A critical review. *Biophysical Journal*, 72(3):1047–1069, 1997.
- [113] John G. Kirkwood. Statistical mechanics of fluid mixtures. *Journal of Chemical Physics*, 3(5):300–313, 1935.
- [114] Yuji Sugita, Akio Kitao, and Yuko Okamoto. Multidimensional replica-exchange method for free-energy calculations. *Journal of Chemical Physics*, 113(15):6042–6051, 2000.
- [115] P. Procacci and G. Guarnieri. Sampl6 blind predictions of water-octanol partition coefficients using nonequilibrium alchemical approaches. *Journal of Computer-Aided Drug Design*, 34:371–384, 2020.
- [116] Hui Xiong, Alejandro Crespo, Marcelo Marti, Dario Estrin, and Adrian Roitberg. Free energy calculations with non-equilibrium methods: Applications of the jarzynski relationship. *Theoretical Chemistry Accounts*, 116:338–346, 2006.
- [117] Rhonald C. Lua and Alexander Y. Grosberg. Practical applicability of the jarzynski relation in statistical mechanics: a pedagogical example. *Journal of Physical Chemistry B*, 109(14):6805–6811, 2005.

- [118] Jeff Gore, Felix Ritort, and Carlos Bustamante. Bias and error in estimates of equilibrium free-energy differences from nonequilibrium measurements. *Proceedings of the National Academy of Sciences*, 100(22):12564–12569, 2003.
- [119] Ignacia Echeverria and L. Mario Amzel. Helix propensities calculations for amino acids in alanine based peptides using jarzynski’s equality. *Proteins: Structure, Function, and Bioinformatics*, 78(5):1302–1310, 2010.
- [120] C Dellago, G Bolhuis P, and L Geissler, P. Transition path sampling methods. *Computer Simulations in Condensed Matter Systems: From Materials to Chemical Biology*, 1:349–391, 2006.
- [121] Sean Sun. Equilibrium free energies from path sampling of nonequilibrium trajectories. *Journal of Chemical Physics*, 118:5769–5775, 2003.
- [122] F. Marty Ytreberg and Daniel M. Zuckerman. Single-ensemble nonequilibrium path-sampling estimates of free energy differences. *Journal of Chemical Physics*, 120, 2004.
- [123] Aryeh Warmflash, Prabhakar Bhimalapuram, and Aaron R Dinner. Umbrella sampling for nonequilibrium processes. *Journal of Chemical Physics*, 127(15):154112, oct 2007.
- [124] Roland Assaraf, Michel Caffarel, and Anatole Khelif. Diffusion monte carlo methods with a fixed number of walkers. *APS Physics*, 61(4566):4566–4575, 2000.
- [125] Aaron R. Dinner, Jonathan C. Mattingly, Jeremy O.B. Tempkin, Brian Van Koten, and Jonathan Weare. Trajectory stratification of stochastic dynamics. *SIAM Review*, 60(4):909–938, 2018.
- [126] Mathias Rousset and Gabriel Stoltz. Equilibrium sampling from nonequilibrium dynamics. *Journal of Statistical Physics*, 123, 2006.
- [127] Daniel M. Zuckerman and Lillian T. Chong. Weighted ensemble simulation: Review of methodology, applications, and software. *Annual Review of Biophysics*, 46(1):43–57, 2017.
- [128] Badi’ Abdul-Wahid, Haoyun Feng, Dinesh Rajan, Ronan Costauuec, Eric Darve, Douglas Thain, and Jesús A. Izaguirre. Awe-wq: Fast-forwarding molecular dynamics using the accelerated weighted ensemble. *Journal of Chemical Information and Modeling*, 54:3033–3043, 2014.
- [129] Bin W. Zhang, David Jasnow, and Daniel M. Zuckerman. Efficient and verified simulation of a path ensemble for conformational change in a united-residue model of calmodulin. *Proceedings of the National Academy of Sciences*, 104:18043–18048, 2007.
- [130] Matthew C. Zwier, Adam J. Pratt, Joshua L. Adelman, Joseph W. Kaus, Daniel M.

- Zuckerman, and Lillian T. Chong. Efficient atomistic simulation of pathways and calculation of rate constants for a protein–peptide binding process: Application to the mdm2 protein and an intrinsically disordered p53 peptide. *Journal of Physical Chemistry Letters*, 7:3440–3445, 2016.
- [131] Upendra Adhikari, Barmak Mostofian, Jeremy Copperman, Sundar Raman Subramanian, Andrew A. Petersen, and Daniel M. Zuckerman. Computational estimation of microsecond to second atomistic folding times. *Journal of the American Chemical Society*, 141:6519–6526, 2019.
- [132] Tony Lelièvre, Mathias Roussel, and Gabriel Stoltz. *Free Energy Computations: A Mathematical Perspective*. Imperial College Press, 01 2010.
- [133] <https://github.com/ADicksonLab/wepy>.
- [134] <https://github.com/ADicksonLab/wepy-activity>.
- [135] David Aristoff and Daniel M. Zuckerman. Optimizing weighted ensemble sampling of steady states. *Multiscale Modeling and Simulation*, 18:646–673, 2018.
- [136] David Aristoff. An ergodic theorem for weighted ensemble. *arXiv*, 2019.
- [137] <https://openmmtools.readthedocs.io/en/0.18.1/>.
- [138] Peter Eastman, Jason Swails, John D. Chodera, Robert T. McGibbon, Yutong Zhao, Kyle A. Beauchamp, Lee-Ping Wang, Andrew C. Simmonett, Matthew P. Harrigan, Chaya D. Stern, Rafal P. Wiewiora, Bernard R. Brooks, and Vijay S. Pande. Openmm 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Computational Biology*, 13(7):1–17, 07 2017.
- [139] Gerhard Hummer and Attila Szabo. Free energy reconstruction from nonequilibrium single-molecule pulling experiments. *Proceedings of the National Academy of Sciences*, 98(7):3658–3661, 2001.
- [140] Alan M. Ferrenberg and Robert H. Swendsen. Optimized monte carlo data analysis. *Physical Review Letters*, 63:1195–1198, Sep 1989.
- [141] Mathieu Bastian, Sebastien Heymann, and Mathieu Jacomy. Gephi: An open source software for exploring and manipulating networks, 2009.
- [142] Robert M. Turner, Thomas Speck, and Juan P. Garrahan. Meta-work and the analogous jarzynski relation in ensembles of dynamical trajectories. *Journal of Statistical Mechanics*, 2014, 2014.
- [143] Dynamic order-disorder in atomistic models of structural glass formers. *Science (New*

- York, N. Y.), 323(5919):1309–1313, 2009.
- [144] J P Garrahan, Robert L Jack, V Lecomte, E Pitard, K Van Duijvendijk, and F Van Wijland. First-order dynamical phase transition in models of glasses: an approach based on ensembles of histories. *Journal of Physics A: Mathematical and Theoretical*, 42(7):29, 2010.
- [145] Antonia S.J.S. Mey, Phillip L. Geissler, and Juan P. Garrahan. Rare-event trajectory ensemble analysis reveals metastable dynamical phases in lattice proteins. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 89:032109, 2014.
- [146] Michael R. Shirts, Eric Bair, Giles Hooker, and Vijay S. Pande. Equilibrium free energies from nonequilibrium measurements using maximum-likelihood methods. *Physical Review Letters*, 91(14):1–4, 2003.
- [147] R. P. Feynman. Space-time approach to non-relativistic quantum mechanics. *Review of Modern Physics*, 20:367–387, Apr 1948.
- [148] Samuel D. Lotz and Alex Dickson. Multiple ligand unbinding pathways and ligand-induced destabilization revealed by wexplore. *Biophysical Journal*, 112:620–629, 2018.
- [149] Riccardo Capelli, Paolo Carloni, and Michele Parrinello. Exhaustive search of ligand binding pathways via volume-based metadynamics. *Journal of Physical Chemistry Letters*, 10(12):3495–3499, 2019.
- [150] Daria B. Kokh, Marta Amaral, Joerg Bomke, Ulrich Grädler, Djordje Musil, Hans-Peter Buchstaller, Matthias K. Dreyer, Matthias Frech, Maryse Lowinski, Francois Vallee, Marc Bianciotto, Alexey Rak, and Rebecca C. Wade. Estimation of drug-target residence times by -random acceleration molecular dynamics simulations. *Journal of Chemical Theory and Computation*, 14(7):3859–3869, 2018.
- [151] Yong Zhang and Gregory A. Voth. A combined metadynamics and umbrella sampling method for the calculation of ion permeation free energy profiles. *Journal of Chemical Theory and Computation*, 7(7):2277–2283, 2011.
- [152] P. L. Chau. Water movement during ligand unbinding from receptor site. *Biophysics Journal*, 87(1):121–128, 2004.
- [153] Partyush Tiwary, Jagannath Mondal, Joseph A. Morrone, and B. J. Berne. Role of water and steric constraints in the kinetics of cavity–ligand unbinding. *Proceedings of the National Academy of Sciences*, 112(39):12015–12019, 2015.
- [154] Manuela Maurer and Chirs Oostenbrink. Water in protein hydration and ligand recognition. *Journal of Molecular Recognition*, 32(12), 2019.

- [155] Valerio Rizzi, Luigi Bonati, Narjes Ansari, and Michele Parrinello. The role of water in host-guest interaction. *Nature Communications*, 12(93), 2021.
- [156] Luigi Bonati, Valerio Rizzi, and Michele Parrinello. Data-driven collective variables for enhanced sampling. *Journal of Physical Chemistry Letters*, 11:2998–3004, 2020.
- [157] Aykut Erbas, Monica Olvera de la Cruz, and John F. Marko. Effects of electrostatic interactions on ligand dissociation kinetics. *Physical Review E*, 91(2-1), 2018.
- [158] Rob Hall, Tom Dixon, and Alex Dickson. On calculating free energy differences using ensembles of transition paths. *Frontiers in Molecular Biociences*, 7, 2020.
- [159] Michail Papadourakis, Stefano Bosisio, and Julien Michel. Blinded predictions of standard binding free energies: lessons learned from the sampl6 challenge. *Journal of Computer Aided Molecular Design*, 32:1047–1058, 2018.
- [160] Lin F. Song, Nupur Bansal, Zheng Zheng, and Kenneth M. Merz Jr. Detailed potential of mean force studies on host-guest systems from the sampl6 challenge. *Journal of Computer Aided Molecular Design*, 32:1013–1026, 2018.
- [161] Jian Yin, Niel M. Henriksen, David R. Slochower, Michael R. Shirts, Michael W. Chiu, David L. Mobley, and Michael K. Gilson. Overview of the sampl5 host-guest challenge: Are we doing better? *Journal of Computer Aided Molecular Design*, 31(1):1–19, 2016.
- [162] Masafumi Yano, Takeshi Yamamoto, Shigeki Kobayashi, and Masunori Matsuzaki. Role of ryanodine receptor as a ca²⁺ regulatory center in normal and failing hearts. *Journal of Cardiology*, 53(1):1–7, 2009.
- [163] Soohyung Park and Wonpil Im. Two dimensional window exchange umbrella sampling for transmembrane helix assembly. *Journal of Chemical Theory and Computation*, 9:13–17, 2013.
- [164] Alex Dickson, Logan S. Ahlstrom, and Charles L. Brooks III. Coupled folding and binding with 2d window-exchange umbrella sampling. *Journal of Computational Chemistry*, 37:587–594, 2015.
- [165] Zoe Cournia, Bryce Allen, and Woody Sherman. Relative binding free energy calculations in drug discovery: Recent advances and practical considerations. *Journal of Chemical Information and Modeling*, 57(12):2911–2937, 2017.
- [166] A. Lyubartsev, A. Martsinovski, S. Shevkunov, and P. Vorontsov-Velyaminov. New approach to monte carlo calculation of the free energy: Method of expanded ensembles. *Journal of Chemical Physics*, 96(3):1776–1783, 1992.
- [167] A. Barducci, G. Bussi, and M. Parrinello. Well-tempered metadynamics: a smoothly

- converging and tunable free-energy method. *Phys. Rev. Lett.*, 100(2), 2008.
- [168] Nupur Bansal, Zheng Zheng, David S. Cerutti, and Kenneth M Merz. On the fly estimation of host-guest binding free energies using the movable type method: participation in the sampl5 blind challenge. *Journal of Computer-Aided Molecular Design*, 1(47):47–60, 2017.
- [169] Wanli You, Zhiye Tang, and Chia en A. Chang. Potential mean force from umbrella sampling simulations: What can we learn and what is missed? *Journal of Chemical Theory and Computation*, 15(4):2433–2443, 2019.
- [170] Giovanni Bussi and Alessandro Laio. Using metadynamics to explore complex free-energy landscapes. *Nat. Rev. Phys.*, 2:200–212, 2020.
- [171] Anita de Ruiter and Chris Oostenbrink. Free energy calculations of protein-ligand interactions. *Current Opinions in Chemical Biology*, 15(4):547–552, 2011.
- [172] C.D. Christ and T. Fox. Accuracy assessment and automation of free energy calculations for drug design. *Journal of Chemical Information and Modeling*, 54(1):108–120, 2013.
- [173] D.L. Mobley and M.K. Gilson. Predicting binding free energies: frontiers and benchmarks. *Ann. Rev. Biophys.*, 46:531–538, 2017.
- [174] D.L. Mobley and P.V. Klimovich. Perspective: Alchemical free energy calculations for drug discovery. *Journal of Chemical Physics*, 137(230901), 2012.
- [175] Matthew C. Zwier, Joshua L. Adelman, Joseph W. Kaus, Adam J. Pratt, Kim F. Wong, Nicholas B. Rego, Ernesto Suárez, Steven Lettieri, David W. Wang, Michael Grabe, Daniel M. Zuckerman, and Lillian T. Chong. Westpa: An interoperable, highly scalable software package for weighted ensemble simulation and analysis. *Journal of Chemical Theory and Computation*, 11(2):800–809, 2015.
- [176] https://gitlab.com/ADicksonLab/cutoff-revo_tutorial.
- [177] Sunhwan Jo, Taehoon Kim, Vidyashankara G. Iyer, and Wonpil Im. Charmm-gui: A web-based graphical user interface for charmm. *Journal of Computational Chemistry*, 29(11):1859–1865, 2008.
- [178] B. R. Brooks, C. L. Brooks III, A. D. Mackerell Jr., L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caffisch, L. Caves, Q. Cui, A.R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, and M. Karplus. Charmm: The biomolecular simulation program. *Journal of Computational Chem-*

istry, 30(10):1545–1614, 2009.

- [179] J. Lee, X. Cheng, J.M. Swails, M.S. Yeom, P.K. Eastman, J.A. Lemkul, S. Wei, J. Buckner, J.C. Jeong, Y. Qi, S. Jo, V.S. Pande, D.A. Case, C.L. Brooks III, A.D. MacKerell Jr, J.B. Klauda, and W. Im. Charmm-gui input generator for namd, gromacs, amber, openmm, and charmm/openmm simulations using the charmm36 additive force field. *Journal of Chemical Theory and Computation*, 12(1):405–413, 2015.
- [180] Steven Murkli, John N. McNeill, and Lyle Isaacs. Cucurbit[8]uril guest complexes: blinded dataset for the sampl6 challenge. *Supramolecular Chemistry*, 31, 2018.
- [181] <https://github.com/samplchallenges/SAMPL9>.
- [182] D. Aristoff, J. Copperman, G. Simpson, R.J. Webber, and D.M. Zuckerman. Weighted ensemble: Recent mathematical developments. *arXiv*, 2022.

APPENDIX A: ENHANCED JARZYNSKI FREE ENERGY CALCULATIONS USING WEIGHTED ENSEMBLE

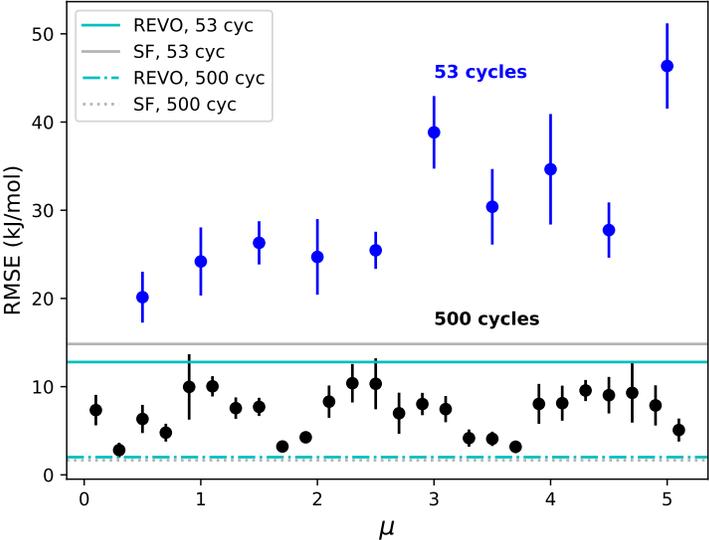


Figure A.1: Error in $\Delta\Delta F$ for importance resampling for multiple amplification values (μ). Error is shown for $\epsilon = 20$ for 1000-trajectory sets using 53 and 500 resampling cycles. Error bars show uncertainty calculated using three sets of trajectories. Horizontal lines represent the equivalent values obtained for straightforward and REVO for 1000 trajectories (error not shown). For 53 cycle simulations, we found that values of $\mu < 1$ perform best, with the RMSE gradually increasing as the value of μ increased. For 500 cycle simulations, the RMSE stayed within a smaller range with smaller error bars.

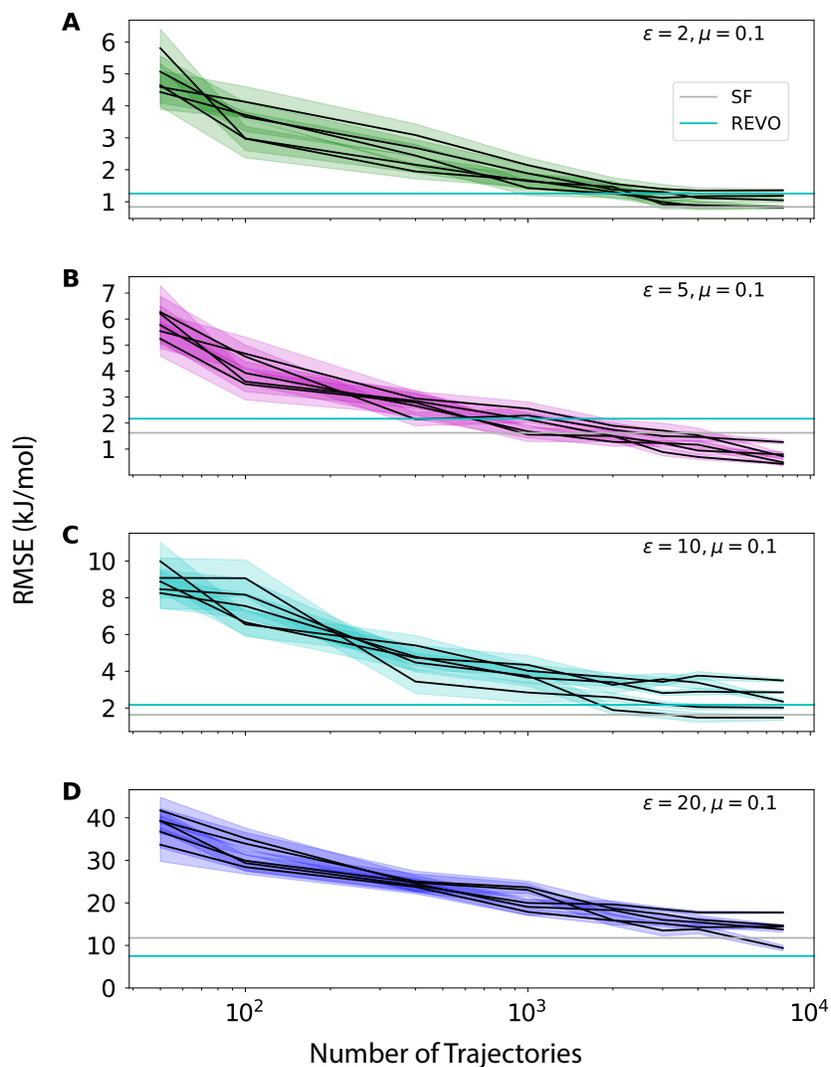


Figure A.2: Error per dataset. The convergence of the unbinding free energy for $\epsilon =$ (A) 2, (B) 5, (C) 10, and (D) 20 kcal/mol for five separate sets of IMP simulations as a function of the total number of walkers used to calculate each free energy value for all values of ϵ . For IMP simulations, $\mu = 0.1$. The horizontal lines represent the value (error not shown) obtained from 8000 trajectories for straight forward and REVO simulations.

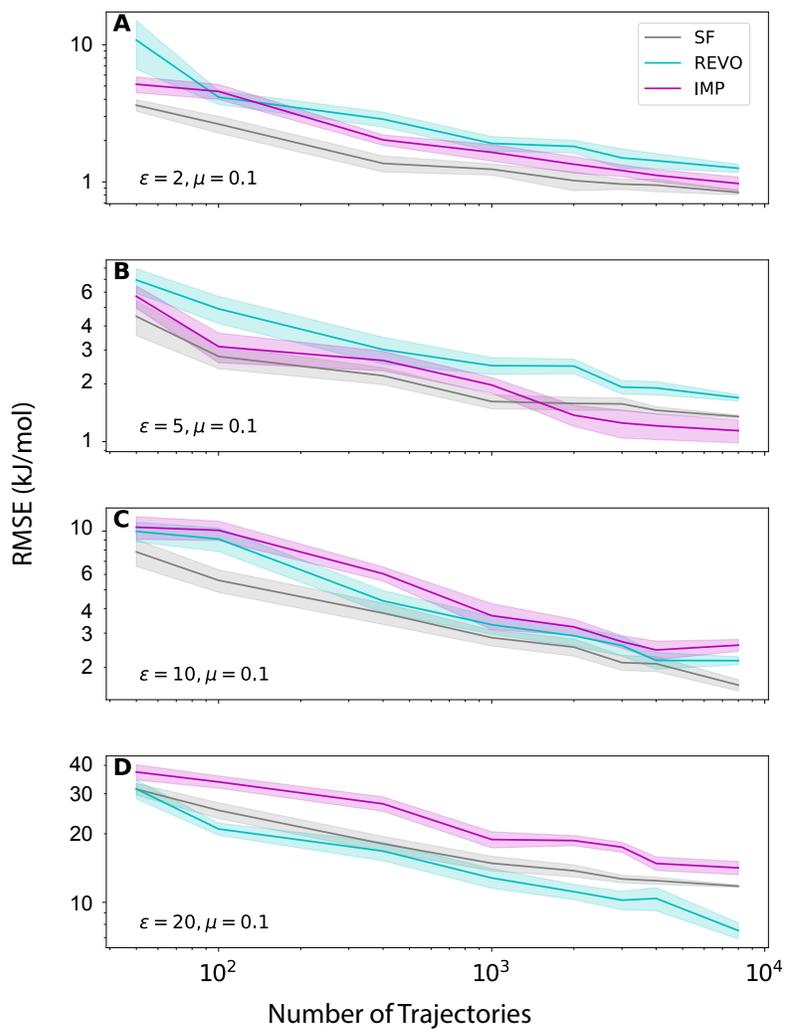


Figure A.3: Error vs Time. The convergence of the unbinding free energy for $\epsilon =$ (A) 2, (B) 5, (C) 10, and (D) 20 kcal/mol for straightforward, REVO, and IMP as a function of the total number of walkers used to calculate each free energy value for all values of ϵ . For IMP simulations, $\mu = 0.1$. For IMP, files were randomly selected from all five sets of data shown in A.2.

APPENDIX B: LOCAL ION DENSITIES CAN INFLUENCE TRANSITION PATHS OF MOLECULAR BINDING

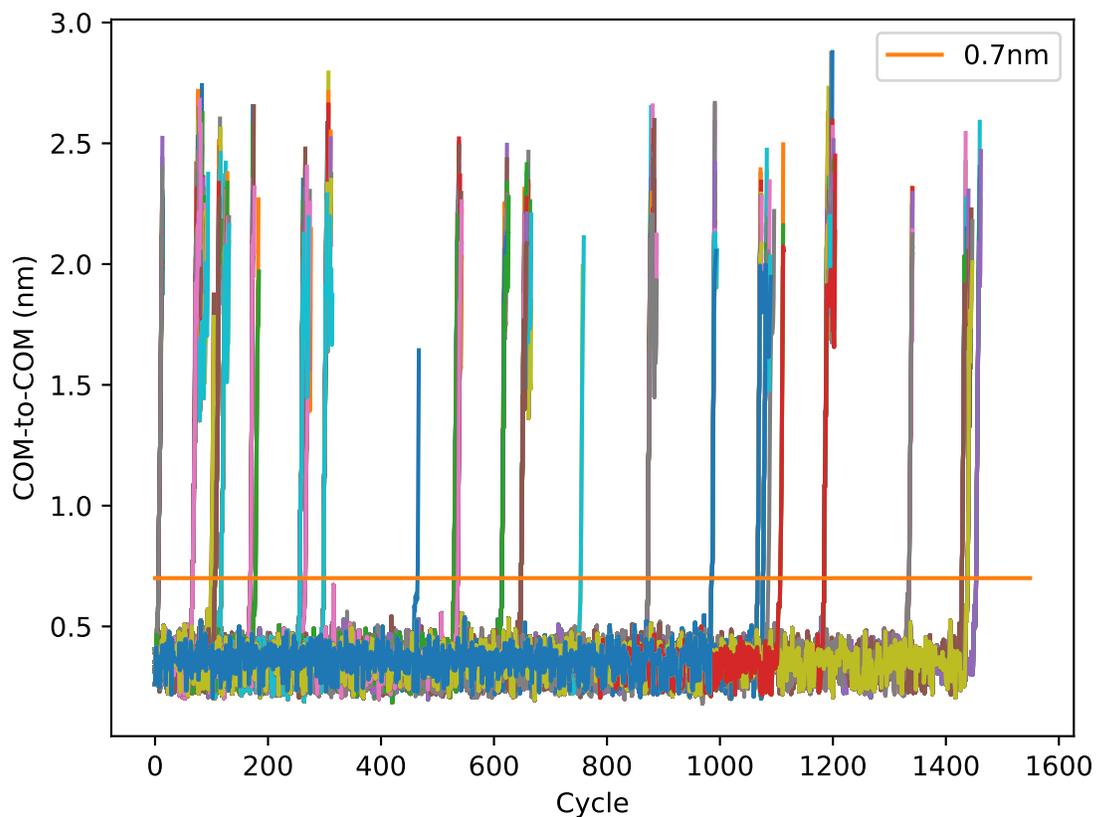


Figure B.1: Center of Mass to Center of Mass Distances. The COM-to-COM distance between the host and guest is shown for several hundred unbinding events for OA-G6. Rebinding only occurs below 0.7nm (orange line).

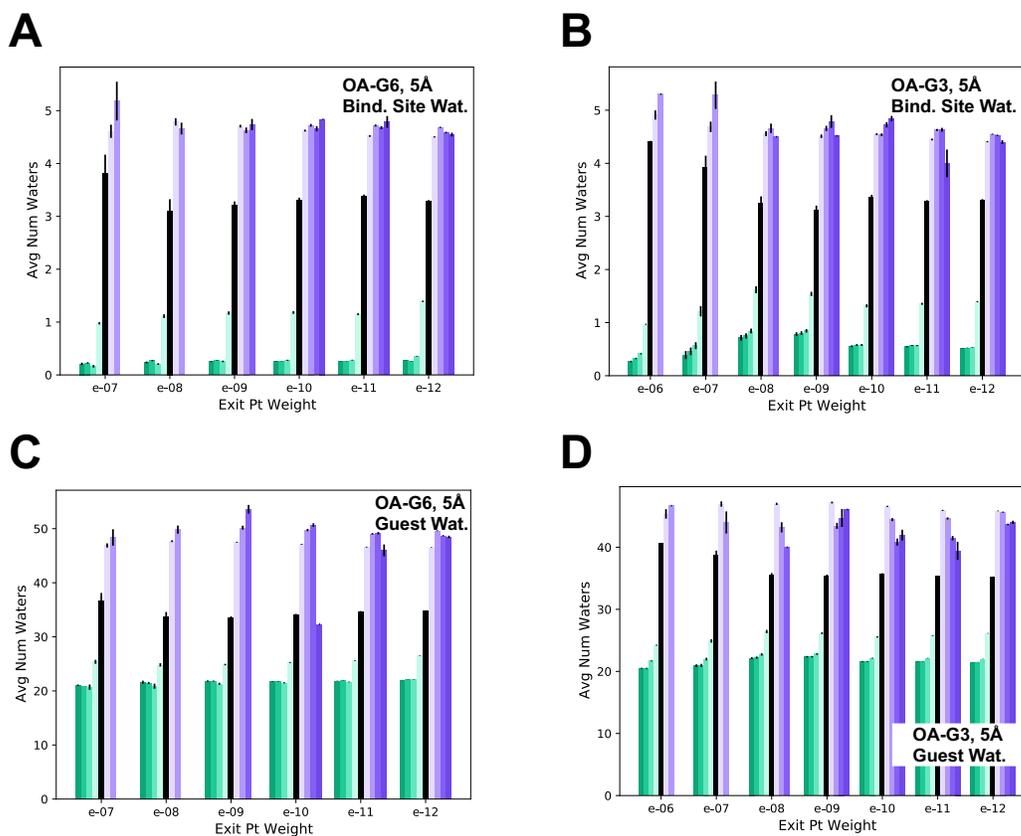


Figure B.2: 5Å Analysis of Water Based Features. Molecule counts for waters with results organized by exit point probability. The legend in Fig. B.3 applies to all four plots. The average total water count (5Å) in the binding site of the host for A) OA-G6 and B) OA-G3. The average total water count (5Å) around the guest for C) OA-G6 and D) OA-G3.

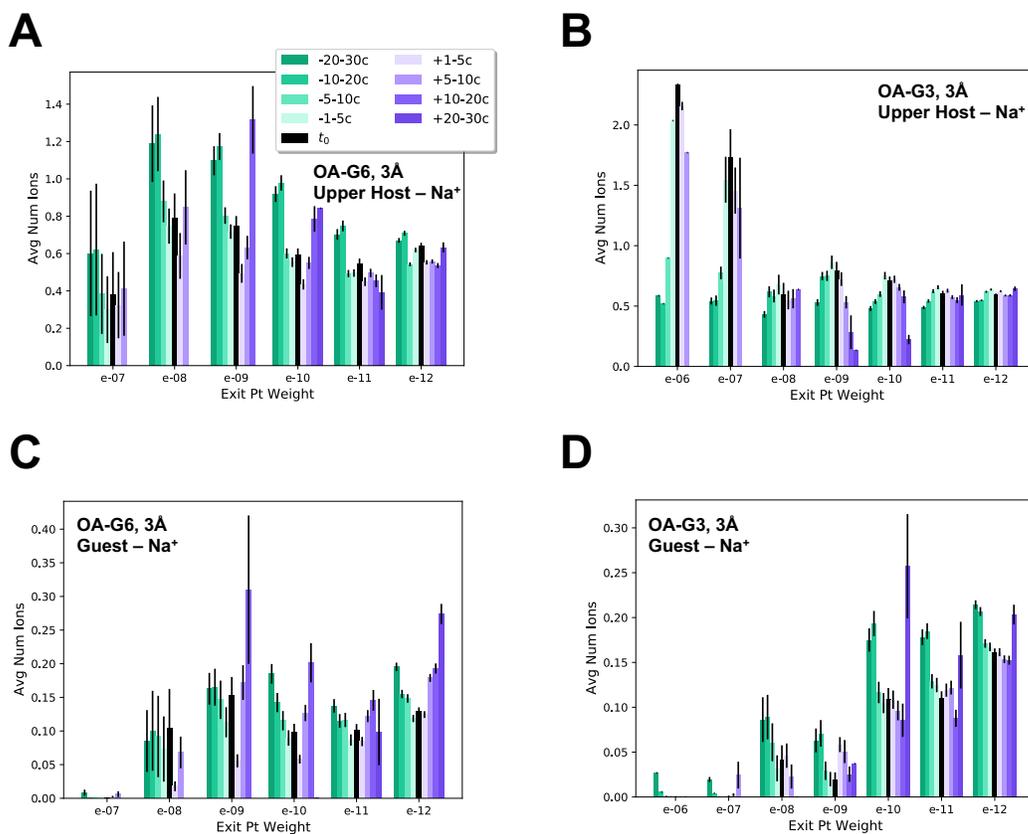


Figure B.3: 3Å Analysis of Ion Features. Molecule counts for Na^+ ions with results organized by exit point probability. The legend in A applies to all four plots. The average total ion count (3Å) around the upper negative charges of the host for A) OA-G6 and B) OA-G3. The average total ion count (3Å) around the guest for C) OA-G6 and D) OA-G3.

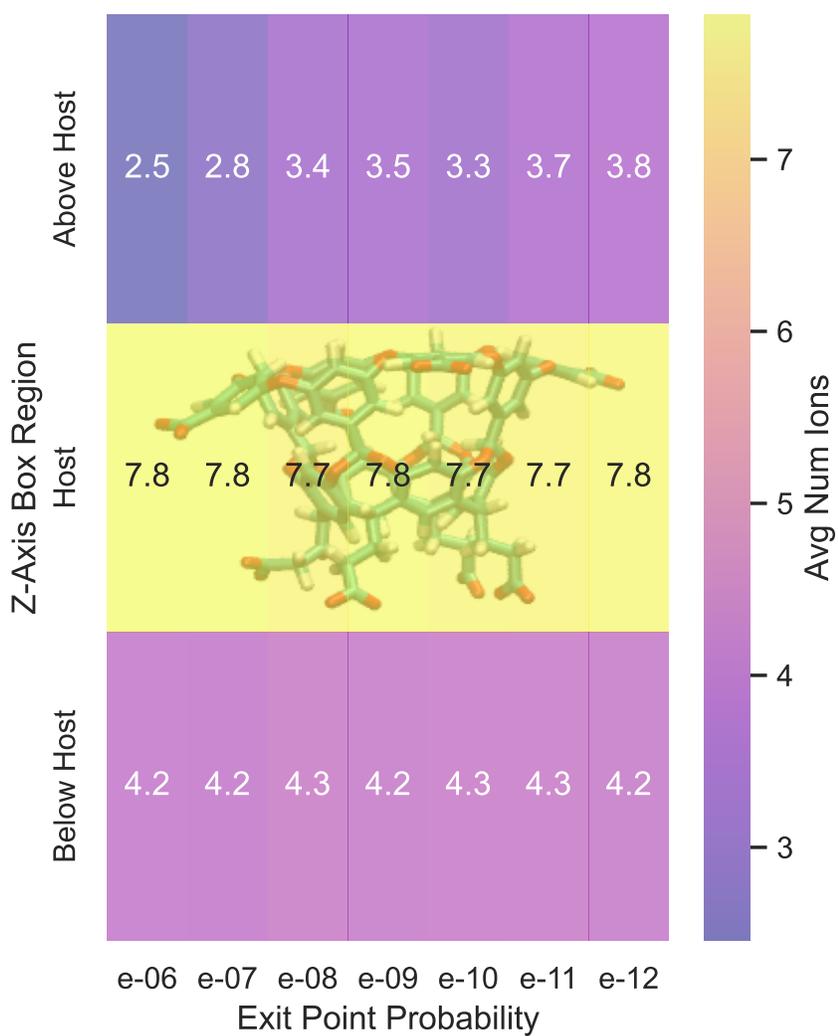


Figure B.4: Z-Axis Ion Density. The average number of ions in the region of space above the host, at host level, and below the host for the t_0 adjacent cycles ($[t_0 - 3, t_0 + 3]$) for all OA-G3 exit points. Corresponding errors by weight are as follows:
 $\pm 0.78401176, 0.77167242, 0.52423166, 0.35946138, 0.21856345, 0.14066211, 0.07147705$.

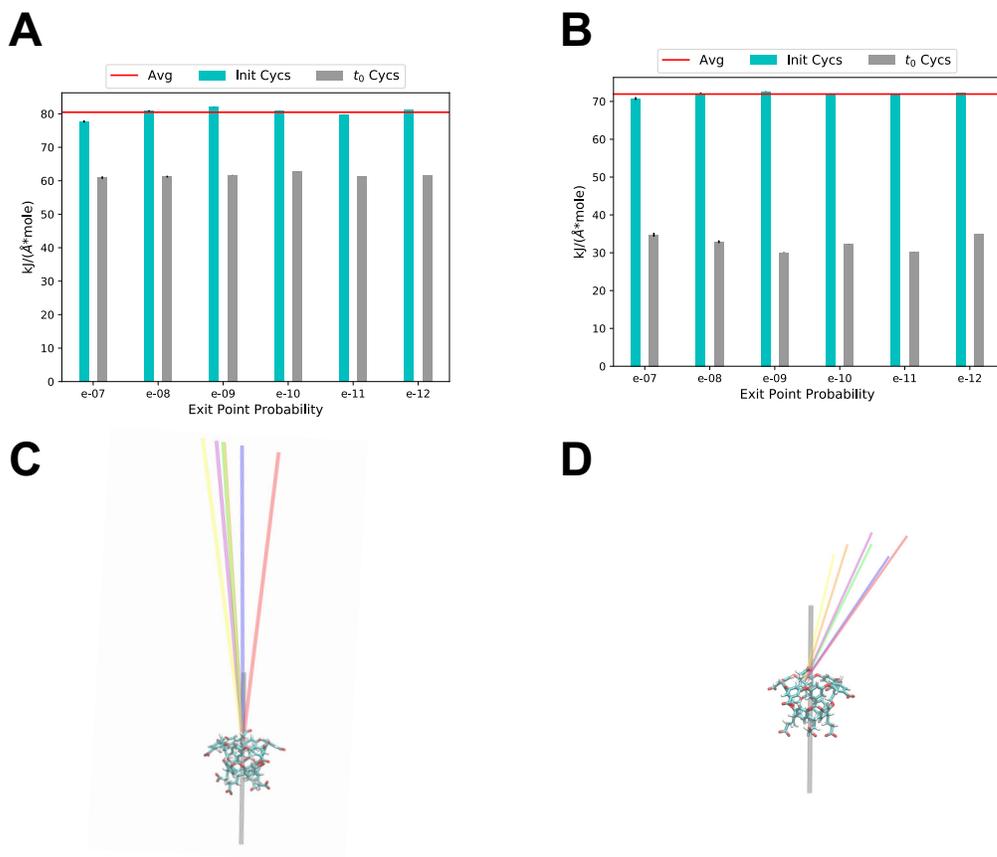


Figure B.5: Electrostatic Analysis. A) The magnitude and B) the z-axis contribution of the electrostatic force for the initial cycles (cyan) and t_0 cycles. The average of the initial cycles is shown as a horizontal red line. C) The force vectors for the initial cycles. D) The force vectors at t_0 . For both C and D, the gray bar is the Z axis. Red, orange, yellow, green, blue and violet correspond to 10^{-7} to 10^{-12} exit point probabilities.

APPENDIX C: QUALITY OVER QUANTITY: SAMPLING HIGH PROBABILITY RARE EVENTS WITH THE WEIGHTED ENSEMBLE ALGORITHM

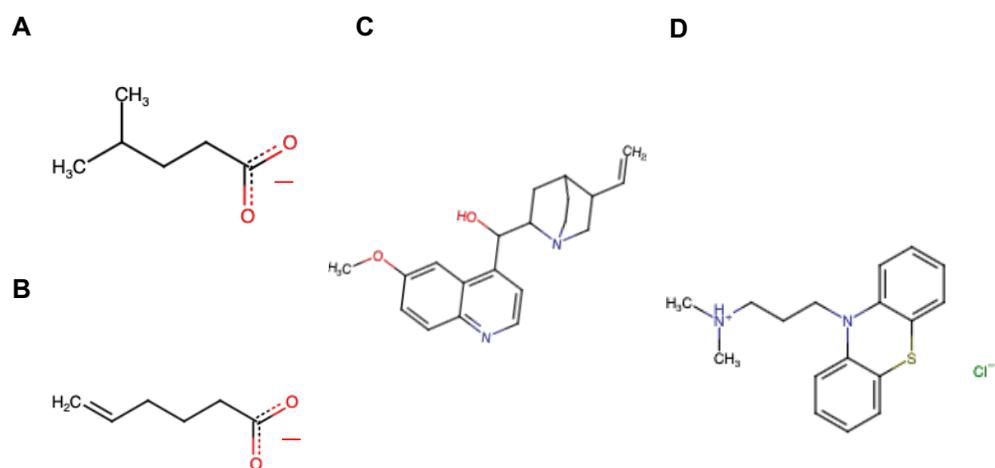


Figure C.1: Structures of the four ligands used. The four ligands used in the simulation where A) is G6, or 4-methyl pentanoic acid, B) is G3 from the OA-G3 system, or 5-hexanoic acid, C) is G3, or quinine, from the CB8-G3 system and D) is PMZ, or promazine hydrochloride.

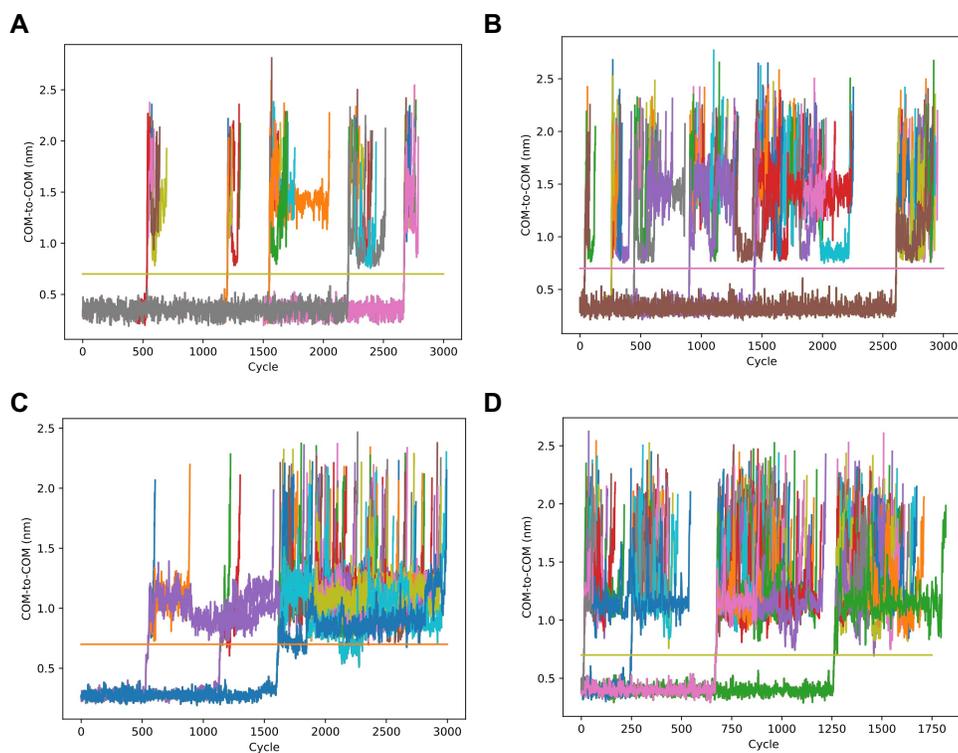


Figure C.2: d_{COM} distances. Host to guest d_{COM} distance used to determine cutoff values for all four systems used. For all systems, a subset of unbinding events for one run is plotted. The horizontal line represents 0.7 nm. A) COM distances for OA-G6 B) OA-G3 C) CB8-G3 D) and β -CD-PMZ.

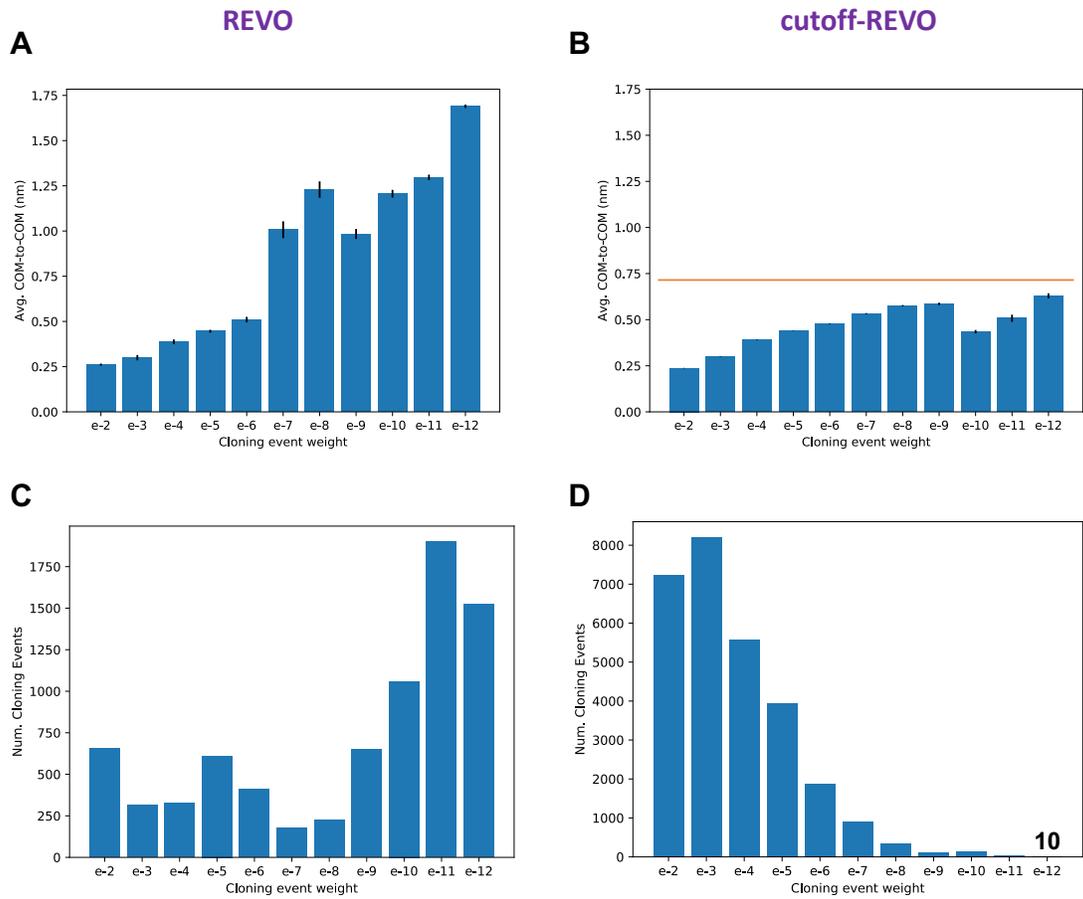


Figure C.3: Cloning event d_{COM} distances and weights. The average cloning d_{COM} distance organized by event weights across five simulations is shown for β -CD-PMZ for A) REVO and B) cutoff-REVO. The orange horizontal line in B represents the cloning cutoff (0.7 nm). The total number of cloning events per weight bin across five simulations is shown for C) REVO and D) cutoff-REVO.

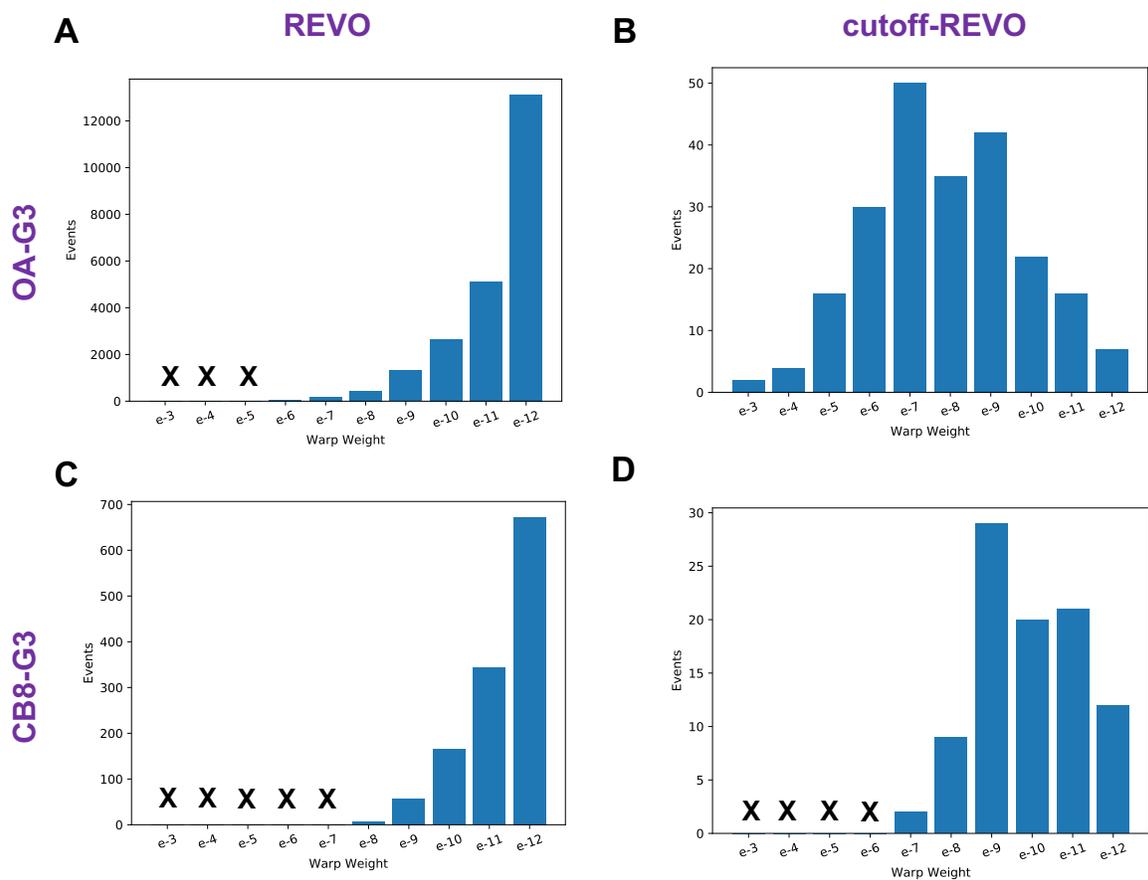


Figure C.4: Warped walker weights for both methods. Walker weights at the time of unbinding for standard REVO (left) and cutoff-REVO (right) for A and B) the OA-G3 system (22901 events total for REVO and 224 total for cutoff-REVO) and for C and D) the CB8-G3 system (1246 events total for REVO and 93 total for cutoff-REVO).

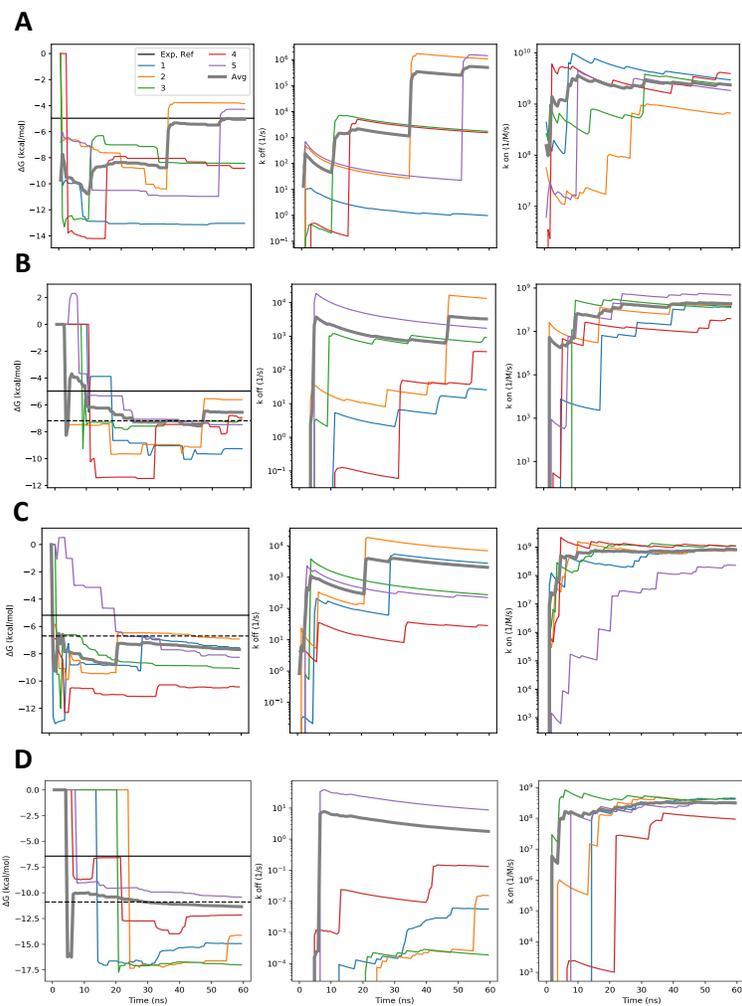


Figure C.5: Rates and free energies for standard REVO. The free energy, k_{off} , and k_{on} for all systems. Legend in A applies to all plots. Dashed lines in B, C, and D are computational reference values. A) β -CD-PMZ. B) OA-G6. C) OA-G3. D) CB8-G3. The X axis represents trajectory length and not aggregate simulation time.

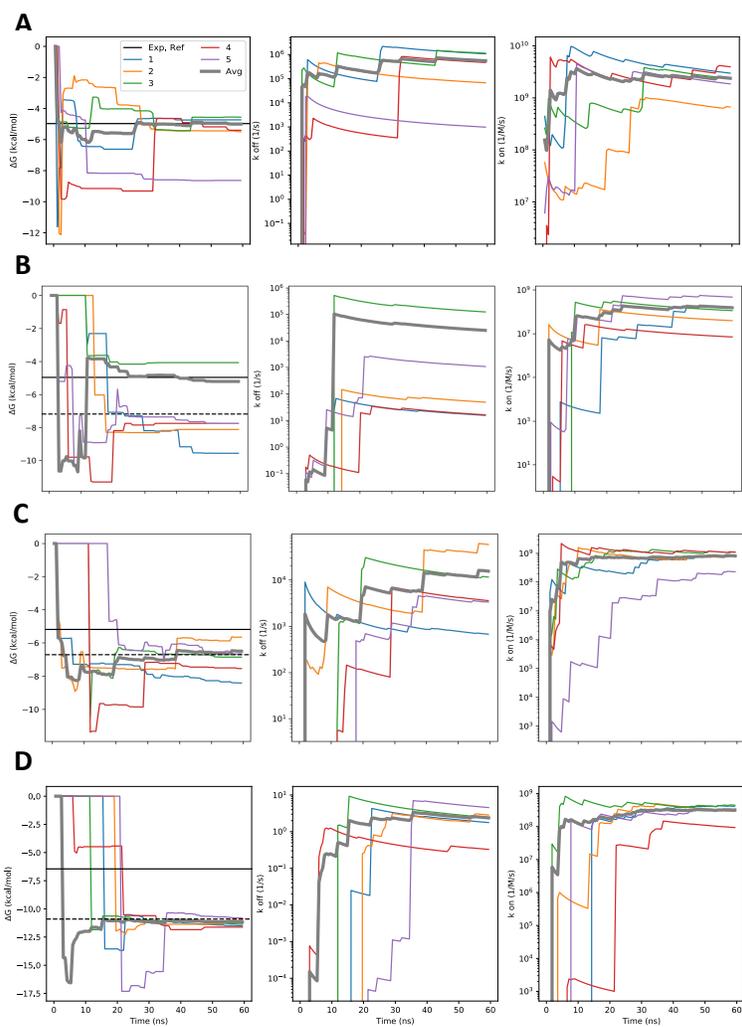


Figure C.6: Rates and free energies for cutoff-REVO. The free energy, k_{off} , and k_{on} for all systems. Legend in A applies to all plots. Dashed lines in B, C, and D are computational reference values. A) β -CD-PMZ. B) OA-G6. C) OA-G3. D) CB8-G3. The X axis represents trajectory length and not aggregate simulation time.