UNIVERSITY OF CALIFORNIA, SAN DIEGO

Improvements and Applications of Alchemical Free Energy, Constant pH and Accelerated Molecular Dynamics Calculations in the AMBER Molecular Dynamics Suite.

A dissertation submitted in partial satisfaction of the requirements for the degree of Doctor of Philosophy

in

Chemistry with a Specialization in Multi-Scale Biology

by

Daniel Janson Mermelstein

Committee in Charge:

>   Professor J. Andrew McCammon, Co-Chair
>   Professor Ross C. Walker, Co-Chair
>   Professor Michael K. Gilson
>   Professor Katja Lindenberg
>   Professor Andrew McCulloch
>   Professor Wei Wang

2018

The Dissertation of Daniel Janson Mermelstein is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____
                                                    Co-Chair

_____
                                                    Co-Chair

University of California, San Diego

2018

DEDICATION

To my Mother and Father: Thank you for always being there for me whenever I needed you.

Everything I accomplish reflects as much on you and how you raised me as it does on my own

virtues.

To Liza and Eric: Thank you for your numerous words and acts of encouragement during my

Ph.D. You've been great mentors throughout, and made this journey much easier and more

enjoyable.

To Myles: Hopefully, the work I've done during my Ph.D. will help create a brighter future for

you.

To Grandma Claire: Thank you for always supporting me in all my endeavors. You inspire all of

us every day. I will always cherish your holiday cards and our frequent phone updates.

To Grandma Mimi, Grandpa Steve, and Grandpa Bubbi: I wish you could have been here for this

journey. I know you would have been proud. I miss you all very much.

EPIGRAPH

There's always money in the banana stand

George Bluth, Sr., <u>Top Banana, Arrested Development</u>

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

Å – Angstroms
ΔG – Gibb's Free Energy
ABFE – Absolute Binding Free Energy
AD – Alzheimer's Disease
AFE – Alchemical Free Energy
AMBER – AMBER Molecular Dynamics Suite
aMD – Accelerated Molecular Dynamics
APBS – Adaptive Poisson-Boltzmann Solver
Apo – Ligand Free
Asp – Aspartic Acid
BACE-1 – Beta Secretase-1
BAR – Bennett's Acceptance Ratio
Cl – Chloride
CpHMD – Constant pH Molecular Dynamics
Dih – Dihedral
DPDP – Double Precision Double Precision
DPFP – Double Precision Floating Precision
EEL – Coulombic Contribution to Potential Energy
FEP – Free Energy Perturbation
FESetup – Free Energy Setup
FEW – Free Energy Workflow
GAFF – General Amber Forcefield
GB – Generalized Born
GPU – Graphics Processor Unit
GROMACS - Groningen Machine for Chemical Simulations
H-mass repartitioning – Hydrogen Mass Repartitioning
MBAR – Multistate Bennett's Acceptance Ratio
Holo – Protein with Ligand Bound
Kcal/mol – Kilocalories per Mole
MC – Monte Carlo
MD – Molecular Dynamics
ms – Millisecond
Na – Sodium
NaMD – Nanoscale Molecular Dynamics
ns – Nanosecond
OpenMM – Open Source High-performance Molecular Mechanics
PDB – Protein Databank
PE – Potential Energy
pH – Power of Hydrogen
pK_a – Logarithm of Acid Dissociation Constant
PME – Particle Mesh Ewald
RBFE – Relative Binding Free Energy
SPFP – Single Precision Floating Precision

Syk – Spleen Tyrosine Kinase
TI – Thermodynamic Integration
Tyr – Tyrosine
VDW – Van der Waals
WHAM – Weighted Histogram Analysis Method
WT – Wildtype

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

Pecora de Barros, Dr. Giulia Palermo, Dr. Jeffrey Wagner, Dr. Joseph Kaus, Kartik Lakshmi, Professor Lane Votapka, Mitchell Hurwitz, Dr. Mindy Huang, Dr. Nuo Wang, Dr. Olivia Pierce, Sara Carmigniani, Sarah Kochanek, Sasha Heyneman, Terra Sztain, and Professor Yinglong Miao for many inciteful scientific conversations, helpful mentoring moments, and for keeping me entertained during my Ph.D.

Last, but certainly not least, I would like to thank my entire family, especially Liza, Eric, Myles, Grandma Claire, Grandma Mimi, Nancy, Marc Chytilo, Sam, Barbara and Bob, Mara, Dan Golding, Dylan, Ariana, Karen, Roger, Lee, Renee, Lynne, Mark Blossom, and everyone else who offered words and acts of kindness and encouragement during this journey. And of course, most importantly, my parents. You have been my role models from day 1. You have shown unwavering support and kindness. You have always believed in me, and treated me as your equal. I could not have done this without you.

Chapter 2 is a modified reprint of the material as it will appear in "Mermelstein, D. J.**,** Lin, C., Nelson, G., Kretsch, R., McCammon, J. A., Walker, R. C. (2018) Fast and Flexible GPU Accelerated Binding Free Energy Calculations within the AMBER Molecular Dynamics Package, *Journal of Computation Chemistry. DOI: 10.1002/jcc.25187*." The dissertation author was the primary investigator and author of this paper.

Chapter 3 is a modified reprint of the material as it will appear in "Mermelstein, D. J.**,** McCammon, J. A., Walker, R. C. (2018) pH dependent conformational dynamics of Beta-secretase 1: a molecular dynamics study *submitted*.". The dissertation author was the primary investigator and author of this paper.

Chapter 4 is a modified reprint of the material as will appear in "Feher, V. A., Schiffer, J. M., Mermelstein, D. J., Mih, N., Pierce, L.C.T., McCammon, J.A., Amaro, R. E. Mechanisms for

Benzene Dissociation through the Excited State of T4 Lysozyme L99A, *submitted*". The dissertation author was one of two primary authors of the paper. The dissertation author made key contributions to the analysis of the data.

VITA

2011-2014	Undergraduate Research Fellow Van Wynsberghe Lab, Department of Chemistry, Hamilton College
2012	Summer Undergraduate Research Fellow Kinnel Lab, Department of Chemistry, Hamilton College
2013	Summer Undergraduate Research Fellow Tran Lab, Wadsworth Labs, New York State Department of Health
2014	Bachelor of Arts in Chemistry, Minor in Mathematics Hamilton College
2014 – 2018	Graduate Student Researcher, McCammon and Walker Labs University of California, San Diego
2018	Doctor of Philosophy in Chemistry with a Specialization in Multi-Scale Biology University of California, San Diego

PUBLICATIONS

Mermelstein, D. J**,** Lin, C., Nelson, G., Kretsch, R., McCammon, J. A., Walker, R. C. (2018) Fast and Flexible GPU Accelerated Binding Free Energy Calculations within the AMBER Molecular Dynamics Package, *Journal of Computation Chemistry. 10.1002/jcc.25187*."

Mermelstein, D. J**,** McCammon, J. A., Walker, R. C. (2018) pH dependent conformational dynamics of Beta-secretase 1: a molecular dynamics study *submitted*.

Feher, V. A., Schiffer, J. M., Mermelstein, D. J., Mih, N., Pierce, L.C.T., McCammon, J.A., Amaro, R. E. Mechanisms for Benzene Dissociation through the Excited State of T4 Lysozyme L99A, *submitted*

Case, D. A.; Cerutti, D. S.; Cheatham III, T. E.; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; Greene, D.; Homeyer, N.; Izadi, S.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Madej, B.; Mermelstein, D.; Merz, K. M.; Monard, G. H.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Simmerling, C. L.; Botello-Smith, W. M.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; Xiao, L.; York, D. M.; Kollman, P. A., AMBER 2017, University of California, San Francisco.

AMBER 16. Case, D. A.; Cerutti, D. S.; Cheatham III, T. E.; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; Greene, D.; Homeyer, N.; Izadi, S.; Kovalenko, A.; Lee, T. S.;

LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Madej, B.; Mermelstein, D.; Merz, K. M.; Monard, G. H.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Simmerling, C. L.; Botello-Smith, W. M.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; Xiao, L.; York, D. M.; Kollman, P. A., AMBER 2016, University of California, San Francisco.

ABSTRACT OF THE DISSERTATION

Improvements and Applications of Alchemical Free Energy, Constant pH and Accelerated

Molecular Dynamics Calculations in the AMBER Molecular Dynamics Suite

by

Daniel Janson Mermelstein

Doctor of Philosophy in Chemistry with a Specialization in Multi-Scale Biology

University of California, San Diego, 2018

Professor J. Andrew McCammon Co-Chair

Professor Ross C. Walker, Co-Chair

The field of molecular dynamics is rapidly advancing as new theoretical techniques, software optimizations and computer architectures are unveiled seemingly daily. In this dissertation, I first review the best practices and recent developments in molecular dynamics based alchemical free energy calculations in the AMBER molecular dynamics suite. I then present a graphics processor enabled implementation of alchemical free energy calculations with performance 360 times that of the existing CPU implementation, while maintaining equivalent accuracy through the judicious use of a combination of floating point and fixed precision. Next, I

discuss the application of constant pH molecular dynamics to investigate the role of water in beta secretase-1 catalysis. This protein has a known pH dependence and is a key target in the fight against Alzheimer's Disease. I present a hypothesis for the role of the flap region in regulating beta secretase-1 catalysis. Finally, I have investigated benzene egress from the binding pocket of the L99A mutant of T4 lysozyme using accelerated molecular dynamics. It is found that benzene exits the binding pocket by a multistep process from the buried cavity to ultimate release through an opening between the F/G, H and I helices.

Chapter 1 : A review of alchemical free energy calculations and their applications in the

AMBER molecular dynamics suite: current approaches, success stories, and future directions

Introduction

In the two years that have elapsed since January 2016, there have been as many new

publications with the topic "Molecular Dynamics" and "Drug Design" or "Drug Discovery" as in

any five-year period before 2016. Clearly, we have begun to reach the era in which molecular

dynamics calculations are fast enough and reliable enough to be used routinely. Much of the

increased interest in molecular dynamics simulations is due to the use of graphics processors

(GPUs) for MD simulations. These have allowed access to previously inaccessible timescales.

Among open source and academic molecular dynamics suites, AMBER has traditionally led the

way in the development of GPU accelerated tools[1–3], although notable strides have been made in

NaMD,[4] GROMACS[5–8], OpenMM[9], CHARMM[10], Desmond[11], and ACEMD[12] as well.

Recently, alchemical free energy (AFE) and constant pH molecular dynamics (CpHMD) were

ported to GPUs in the AMBER molecular dynamics suite.[13,14] With these recent developments,

this seems an opportune time to review the current capabilities of the AMBER molecular

dynamics suite with regards to drug discovery. As these new developments (AFE and CpHMD)

are focused on the calculation of thermodynamic quantities, this review will emphasize tools

used to calculate thermodynamic properties of use in drug design. The topics covered in this

review in detail include AMBER's GPU enabled alchemical free energy, constant pH MD toward

the calculation of pH dependent free energy, and free energies of solvation using alchemical free

energy. For each of these topics we will present a summary of the key theoretical developments

in AMBER, followed by recent examples of their application to drug design efforts. Additionally, since each of these techniques is built upon alchemical free energy, a review of the most up to date practices will first be presented.

Best practices in molecular dynamics based alchemical free energy calculations

The development of accurate tools to calculate thermodynamic properties of biomolecular systems has been a goal of the computational chemistry community almost since the first MD simulations of a protein were run.[15]   One such tool rapidly growing in popularity is MD based AFE. Since the first application of AFE calculations to model ligand-binding systems,[16] there has been a steady stream of theoretical milestones reached towards the applicability of these calculations to systems of biological and physical relevance. Here, we will attempt to outline the most important theoretical developments for AFE, what the current accepted practices are, and how these practices apply to AMBER.

Analysis methods

The historically most used equilibrium AFE analysis methods include free energy perturbation (FEP), developed initially by Zwanzig,[17] thermodynamic integration (TI), developed initially by Kirkwood,[18] Bennett acceptance ratio (BAR) from Charles Bennett,[19] the expansion to more than two simultaneous states multistate Bennett acceptance ratio (MBAR) by Shirts and Chodera,[20] and the weighted histogram analysis method (WHAM) by Kumar et al[21] based on the work of Ferrenberg and Swendsen.[22] While each of these methods can be shown to be equivalent, in practice the results obtained vary due to a combination of limited sampling time and unique errors intrinsic to the method. For example, WHAM, which requires discretization of

samples into bins according to the degree of freedom of interest, will have an error associated

with discretizing the values of the degree of freedom. Currently, BAR/MBAR and TI see the

most use in the literature. WHAM can only ever be as precise as MBAR, never superior, and so

has fallen out of use for AFE calculations of late. FEP tends to suffer from difficulty in overlap

between neighboring states leading to poor convergence relative to other estimators. A complete

review of why this is the case can be found in the book by Chipot and Pohorille.[23]

Comparison of TI and BAR/MBAR in AMBER

Both TI and MBAR are supported in the GPU accelerated AFE calculations.[13,24] TI tends

to be faster than MBAR for a variety of reasons, most notably the extra energies needed for

MBAR compared to TI. Our estimates indicate a 10-20% performance difference between TI and

MBAR on a GPU (unpublished work). This does not account for analysis time, which in general

is slower for MBAR than TI, although the absolute difference likely pales in comparison to the

cost of running the simulations. It is much easier to add new lambda windows to a TI simulation

than BAR/MBAR, because BAR/MBAR requires rerunning other simulations to get the adjusted

value accounting for all states.[25] Additionally, TI, relying on integration of the derivative of the

potential energy with respect to the progress variable lambda (often referred to as dV/dL) is

subject to integration error if an integrator is not sufficiently sensitive to the curvature of

dV/dL.[25–27]. However, using Simpson's rule or cubic spline seems to significantly reduce the

integration error.[25] MBAR (or BAR if MBAR is not supported in the software suite of choice)

offers several analysis advantages. It is possible to calculate an overlap matrix, a direct

representation of overlap between adjacent states. This is extremely useful for predicting when

additional lambda windows need to be added for convergence.[28] MBAR (not BAR) is also

proven to be the minimal variance estimator for a set of states,[20] and results have lower bias and smaller variance than analogous TI calculations when the phase-space overlap between end states of a transformation is small.[29] In summary, if performance over a large set of simulations is key, then TI is often a good choice, particularly for cases such as charge change without the use of softcore potentials where dV/dL is expected to be linear. If performance is not as important, MBAR is a safer tool and often can be run in conjunction with TI, which allows for better error checking as the results should be similar between the two analysis methods.

Analysis tools

One particularly useful tool for analysis of AFE simulations is the alchemical analysis library pymbar[30] and associated toolkit alchemlyb and predecessor alchemical-analysis.py.[31] These tools allow for analysis of virtually every type of simulation data (dV/dL for TI, multiple state information for MBAR, etc.), and work for most major simulation engines. Of note is the ability to generate the above-mentioned overlap matrix between all lambda windows for MBAR simulations, and ability to test for sufficiently uncorrelated samples for any type of simulation.

Multi lambda window approach

It is well established that subdividing the change associated with each leg of a thermodynamic cycle into many smaller changes is necessary to enhance overlap between the end states.[23,32] This is generally accomplished by coupling the thermodynamic leg to a variable lambda, which varies from 0 to 1. Each smaller individual simulation then has a unique value of lambda, and is referred to as a lambda window. However, it is also often necessary to equilibrate

4

each individual lambda window prior to production to avoid kinetically trapped states which may result from conformational changes between different values of lambda.[33]

In AMBER, the dynamic lambda functionality allows for a single equilibration simulation to be run with a steadily increasing value of lambda and to generate initial conformations at each value of lambda. While this is likely to be inferior as far as sampling is concerned to running minimization, heating and pressure equilibration separately for each lambda window, the increased computational cost of running extra minimization, heating and pressure simulations for potentially dozens of lambda windows means dynamic lambda is often worthwhile from a cost-benefit perspective.

Softcore potentials

One of the earliest practical problems encountered in AFE calculations was the frequent instability of simulations at endpoints, also referred to as the endpoint problem or the endpoint catastrophe. One way of surmounting the nonbonded instabilities is by using "softcore" potentials for atoms that are appearing or disappearing. The use of a "softcore" potential for the annihilation/decoupling or creation/recouping of atoms has become standard practice, and is present in sander, PMEMD and PMEMD.cuda. The form present in all three AMBER executables for Lennard-Jones parameters is

$$V_{dissappearing} = 4\varepsilon(1-\lambda)\left[\frac{1}{\left[\alpha\lambda+\left(\frac{r_{ij}}{\sigma}\right)^6\right]^2} - \frac{1}{\alpha\lambda+\left(\frac{r_{ij}}{\sigma}\right)^6}\right] \quad (1)$$

And

$$V_{appearing} = 4\varepsilon\lambda\left[\frac{1}{\left[\alpha(1-\lambda)+\left(\frac{r_{ij}}{\sigma}\right)^6\right]^2} - \frac{1}{\alpha(1-\lambda)+\left(\frac{r_{ij}}{\sigma}\right)^6}\right] \qquad (2)$$

As originally presented by Shirts and Pande.[34] Where λ is the value of the progress variable lambda, ε is the depth of the potential well, r_{ij} is the distance between the two atoms i and j, σ is the finite distance at which the inter particle potential is zero, and α is a scaling parameter. α is typically set to 0.5.

One step vs three step protocol

There are analogous equations to (1) and (2) for electrostatic terms[35] allowing for a "one-step" transformation rather than separate charge and LJ changes:

$$V_{dissappearing} = (1-\lambda)\frac{q_i q_j}{4\pi\varepsilon_0\sqrt{\beta\lambda+r_{ij}^2}} \qquad (3)$$

And

$$V_{appearing} = \lambda\frac{q_i q_j}{4\pi\varepsilon_0\sqrt{\beta(1-\lambda)+r_{ij}^2}} \qquad (4)$$

Where q_i and q_j are the charges on atoms i and j respectively, ε_0 is the permittivity of free space, r_{ij} is the distance between atoms i and j, λ is the value of the progress variable lambda, and β is a scaling parameter. β is typically set to 12Å^2. In this one-step protocol, electrostatic nonbonded terms and Lennard-Jones terms are decoupled simultaneously. While this does reduce by potentially two thirds the number of simulations needed, in practice this is not commonly done. Instead, it is common to first turn off the charges on atoms unique to region 1, then change LJ parameters with no charge, then re-add the charges to these new atoms. This is common in practice because simulations involving simultaneous charge and LJ parameter changes are often unstable. This is due to large forces which result from incidental clashes as the repulsive terms

become very small near the endpoints while the attractive coulomb forces still present draw the atoms closer to one another. It is possible to have atoms wind up almost on top of one another, which in turn results in very large forces which crash the simulation.[36,37] Additionally, charge only transformations can often be performed linearly without the use of softcore, and tend to converge comparatively quickly. However, there are some cases in which the one step protocol will be more efficient.[38] A more thorough study utilizing GPU AFE would be an ideal next step to identify the short comings of the one-step transformations, perhaps with the goal of examining dynamics to see when and why the one step transformation fails, and whether a specific type of transformation is responsible for most of the simulation crashes. Note, the use of softcore potentials and designation of a part of the system as softcore will also have implications in the treatment of bonded terms (see below).

Topology

There are a variety of issues relating to the treatment of bonded terms including topology and endpoint treatment of bonded terms. Topology refers to how the endpoints of a thermodynamic cycle's legs are mapped to one another. There is no consensus as to which topology is better, as long as they are treated correctly.[39–41] See Shirts and Mobley[41] for a more complete discussion. AMBER supports both single and dual topologies in sander, PMEMD and PMEMD.cuda. However, if one chooses to examine an individual leg of a thermodynamic cycle, the choice of topology becomes relevant because there are terms which normally would cancel which must be accounted for. See the two papers by Boresch and Karplus for a thorough discussion.[39,40]

Bonded terms

The treatment of bonded terms can have major implications for the stability of a simulation, particularly at endpoints (lambda=0 and lambda=1). The two choices of bonded terms are ideal gas atom end state and ideal gas molecule end state. The difference is if bonded terms unique to one endpoint are scaled by lambda or not. In the ideal gas atom end state, bonded terms unique to one endpoint are scaled. This results in the breaking of these bonds. In the ideal gas molecule end state, bonded terms unique to one endpoint are NOT scaled. These bonds are then not broken. The free energy change associated with these bonded terms then cancels from the complete thermodynamic cycle. Analysis by Boresch and Karplus[39,40] demonstrated that the endpoint catastrophe could be avoided by utilizing an ideal gas molecule end state rather than an ideal gas atom end state, allowing for direct sampling rather than extrapolation of the endpoint lambda windows (lambda=0 and lambda=1). This is implemented in the PMEMD and PMEMD.cuda modules of AMBER, but not in the sander module.

Charge perturbations

As computational power increases, a natural direction for the application of AFE is towards more complex systems. One such complexity that may increasingly be encountered is the perturbation of charged ligands. The difficulty in charge perturbations arises from the use of periodic boundary conditions[42] to describe certain long range electrostatic effects. Periodic boundary conditions are subject to potential finite size effects as the size of the individual periodic unit cell shrinks relative to the size of the entire system of interest. The magnitude of these artifacts was recently investigated by Lin et al, who concluded that it is necessary to use the same size periodic box for each simulation.[43] The source of these artifacts is completely

explained by Hünenberger and McCammon.[44] An algorithm for the analytical correction of finite size terms has been developed by Rocklin et al., and should be applied whenever the net charge is changed between end states.[45]

Automation of calculations

With the increased computational power afforded by GPUs, as well as the various analysis developments of late, the bottleneck of AFE calculations has shifted to setup. AFE calculations often require complicated topologies and input files, as well as on the order of 100 simulations to be run for each thermodynamic cycle leg between equilibration, numerous lambda windows, and multiple replicates. Additionally, virtual screening efforts ideally could include dozens of potential molecules, each with a minimum of two topologies but often six if using the three-step approach described above. Therefore, automation whenever possible is highly recommended. In AMBER, there are currently two highly used options: Free energy workflow (FEW)[46,47] and Free energy setup (FESetup).[48] Both can be used to generate input and parameter/topology files for a large set of ligands with the ability for the user to override default parameters, as well as run minimization, equilibration and/or production directly. However, there are advantages and disadvantages to both. FEW allows for automated setup of membrane systems.[47] However, the ligand topologies are restricted to single step calculations rather than separate charge and LJ parameter changes. Also, FEW is currently restricted to sander input format, although that will be changing shortly (personal communication). FESetup does allow for some three step protocols, but it forces the starting configurations for systems generated in this way to be the same for all three steps, which in some cases is incorrect.

9

There are a variety of useful reviews written in the last 10 years that go beyond the scope of this work but merit attention. These include reviews of the statistical mechanics underlying the analysis methods,[49–51] the techniques available for calculating changes in free energy,[52] and the current successes and limitations of alchemical free energy calculations[53–55].

Free energy of binding for ligand protein systems

A natural application of AFE is the calculation of ligand protein binding affinities. There are two types of binding free energy calculations in common use: relative binding free energy (RBFE) and absolute binding free energy (ABFE). An example thermodynamic cycle for protein-ligand binding, as first applied by Wong and McCammon,[56] is shown in figure 1 for both RBFE and ABFE.
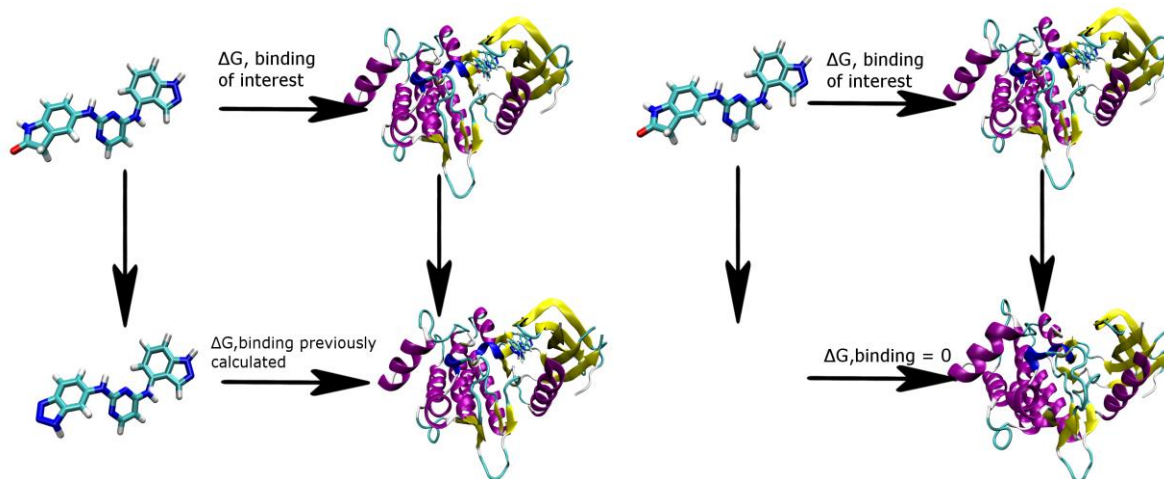
Figure 1.1: Example relative binding free energy (left) and absolute binding free energy (right) of ligand GTC000107A to spleen tyrosine kinase (Syk).

Ligand GTC000107A was taken from the GSK Syk database available through the D3R project.[57] The ABFE calculation is significantly more difficult due to the added phase space needed to be explored by the ligand as it is completely decoupled from its environment. In the RBFE case, the binding affinity of GTC000112A, a similar ligand, is already known and can be used to simplify the calculation.

Relative binding free energy

RBFE calculations take advantage of previously calculated or experimentally measured binding free energies to reduce the size of the alchemical change that must be performed (figure 1). They have been an ideal candidate for computational drug design due to the ability to simultaneously set up and run an entire set of similar ligands with just one initial experimental value.[58] There have been a variety of reviews published in the last five years on the subject of relative binding free energy calculations[55,59–66], many of them published in the last year,[55,59,60,62–64] indicative of the renewed interest in these calculations. These reviews have focused on topics such as benchmarks[55,61,62,65], practical considerations for computational projects,[59] methodological developments,[60,64,66] and the role of experimental collaboration.[63] RBFE with AFE in AMBER has been used for a variety of purposes, including protein-ligand binding,[61,65,67–71] protein mutation energetic calculations,[72–75] benchmarking forcefields and sampling protocols,[76] and method development.[77,78]

Absolute binding free energy calculations

There are several notable differences in absolute free energy of binding calculations from RBFE calculations. The most significant is that the ligand needs to explore potentially all of phase space once it is decoupled from the protein.[79] The most significant developments recently in ABFE have focused on the use of positional and orientational restraints to reduce this sampling requirement.[80,81] Despite impressive theoretical developments in ABFE restraints sampling as well as general improvements to alchemical free energy calculations, the usage of ABFE in practice remains sparse.[82] There are some examples of applications, primarily to ligand binding free energy.[76,83–85]

We have reviewed some of the best practices, new developments, and recent applications of AFE based binding free energy calculations in the AMBER molecular dynamics suite. Following this review, we will describe the newly developed GPU accelerated AFE in chapter 2 of this document.

Free energies of solvation

Solvation free energy calculations have been used to calculate partition coefficients,[86] develop solvation prediction models for use in drug design,[87] and calculate ionic hydration energies to help settle long standing experimental debates,[88] among other uses.[43,89–91] One particularly important use of solvation free energies is in the parametrization and testing of force fields.[92–98] To this end, the FreeSolv database was designed and has been updated several times, most recently in 2017.[93] The goal of FreeSolv is for it to be an ongoing effort to characterize rigorously the success or lack thereof for existing force fields and best practices in AFE. The calculations in the recent FreeSolv update were run using the Gromacs molecular dynamics suite, version 4.6.7.[5–8] The ability to perform gas phase alchemical free energies has been added to AMBER's PMEMD module (see Supporting Information, below). This, combined with our recently developed GPU AFE code,[13] allows for the rapid calculation of solvation free energies in the AMBER molecular dynamics suite.

pH dependent free energies

Theoretical developments

constant pH molecular dynamics in AMBER

An overview of the theory underlying constant pH MD can be found in Chapter 3 of this thesis, where CpHMD was applied to beta secretase-1 to examine the pH dependent dynamics. This section will focus solely on pH dependent free energy calculations.

pH dependent free energies

Kim and McCammon recently reviewed progress in the calculation of pH dependent binding free energies.[99] For the sake of brevity and not repeating what has already been done, this section will only point out the most relevant papers related to the theory and practical application of pH dependent free energy calculations. pH dependent interactions have been studied for protein-protein interactions,[100,101] RNA,[102] DNA,[103] and protein ligand binding.[104–106] Following on the various developments in constant pH MD (CpHMD),[107–114] Kim et al.[115] and Lee et al.[116] independently applied the work of Wyman[117,118] and Tanford[119] to the calculation of the pH dependent component of binding free energy. Additionally, the work by Kim et al. included an equation for the case in which residues are not necessarily titrating independently of one another. However, these calculations can require potentially hundreds of nanoseconds of simulation time to converge. This has traditionally made these calculations in explicit solvent very difficult. Recently, the Walker lab[14] developed a GPU enabled constant pH molecular dynamics algorithm. This code has made the calculation of pH dependent free energies realistic for small molecule-protein systems.

Applications of pH dependent binding free energy

Due to the relatively novel nature of CpHMD as a simulation technique, as well as the computational cost associated with pH dependent free energies, there have not been many applications to pH dependent binding affinities yet.[104–106,115] There have been a number of attempts to study the effect of pH on binding affinity by running cMD with fixed protonation states in the predicted states for several different pHs,[120–128] likely due to the aforementioned computational cost of running CpHMD. However, this strategy may fail when the pK_a of a residue is significantly different from the standard value, as empirical pK_a prediction tools, such as propka, are still an area of active development.[129,130] For example, when residues are sufficiently buried, or else in a coupled titration environment with another residue, or are directly involved in interactions with a binding ligand. It is expected that as GPUs continue to improve, and these methods begin to circulate and find acceptance among the greater community, that applications of pH dependent binding affinity calculations will increase, particularly in a pharmaceutical setting.

Accelerated MD and possible applications to drug design

While there has been great progress in the development of binding free energy calculations, one case in which they sometimes struggle is when there is a major protein conformational change which occurs beyond the timescale of the AFE calculation. One method to address this shortcoming is accelerated molecular dynamics (aMD).[131] In chapter 4, I discuss the application of aMD to a mutant form of T4 lysozyme to elucidate the mechanism by which benzene exits the interior cavity.

14

Conclusions and future directions with regards to AMBER's current free energy capabilities

The AMBER molecular dynamics suite can now be used to calculate a wide variety of free energies beyond relative free energy of binding. The advent of GPU accelerated AFE has pushed the boundaries of feasibility considerably. Constant pH molecular dynamics in AMBER with GPU acceleration represents a major step forward particularly towards the goal of realistic theoretically rigorous modeling of protein scale systems. Additionally, AMBER's ability to calculate solvation free energy appears equal to Gromacs in precision and accuracy. Together, these developments mark a major step forward in not just AMBER's development, but for the entire free energy community. In this work I also discuss the use of CpHMD to probe the water dynamics of BACE-1, and offered a potential explanation for the pH dependence of BACE-1 catalytic activity. Finally, the utility of aMD in sampling rare events is demonstrated. This tool holds potential value in computing binding affinities for protein systems with suspected large conformational changes. Future directions of the AMBER molecular dynamics suite free energy tools include PI-TI[132], development of a user interface, graphical and command line, greater incorporation of pH dependent molecular dynamics in free energy calculations, including adding support for continuous state constant pH MD.[109,133]

Chapter 2 : Fast and Flexible GPU Accelerated Binding Free Energy Calculations within the

AMBER Molecular Dynamics Package

Abstract

Alchemical free energy calculations (AFE) based on molecular dynamics (MD) simulations are key tools in both improving our understanding of a wide variety of biological processes and accelerating the design and optimization of therapeutics for numerous diseases. Computing power and theory have, however, long been insufficient to enable AFE calculations to be routinely applied in early stage drug discovery. One of the major difficulties in performing AFE calculations is the length of time required for calculations to converge to an ensemble average. CPU implementations of MD based free energy algorithms can effectively only reach tens of nanoseconds per day for systems on the order of 50,000 atoms, even running on massively parallel supercomputers. Therefore, converged free energy calculations on large numbers of potential lead compounds are often untenable, preventing researchers from gaining crucial insight into molecular recognition, potential druggability, and other crucial areas of interest. Graphics Processing Units (GPUs) can help address this. Presented here is a seamless GPU implementation, within the PMEMD module of the AMBER molecular dynamics package, of thermodynamic integration (TI) capable of reaching speeds of >140 ns/day for a 44,907-atom system, with accuracy equivalent to the existing CPU implementation in AMBER. The implementation described here is currently part of the AMBER 18 beta code and will be an integral part of the upcoming version 18 release of AMBER.

Introduction

Computational chemists have for some time been seeking to correctly and universally predict the answer to the question "How well will ligand X bind to protein Y?" Couched in this is the implicit assumption that our predictive tool will be able to answer this question in a reasonable amount of time. For years, computational chemists have been forced to strike a difficult balance between speed and accuracy, often being forced to rely on faster, less accurate methods because (TI) and related Free Energy Perturbation (FEP) methods were simply too computationally intensive. However, the ceiling on the applicability of docking and MM/PBSA for use in making quantitative predictions of ligand binding, possibly the most relevant aspect of free energy in the pharmaceutical industry, has quickly become apparent.[82] This has led to a renewed interest in improving slower but potentially more accurate methods such as TI. One of the primary difficulties which TI suffers from is a severe sampling limitation.[64] This limitation is the result of several factors including the difficulty in sampling the relevant states,[23] potentially dozens of unphysical intermediate states that must be simulated even for systems as small as host-guest systems,[135] and the complexity and cost of the underlying molecular dynamics based algorithm. Thus, either hundreds of CPU cores and/or weeks of simulation time are required to obtain results, within acceptable error limits, for a single ligand-protein system in a drug design setting. This is untenable in an industrial setting where high-throughput is integral to success. The purpose of this work is to help address the throughput and finite sampling problem in TI using cost effective consumer hardware.

While there have been many different approaches to addressing the issue of sufficient sampling in TI calculations[25,27,34] the most straightforward manner of improving on the sampling is to write faster code. This paper details a substantially faster implementation of alchemical

17

transformations using cost effective, NVIDIA GeForce graphics processors (GPUs). This code is fully implemented within the AMBER molecular dynamics package,[136–138] and takes direct advantage of the existing, highly efficient, GPU support.[1–3] The implementation described here will be released as an integral part of the GPU accelerated PMEMD program in the upcoming AMBER 18 software release scheduled for Q2 2018. It is currently part of the AMBER 18 Beta code and a patch against AMBER 17 is available by request from the corresponding author.

Theory and Methods

Model Calculations

The goal of this paper is to demonstrate the efficacy and numerical precision of our new GPU TI implementation in AMBER. We have chosen two types of calculations for this purpose (Figure 2.1). The first, a free energy of solvation, the second, a binding affinity calculation.

A.

B.
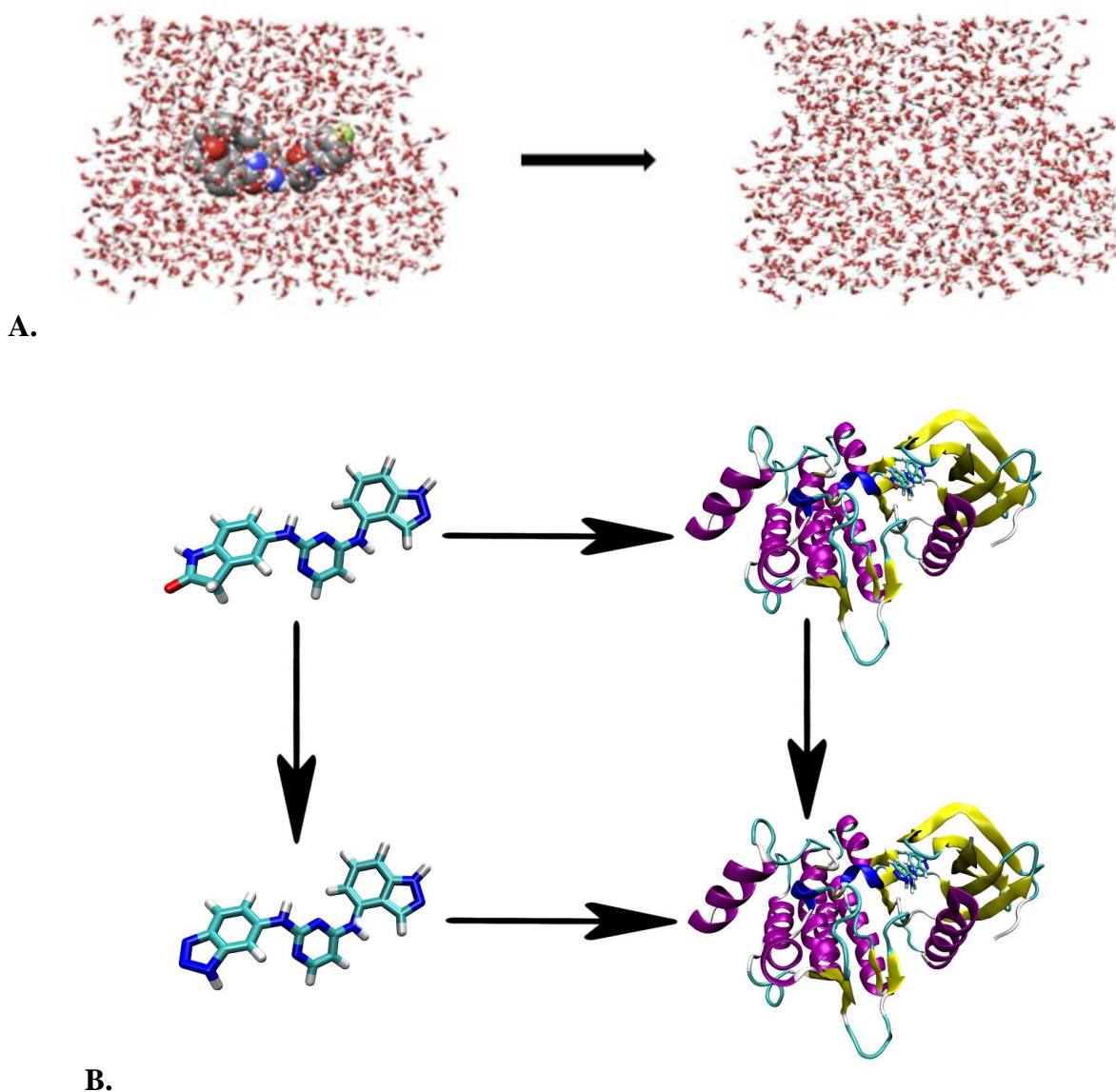
Figure 2.1: Perturbation cycle for sample calculations performed to test the accuracy and performance of our new GPU implementation of TI.

Each molecule represents a thermodynamic endpoint of a calculation. **Top**: Solvation free energy of N1-((2S,3S,5R)-3-AMINO-6-(4-FLUOROPHENYLAMINO)- 5-METHYL-6-OXO-1-PHENYLHEXAN-2-YL)-N3,N3- DIPROPYLISOPHTHALAMIDE, a promising BACE-1 inhibitor.[105]   **Bottom**: Relative free energy of binding of ligands GTC000107A and GTC000112A to spleen tyrosine kinase (Syk) from the GSK Syk database available through the D3R project.[57] The transformation includes 3 deletions and two atom changes from GTC000107A to GTC000112A.

Solvation free energy:

For our present purpose of demonstrating the precision of our GPU TI code compared to the existing CPU based TI implementation it was important to test calculations that were difficult to converge. In the original implementation of TI in pmemd,[139] the solvation free energy calculation was the most difficult to converge, despite free energies of solvation (ΔG_{solv}) having been one of the first types of free energy calculations attempted.[140] In that instance, ΔG_{solv} was defined as decoupling the ligand from the water over the course of one simulation (Figure 2.1A). We chose a potential inhibitor of beta secretase-1 (BACE-1), N1-((2S,3S,5R)-3-AMINO-6-(4-FLUOROPHENYLAMINO)- 5-METHYL-6-OXO-1-PHENYLHEXAN-2-YL)-N3,N3-DIPROPYLISOPHTHALAMIDE as our test system here.[141]

Relative free energy of binding

Relative free energy of binding (RFEB) calculations are the most widely used application of thermodynamic integration given their utility in early stage (lead discovery and lead optimization) small molecule drug discovery.[16,49] For this very reason we have chosen to demonstrate an RFEB calculation in this paper. Our example thermodynamic cycle for RFEB is shown in figure 2.1B for the spleen tyrosine kinase (Syk) system. Syk was chosen because it has a stable, well-defined binding pocket with a crystal structure. Ligands GTC000107A and GTC000112A were chosen from the D3R database because they were shown to have similar binding poses[57] These similar binding poses meant that convergence issues were less likely to dominate the results allowing us to better test the GPU implementation and to compare the differences in the precision model between the CPU and GPU implementations. As is typical in

20

RFEB calculations, alchemical transformations from GTC000107A to GTC000112A were performed for the ligands bound to the protein and the ligands free in solution.

Docking of ligands to Syk

Ligands were docked to crystal structure 1XBA[21] using AutoDock Vina.[22] Gaps in the crystal structure (residues 360-362, 393-394 and 405-406) were filled using Modeller 9v2.[23] Side chains for residues 402, 420, and 448 were treated as flexible, while the rest of the protein was treated as rigid. Ligands were fully flexible.

Simulation details

All the simulations follow a similar protocol; any details that are specific to a particular system will be described below

System preparation and equilibration

The leap module in AMBER 16 was used to parametrize all systems. Protein models used the Amber ff14SB forcefield,[142] with theTIP3P[134] model for water. Neutralizing counterions, sodium or chloride, were added to each system as needed using TIP3P ions with parameters from Joung and Cheatham.[143,144] Ligands were parametrized using the second generation generalized Amber forcefield (GAFF2)[95] for the bonded and van der Waals parameters. Partial charges for ligands were obtained using RESP[145] fitting for the electrostatic potentials, calculated using Gaussian[146] at the Hartree − Fock/6-31G* level of theory. A cubic periodic box was used with a minimum distance of 15 Å between any box edge and any solute atom. All systems were minimized for 1000 cycles of steepest descent followed by 1000 cycles

of conjugate gradient. Solute atoms were restrained with a restraint weight of 10 kcal/(mol*Å)^2.

Minimization was followed, for all values of λ, by 100ps of heating at constant volume and then

1 ns of equilibration at constant pressure. Temperature was regulated via a Langevin thermostat

set to a target temperature of 300 K and a collision frequency of 5.0 ps^{-1}. Pressure was regulated

using a Monte Carlo barostat with a target pressure of 1.0 atm and pressure relaxation time of 2.0

ps.

Production

All production simulations were run in the NPT ensemble with a Langevin thermostat set

to 300 K with collision frequency of 5.0 ps^{-1}, and a Monte Carlo barostat at a target pressure of

1.0 atm and pressure relaxation time of 2.0 ps. The direct space cutoff was set to 10 Å for both

van der Waals and electrostatics. Long range electrostatics we handled via the Particle Mesh

Ewald (PME) method[42] with a FFT grid spacing of ~ 1 point per angstrom. Both the solvation

free energy calculation and the RFEB calculation were run with 11 equally spaced lambda

windows ranging from 0 to 1. The default values for scalpha (0.5) and scbeta (12.0) were used.

Energies were printed every 0.5 ps. CPU simulations were run using pmemd.MPI from AMBER

16 with 12 Xeon cores with a single socket running at 2.50 GHz on an Intel Haswell standard

compute node on XSEDE Comet.[147] Pmemd.MPI was compiled with MVAPICH2 2.1. GPU

simulations were run on NVIDIA GeForce Titan-X GPUs with Pascal architecture using the

current AMBER 18 development tree with our GPU TI support incorporated. The GPU code

was compiled for the default SPFP precision model using CUDA 8.0 and NVIDIA driver version

367.57. Both pmemd.MPI and PMEMD.cuda were compiled using the gnu compiler in gnutools

2.69. Simulations were run with a 1 fs time step (except for timing comparisons using H-mass repartitioning). The solvation free energy system was simulated for 15 ns. The RFEB complex and solvated systems were simulated for 10 ns. The first 5 ns of each simulation was discarded for equilibration purposes. The VDW change of the RFEB complex system required an additional 10 ns of simulation to converge. Each simulation was replicated three times with unique random seeds in each case.

Single step alchemical change versus separate charge/VDW changes

In transformations involving charge changes, it is possible in AMBER to use a softcore coulomb potential alongside the softcore Lennard-jones potential. To test both approaches for reproducibility between the two codes, the solvation free energy calculation was run with softcore electrostatics, while the RFEB was run in separate steps, turning charges off linearly before performing VDW changes.

H-Mass Repartitioning

Hydrogen masses bound to heavy atoms were repartitioned to 3.024 Daltons using ParmEd to allow for a 4-fs time step.[148] Some short simulations were run with identical input parameters, except the additional use of SHAKE to restrain the bonds between hydrogen and heavy atoms. To obtain the same number of data points, print and write frequencies were quadrupled for these simulations.

Performance comparison between CPU and GPU

All CPU simulations were run using the pmemd.MPI module om 12 Intel Xeon E5-2680v3 cores at 2.50 GHz on a single socket of an Intel Haswell standard compute node on XSEDE Comet. All GPU simulations were run on NVIDIA GeForce Titan-X (Pascal) GPUs with driver 367.57 and CUDA 8.0.

Analysis

All integrations were carried out using a cubic spline over 11 lambda windows: 0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0. The Alchemical analysis python package[31] was used to calculate our free energy estimates and associated errors. Autocorrelation times were estimated using pymbar[30]. Uncertainties for average values are the standard deviation of the replicate calculations.

Results and Discussion

Numerical comparison

Previous work by several groups has shown the CPU implementation of TI to be already capable of predicting experimental free energies.[67,72,73,149] Thus for the purposes of validating our GPU implementation we consider values that agree within statistical error to the CPU implementation to indicate success.

The results of the solvation free energy and RFEB calculations are shown below (Table 1). The average results of the CPU and GPU simulations agree to within the standard deviation of three replicates.

Table 2.1: Numerical comparison of free energy of solvation and RFEB calculations on CPU and GPU.

All values are the average ± the standard deviation of three replicates. RFEB – complex is the sum of the simulations in which the ligands are bound to the protein. RFEB – solution is the sum of the simulations in which the ligands are free in solution. Values for individual charge transformation and Lennard-Jones transformation simulation can be found in SI table 1.

Calculation	code	Relative Gibbs free energy change (kcal/mol)
Solvation free energy	CPU	53.0 ± 2.4
	GPU	53.8 ± 2.4
RFEB - complex	CPU	27.50 ± 0.46
	GPU	27.36 ± 0.55
RFEB - solution	CPU	28.95 ± 0.12
	GPU	28.86 ± 0.05
RFEB ($\Delta\Delta G$)	CPU	-1.45 ± 0.48
	GPU	-1.50 ± 0.55

The results agree between the CPU and GPU codes but, as shown with the performance numbers below, the GPU code obtained the result in approximately 1/30th of the time it took to run the calculations on all CPUs cores within the node. For the RFEB calculation, the charge and VDW changes were carried out separately. These calculations also agreed to within statistical error (table 2.3).

Since the values measured here are ensemble averages using a Langevin thermostat that introduces random friction forces to control the temperature, it is still potentially possible that while the two codes agree within statistical error that we are using an incorrect potential in the

GPU code due to some subtle implementation bug. Therefore, it is important from a code

validation perspective to directly compare the potentials (both energy and gradients) between the

CPU and GPU codes by comparing the energies prior to including the random friction forces.

Table 2 below compares initial energies for the same starting structure between the CPU TI

(DPDP precision model) and GPU TI (SPFP precision model).

Table 2.2: Numerical comparison of energies at step 0 of our RFEB trajectories.
Results are for Lennard-Jones parameter change with softcore VDW potential. Electrostatic non-bonded includes both direct space and PME reciprocal space terms. "Bonded terms" includes bonds, angles, dihedrals, and 1-4 adjustment terms. DVDL, the derivative of potential with respect to lambda, is the sum of the contributions from all potential terms (bonded, VDW, and electrostatic). Forces were also equivalent to the same numerical precision, but are not shown due to the sheer number of atoms for which force is calculated.

Term	CPU (kcal/mol)	GPU (kcal/mol)
VDW non-bonded	16209.8375	16209.8375
Electrostatic non-bonded	-171344.5200	-171344.5191
Bonded terms	19937.6161	19937.6162
DVDL	-35.4512	-35.4518

Energy and DVDL values agree to at least 5 significant figures which agrees with previous

comparisons of conventional MD simulations between the DPDP (CPU) and SPFP (GPU)

precision models.

Performance comparison

The use of GPUs has allowed for a greater than 30X average performance increase over a

single socket 12 Intel Xeon E5-2680v3 core node. (Figure 2.2).
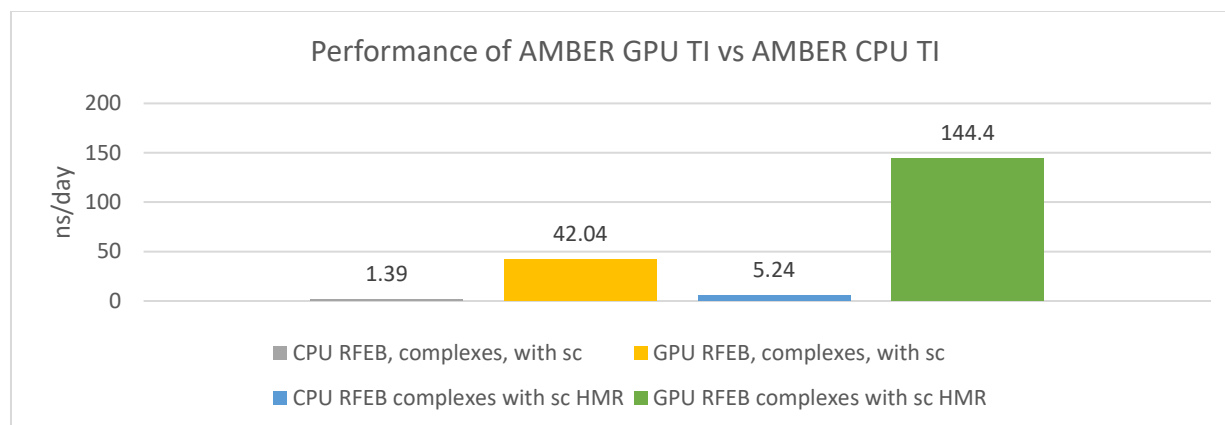
Figure 2.2: comparison of average performance of CPU and GPU TI code for a protein-ligand binding calculation.

CPU code was run on 12 Intel Xeon E5-2680v3 cores, while the GPU code was run on a single NVIDIA GeForce Titan-X [Pascal] GPU. Numbers are from a VDW transformation from the lambda = 0 window. There was a total of 44,907 atoms, 81 of which were defined as TI atoms, with 3 being softcore. Charges on the softcore atoms were turned off. The GPU code performance did not scale as well with HMR compared with the CPU code (3.43X vs 3.77X) because of the higher energy printing frequency (every 125 steps) which more negatively affects GPU performance than CPU performance.

Hydrogen mass repartitioning enabled the use of a 4-fs time step. We noticed no issues in using 4-fs as a time step, in keeping with work done by others.[148,150] Overall, we have achieved a greater than 360X performance enhancement for RFEB protein-ligand complex systems over a single CPU core, or greater than 30X performance improvement over a CPU node. With our code, these calculations can now be carried out on the order of hours on a single cost-efficient GeForce GPU, instead of weeks on a single node or having to utilize large number of CPU nodes with expensive interconnects. For way of reference the node used in this work can be purchased for approximately $2800 with a single GeForce Titan-X [Pascal] GPU. Our recommended configuration would, at the time of writing, be a system containing 2 x E5-2620V4 CPUs, 64GB of memory and 4 x NVIDIA GeForce 1080TI GPUs. This configuration currently costs between $5000 and $6500 and can run 4 lambda windows at once, one on each GPU, obtaining performance equivalent to that shown above for each window simultaneously.

Conclusions

We have demonstrated GPU enabled TI in the AMBER molecular dynamics suite that will form the officially supported GPU TI implementation to be released in the upcoming AMBER v18. Our implementation compared with the most recent version of TI in pmemd[136] is over 360X faster for protein-ligand binding than a single CPU core, 30X faster compared with a single node if using a single GPU per node and 120X faster per node if one considers that 4 GeForce GPUs can be added to a single node in a very cost effective manner. These performance improvements come without sacrificing precision or accuracy. The input format is identical to CPU pmemd, making the transition seamless for existing users. MBAR support has also implemented, and will be discussed in follow up publications. The code described in this manuscript is available by request from the corresponding author as a patch against the release version of AMBER 17 and applicable updates as of Sept 1st, 2017 and forms part of the AMBER 18 release scheduled for April 2018.

Acknowledgements

Interfaces training program for multi scale biology from the National Institutes of Health (NIH).

Work in the JAM group is supported in part by NIH, NBCR, and the NSF supercomputer

centers. RCW's contributions were funded by royalties received from licensing of the AMBER

software by UCSF. No contributions to this work, monetary or otherwise, were made by

NVIDIA Corp.

Supporting Information
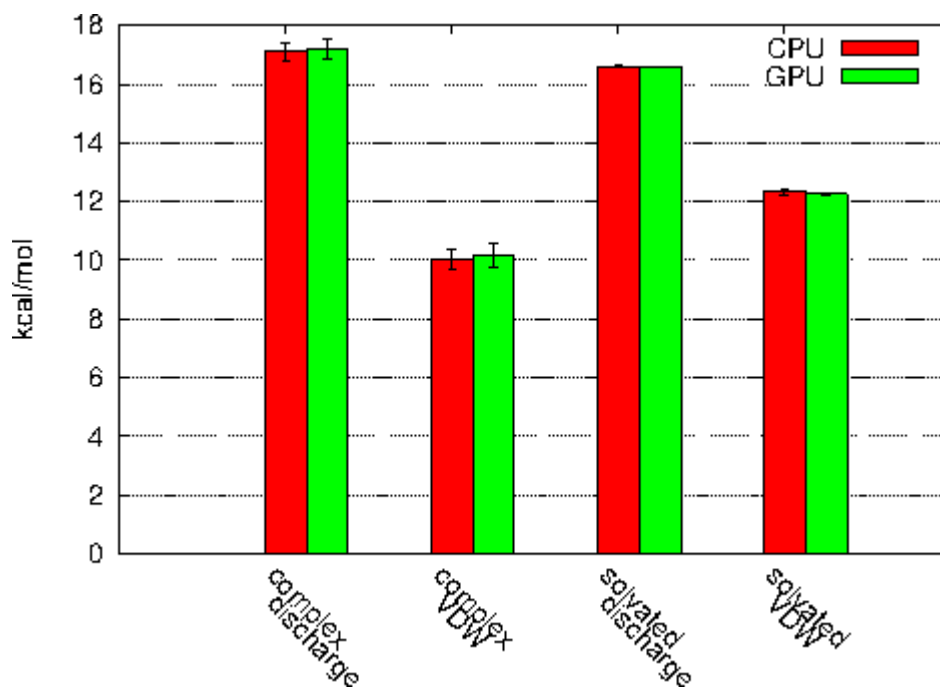
Figure 2.3: breakdown of RFEB calculation into charge change and Lennard Jones change.

All calculated numbers agree within error (Table 2.3).

Table 2.3: numerical values associated with figure 2.3

Calculation		Step	Relative Gibbs free energy change (kcal/mol)
ligand binding – complex	cpu	discharge	17.11 ± 0.29
	gpu	discharge	17.21 ± 0.36
	cpu	vdw	10.38 ± 0.36
	gpu	vdw	10.15 ± 0.42
ligand binding - ligands	cpu	discharge	16.61 ± 0.01
	gpu	discharge	16.59 ± 0.02
	cpu	vdw	12.34 ± 0.11
	gpu	vdw	12.27 ± 0.03

```
| TI region  1


NSTEP =         0   TIME(PS) =       0.000  TEMP(K) =      0.00  PRESS =      0.0
Etot    =   -135197.0654  EKtot    =          0.0000  EPtot      =   -135197.0654
BOND    =      1096.8230  ANGLE    =       2447.0242  DIHED      =       3541.2099
1-4 NB  =      1061.6433  1-4 EEL  =      11790.9158  VDWAALS    =      16209.8375
EELEC   =   -171344.5191  EHBOND   =          0.0000  RESTRAINT  =          0.0000
EKCMT   =         0.0000  VIRIAL   =          0.0000  VOLUME     =     448410.7169
                                                      Density    =          1.0173
DV/DL   =       -35.4518
 ------------------------------------------------------------------------------

 Softcore part of the system:       3 atoms,        TEMP(K)    =        215.86
SC_Etot=        18.2619  SC_EKtot=          1.9303  SC_EPtot   =        16.3315
SC_BOND=         4.5587  SC_ANGLE=          1.7758  SC_DIHED   =         9.9882
SC_14NB=         0.0088  SC_14EEL=          0.0000  SC_VDW     =         0.0000
SC_EEL =         0.0000
SC_RES_DIST=     0.0000  SC_RES_ANG=        0.0000  SC_RES_TORS=         0.0000
SC_EEL_DER=      0.0000  SC_VDW_DER=        0.0000  SC_DERIV   =         0.0000
 ------------------------------------------------------------------------------


| TI region  2


NSTEP =         0   TIME(PS) =       0.000  TEMP(K) =      0.00  PRESS =      0.0
Etot    =   -135197.0654  EKtot    =          0.0000  EPtot      =   -135197.0654
BOND    =      1096.8230  ANGLE    =       2447.0242  DIHED      =       3541.2099
1-4 NB  =      1061.6433  1-4 EEL  =      11790.9158  VDWAALS    =      16209.8375
EELEC   =   -171344.5191  EHBOND   =          0.0000  RESTRAINT  =          0.0000
EKCMT   =         0.0000  VIRIAL   =          0.0000  VOLUME     =     448410.7169
                                                      Density    =          1.0172
```
VDW GPU

Figure 2.4: output initial energies for both Discharge and VDW changes in RFEB calculation.

```
| TI region  1


NSTEP =         0   TIME(PS) =      0.000  TEMP(K) =     0.00  PRESS =     0.0
Etot   =  -135197.0665  EKtot   =        0.0000  EPtot      =  -135197.0665
BOND   =     1096.8230  ANGLE   =     2447.0242  DIHED      =     3541.2099
1-4 NB =     1061.6433  1-4 EEL =    11790.9157  VDWAALS    =    16209.8375
EELEC  =  -171344.5200  EHBOND  =        0.0000  RESTRAINT  =        0.0000
EKCMT  =        0.0000  VIRIAL  =        0.0000  VOLUME     =   448410.7169
                                                 Density    =        1.0173
DV/DL  =       -35.4512
Ewald error estimate:   0.2740E-04
 ------------------------------------------------------------------------------

 Softcore part of the system:      3 atoms,       TEMP(K)    =       215.86
SC_Etot=        18.2607  SC_EKtot=        1.9303  SC_EPtot   =       16.3304
SC_BOND=         4.5587  SC_ANGLE=        1.7758  SC_DIHED   =        9.9882
SC_14NB=         0.0088  SC_14EEL=        0.0000  SC_VDW     =       -0.0012
SC_EEL =         0.0000
SC_RES_DIST=     0.0000  SC_RES_ANG=      0.0000  SC_RES_TORS=        0.0000
SC_EEL_DER=      0.0000  SC_VDW_DER=      0.0000  SC_DERIV   =        0.0000
 ------------------------------------------------------------------------------


| TI region  2


NSTEP =         0   TIME(PS) =      0.000  TEMP(K) =     0.00  PRESS =     0.0
Etot   =  -135197.0665  EKtot   =        0.0000  EPtot      =  -135197.0665
BOND   =     1096.8230  ANGLE   =     2447.0242  DIHED      =     3541.2099
1-4 NB =     1061.6433  1-4 EEL =    11790.9157  VDWAALS    =    16209.8375
EELEC  =  -171344.5200  EHBOND  =        0.0000  RESTRAINT  =        0.0000
EKCMT  =        0.0000  VIRIAL  =        0.0000  VOLUME     =   448410.7169
                                                 Density    =        1.0172
DV/DL  =       -35.4512
Ewald error estimate:   0.2740E-04
 ------------------------------------------------------------------------------
```
VDW CPU

Figure 2.4 continued: output initial energies for both Discharge and VDW changes in RFEB calculation.

```
| TI region  1


 NSTEP =        0   TIME(PS) =      0.000  TEMP(K) =      0.00  PRESS =      0.0
 Etot    =   -135160.4071  EKtot    =        0.0000  EPtot    =   -135160.4071
 BOND    =      1037.8846  ANGLE    =     2479.9287  DIHED    =      3510.4997
 1-4 NB  =      1042.1221  1-4 EEL  =    11861.8884  VDWAALS  =     15884.1778
 EELEC   =   -170976.9085  EHBOND   =        0.0000  RESTRAINT =        0.0000
 EKCMT   =         0.0000  VIRIAL   =        0.0000  VOLUME   =    448671.1309
                                                     Density  =         1.0167
 DV/DL   =        -0.4194
 Ewald error estimate:   0.1163E-03
 ------------------------------------------------------------------------------


| TI region  2


 NSTEP =        0   TIME(PS) =      0.000  TEMP(K) =      0.00  PRESS =      0.0
 Etot    =   -135160.4071  EKtot    =        0.0000  EPtot    =   -135160.4071
 BOND    =      1037.8846  ANGLE    =     2479.9287  DIHED    =      3510.4997
 1-4 NB  =      1042.1221  1-4 EEL  =    11861.8884  VDWAALS  =     15884.1778
 EELEC   =   -170976.9085  EHBOND   =        0.0000  RESTRAINT =        0.0000
 EKCMT   =         0.0000  VIRIAL   =        0.0000  VOLUME   =    448671.1309
                                                     Density  =         1.0167
 DV/DL   =        -0.4194
 Ewald error estimate:   0.1163E-03
 ------------------------------------------------------------------------------
```
Discharge CPU
```
| TI region  1


 NSTEP =        0   TIME(PS) =      0.000  TEMP(K) =      0.00  PRESS =      0.0
 Etot    =   -135160.3598  EKtot    =        0.0000  EPtot    =   -135160.3598
 BOND    =      1037.8846  ANGLE    =     2479.9287  DIHED    =      3510.4997
 1-4 NB  =      1042.1221  1-4 EEL  =    11861.8886  VDWAALS  =     15884.1778
 EELEC   =   -170976.8614  EHBOND   =        0.0000  RESTRAINT =        0.0000
 EKCMT   =         0.0000  VIRIAL   =        0.0000  VOLUME   =    448671.1309
                                                     Density  =         1.0167
 DV/DL   =        -0.4202
 ------------------------------------------------------------------------------

| TI region  2


 NSTEP =        0   TIME(PS) =      0.000  TEMP(K) =      0.00  PRESS =      0.0
 Etot    =   -135160.3598  EKtot    =        0.0000  EPtot    =   -135160.3598
 BOND    =      1037.8846  ANGLE    =     2479.9287  DIHED    =      3510.4997
 1-4 NB  =      1042.1221  1-4 EEL  =    11861.8886  VDWAALS  =     15884.1778
 EELEC   =   -170976.8614  EHBOND   =        0.0000  RESTRAINT =        0.0000
 EKCMT   =         0.0000  VIRIAL   =        0.0000  VOLUME   =    448671.1309
                                                     Density  =         1.0167
 DV/DL   =        -0.4202
 ------------------------------------------------------------------------------
```
Discharge GPU

Figure 2.4 continued: output initial energies for both Discharge and VDW changes in RFEB calculation.

Chapter 3 : pH dependent conformational dynamics of Beta-secretase 1: a molecular dynamics study

Abstract

Beta-secretase 1 (BACE-1) is an aspartyl protease implicated in the overproduction of β-amyloid fibrils responsible for Alzheimer's disease. The process of β-amyloid genesis is known to be pH dependent, with an activity peak between solution pH of 3.5 and 5.5. We have studied the pH dependent dynamics of BACE-1 to better understand the pH dependent mechanism. We have implemented support for graphics processor unit (GPU) accelerated constant pH molecular dynamics within the AMBER molecular dynamics software package and employed this to determine the relative population of different aspartyl dyad protonation states in the pH range of greatest β-amyloid production, followed by conventional molecular dynamics to explore the differences among the various Aspartyl dyad protonation states. We observed a difference in dynamics between double protonated, mono-protonated, and double deprotonated states over the known pH range of higher activity. These differences include Tyr71-Aspartyl dyad proximity, water accessibility in the active site, and active site water lifetime. This work suggests that Tyr71 may play a role in stabilizing catalytic water in the Aspartyl dyad active site, enabling BACE-1 activity.

Introduction

Alzheimer's Disease (AD) is characterized by a degradation of cognitive function and memory loss.[151] These phenotypes are thought to be caused by the over accumulation of a cytotoxic form of amyloid-beta peptide, Aβ-42.[152,153] Aβ-42 is one of a variety of amyloid beta

peptides that are produced via the amyloidogenic pathway.[154] When APP is first cleaved by β-secretase (BACE-1) rather than α-secretase, Aβ-42 is formed preferentially over other non-oligomerizing amyloid-beta peptides.[155,156] This, coupled with BACE-1 mice knockout studies in which Aβ-42 production was abolished with no other apparent phenotypic changes,[157,158] have led to BACE-1 being an extremely promising protein target in the fight against AD.[159,160]

BACE-1 is a 501 amino acid transmembrane aspartyl protease.[161,162] As a member of the aspartyl protease family, it has the characteristic aspartyl dyad at the center of its active site.[163] BACE-1 activity has a well-known pH dependence, with activity declining sharply below pH 3.5 and above pH 5.5.[164] BACE-1 is thought to act by some variant of a general acid-base mechanism, possibly involving a bridging catalytic water.[165,166] In the case of BACE-1, the aspartyl dyad (Asp dyad) acts as the general acid and general base. Asp 32, with an experimental pK_a of 5.2,[166] would be the general acid, while Asp 228 (experimental pK_a of 3.5)[166] would be the general base. The exact role of water and in turn how solution pH regulates BACE-1 activity are still subjects of study.[105,164,166,167] Toulokhonova *et al.* proposed a mechanism in which a bridging water allows for the formation of a tetrahedral intermediate and with an unknown pH dependent conformational change as the rate limiting step.[166] Shimizu *et al.* proposed that at acidic pH water is prevented from entering the active site and catalyzing peptide cleavage while at basic pH activity is prevented by a conformational change of BACE-1.[164] Both hypotheses involved the motion of a β-hairpin loop consisting of residues Tyr68 through Glu77, more commonly referred to as the "flap" region. Molecular dynamics and crystallography studies have demonstrated the flap region's flexibility, shown below in figure 3.1.[164,168–170]
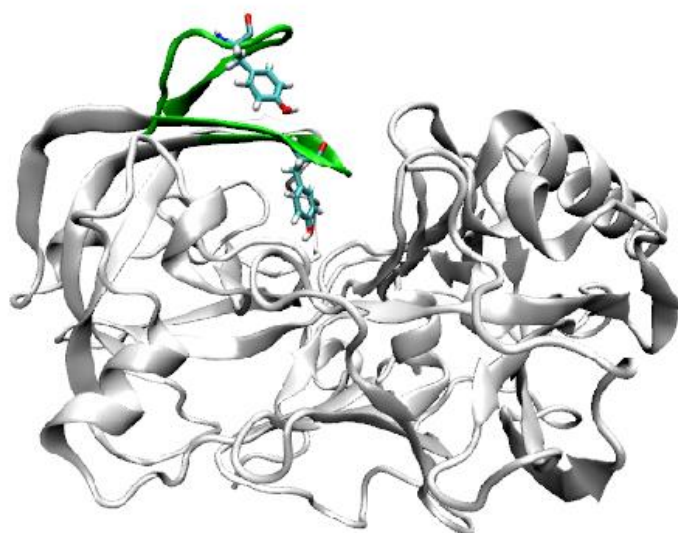
Figure 3.1: example flap motion in BACE-1.

Previous work by Kim *et al.*[105] as well as Ellis and Shen[167] both using constant pH

molecular dynamics has shown that the flap region dynamics are vastly different at different

pHs. Of relevance to the question of mechanism, pH control, and the role of water in BACE-1

catalysis, Ellis and Shen demonstrated that Tyr 71 can hydrogen bond to BACE-1 inhibitors

resulting in openings small enough to potentially occlude water.[167] It is possible that the flap is

capable of occluding water without the presence of inhibitor, in which case water would be a

major limiting factor in catalysis. By calculating how much water is present in the active site at

different pHs, and how long said water remains in the active site, we can determine if water is

limiting to BACE-1 activity. From this information, we may be able to ascertain whether a

conformational change is responsible for BACE-1 regulation, or whether BACE-1 pH

dependence is due entirely to the electrostatic environment of the active site.

To explore water residence times across a range of solution pHs, we have employed two

variations of molecular dynamics simulations (MD): conventional MD (cMD) and constant pH

36

MD (CpH-MD). MD simulations allow the direct observation of protein sized systems as they propagate forward in time. This enables us to observe and account for all the binding poses and conformational changes that can occur. However, cMD requires a constant protonation state that is set prior to the simulation. For BACE-1, which has a known pH dependence, ignoring changes in protonation state would potentially miss very relevant information. CpH-MD can be used to probe the effect of the surrounding environment on the pK_a of a given residue by accounting for the effect of conformational flexibility on pK_a.[108] By accounting explicitly for the possibility of multiple protonation states, we can vastly improve our understanding of pH dependent conformational changes and the associated changes in electrostatic environment. Previous studies have attempted to use constant pH[105,167] or other forms of molecular dynamics[168,171–175] to study BACE-1 dynamics. To our knowledge, this is the first study to use explicit solvent CpH-MD to quantitatively examine the effect of pH on water lifetimes in BACE-1, and to attempt to correlate this with a conformational change in the protein towards the understanding of the source of the pH dependence of BACE-1 activity.

Methods

Calculating pK$_a$s using CpH-REMD

In CpH-MD, protonation states of the residues of interest are allowed to change over the molecular dynamics simulations, sampling from a semi-grand canonical ensemble.[107] In the particular variety of CpH-MD employed in this work, modified from the method of Mongan *et al*.[108] and implemented on CPUs within the AMBER molecular dynamics suite[136–138] by Swails *et al*.[111] and subsequently extended to GPUs[1–3] as part of this work by us, dynamics propagate

from an initial set of protonation states using explicit solvent conventional MD. In this method, dynamics are interrupted, and for each residue being titrated a change of protonation state is attempted sequentially. Protonation state changes are attempted using generalized Born implicit solvent. Acceptance is decided by the following Monte Carlo (MC) criteria (1):[108]

$$\Delta G_{trans} = k_b T \left(pH - pK_{a,ref} \right) \ln(10) + \Delta G_{elec} - \Delta G_{elec,\ ref} \quad (1)$$

where k_b is the Boltzmann constant, T is the temperature in Kelvin, pH is the solvent pH, $pK_{a,ref}$ is an experimentally measured pK_a value for a simpler form of the amino acid, with the sequence acetyl–amino acid–methyl amine, and $\Delta G_{elec,\ ref}$ is the precomputed free energy of changing the protonation state of the reference compound. ΔG_{elec} is the calculated free energy of changing the pronation state at the current simulation condition using generalized Born. Following titration, the water molecules must be allowed to relax before protein dynamics can continue. This is accomplished by holding the solute position constant and running dynamics on the water molecules. Replica exchange was employed along the pH coordinate. After every cycle described above, adjacent replicas attempt to exchange solution pH per the following MC criteria (2):[110]

$$P_{i \to j} = \min \left\{ 1, \exp \left[\ln 10 \left(N_i - N_j \right) \left(pH_i - pH_j \right) \right] \right\} \quad (2)$$

Where N_i and N_j are the number of titratable residues that are currently protonated in replica i and j, respectively, and pH_i and pH_j are solvent pH in replica i and j, respectively. The result of such a simulation is a protonation fraction for each residue at each pH. These values can then be fit to the Hill equation (3) to generate predicted pK_a values:

$$\text{fraction protonated}(pH) = \frac{1}{\left(1 + 10^{(n*(pKa-pH))} \right)} \quad (3)$$

Where n is the cooperativity coefficient, fraction protonated is the percentage of time a residue spent in the protonated state at a given pH, and pH is the solvent pH.

Conventional MD for dynamics

While it is possible to use CpH-MD to study dynamics, as even in explicit solvent it satisfies the weaker detailed balance criteria,[176] the solvent relaxation time requirement makes extracting time dependent information, e.g. water residence or flap motion, from CpH-MD very difficult in practice. Conventional MD (cMD) is useful for generating such time dependent quantitative information, but requires a set protonation state. By using CpH-MD to calculate the relative populations of each set of protonation states at each pH, and then running cMD on each of these sets of protonation states that are present at our pH range of interest, we can calculate residence times and flap motion for the states of interest. To test the differences at, below, and above the pH range of high activity, we first need to determine the different protonation states present within pH 3.5 – 5.5, below 3.5, and above 5.5. Then we can run conventional MD to extract flap-dyad distance and residence time, and correlate this with how frequently each protonation state occurs below, above, and within the pH range of interest.

Choice of titratable residues

Propka[130] was used to determine the predicted pK_a of all potentially titratable residues at pH 3.5 and pH 5.5. From this, all residues with a predicted difference in pK_a between pH 3.5 and pH 5.5 of greater than 0.5 pK_a units were selected to be titratable in our CpHMD simulations.

System preparation and generation of production coordinates

The X-ray crystallographic structures of BACE-1 in complex with the inhibitor N-[(1S,1R)-ben- zyl-3-(cyclopropylamino)-2-hydroxypropyl]-5-[methyl(methylsulfonyl)]amino-

N'-[(1R)- 1-phenylethyl]isophthalamide (PDB ID 2B8L)[177] was used to build the starting structure for all simulations. The apo structure of BACE-1 was generated by removing the bound inhibitor from the refined 2B8L structure. The mutations that were added to the protein for crystallographic purposes were corrected to the original sequence.[177] Residues from Gly158 to Ser169 were not resolved in this structure. This loop was constructed using homology modeling from the Structure Prediction Wizard module of Schrödinger's Prime program.[178–180] First, the FASTA sequence of the protein including the missing loop region was obtained from UniProt.[181] Then, utilizing the homologs found by the BLAST search algorithm[182] a chimera model containing the missing loop region was built. Finally, the homology-modeled loop region was energy-refined for relaxation using the Refine Loops panel of the Prime program.[180]

The leap module in AMBER 16[137] was used to parameterize Apo BACE-1. the Amber ff14SB forcefield[142] was used for protein parameters. TheTIP3P[134] model was used for water. 18 sodium and eight chlorine TIP3P ions with Joung and Cheatham parameters[143,144] were added to generate a neutralized system with a 0.1M ion concentration. A 0.1M concentration was needed to match the ion concentration for which the GB $pK_{a,ref}$ and $\Delta G_{elec,ref}$ were parameterized for. A cubic periodic box was used with a minimum distance of 10 Å between any box edge and any solute atom. Disulfide bonds were added manually in leap between Cys 155 and Cys 359, Cys 217 and Cys 382, and Cys 269 and Cys 319. Residues Asp 32, Asp106, Asp138, Asp223, Asp 228, Glu116, Glu265, Glu339, His45 and Tyr 71 were selected for titration. Initial protonation states were selected by using PROPKA.[130,183] A modified prmtop and cpin file with generalized Born (GB) parameters $GB^{OBC,I}$ from Onufriev, Bashford and Case[184,185] were generated using cpinutils.py in AmberTools16.[136]

Minimization was carried out over four steps. First, 2000 cycles of steepest descent were performed with a restraint weight of 5 kcal/(mol*$Å^2$) on all non-hydrogen atoms. Second, 5000 cycles of steepest descent were performed with a restraint weight of 5 kcal/(mol*$Å^2$) on all non-hydrogen protein atoms. Third, 5000 cycles of steepest descent were performed with a restraint weight of 5 kcal/(mol*$Å^2$) on all atoms except for carbons and nitrogens. Fourth, 25000 cycles of steepest descent were performed with no restraint. Following minimization, the system was heated gradually to 300K over 250 ps using a Langevin thermostat with a collision frequency of 2.0 ps^{-1}. All protein atoms were restrained with a weight of 5 kcal/(mol*$Å^2$). The system was then pressure equilibrated to 1 atm over 2 ns with a Langevin thermostat set to 300 K with collision frequency of 5.0 ps^{-1} and a Berendsen barostat with a pressure relaxation time of 1.0 ps. Finally, the system was run under NVT conditions for 100 ns of equilibration with a Langevin thermostat, collision frequency of 2.0 ps^{-1} and a target temperature of 300 K. Particle Mesh Ewald[42] was used for long range electrostatic forces, with direct force calculation truncated after 8.0Å.

Constant pH REMD

CpH REMD simulations were run using our GPU implementation of CpH-MD in PMEMD.cuda.MPI within AMBER 16.[137] The only modification beyond the code released as part of AMBER 16 was to allow for coupled titrations for residues which were separated by more than 2Å such as in this case where the bridging water caused Asp 32 and Asp 228 to be separated by on average 5Å. 18 total pH replicates were run, spanning from -6 to 12 by single pH units. Each of three replicates was run for 60 ns of dynamics (excluding solvent relaxation). However, pK_a values appear converged after only 20 ns (Figure 3.3). Simulations were run with

NVT at a target temperature of 300K using the Langevin thermostat with a collision frequency of 2.0 ps^{-1}. Dynamics was propagated for 200 fs, at which point titration was attempted on all titratable residues. In the case of a successful titration on any of the residues, 200 fs of solvent relaxation was performed. Following solvent relaxation, replica exchange was attempted. Explicit solvent was used for the dynamics (icnstph=2), while generalized Born implicit solvent was used for the titration attempts, with igb=2 which corresponds to GB^{OBC,I}.[185] A salt concentration of 0.1M was used to match the parametrization of the reference pK$_a$ values. Protonation state population data was recorded after every set of titration attempts. Energies and coordinates were recorded every 10 ps.

Production MD on four protonation states

System preparation for cMD was identical to CpH-REMD, except that instead of generating a modified prmtop and cpin file, the initial prmtop was modified to have the desired protonation states for each of the four combinations of Asp dyad protonation. The four protonation states were Asp 32 protonated, Asp 228 deprotonated; Asp 32 deprotonated, Asp 32 protonated; both Asp 32 and Asp 228 protonated; and both Asp 32 and Asp 228 deprotonated. Prmtop modification was accomplished using ParmEd in AmberTools17[136]. For each protonation state, three replicate cMD simulations were run for 80 ns with identical NVT conditions to CpH-REMD simulations, for a total of 12 cMD simulations. Energies and coordinates were recorded every 10 ps.

Data analysis

Cphstats in amber was used to reconstruct trajectories by pH following replica exchange, as well as calculated protonation fractions of all residues at each pH. This data was fit to the Hill equation (3) using Gnuplot.[186] Cpptraj in AmberTools[136] was used to calculate the dyad-Tyr 71 distance over the course of the cMD trajectories as well as generate water-dyad distances for residence times. The center of mass of the aspartyl dyad titratable protons was used for the dyad location when measuring water-dyad and Tyr71-dyad distances. Trajectories were visualized in VMD version 1.9.3a6.[187]

Calculating hydration numbers and water residence time

Hydration numbers and residence times were calculated using an in-house python script. Water molecules were considered to be within the first hydration shell of the aspartyl dyad if their oxygen atom was within 3.5Å from the center of mass of the aspartyl dyad titratable protons. Only the closest 10 water molecules were considered for computational simplicity. There was no noticeable effect of including the 11th closest water molecule on the hydration number compared to just the 10 closest water molecules (Table 3.4).

Results and Discussion

Calculation of pK_as

Constant pH MD was used to generate ensemble averages of protonation fractions from pH -6 through pH 12 in units of single pH units for 10 residues. Residues titrated included the aspartyl dyad, the Tyr 71, and seven others which were hypothesized to vary significantly over the active pH range. The predicted pKa values are shown in table 1.

Table 3.1: Predicted pka values for titrated residues.
Uncertainties are derived from the fit to the Hill equation.

Residue	Calculated pKa
Asp 32	3.16 ± 0.08
Asp106	3.35 ± 0.01
Asp138	2.53 ± 0.02
Asp223	3.50 ± 0.002
Asp228	-1.17 ± 0.08
Glu116	3.26 ± 0.04
Glu265	5.30 ± 0.004
Glu339	4.04 ± 0.02
His45	7.26 ± 0.01
Tyr 71	11.80 ± 0.001

The pKa of the aspartyl dyads of note. The experimentally calculated values of Asp 32 and Asp 228 are 5.2 and 3.5 respectively.[166] Our macroscopic pK_as were correctly ordered, but differ by 2.1 and 4.6, respectively. This variation is likely due to the bridging water which is removed during the implicit solvent titration attempts. The other key titratable residues had pK_a values near their normal ranges. The pK_a of Tyrosine 71 was a bit elevated from the typical value of 9.1. Because the calculated pK_a's of Asp 32 and Asp 228 appear to be artifactually perturbed by the missing water molecule, we used the experimental values of those resides, together with the calculated values of the other residues, to guide the chosen protonation states for the cMD simulations.

Protonation states relevant to the dyad motion and hydration

To understand the Asp flap dynamics over time at different pHs, we first needed to determine which protonation states were likely to be present at each pH. Of the titratable residues, only the Asp dyad and Tyr 71 were found to influence the dyad motion. However, Tyr 71 had a calculated pK_a of 11.8, far above that of the Asp dyad, and did not change protonation state below the upper bound of pH range of high activity (pH 5.5). Therefore, we assumed Tyr 71 would be monoprotonated in all cases. As such, there were only four combinations of protonation states whose dynamics were of interest. These were Asp 32 protonated Asp 228 deprotonated (32p 228d), Asp 32 deprotonated Asp 228 protonated (32d 228p), both Asp protonated (32p 228p), and both Asp deprotonated (32d 228d).

Differences in water residence between different protonation states.

In the hypothesized general acid base mechanism for BACE-1, a bridging water facilitates a proton transfer allowing for cleavage of substrate.[165,166] One hypothesis for how BACE-1 activity is regulated is that water is prevented from entering and leaving the active site. To examine this, we have calculated the residence time of water for each of the four protonation states. This is shown in table 2 below:

Table 3.2: hydration number, active water lifetime and residence time for water molecules in all four combinations of protonated and deprotonated for the Asp dyad residues.
Active water lifetime is defined as the continuous length of time the water closest to the dyad remains the closest water. Residency is defined as being within 3.5 A of the Asp dyad, which is defined as the oxygen of the water molecule to the center of mass of the four total titratable protons of Asp 32 and Asp 228.

State	Number of water present in first Hydration Shell	Active water lifetime (ns)	Residence Time for all water(ns)
32p 228d	2.57 ± 1.4	0.16 ± 0.17	1.25 ± 0.15
32d 228p	3.89 ± 1.8	0.20 ± 0.23	1.14 ± 0.18
32p 228p	4.35 ± 1.8	0.07 ± 0.08	0.87 ± 0.09
32d 228d	6.16 ± 1.7	0.03 ± 0.05	1.56 ± 0.18

The data in table 2 suggests that in the 32p 228d state, water was more often occluded from entering the active site. However, waters that moved closest to the active site center tended to remain the closest water for two or three times as long as in the 32p228p or 32d228d states. This is intriguing given that 32p228d is the most populated state in the pH range of highest activity, and therefore a water molecule should remain available to participate in catalysis in this state. Additionally, titratable water-dyad distances do not appear to vary greatly between different protonation states (table 3), indicating this behavior is likely not due to water simply avoided the dyad and being kinetically trapped in the 32p228d case.

Table 3.3: Average distance of the center of mass of the aspartyl dyad protons to the oxygen of the closest water.
While the relative variance for 32d228p is almost twice that of the other protonation states, this is likely an artifact of the naturally high deviations that accompany water measurements, as 32p228d did not show similar results.

Protonation state	Distance of Dyad to closest H_2O (Å)
32d 228p	1.92 ± 0.16
32p 228d	1.78 ± 0.07
32p 228p	1.72 ± 0.08
32d 228d	1.86 ± 0.08

Tyr flap – Asp dyad distance as a possible explanation for unique water behavior among protonation states

Given the differences in water access, we next examined the distance of the Tyr 71 flap to Asp dyad to test for a possible cause for these differences. It is possible that water could be excluded from the active site or prevented from leaving the active site if the flap moves close enough to the dyad. The distance between Tyr 71 to the Asp dyad as a function of time for all four protonation states is shown in Fig 2.
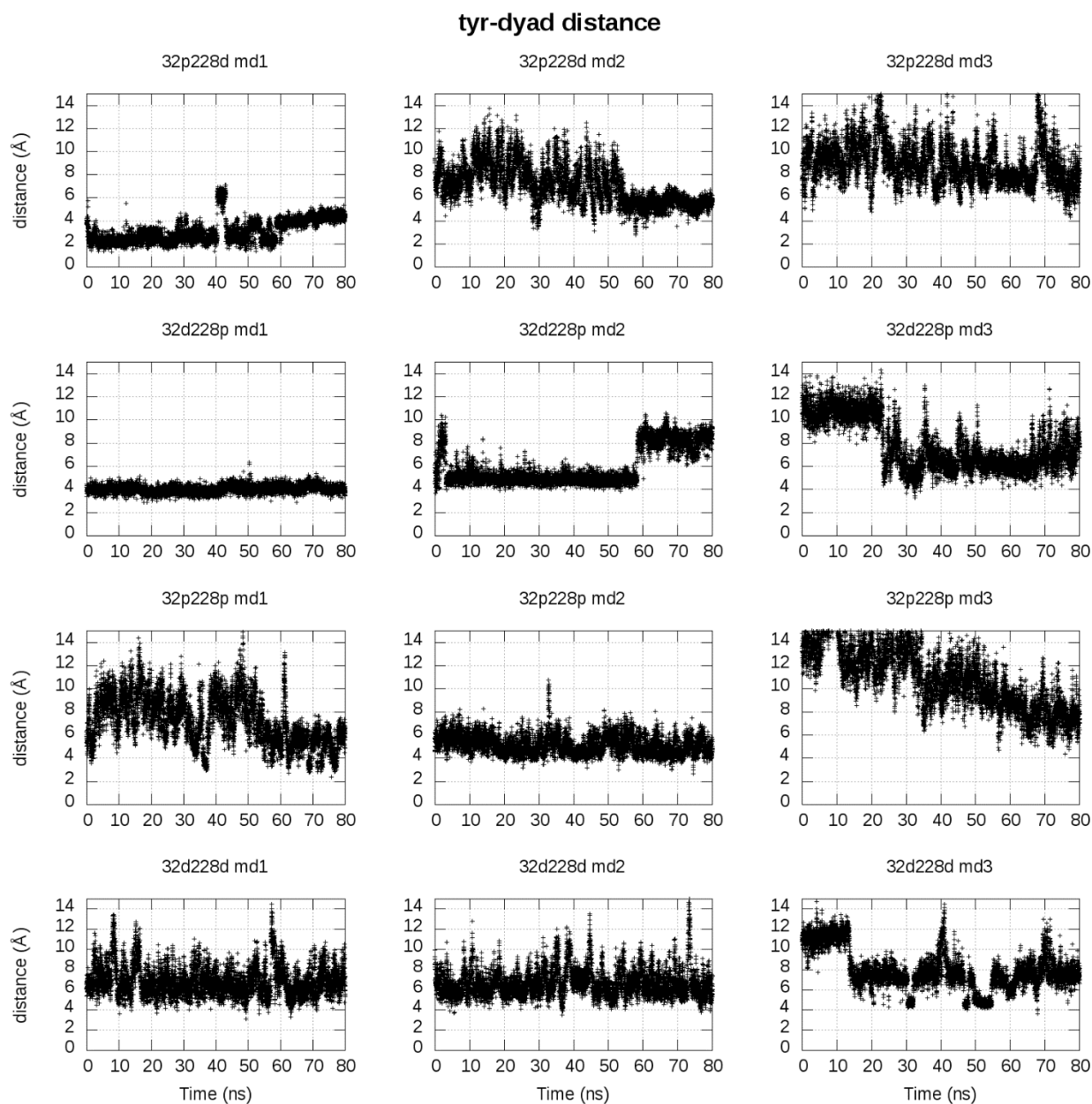
Figure 3.2: Distance of Tyr 71 to the Asp dyad in 32p228d (top), 32d228p (second from top), 32p228p (second from bottom), and 32d228d (bottom).

Md1, md2 and md3 each represent 80 ns trajectories starting from different random seeds post equilibration. Distance is defined as hydrogen of the Tyr 71 side chain to the center of mass of the carboxylate oxygen atoms of the Asp dyad side chains.

In both monoprotonated states, the flap spent more time near the dyad. Additionally, in the 32p

228d state, the flap effectively closed (distance of < 2Å). While the actual protein motion is some

48

combination of these four states and distances, clearly if any of the four states are going to have limited access to water or substrate sterically, it would be 32p 228d. However, 32p 228d is the highest activity state. This suggests that Tyr 71 flap closure may play a role in stabilizing the catalytically relevant water molecule, anchoring it to allow for catalysis.

Conclusions

Based on our CpH simulations and knowledge of the experimental pK_as, 32p 228d should be the most prevalent at the pH of greatest in vivo activity. This state appears to have relatively little access to water, compared to 32p 228p and 32d 228d. However, in the 32p 228d state, the water that is in the active site remains proximal to the active site for longer. It is beyond the scope of this work to attempt to measure the kinetics and timescale of the proton transfer, but it is also possible that the water is not able to stay long enough in the fully protonated or fully deprotonated states for transfer to occur. The stabilization of water in 32p 228d may be due to not just to electrostatics, but a conformational change in which the Tyr71 flap closes and interacts with the aspartyl dyad. It seems possible that the extra time which water can spend engaged with the dyad due to this anchoring by Tyr 71 may allow for optimal proton transfer. Given these results as well as previous studies which indicate the importance of Tyr 71 in BACE-1[105,167,170], focus on inhibitors that are capable of disrupting the Tyr 71-Asp dyad interaction may aid in combating AD via BACE-1 inhibition.

Acknowledgements

secretase 1: a molecular dynamics study. *submitted*.". The dissertation author was the primary investigator and author of this paper.

Supporting Information

Table 3.4: difference in average hydration number when counting 10 closest waters vs counting 11 closest waters for all four protonation states.

State	Number of water present in first Hydration shell counting 10 closest waters	Number of water present in first Hydration shell counting 11 closest waters
32p 228d	2.43	2.43
32d 228p	3.74	3.74
32p 228p	4.39	4.39
32d 228d	6.52	6.52

Table 3.5: Cooperativity constant, n in the Hill equation.
A value of 1.0 indicates no cooperativity, and a perfect fit to the Hill equation.

Residue	Hill coefficient
Asp 32	0.82 ± 0.02
Asp106	0.73 ± 0.01
Asp138	0.83 ± 0.03
Asp223	0.71 ± 0.02
Asp228	0.71 ± 0.09
Glu116	0.73 ± 0.04
Glu265	0.95 ± 0.01
Glu339	0.93 ± 0.04
His45	1.37 ± 0.04
Tyr 71	1.06 ± 0.01
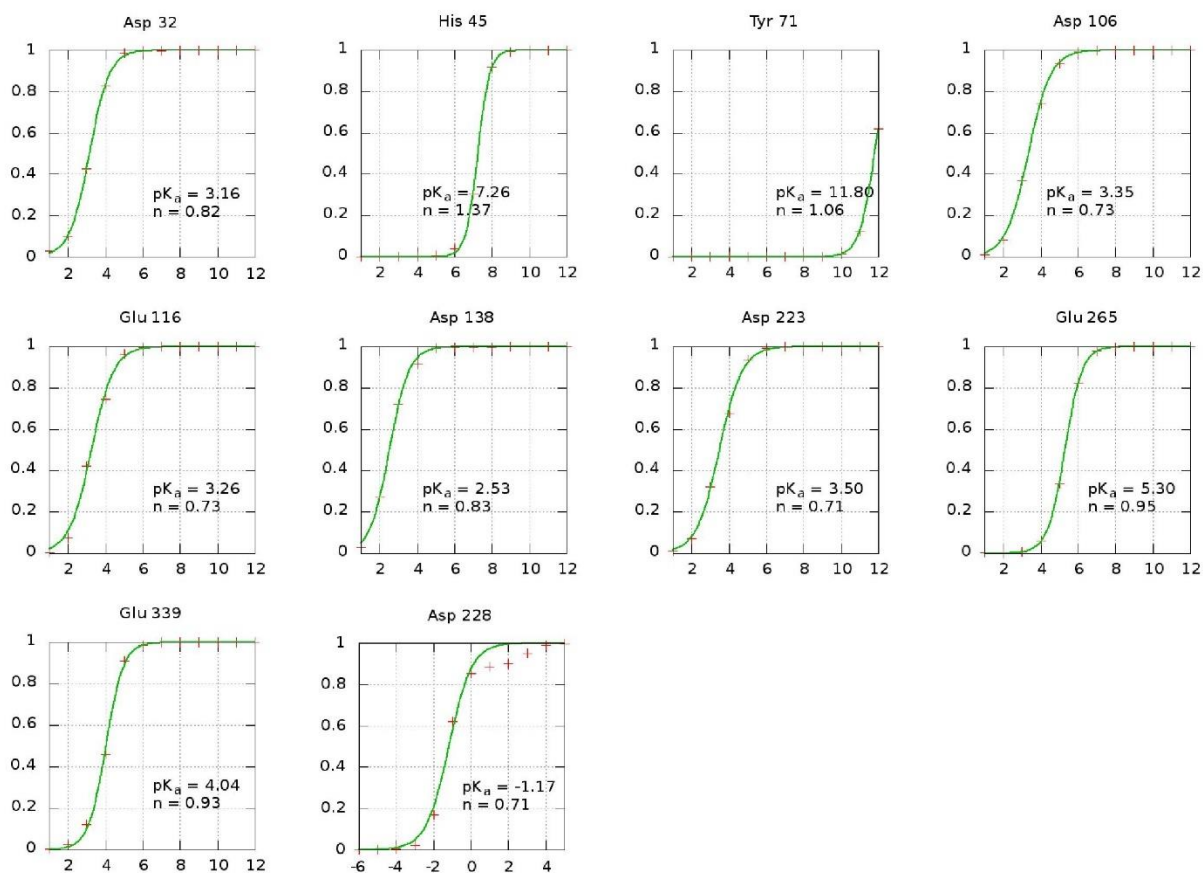
Figure 3.3: Hill equation plots for titrated residues.

Table 3.6: average distance of Tyr 71 to Asp dyad for all four protonation states.
Error reported is the standard deviation.

Protonation state	Average Tyr71-dyad Distance (Å)
32d 228p	6.4 ± 2.2
32p 228d	6.9 ± 2.4
32p 228p	8.4 ± 3.2
32d 228d	7.1± 1.7

Chapter 4 : Mechanisms for Benzene Dissociation through the Excited State of T4 Lysozyme

L99A

Abstract

The atomic-level mechanisms that coordinate ligand release from protein pockets are only known for a handful of proteins. Here, we report the mechanisms for benzene dissociation from the buried cavity of the T4 lysozyme L99A mutant observed by accelerated molecular dynamics. Remarkably, the benzene is released through a previously characterized sparsely populated room temperature excited state of the mutant and explains the coincidence for experimentally measured benzene off rate and apo protein slow timescale NMR relaxation rates between ground and excited states. The path observed for benzene egress is a multistep ligand migration from the buried cavity to ultimate release through an opening between the F/G, H and I helices and requires many cooperative multi-residue rearrangements within the C-terminal domain core. These rearrangements are similar to those observed along the ground state to excited state transitions characterized by Anton molecular dynamic simulations. Analysis of the residue motion cooperativity as well as the molecular properties of the residues lining the egress path suggest that protein polarity in addition to cooperative residue motions plays an important role in the release mechanism.

Introduction

While receptor flexibility has been acknowledged as important for ligand binding and molecular recognition[188–190], its importance is less emphasized in studies of ligand release, particularly in cases where the ligand release occurs on timescales longer than a few hundred microseconds. Recently, the development of new molecular dynamics simulation methodologies and faster calculations enabled through GPU or Anton technologies has allowed for a more detailed atom-level investigation of ligand release mechanisms [191–198].

The T4 lysozyme L99A mutant is a popular model system for studying excited states of proteins and the thermodynamics of ligand binding to buried cavities due to some interesting chemical features[199–203]. The hydrophobic cavity in the L99A mutant is greater than 100 Å^3 and is buried more than 5Å from the protein surface[199,204]. Despite this large cavity, the backbone of the crystal structure of L99A is identical to the backbone of the crystal structure of the wild-type protein (Figure 4.1). This backbone overlap proved to be misleading, as two experimentally-detected dynamic motions occur in this mutant that have not been previously witnessed in the wild-type protein: the formation of a sparsely populated excited state[205–207] and the binding of ligands to the buried cavity. Strikingly, a clear path from solvent to the buried cavity is absent from the crystal structure[200,208–217].
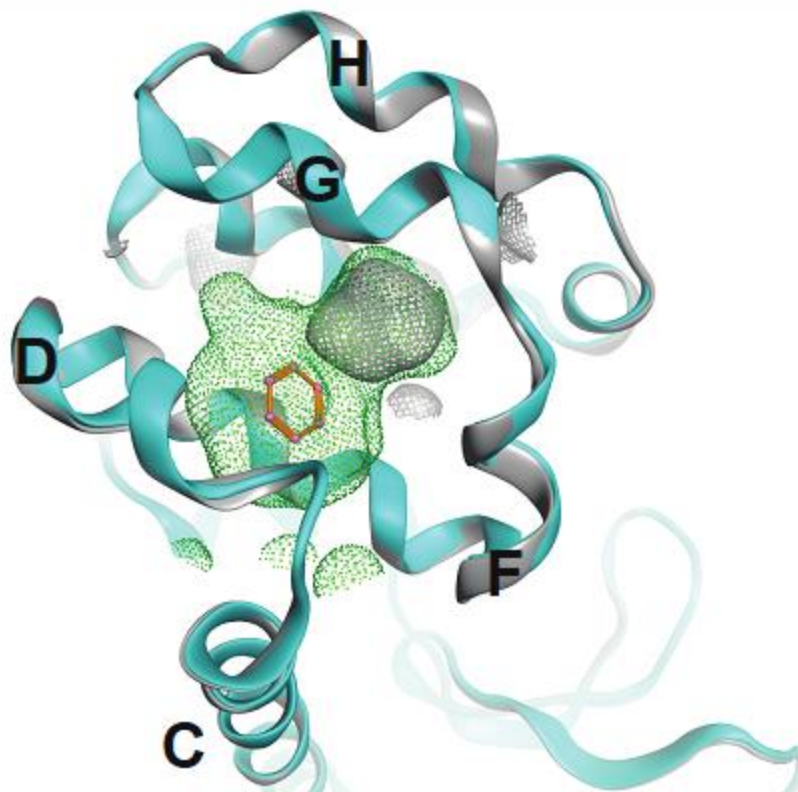
Figure 4.1: Crystallographic comparison of T4 lysozyme WT* (PDB: 1L63, gray) and L99A benzene bound mutant (PDB: 181L, cyan and green cavity).

There are few noticeable differences between the two structures apart from cavity expansion.

Here we have performed accelerated molecular dynamics (aMD) simulations of the L99A mutant of T4 lysozyme with benzene bound to elucidate the protein motions necessary for ligand dissociation. This methodology was chosen because the benzene off rate is longer than can be simulated using conventional molecular dynamic simulations and aMD has previously been shown to accurately recapitulate experimentally measured long-lived conformational phenomena such as torsional populations and water-protein exchange[131,218]. Through this atomic-level mechanistic approach we witness and characterize benzene dissociation through a surface between the F/G, H and I helices (the "FGHI surface"). Moreover, chemical shift calculations [219]

and rotamer measurements confirm benzene dissociates from L99A through the NMR-characterized excited state and not the ground state. Poisson-Boltzmann calculations for cavity opening surfaces between the D and F/G helices (DG), where benzene binds, and between the FGHI surface, where benzene exits, suggest that polar surfaces impact the direction of ligand ingress and egress.

Results

Benzene leaves through a transient surface opening in L99A excited state

Accelerated molecular dynamic simulations were carried out for the L99A cavity mutant (apo)[204], benzene bound to the L99A cavity mutant (holo)[220], and WT* T4 lysozyme (PDB = 1L63)[221].

Figure 4.2: **Benzene egress along a multistate pathway.**

Superposition of benzene (orange) sampling three substates labeled S1, S2 and S3 and the solvent on its path of egress. L99A ribbon representation shows the structure of L99A after benzene exits.

In one of three 500ns trajectories for holo T4 L99A, benzene unbinds from the protein through a transient surface opening between the F/G, H, and I helices, in a succession of steps, labeled S1, S2, and S3 in Figure 4.2. These substrates trace the same path that the aromatic ring of F114 travels in the transition from ground to excited state of L99A, albeit in the opposite direction. This conserved path is illuminated through internal distance measurements and protein alignments (Figure 4.6).

Dynamic motions surrounding the buried cavity result in productive and non-productive excursions from the ground state to the excited state/benzene egress

The concerted protein motions and final structural state through which benzene leaves L99A recapitulate the hallmarks of the L99A ground state to excited state transition and the previously published structure for the excited state. Figure 4.3 illustrates a comparison of measurements for our holo L99A simulations compared to published characteristic rotamer fluctuations, H-bond distances and back calculated chemical shifts for key residues involved in the ground state to excited state transition of apo T4 L99A [193,199,207,222]. As benzene egresses (Figure 4.3b), concerted motions of buried hydrophobic side chain rotations and secondary
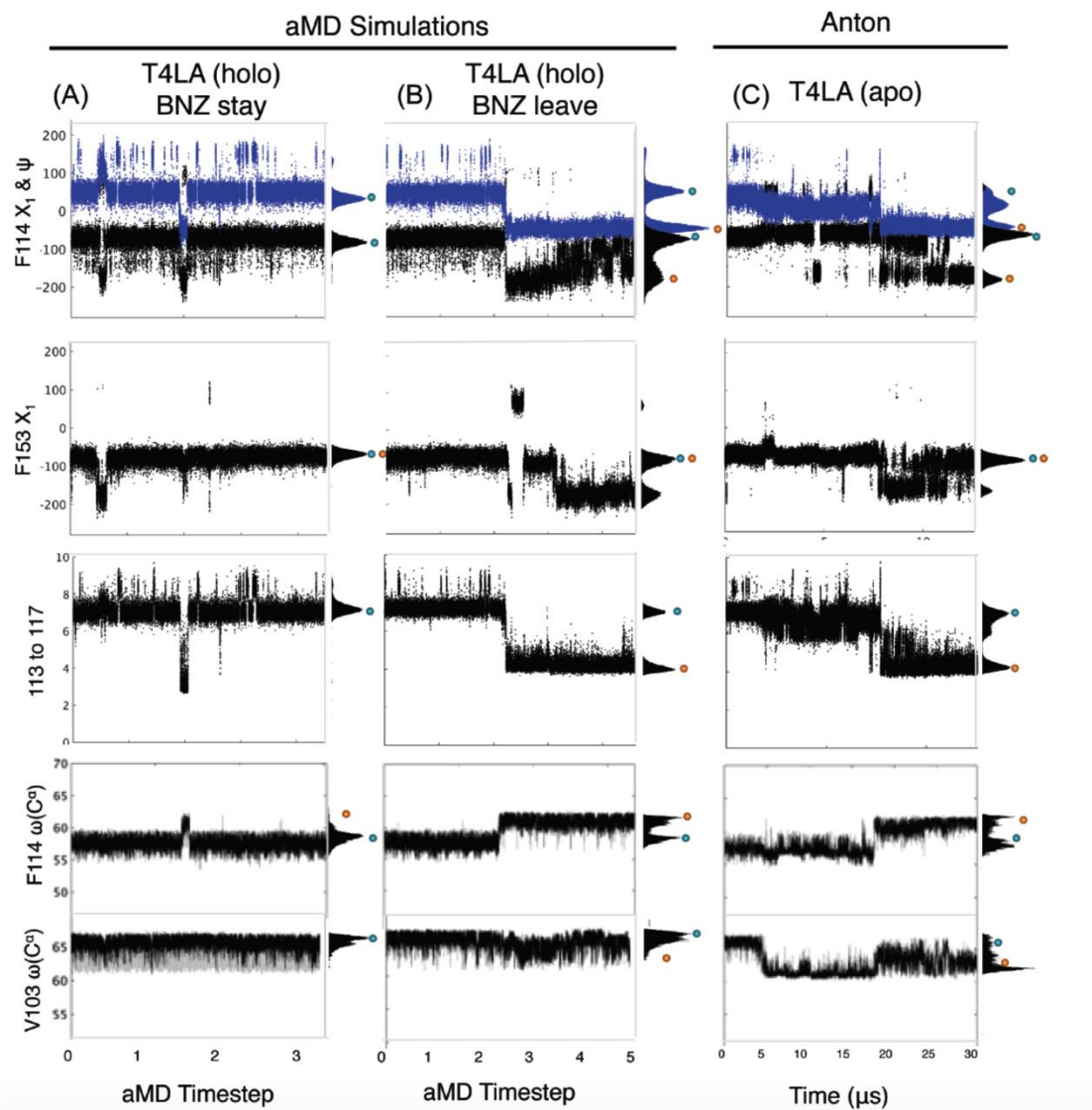
Figure 4.3: comparison of aMD and Anton L99A simulations.

Rotamer changes, hydrogen bond distances, and back-calculated chemical shifts are shown over aMD simulation trajectories for (a) stable benzene holo L99A simulation, (b) for the benzene egress simulation, and compared to the previously reported ground to excited state L99A simulation (c) performed on Anton [199].

structure changes for the F and G-helices give rise to similar substates as the apo transitions

(Figure 4.3c). The F-helix unwinds (Figure 4.12) and then refolds to form a single long helix

with the G helix (referred to as F/G helix). At aMD timestep 2.2×10^4 the F/G helix rewinds and

L121, L133 and F153 side chains rotate to open an adjacent cavity closer to the H and J helices,

58

similar to substate 2. Concurrent with this benzene relocation, the F114 ring relaxes into a more buried position in the L99A pocket, the excited state position. These visual observations are supported by the torsional angles measured for F114 and F153, and the shortened amide nitrogen to carbonyl oxygen i to i+4 distance for G110-F114, A112-N116, and G113-S117 (Figure 4.3b, 4.13, 4.14).

Non-productive excursions from the ground state structure of L99A, in which benzene remains in the buried L99A cavity, are also seen in the aMD simulations of apo and holo L99A (Figure 4.3b, 4.12), These fluctuations are consistent with an array of experimental observables for these residues: low NMR order parameters[223], lower hydrogen exchange protection factors relative to WT[224,225], high crystallographic B-factors[226], and propensity to deform upon ligand binding to the nearby cavity[214,220,227]. These excursions are largely unproductive because the residue transitions are not concerted as they are when benzene egresses (Figure 4.3b).

Polarity of mobile defects prescribe whether benzene stays or leaves

Upon characterization of the different states of benzene, we were interested to know why benzene leaves through this path rather than the more direct path(s) observed for benzene binding.[205,228] We had anticipated that egress would be governed solely by the opening of a path with an appropriate shape and volume to the surface sufficient for benzene to pass. However, our results illustrate that multiple paths are available for benzene egress (Figure 4.7), as was also seen in the apo L99A simulation to the excited state (Figure 4.8)[199]. Interestingly, benzene only egresses from one path in our simulations. It is possible that egress through the more direct path(s) would be observed in longer simulations.

Surprisingly, the opening where benzene exits from, between the F/G, H and I helices (FGHI), is narrower than the openings between the D and F/G helices (DG) (Figure 4.8). In fact,

cavity opening at the DG surface is large enough to completely expose benzene to solvent. Thus, cavity volume and mobile defects alone cannot account for the preference of benzene to egress through the FGHI pocket.

Electrostatic changes of the protein surface over the trajectory, characterized by the Adaptive Poisson-Boltzmann Solver (APBS) method, reveal that polarity plays a role in benzene binding and release[229,230]. Snapshots from the trajectory, with the cavity openings simulated to both bind benzene (DG) and release it (FGHI), were analyzed with APBS. These results demonstrate that the FGHI cavity is more polar and positively charged as compared to the hydrophobic DG cavity (Figure 4.3, 4.9, 4.10, 4.11). The slight positive charge of the FGHI pocket is due to local helix unwinding that exposes backbone amides at the pocket surface. It is therefore of interest that the mere opening of a path between the core and surface is not sufficient for benzene egress[231]. The change in surface charge of the buried cavity is accompanied by, or perhaps due to, a shift in the location of the cavity closer to the H and J helices.
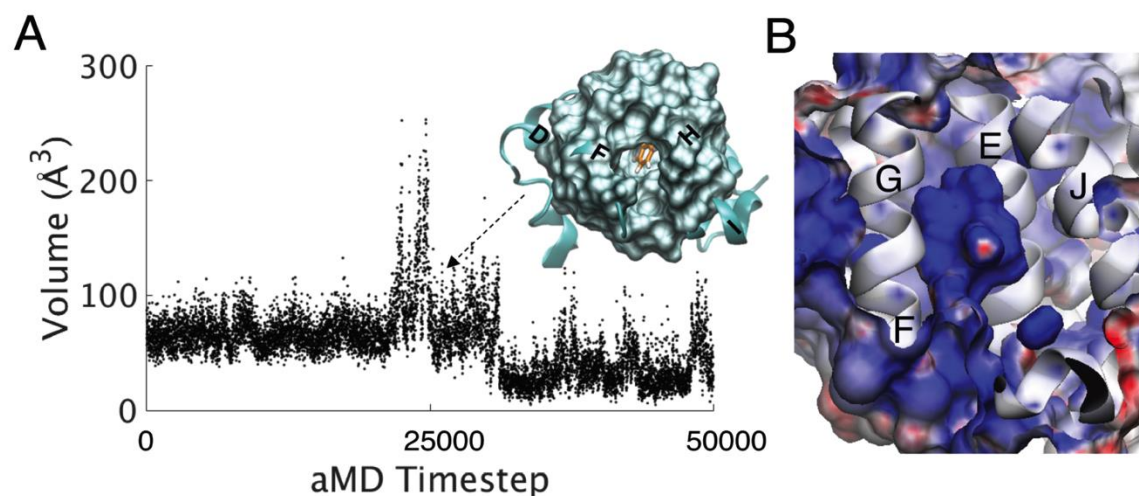
Figure 4.4: Mobile defects provide benzene a more polar cavity on the path of egress.

(A) The measured buried pocket volume throughout the simulation, calculated using POVME2.0[232]. The pocket at the F, H and I helices is observed to open to solvent at ~ 25,000 aMD timesteps. (B) The pocket opening between the F, H and I helices from the buried cavity to solvent occurs when benzene is in the S2 state (Figure 4.2). Blue indicates a positive charge and red indicates a negative charge. Helices are shown in white.

Wild-type T4 lysozyme samples similar yet attenuated plasticity relative to the cavity-expanded

mutant

Accelerated MD simulations on WT* lysozyme were conducted for comparison to aMD simulations on apo L99A and benzene bound L99A. Concerted motions seen in WT* simulations are reminiscent of those seen for the benzene-L99A complex (Figure 4.5A/B). Unlike in L99A, fluctuations within the WT* protein quickly relax back to the "ground" state. There is not a large enough cavity to accommodate F114 in WT* as there is in L99A. In wild type, these concerted motions of internal and backbone dihedral angles serve to open the groove between the F-G, H, and I helices, the same groove that binds the peptide portion of the peptidoglycan substrate of T4 lysozyme (Figure 4.5C/D).
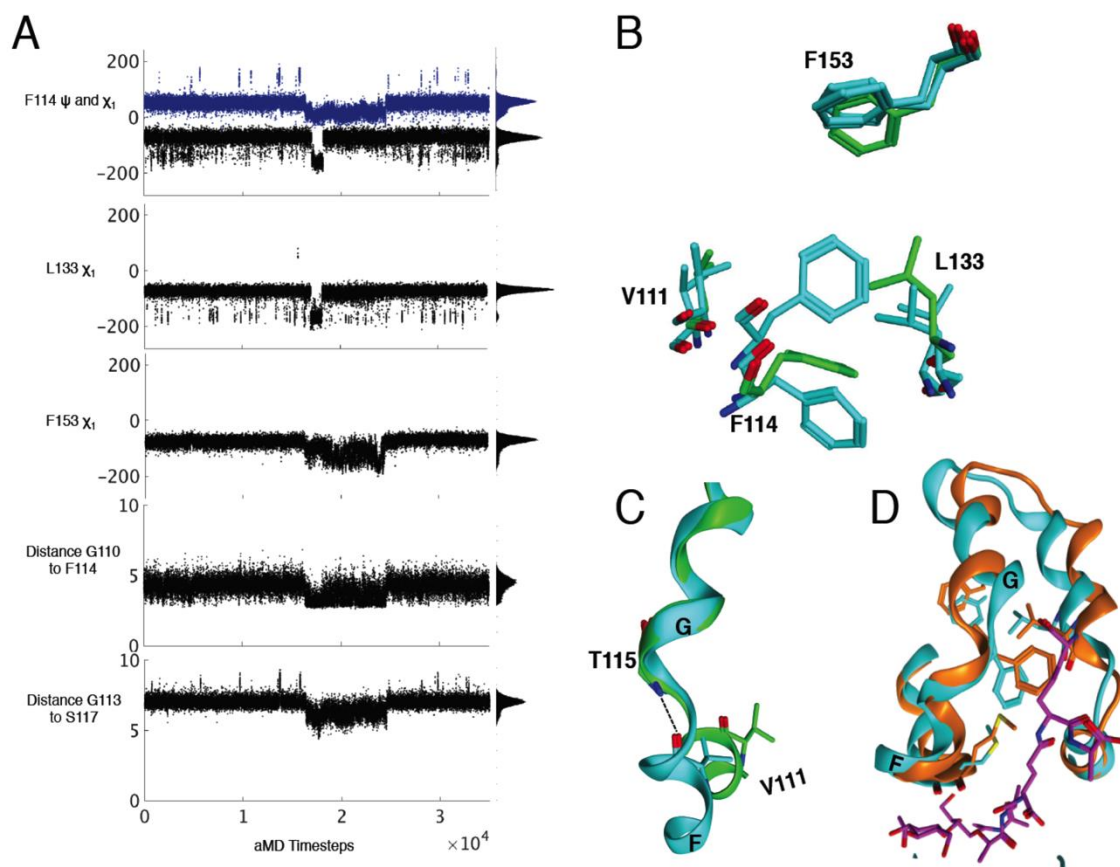
Figure 4.5: Wild type T4 lysozyme transition from ground state to a high energy-like state.

(A) Concerted motions and backbone breakages occurring at 17000 aMD timesteps. (B/C) Structures from the crystal structure (green) compared to the high energy-like state (cyan) for F114, L133, F153, and V111. (D) Helices F, G, H and I in the ground state (cyan) and WT* "excited" state (orange) with the peptidoglycan substrate (purple).

Discussion

Recently, a number of T4 L99A dynamics studies have investigated the kinetic rates for benzene egress with metadynamics simulations [233] and suggested that the major path of ligand egress occurs through the FGHI pocket in the ground state with adiabatic-biased molecular dynamics[193]. Yet, the rates of benzene leaving and excited state transitioning in the L99A mutant

of T4 lysozyme are experimentally indistinguishable ($\sim$1 ms^{-1})[209,234]. The simulations reported here are the first to demonstrate that these events are coincidental because benzene leaves through the excited state of L99A. Additionally, our results illustrate two cooperative mechanisms for benzene egress through the FGHI surface: first, concerted backbone and side chain motions create a path to the protein surface and second, the final substate's surface polarity allows benzene to partition into the aqueous environment. Thus, these results suggest that the benzene off pathway requires a greater degree of cooperativity than the benzene on pathway[208,231].

The FGHI pocket that benzene leaves from has been previously observed in several instances: in the crystal structure of the L99G mutant (Figure 4.11),[235] in crystal structure of L99A bound to a congeneric series of ligands[210], in the L99A excited state from long-timescale simulations performed on the Anton supercomputer[199], and in biased MD simulations of benzene dissociation.[193] The polarity of this pocket is supported by the measured lower free energy of L99A binding iodobenzene as compared to L99A binding benzene.[236] We suggest that pocket polarity could lead to a slower off-rate for iodobenzene than for benzene, due to greater energetic stability of the iodide moiety in the polar FGHI pocket. While further experiments measuring the on and off rates of various benzene analogs could confirm or deny these results, our findings provide a strong case for the importance of surface polarity in the mechanism of ligand release for the L99A mutant.

It does not appear that benzene induces the conformational changes in L99A, as identical motions are seen in the apo L99A simulations[199], and suggests that the experimental off rate is a consequence of the rate required for cooperative motions to develop the sparsely populated higher energy state. Additionally, the L99A mutant does not elicit a completely unique set of

dynamics, but rather appears to exaggerate the breathing and quaking motions[237] intrinsic to WT. In wild-type T4 lysozyme, the dynamics observed in the F helix and surrounding residues may be important for the threading of the peptide portion of peptidoglycan out of the active site cleft as part of the enzymatic mechanism of T4 lysozyme[237]. These results have broad implications for the design of new proteins from wild-type proteins, and for the engineering of dynamical motions in protein interiors.

Finally, we find that in an era where various enhanced sampling methods are being tested for elucidation of rare, long-timescale cooperative conformational events, the accelerated molecular dynamics employed here was suitable for recapitulating a rare conformational event that occurs on the millisecond timescale experimentally. Similarly, Anton simulations captured the rare transition to the excited T4 lysozyme state of the apo protein[199]. It will be of interest to see if other enhanced sampling methods such as adiabatic metadynamics can also capture these rare conformational events that result from cooperative fluctuations in buried protein cavities.

Methods

Accelerated MD simulations

Three systems were set up for accelerated molecular dynamics (aMD) simulations using identical methods described previously[131] with the following exceptions; PDB codes 1L63.pdb, 1L90.pdb, and 181L.pdb were used for the wild type (WT*)[221], L99A apo mutant ground state[238] and L99A benzene bound mutant lysozyme ground state[227], respectively. WT* refers to a sequence difference from WT where residues C54T and C97A mutations were made to promote protein stability *in vitro* and were maintained here *in silico* for alignment with experiments[236]. Proteins were prepared using the Protein Wizard in Maestro module (Schrödinger, Inc.), using

crystallographic pH conditions, 150mM NaCl (30 Cl⁻ and 21 Na⁺ ions) and visual inspection for correct side-chain protonation states.   Benzene for the holo system was parameterized using Jaguar (Schrödinger, Inc.), for geometry optimization and charge calculations, then incorporated into the starting structure MD parameter file using GAFF and XLeap. Each system contained ~33,000 atoms. ~1.5us of simulation was conducted for each system using AMBER 11 PMEMD 2010 and AMBER 12 PMEMD 2012 for the holo system[2,136], TIP4P waters[134] and the AMBER-ff99SB[239] force field; three independently seeded runs of ~500ns were simulated for each system.

50-ns of conventional MD was used as a reference for each system and to calculate boost potentials as described previously[131]. Boost potentials for the three systems were: WT*, E(dih-boost) = 2535 kcal/mol, E(PE-boost) = -89887 kcal/mol; L99A apo, E(dih-boost) = 2540 kcal/mol, E(PE-boost) = -89829 kcal/mol; L99A holo, E(dih-boost) = 2431 kcal/mol, E(PE-boost) = -90597 kcal/mol.

For all simulation analyses and simulation details of the Anton trajectory please see Schiffer et. al [199].

APBS calculations

The PDB2PQR [230] automated pipeline was used to investigate the electrostatics of buried cavities and pocket openings from the aMD trajectories. A pH of 5.5 was chosen to align with the pH from multiple NMR studies.[240,241] The PARSE force field was chosen for the calculation.[242]

Acknowledgements

Figure 4.6: The stochastic path of benzene egress and the concerted motions of benzene leaving and excited state transitioning.

(a,b) Internal distances between the center of mass between the aromatic ring of F114 (a) or the aromatic ring of benzene (b) and the beta carbon of A99 (red), the beta carbon of V87 (blue), and the center of mass of the four helix bundle (black). (d) The locations of A99 (red), V87 (blue), F114 (cyan) and the four-helix bundle (black) are shown for reference. The location of buried cavity is also shown in purple with the C-terminal domain of L99A from ground state apo crystal structure (PDB: 4W51). (e) The conformation of L99A when benzene samples the S2 state.
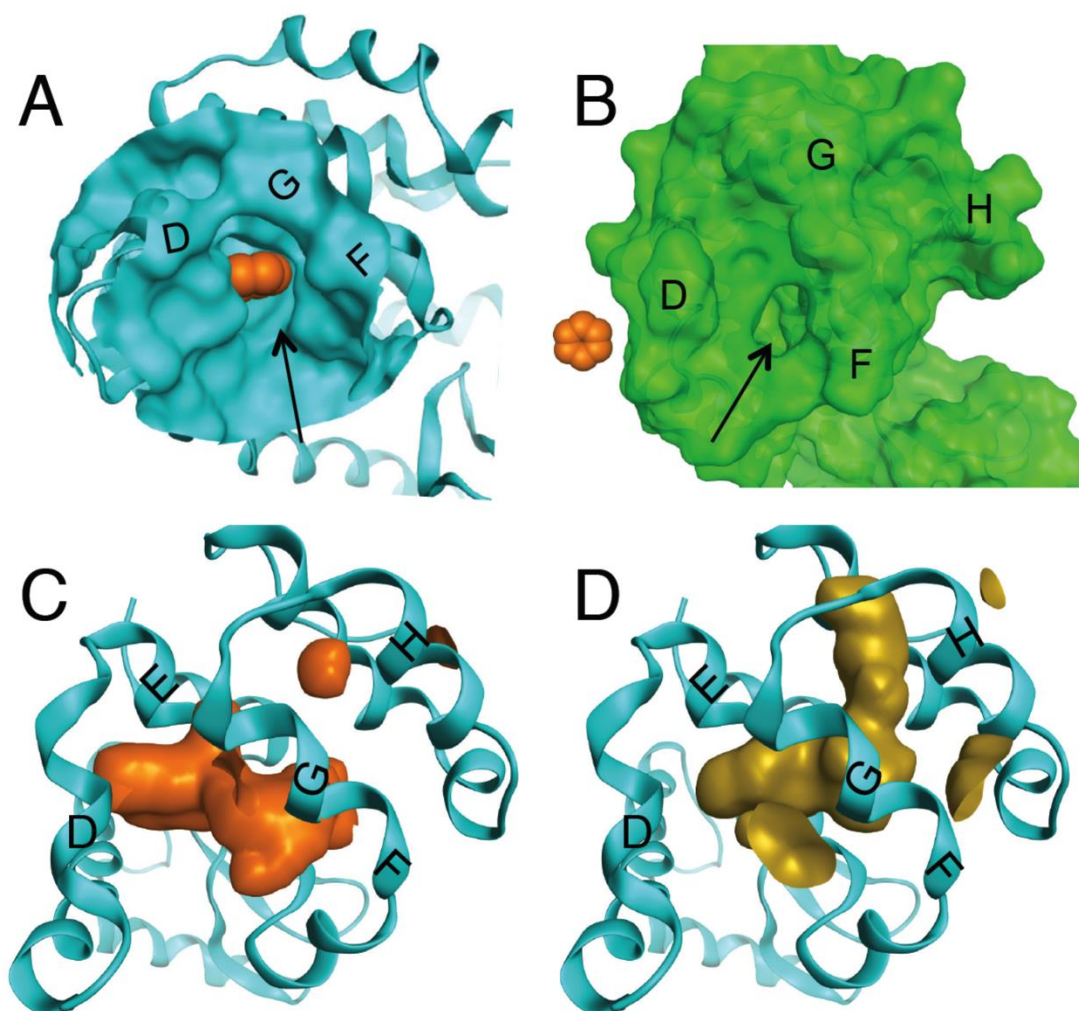
Figure 4.7: Internal flexible motions that result in mobile defects to the protein surface suggest multiple potential sites of ligand egress and ingress, depending on ligand size and nature.

(a,b) A pocket opening between the D, F, and G helices is seen to occur in the aMD trajectory both when benzene is bound (cyan, A) and after benzene has escaped from the buried pocket the L99A is in its apo form (green, B). (c, d) Mobile defects of the buried cavity in the ground state reveal a large conformational change as well as cavity opening to surfaces previously simulated to either exist in the apo state[196] or to allow for the egress or ingress of molecular O_2.[202]
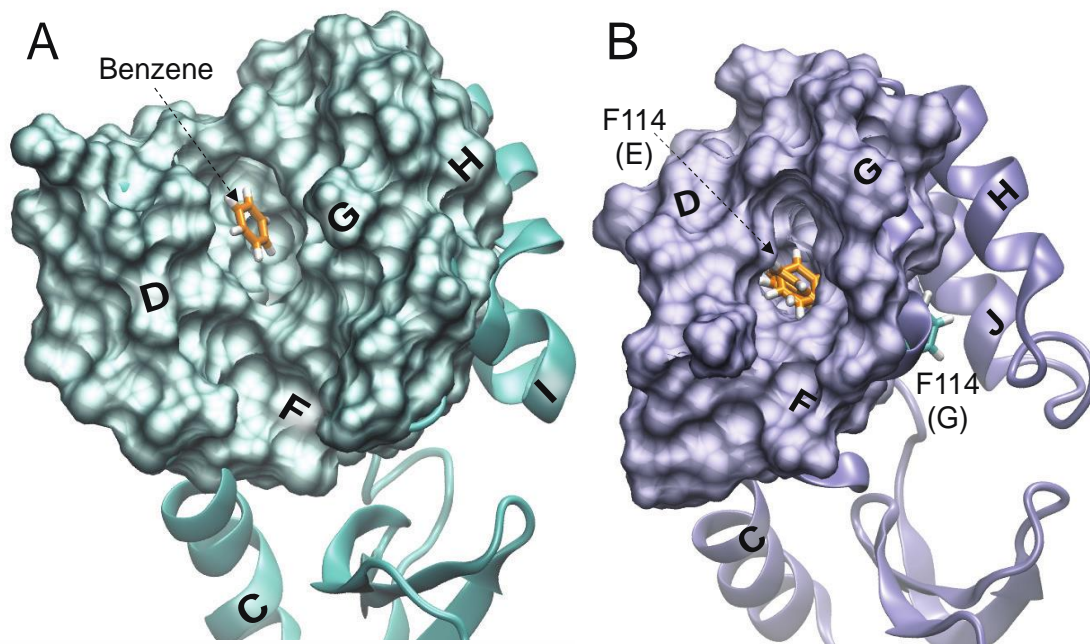
Figure 4.8: A large transient packing defect precedes correlated fluctuations that allow for benzene egress and excited state transitioning.

(a) The structure of L99A (holo) aMD simulation after ~ 24000 timesteps (~ 150 ns). This large cavity opening precedes the shift of benzene from S1 to S2. The surface of the protein surrounding benzene in this state consists primarily of hydrophobic sidechains. The location of benzene is shown (orange) for reference. (b) The structure of the L99A (apo) Anton trajectory at ~ 18 μs, just before the transition to an intermediate state that is characterized by flipping of the F114 phenyl ring into the buried cavity. A few examples of F114 in the excited state (orange) and one example of F114 in the ground state (cyan) are shown for reference.
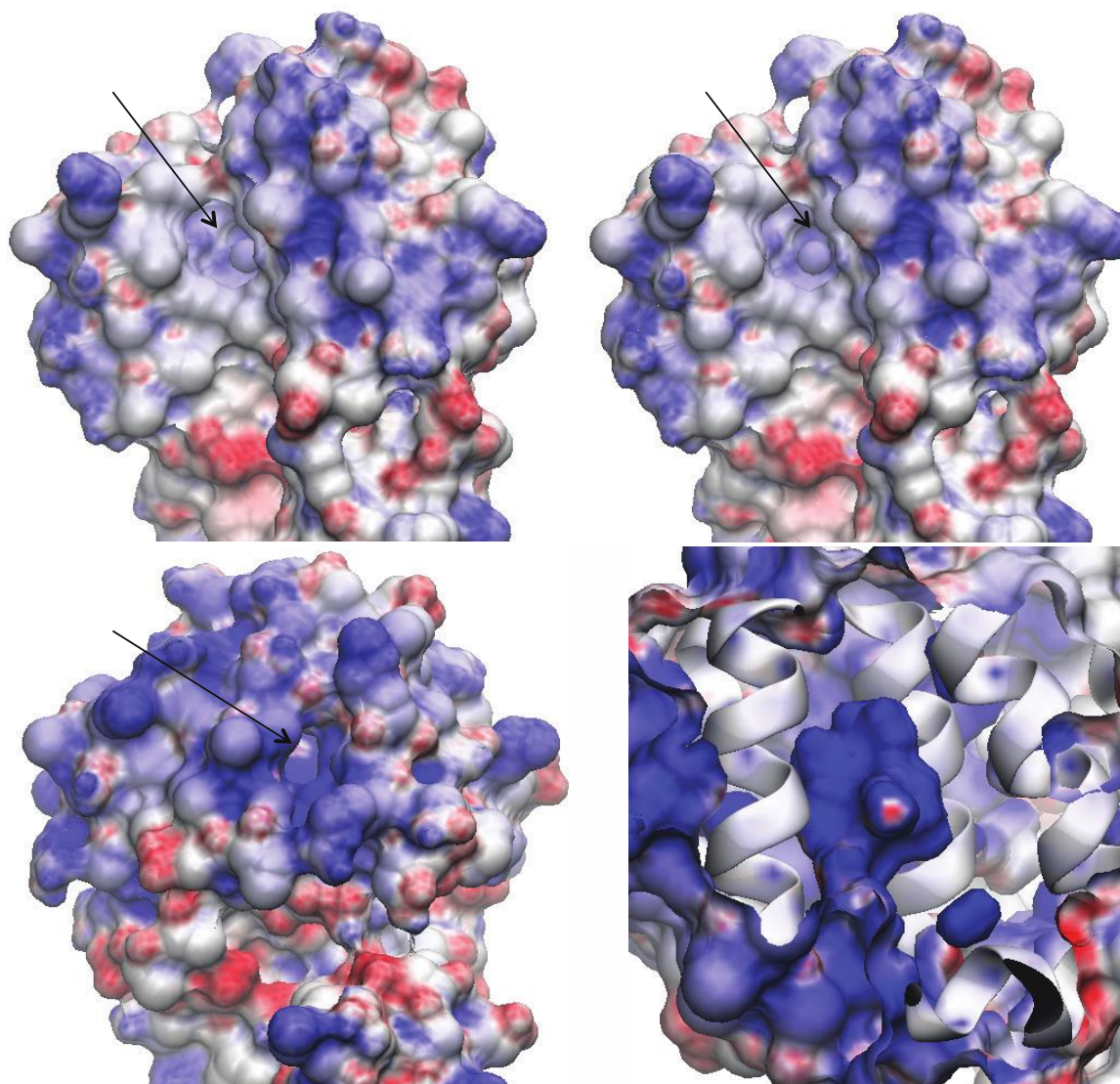
Figure 4.9: Cavity polarity differences in mobile defects between the D/G helices and between the F/G/H helices.

Positively charged surfaces are displayed in blue, negatively charged surfaces are displayed in red, and hydrophobic surfaces are shown in white. The cavity between the D/G helices is largely nonpolar and hydrophobic, whereas the cavity formed between the F/G/H helices is highly cationic and surrounded by charge surfaces on the solvent face of the cavity.
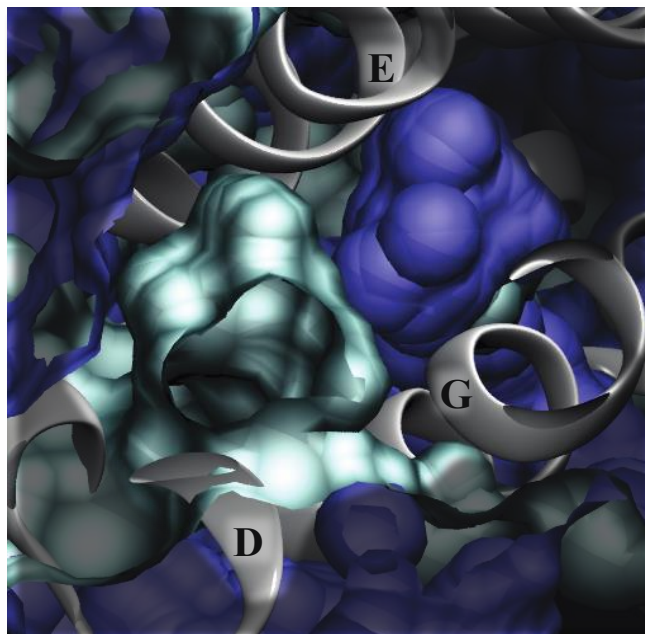
Figure 4.10: Locations of the DG and FGHI pockets differ in the C-terminal domain.

In white cartoon is showing the backbone of the C-terminal domain from the L99A structure with the FGHI pocket. The FGHI pocket surface is shown in blue, whereas the DG pocket surface is shown in cyan. Snapshots from trajectory are aligned to the backbone of residues 75 to 155.

| L99A BNZ leaving state aMD | WT T4 lysozyme aMD processive state | L99G T4 lysozyme Crystal |
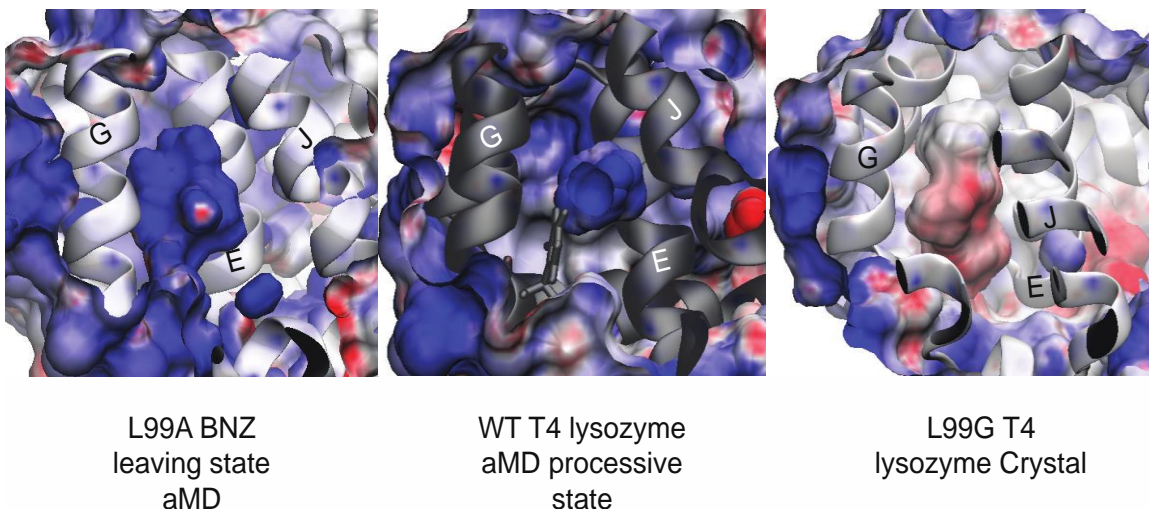
Figure 4.11: Cavity polarity differences between the benzene leaving state in the L99A aMD trajectory, the processive state in the WT aMD trajectory, and the L99G T4 lysozyme crystal structure.

Positively charged surfaces are displayed in blue, negatively charged surfaces are displayed in red, and hydrophobic surfaces are shown in white. The backbone secondary structures are depicted in white for mutants and black for wildtype.

Figure 4.12: Helix cracking and backbone flexibility of the F/G helices during benzene egress.

Snapshots from our aMD trajectory along the path of benzene (orange) egress demonstrate the inherent flexibility of the F and G helices in the C-terminal domain of the L99A cavity mutant (cyan) during the transition. The ground state crystal structure is shown for reference in gray.
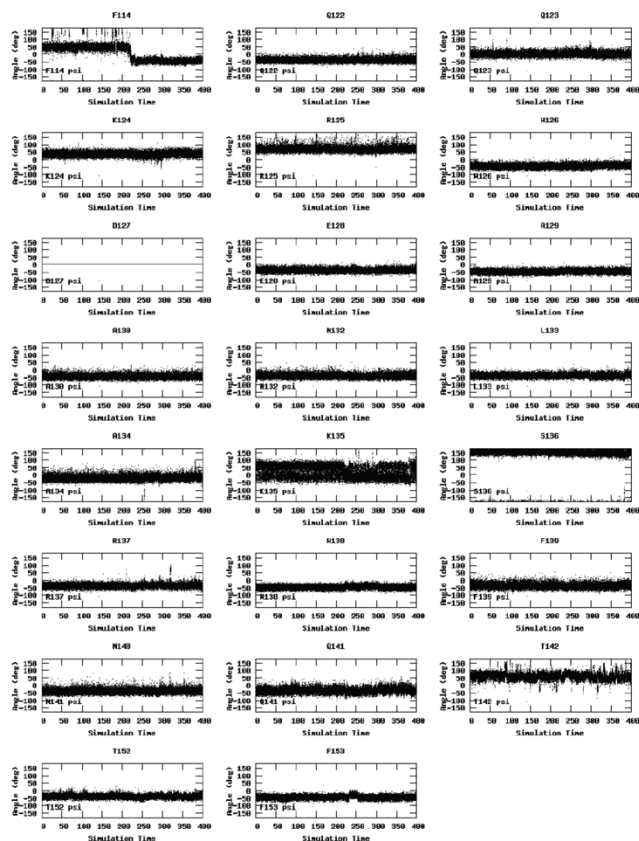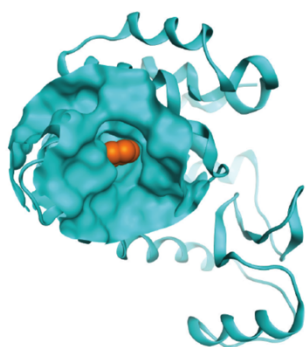
Figure 4.13: Dihedral angle changes enable cavity openings between various helices, including the D and G helices.

An array of dihedral angle changes are mapped over time for the aMD trajectory where benzene egresses at ~25000 time steps.

dist(i, i+4) carbonyl to amide   residue chi1 angle   residue psi angle
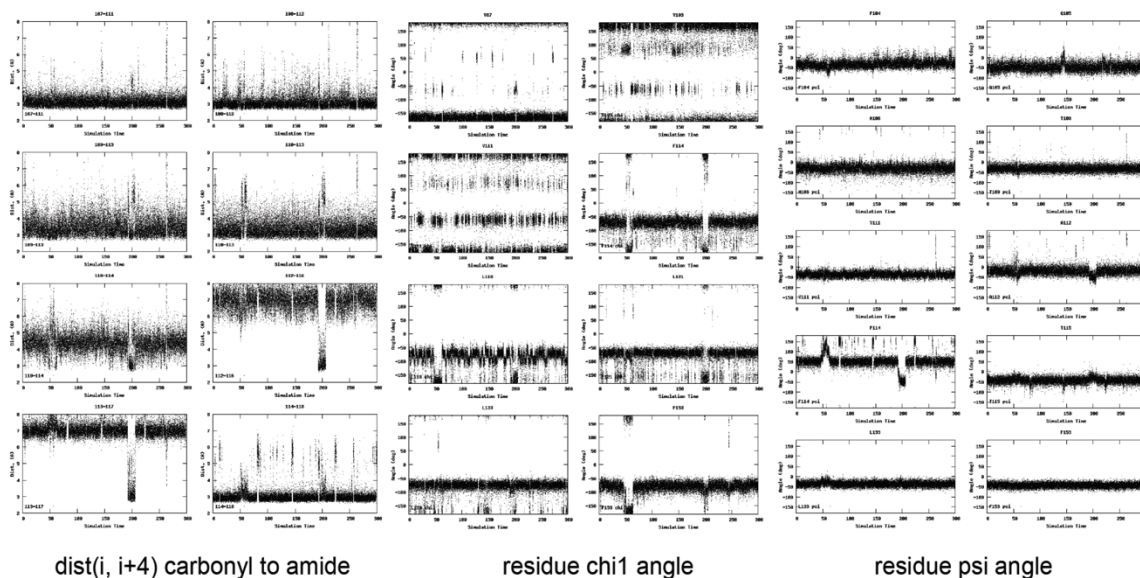
Figure 4.14: Backbone hydrogen bond breakages and side chain and backbone torsion angle changes that accompany non-productive excursions from the ground state.

An array of backbone hydrogen bond distances, χ_1 angle, and ψ angle changes from an aMD trajectory where benzene does not egress but exchanges between the S1 and S2 states of benzene. The difference between a productive and non-productive transition lies in the degree of coincidence for both backbone and side chain torsion angles. In the cases where benzene moves around in the buried cavity but does not leave during our aMD trajectories (column 1, Figure 4.2), conformational changes are less concerted. Two instances of rotamer shifts are witness, at 3,000 aMD timesteps and at 13,000 aMD timesteps. The first non-productive excursion results from shifting of both F114 and F153 χ_1 rotamer flips, but are unaccompanied by any backbone conformational changes or motion of benzene. The second non-productive excursion results from F114 χ_1 rotamer flips as well as folding of the F and G helices into a single helix, as witnessed in the reduction in distance between backbone nitrogen of G113 and the backbone carbonyl oxygen of S117. This second excursion is accompanied by benzene shifting to an S2-like state. However, in this second excursion, F153 and L133 χ_1 rotamer flips do not follow the transition, and benzene returns to its crystallographic position.

74

Appendix

Methods

Alchemical path

One disadvantage of the AMBER solvation free energy code is that it is not currently possible to selectively decouple individual terms (electrostatic nonbonded interactions, Lennard-jones interactions). The result is that solvation free energy calculations must be carried out along a less efficient alchemical path than in Gromacs (figure 1.2)

GROMACS:
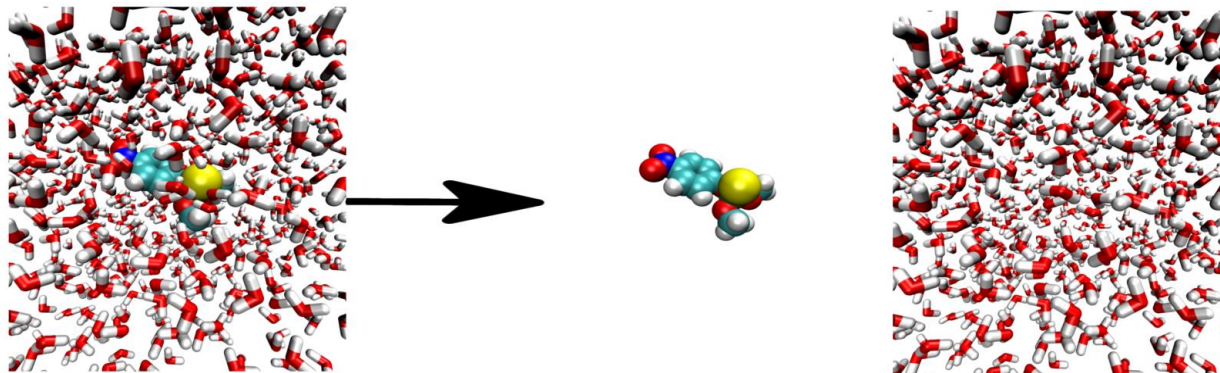
AMBER
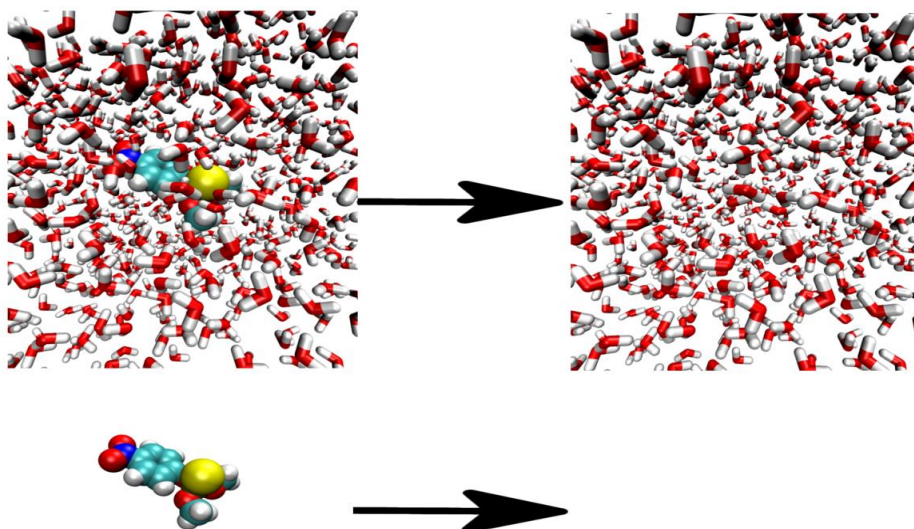
Figure A.1: comparison of GROMACS (top) and AMBER (bottom) solvation free energy alchemical paths.

Each arrow represents one simulation which must be run. In GROMACS, it is possible to selectively decouple only the interactions between ligand atoms and solvent atoms, but leave the ligand-ligand nonbonded interactions unchanged.

However, the additional gas phase calculations achieve over 1000 ns/day of performance and therefore add almost no overhead, and the availability of GPU AFE in AMBER adds a significant performance boost which more than offsets this added cost.

System setup and minimization

Mol2 and frcmod files for methyl hexanoate, butan-1-ol , 1,2,3,4-tetrachloro-5-(3,4-dichlorophenyl)benzene, propanethiol, butan-2-ol, acetaldehyde, benzyl bromide, biphenyl, methylparathion, and nonanal were taken from the FreeSolv database.[93] The leap module in AMBER 16 was used to create all systems. Ligands were parametrized using the same generalized Amber forcefield (GAFF)[95] parameters for the bonded and van der Waals parameters, and then solvated with TIP3P[134] water. A cubic periodic box was used with a minimum distance of 15 Å between any box edge and any solute atom. All systems were minimized for 1000 cycles of steepest descent followed by 1000 cycles of conjugate gradient. Solute atoms were restrained with a restraint weight of 10 kcal/(mol*Å)^2. Minimization was followed, for all values of λ, by 100ps of heating at constant volume and then 1 ns of equilibration at constant pressure. Temperature was regulated via a Langevin thermostat set to a target temperature of 298.15 K and a collision frequency of 5.0 ps^{-1}. For the solvated systems, Pressure was regulated using a Monte Carlo barostat with a target pressure of 1.0 atm and pressure relaxation time of 2.0 ps. This step was unnecessary and therefore skipped for gas phase calculations.

Simulation details

Solvent simulations were run in the NPT ensemble with a Langevin thermostat set to 298.15 K with collision frequency of 5.0 ps^{-1}, and a Monte Carlo barostat at a target pressure of 1.0 atm and pressure relaxation time of 2.0 ps. The direct space cutoff was set to 10 Å for both van der Waals and electrostatics. Long range electrostatics were handled via the Particle Mesh Ewald (PME) method[42] with a FFT grid spacing of ~ 1 point per angstrom. Gas phase simulations were run in the NVT ensemble with a Langevin thermostat set to 298.15 K with collision frequency of 5.0 ps^{-1} The direct space cutoff was set to 9999 Å for both van der Waals and electrostatics, as there were no long-range effects to consider. Charge changing simulations were run with six lambda windows: 0.0, 0.2, 0.4, 0.6, 0.8 and 1.0. Softcore potentials were not used for charge-change simulations. Lennard-Jones changing simulations were only necessary for the solvated ligand and were run with entirely softcore ligands using equations (1) and (2). Lennard-Jones changing simulations were run at 16 different lambda windows: 0.0, 0.5, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8 0.85, 0.9, 0.95, and 1.00. Energies and dV/dL values were printed every 0.5 ps. The default value for scalpha (0.5) was used. Energies were printed every 0.5 ps. Gas phase simulations were run using the current AMBER 18 development tree with serial pmemd on CPU. GPU simulations were run on and NVIDIA GeForce GTX 1080 Ti GPU using the current AMBER 18 development tree with our GPU TI support incorporated. Simulations were run with a 2-fs time step. The RFEB complex and solvated systems were simulated for 10 ns. The first 5 ns of each simulation was discarded for equilibration purposes

A diverse subset of the FreeSolv database was chosen to cover as large a portion of "database space" as possible with regards to atom types, molecular weight and functional groups to best test the accuracy of AMBER's alchemical free energy tools. To improve the comparison, the same protein and ligand forcefields, as well as lambda schedule and equilibration procedures were replicated from the recent FreeSolv update. Table A.1 compares AMBER solvation free energies with those calculated using Gromacs and experimentally.

Table A.1: comparison of our calculated values with those of Duarte et al. and experiment. All values are in kcal/mol.

Molecule	AMBER	Gromacs[93]	Experiment [93]
Methyl hexanoate	-3.40	-3.30	-2.49
Butan-1-ol	-3.27	-3.23	-4.72
1,2,3,4-tetrachloro-5-(3,4-dichlorophenyl)benzene	-1.01	-1.08	-3.04
Propanethiol	-0.19	-0.18	-1.1
Butan-2-ol	-3.17	-3.15	-4.62
Acetaldehyde	-3.44	-3.37	-3.50
Benzyl bromide	-1.92	-1.85	-2.38
Biphenyl	-3.12	-3.14	-2.70
Methylparathion	-10.52	-10.47	-7.19
Nonanal	-2.44	-2.34	-2.07

Discussion

Comparison to Gromacs

Everything in the system setup and simulation was identical to the simulations run in the most recent FreeSolv update, except for two factors: analysis method and alchemical path. Gromacs can perform a direct calculation, in which selectively the non-bonded interactions between solvent and solute are removed, but those between solute and solute remain on (figure A.1, GROMACS). AMBER requires the use of an indirect calculation. In an indirect calculation, all non-bonded interactions are turned off in solution, and then the process is repeated in gas phase (figure A.1, AMBER). This requires additional simulations. All results agreed with the Gromacs calculations within error. There was deviation from experiment, in some cases as large as 1.5 kcal/mol. Additionally, TI results were identical to MBAR results obtained from Gromacs. These facts taken together indicate that in this case, analysis method and alchemical path are not the dominant source of error in these calculations. Factors outside of alchemical path and analysis method, most likely force field accuracy or protonation state,[59] led to deviation from experiment. Statistical uncertainty due to a 10ns simulation time limit also likely added to the deviation from experiment, although it's unclear by how much given that the two sets of simulations quickly converged to the same values. Whether this is true for all solvation free energy calculations or in general ligand sized systems may require further testing, but certainly results are promising. Additionally, for solvation free energy calculations it may be acceptable to rely on TI alone rather than MBAR or MBAR and TI. While this will not always be the case, particularly in larger more complex systems where dV/dL is expected to be very non-linear[23], it would be a major boon as it does provide a 10-20% performance boost

Conclusions and Future work

A robust protocol first presented by Duarte et al.[9] and then adapted here for AMBER has been used to successfully predict the solvation free energy of 10 diverse ligands, with virtually identical precision and accuracy to Gromacs at a fraction of the computational cost. Future work will include collaborating with the curators of FreeSolv and comparison of these results to the same systems parametrized with GAFF 2, to quantify the improvements in GAFF 2 vs GAFF 1.7.

References

(1)   Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C. *J. Chem. Theory Comput.* **2013**, *9* (9), 3878–3888.

(2)   Götz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C. *J. Chem. Theory Comput.* **2012**, *8* (5), 1542–1555.

(3)   Le Grand, S.; Götz, A. W.; Walker, R. C. *Comput. Phys. Commun.* **2013**, *184* (2), 374–380.

(4)   Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781–1802.

(5)   Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. *Comput. Phys. Commun.* **1995**, *91* (1–3), 43–56.

(6)   Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. *J. Comput. Chem.* **2005**, *26* (16), 1701–1718.

(7)   Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindah, E. *SoftwareX* **2015**, *1–2*, 19–25.

(8)   Hess, B.; Kutzner, C.; Van Der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4* (3), 435–447.

(9)   Eastman, P.; Friedrichs, M. S.; Chodera, J. D.; Radmer, R. J.; Bruns, C. M.; Ku, J. P.; Beauchamp, J. A.; Lane, T. J.; Wang, L.; Shukla, D.; Tye, T.; Houston, M.; Stich, T.; Klein, C.; Shirts, M. R.; Pande, V. S. *J. Chem. Theory Comput.* **2013**, *9* (1), 461–469.

(10)   Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4* (2), 187–217.

(11)   Bowers, K. J.; Chow, E.; Xu, H.; Dror, R. O.; Eastwood, M. P.; Gregersen, B. A.; Klepeis, J. L.; Kolossvary, I.; Moraes, M. A.; Sacerdoti, F. D.; Salmon, J. K.; Shan, Y.; Shaw, D. E. In *Proceedings of the ACM/IEEE Conference on Supercomputing (SC06)*; 2006; pp 11–17.

(12)   Harvey, M.; Giupponi, G.; Fabritiis, G. De. *J. Chem. Theory Comput.* **2009**, *5*, 1632.

(13)   Mermelstein, D. J.; Lin, C.; Nelson, G.; Kretsch, R.; McCammon, J. A.; Walker, R. C. *J. Comput. Chem.* **2018**, *DOI: 10.10*.

(14)   Mermelstein, D. J.; McCammon, J. A.; Walker, R. C. *submitted* **2018**.

(15)   McCammon, J. A.; Gelin, B. R.; Karplus, M. *Nature* **1977**, *267* (5612), 585–590.

(16)   Tembe, B. L.; McCammon, J. A. *Comput. Chem.* **1984**, *8* (4), 281–283.

(17)   Zwanzig, R. *J. Chem. Phys.* **1954**, *1420* (1954), 1420–1426.

(18)   Kirkwood, J. G. *J. Chem. Phys.* **1935**, *3* (1935), 300–313.

(19)   Bennett, C. H. *J. Comput. Phys.* **1976**, *22*, 245–268.

(20)   Shirts, M. R.; Chodera, J. D. *J. Chem. Phys.* **2008**, *129* (12).

(21)   Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13* (8), 1011–1021.

(22)   Ferrenberg, A. M.; Swendsen, R. H. *Phys. Rev. Lett.* **1989**, *63* (12), 1195–1198.

(23)   Chipot, C.; Pohorille, A. *SPRINGER SERIES IN CHEMICAL PHYSICS 86 Free Energy Calculations Theory and Applications in Chemistry and Biology*; Chipot, C., Pohorille, A., Eds.; Springer: New York, 2007.

(24)   Lee, T. S.; Hu, Y.; Sherborne, B.; Guo, Z.; York, D. M. *J. Chem. Theory Comput.* **2017**, *13* (7), 3077–3084.

(25)   Bruckner, S.; Boresch, S. *J. Comput. Chem.* **2011**, *32* (7), 1320–1333.

(26)   Queimada, J.; Economou, I. G. *J. Chem. Theory Comput.* **2010**, *6*, 1018–1027.

(27)   Bruckner, S.; Boresch, S. *J. Comput. Chem.* **2011**, *32* (7), 1303–1319.

(28)   Chipot, C. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2014**, *4* (1), 71–89.

(29)   Shirts, M. R.; Pande, V. S. *J. Chem. Phys.* **2005**, *122* (14), 144107.

(30)   Chodera, J. D. *J. Chem. Theory Comput.* **2016**, *12*, 1799–1805.

(31)   Klimovich, P. V.; Shirts, M. R.; Mobley, D. L. *J. Comput. Aided. Mol. Des.* **2015**, *29* (January), 397–411.

(32)   Nguyen, T. H.; Mihn, D. *J. Chem. Theory Comput.* **2016**, *12*, 2154–2161.

(33)   Mobley, D. L.; Chodera, J. D.; Dill, K. A. *J. Chem. Theory Comput.* **2007**, *3* (4), 1231–1235.

(34)   Shirts, M. R.; Pande, V. S. *J. Chem. Phys.* **2005**, *122* (13), 134508.

(35) Steinbrecher, T.; Mobley, D. L.; Case, D. A. *J. Chem. Phys.* **2007**, *127* (21), 214108.

(36) Pitera, J. W.; Van Gunsteren, W. F. *Mol. Simul.* **2002**, *28* (1–2), 45–65.

(37) Anwar, J.; Heyes, D. M. *J. Chem. Phys.* **2005**, *122* (22), 224117.

(38) Steinbrecher, T.; Joung, I.; Case, D. A. *J. Comput. Chem.* **2011**, *32* (15), 3253–3263.

(39) Boresch, S.; Karplus, M. *J. Phys. Chem. A* **1999**, *103* (1), 103–118.

(40) Boresch, S.; Karplus, M. *J. Phys. Chem. A* **1999**, *103* (1), 119–136.

(41) Shirts, M. R.; Mobley, D. L. In *Methods in molecular biology (Clifton, N.J.)*; 2013; Vol. 924, pp 271–311.

(42) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089.

(43) Lin, Y. L.; Aleksandrov, A.; Simonson, T.; Roux, B. *J. Chem. Theory Comput.* **2014**, *10* (7), 2690–2709.

(44) Hünenberger, P. H.; McCammon, J. A. *J. Chem. Phys.* **1999**, *110* (4), 1856–1872.

(45) Rocklin, G. J.; Mobley, D. L.; Dill, K. A.; Hünenberger, P. H. *J. Chem. Phys.* **2013**, *139* (18), 184103-1-184103–184129.

(46) Homeyer, N.; Gohlke, H. *J. Comput. Chem.* **2013**, *34* (11), 965–973.

(47) Homeyer, N.; Gohlke, H. *Biochim. Biophys. Acta - Gen. Subj.* **2015**, *1850* (5), 972–982.

(48) Loeffler, H. H.; Michel, J.; Woods, C. J. *J. Chem. Inf. Model.* **2015**, *55* (12), acs.jcim.5b00368.

(49) Wereszczynski, J.; McCammon, J. A. *Quarterly Reviews of Biophysics*. 2012, pp 1–25.

(50) Pohorille, A.; Jarzynski, C.; Chipot, C. *J. Phys. Chem. B* **2010**, *114* (32), 10235–10253.

(51) Shirts, M. R.; Mobley, D. L.; Chodera, J. D. In *Annual Reports in Computational Chemistry*; 2007; Vol. 3, pp 41–59.

(52) Christ, C. D.; Mark, A. E.; Van Gunsteren, W. F. *Journal of Computational Chemistry*. 2010, pp 1569–1582.

(53) Shirts, M. R.; Klein, C.; Swails, J. M.; Yin, J.; Gilson, M. K.; Mobley, D. L.; Case, D. A.; Zhong, E. D. *J. Comput. Aided. Mol. Des.* **2017**, *31* (1), 147–161.

(54) Gaieb, Z.; Liu, S.; Gathiaka, S.; Chiu, M.; Yang, H.; Shao, C.; Feher, V. A.; Walters, W.

P.; Kuhn, B.; Rudolph, M. G.; Burley, S. K.; Gilson, M. K.; Amaro, R. E. *J. Comput. Aided. Mol. Des.* **2017**, *32* (1), 1–20.

(55)   Mobley, D. L.; Gilson, M. K. *Annu. Rev. Biophys* **2017**, *46*, 357–377.

(56)   Wong, C. F.; McCammon, J. A. *J. Am. Chem. Soc.* **1986**, *108* (13), 3830–3832.

(57)   GSK Syk Database https://drugdesigndata.org/about/datasets/225 (accessed May 24, 2017).

(58)   Wang, L.; Wu, Y.; Deng, Y.; Kim, B.; Pierce, L.; Krilov, G.; Lupyan, D.; Robinson, S.; Dahlgren, M. K.; Greenwood, J.; Romero, D. L.; Masse, C.; Knight, J. L.; Steinbrecher, T.; Beuming, T.; Damm, W.; Harder, E.; Sherman, W.; Brewer, M.; Wester, R.; Murcko, M.; Frye, L.; Farid, R.; Lin, T.; Mobley, D. L.; Jorgensen, W. L.; Berne, B. J.; Friesner, R. A.; Abel, R. *J. Am. Chem. Soc.* **2015**, *137* (7), 2695–2703.

(59)   Cournia, Z.; Allen, B.; Sherman, W. *Journal of Chemical Information and Modeling*. 2017, pp 2911–2937.

(60)   Abel, R.; Wang, L.; Harder, E. D.; Berne, B. J.; Friesner, R. A. *Acc. Chem. Res.* **2017**, *50* (7), 1625–1632.

(61)   Homeyer, N.; Stoll, F.; Hillisch, A.; Gohlke, H. *J. Chem. Theory Comput.* **2014**, *10* (8), 3331–3344.

(62)   Mey, A. S. J. S.; Jiménez, J. J.; Michel, J. *J. Comput. Aided. Mol. Des.* **2017**, *32* (1), 1–12.

(63)   Abel, R.; Mondal, S.; Masse, C.; Greenwood, J.; Harriman, G.; Ashwell, M. A.; Bhat, S.; Wester, R.; Frye, L.; Kapeller, R.; Friesner, R. A. *Current Opinion in Structural Biology*. 2017, pp 38–44.

(64)   Procacci, P. *J. Mol. Graph. Model.* **2017**, *71*, 233–241.

(65)   Mikulskis, P.; Genheden, S.; Ryde, U. *J. Chem. Inf. Model.* **2014**, *54* (10), 2794–2806.

(66)   Perez, A.; Morrone, J. A.; Simmerling, C.; Dill, K. A. *Curr. Opin. Struct. Biol.* **2016**, *36*, 25–31.

(67)   Lee, H. C.; Hsu, W. C.; Liu, A. L.; Hsu, C. J.; Sun, Y. C. *J. Mol. Graph. Model.* **2014**.

(68)   Olsson, M. A.; Söderhjelm, P.; Ryde, U. *J. Comput. Chem.* **2016**, *37* (17), 1589–1600.

(69)   Hsu, C.-J.; Hsu, W.-C.; Lee, D.-J.; Liu, A.-L.; Chia, |; Chang, -Ming; Shih, H.-J.; Wun-, |; Huang, H.; Lee-Chen, G.-J.; Hsiu, |; Hsieh-Li, M.; Lee, G.-C.; Sun, Y.-C. *Chem Biol Drug Des* **2017**, *90*, 272–281.

(70)   Martin, D. P.; Blachly, P. G.; McCammon, J. A.; Cohen, S. M. *J. Med. Chem.* **2014**, *57* (16), 7126–7135.

(71)   Ciordia, M.; Pérez-Benito, L.; Delgado, F.; Trabanco, A. A.; Tresadern, G. *J. Chem. Inf. Model.* **2016**, *56* (9), 1856–1871.

(72)   Park, J.; McDonald, J. J.; Petter, R. C.; Houk, K. N. *J. Chem. Theory Comput.* **2016**, *12* (4), 2066–2078.

(73)   Ucisik, M. N.; Hammes-Schiffer, S. *J. Am. Chem. Soc.* **2015**, *137* (41), 13240–13243.

(74)   Ucisik, M. N.; Hammes-Schiffer, S. *J. Phys. Chem. B* **2017**, *121* (15), 3667–3675.

(75)   Sun, Y. C.; Hsu, W. C.; Hsu, C. J.; Chang, C. M.; Cheng, K. H. *J. Mol. Model.* **2015**, *21* (11), 283.

(76)   Olsson, M. A.; García-Sosa, A. T.; Ryde, U. *J. Comput. Aided. Mol. Des.* **2017**, *32* (1), 1–14.

(77)   Kaus, J. W.; Mccammon, J. A. *J. Phys. Chem. B* **2015**, *119* (20), 6190–6197.

(78)   Wan, S.; Bhati, A. P.; Zasada, S. J.; Wall, I.; Green, D.; Bamborough, P.; Coveney, P. V. *J. Chem. Theory Comput.* **2017**, *13* (2), 784–795.

(79)   Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. *Biophysical Journal*. 1997, pp 1047–1069.

(80)   Boresch, S.; Tettinger, F.; Leitgeb, M.; Karplus, M. *J. Phys. Chem. B* **2003**, *107* (35), 9535–9551.

(81)   Mobley, D. L.; Chodera, J. D.; Dill, K. A. *J. Chem. Phys.* **2006**, *125* (8), 84902.

(82)   Mobley, D. L.; Dill, K. A. *Structure*. 2009, pp 489–498.

(83)   Gumbart, J. C.; Roux, B.; Chipot, C. *J. Chem. Theory Comput.* **2013**, *9* (1), 794–802.

(84)   Aldeghi, M.; Heifetz, A.; Bodkin, M. J.; Knapp, S.; Biggin, P. C. *Chem. Sci.* **2016**, *7* (1), 207–218.

(85)   Liu, W.; Jia, X.; Wang, M.; Li, P.; Wang, X.; Hu, W.; Zheng, J.; Mei, Y. *RSC Adv.* **2017**, *7* (61), 38570–38580.

(86)   Bannan, C. C.; Calabró, G.; Kyu, D. Y.; Mobley, D. L. *J. Chem. Theory Comput.* **2016**, *12* (8), 4015–4024.

(87)   Liu, S.; Cao, S.; Hoang, K.; Young, K. L.; Paluch, A. S.; Mobley, D. L. *J. Chem. Theory*

Comput. **2016**, *12* (4), 1930–1941.

(88) Zhang, H.; Jiang, Y.; Yan, H.; Cui, Z.; Yin, C. *J. Chem. Inf. Model.* **2017**, *57* (11), 2763–2775.

(89) Jia, X.; Wang, M.; Shao, Y.; König, G.; Brooks, B. R.; Zhang, J. Z. H.; Mei, Y. *J. Chem. Theory Comput.* **2016**, *12* (2), 499–511.

(90) Abrams, J. B.; Rosso, L.; Tuckerman, M. E. *J. Chem. Phys.* **2006**, *125* (7), 74115.

(91) Shivakumar, D.; Williams, J.; Wu, Y.; Damm, W.; Shelley, J.; Sherman, W. *J. Chem. Theory Comput.* **2010**, *6* (5), 1509–1519.

(92) Mobley, D. L.; Dumont, E.; Chodera, J. D.; Dill, K. A. *J. Phys. Chem. B* **2007**, *111* (9), 2242–2254.

(93) Duarte Ramos Matos, G.; Kyu, D. Y.; Loeffler, H. H.; Chodera, J. D.; Shirts, M. R.; Mobley, D. L. *Journal of Chemical and Engineering Data*. 2017, pp 1559–1569.

(94) Oostenbrink, C.; Villa, A.; Mark, A. E.; Van Gunsteren, W. F. *J. Comput. Chem.* **2004**, *25* (13), 1656–1676.

(95) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J Comput Chem* **2004**, *25*, 1157–1174.

(96) Jämbeck, J. P. M.; Lyubartsev, A. P. *J. Phys. Chem. B* **2014**, *118* (14), 3793–3804.

(97) Fennell, C. J.; Wymer, K. L.; Mobley, D. L. *J. Phys. Chem. B* **2014**, *118* (24), 6438–6446.

(98) Zhang, J.; Tuguldur, B.; Spoel, D. Van Der. *J. Chem. Inf. Model.* **2015**, *55*, 1192–1201.

(99) Kim, M. O.; McCammon, J. A. *Biopolymers* **2016**, *105* (1), 43–49.

(100) Mason, A. C.; Jensen, J. H. *Proteins Struct. Funct. Genet.* **2008**, *71* (1), 81–91.

(101) Jensen, J. H. *curr. pharm. biotech.* **2008**, *9*, 96–102.

(102) Goh, G. B.; Laricheva, E. N.; Brooks, C. L. *J. Am. Chem. Soc.* **2014**, *136* (24), 8496–8499.

(103) Misra, V. K.; Hecht, J. L.; Yang, A. S.; Honig, B. *Biophys. J.* **1998**, *75* (5), 2262–2273.

(104) Ellis, C. R.; Tsai, C.-C. C.; Hou, X.; Shen, J. *J. Phys. Chem. Lett.* **2016**, *7* (6), 944–949.

(105) Kim, M. O.; Blachly, P. G.; McCammon, J. A. *PLoS Comput. Biol.* **2015**, *11* (10), 1–28.

(106) Harris, R. C.; Tsai, C.-C.; Ellis, C. R.; Shen, J. *J. Phys. Chem. Lett.* **2017**, No. 8, 4832–4837.

(107) Baptista, A.; Vitor, T.; Claudio, S. *J Chem Phys.* **2002**, *117*, 4184–4200.

(108) Mongan, J.; Case, D. A.; McCammon, J. A. *J. Comput. Chem.* **2004**, *25* (16), 2038–2048.

(109) Lee, M. S.; Salsbury, F. R.; Brooks III, C. L. *Proteins Struct. Funct. Genet.* **2004**, *56* (4), 738–752.

(110) Swails, J. M.; Roitberg, A. E. *J. Chem. Theory Comput.* **2012**, *8* (11), 4393–4404.

(111) Swails, J. M.; York, D. M.; Roitberg, A. E. *J. Chem. Theory Comput.* **2014**, *10* (3), 1341–1352.

(112) Lee, J.; Miller, B. T.; Damjanović, A.; Brooks, B. R. *J. Chem. Theory Comput.* **2014**, *10* (7), 2738–2750.

(113) Khandogin, J.; Brooks, C. L. *Biophys. J.* **2005**, *89* (1), 141–157.

(114) Itoh, S. G.; Damjanović, A.; Brooks, B. R. *Proteins Struct. Funct. Bioinforma.* **2011**, *79* (12), 3420–3436.

(115) Kim, M. O.; Blachly, P. G.; Kaus, J. W.; McCammon, J. A. *J. Phys. Chem. B* **2015**, *119* (3), 861–872.

(116) Lee, J.; Miller, B. T.; Brooks, B. R. *Protein Sci.* **2016**, *25* (1), 231–243.

(117) Wyman, J. *Adv. Protein Chem.* **1948**, *4*, 407–531.

(118) Wyman, J. *J. Mol. Biol.* **1965**, *11*, 631–644.

(119) Tanford, C. In *Advances in Protein Chemistry*; 1970; Vol. 24, pp 1–95.

(120) Sprenger, K. G.; Prakash, A.; Drobny, G.; Pfaendtner, J. *Langmuir* **2018**, *34*, 1199–1207.

(121) Martins De Oliveira, V.; De, V.; Contessoto, G.; Bruno Da Silva, F.; Lucas, D.; Caetano, Z.; Jurado De Carvalho, S.; Barbanti, V.; Leite, P. **2018**, *114*, 65–75.

(122) Dadou, S.; El-Barghouthi, M.; Alabdallah, S.; Badwan, A.; Antonijevic, M.; Chowdhry, B. *Mar. Drugs* **2017**, *15* (10), 298.

(123) Paloni, M.; Cavallotti, C. *ACS Omega* **2017**, *2* (10), 6464–6472.

(124) Vergara-Jaque, A.; Comer, J.; Sepúlveda-Boza, S.; Santos, L. S.; Mascayano, C.;

Sandoval-Yáñez, C. *Int. J. Polym. Mater. Polym. Biomater.* **2017**, *66* (10), 485–494.

(125) Peng, Y.; Alexov, E. *Proteins Struct. Funct. Bioinforma.* **2017**, *85* (2), 282–295.

(126) Sensoy, O.; Atilgan, A. R.; Atilgan, C. *Phys. Chem. Chem. Phys.* **2017**, *19* (8), 6064–6075.

(127) Abdizadeh, H.; Atilgan, A. R.; Atilgan, C. *J. Phys. Chem. B* **2017**, *121* (18), 4778–4789.

(128) Martínez-Muñoz, A.; Bello, M.; Romero-Castro, A.; Rodríguez-Fonseca, R. A.; Rodrigues, J.; Sánchez-Espinosa, V. A.; Correa-Basurto, J. *J. Mol. Graph. Model.* **2017**, *76*, 330–341.

(129) Sun, Z.; Wang, X.; Song, J. *J. Chem. Inf. Model.* **2017**, *57* (7), 1621–1639.

(130) Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. *J. Chem. Theory Comput.* **2011**, *7* (2), 525–537.

(131) Pierce, L. C. T.; Salomon-Ferrer, R.; Augusto F. De Oliveira, C.; McCammon, J. A.; Walker, R. C. *J. Chem. Theory Comput.* **2012**, *8* (9), 2997–3002.

(132) Giese, T. J.; York, D. M. *J. Chem. Theory Comput. Artic. ASAP* **2018**, doi:10.1021/acs.jctc.7b01175.

(133) Wallace, J. A.; Shen, J. K. *J. Chem. Theory Comput.* **2011**, *7* (8), 2617–2629.

(134) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79* (2), 926.

(135) Yin, J.; Henriksen, N. M.; Slochower, D. R.; Shirts, M. R.; Chiu, M. W.; Mobley, D. L.; Gilson, M. K. *J. Comput. Aided. Mol. Des.* **2017**, *31* (1).

(136) Case, D. A.; Cerutti, D. S.; Cheatham III, T. E.; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; Greene, D.; Homeyer, N.; Izadi, S.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Madej, B.; Mermelstein, D.; Merz, K. M.; Monard, G. H.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Simmerling, C. L.; Botello-Smith, W. M.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; Xiao, L.; York, D. M.; Kollman, P. A. 2017.

(137) Case, D. A.; Cerutti, D. S.; Cheatham III, T. E.; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; Greene, D.; Homeyer, N.; Izadi, S.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Madej, B.; Mermelstein, D.; Merz, K. M.; Monard, G. H.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Simmerling, C. L.; Botello-Smith, W. M.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; Xiao, L.; York, D. M.; Kollman, P. A.

2016.

(138) Salomon-Ferrer, R.; Case, D. A.; Walker, R. C. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2013**, *3* (2), 198–210.

(139) Kaus, J. W.; Pierce, L. T.; Walker, R. C.; McCammon, J. A. *J. Chem. Theory Comput.* **2013**, *9* (9), 4131–4139.

(140) Postma, J. P. M.; Berendsen, H. J. C.; Haak, J. R., Thermodynamics of cavity formation in water: a molecular dynamics study, F. S. C. S. *Faraday Symp. Chem. Soc.* **1982**, *17*, 55–67.

(141) Yang, W.; Lu, W.; Lu, Y.; Zhong, M.; Sun, J.; Thomas, A. E.; Wilkinson, J. M.; Fucini, R. V.; Lam, M.; Randal, M.; Shi, X. P.; Jacobs, J. W.; McDowell, R. S.; Gordon, E. M.; Ballinger, M. D. *J. Med. Chem.* **2006**, *49* (3), 839–842.

(142) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. *J. Chem. Theory Comput.* **2015**, *11* (8), 3696–3713.

(143) Joung, I. S. *J. Chem. Phys. B* **2009**, *113*, 13279–13290.

(144) Joung, I. S.; Cheatham, T. E. *J. Phys. Chem. B* **2008**, *112* (30), 9020–9041.

(145) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117* (19), 5179–5197.

(146) M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Had, 2009. **2009**.

(147) Comet XSEDE User Guide https://portal.xsede.org/sdsc-comet (accessed May 24, 2017).

(148) Hopkins, C. W.; Le Grand, S.; Walker, R. C.; Roitberg, A. E. *J. Chem. Theory Comput.* **2015**, *11* (4), 1864–1874.

(149) Su, P. C.; Johnson, M. E. *J. Comput. Chem.* **2016**.

(150) Henriksen, N. M.; Fenley, A. T.; Gilson, M. K. *J. Chem. Theory Comput.* **2015**, *11* (9), 4377–4394.

(151) McKhann, G.; Drachman, D.; Folstein, M.; Katzman, R. *Neurology* **1984**, *34* (7), 939.

(152) Hardy, J.; Selkoe, D. J. *Science* **2002**, *297* (5580), 353–356.

(153) Mattson, M. P. *Nature* **2004**, *430* (7000), 631–639.

(154) Vassar, R. *J. Mol. Neurosci.* **2004**, *23* (1–2), 105–114.

(155) Cole, S. L.; Vassar, R. *Mol. Neurodegener.* **2007**, *2* (1), 22.

(156) Economou, N. J.; Giammona, M. J.; Do, T. D.; Zheng, X.; Teplow, D. B.; Buratto, S. K.; Bowers, M. T. *J. Am. Chem. Soc.* **2016**, No. 138 (6), 1772–1775.

(157) Cai, H.; Wang, Y.; McCarthy, D.; Wen, H.; Borchelt, D. R.; Price, D. L.; Wong, P. C. *Nat. Neurosci.* **2001**, *4* (3), 233–234.

(158) Luo, Y.; Bolon, B.; Kahn, S.; Bennett, B. D.; Babu-Khan, S.; Denis, P.; Fan, W.; Kha, H.; Zhang, J.; Gong, Y.; Martin, L.; Louis, J.-C.; Yan, Q.; Richards, W. G.; Citron, M.; Vassar, R. *Nat. Neurosci.* **2001**, *4* (3), 231–232.

(159) Ghosh, A. K.; Tang, J. *ChemMedChem* **2015**, *10* (9), 1463–1466.

(160) Vassar, R.; Kovacs, D. M.; Yan, R.; Wong, P. C. *J. Neurosci.* **2009**, *29* (41), 12787–12794.

(161) Yan, R.; Vassar, R. *Lancet Neurol.* **2014**, *13* (3), 319–329.

(162) Kandalepas, P. C.; Vassar, R. *J. Neurochem.* **2012**, *120* (SUPPL. 1), 55–61.

(163) Hamada, Y.; Kiso, Y. *Biopolymers* **2016**, *106(4)*, 563–579.

(164) Shimizu, H.; Tosaki, A.; Kaneko, K.; Hisano, T.; Sakurai, T.; Nukina, N. *Mol. Cell. Biol.* **2008**, *28* (11), 3663–3671.

(165) Dunn, B. M. *Chem. Rev.* **2002**, *102* (12), 4431–4458.

(166) Toulokhonova, L.; Metzler, W. J.; Witmer, M. R.; Copeland, R. A.; Marcinkeviciene, J. *J. Biol. Chem.* **2003**, *278* (7), 4582–4589.

(167) Ellis, C. R.; Shen, J. *J. Am. Chem. Soc.* **2015**, *137* (30), 9543–9546.

(168) Gorfe, A. A.; Caflisch, A. *Structure* **2005**, *13* (10), 1487–1498.

(169) Hong, L.; Tang, J. *Biochemistry* **2004**, *43* (16), 4689–4695.

(170) Spronk, S. A.; Carlson, H. A. *Proteins Struct. Funct. Bioinforma.* **2011**, *79* (7), 2247–2259.

(171) Hernandez-Rodriguez, M.; Correa-Basurto, J.; Gutierrez, A.; Vitorica, J.; Rosales-Hernandez, M. C. *Eur. J. Med. Chem.* **2016**, *124*, 1142–1154.

(172) Subramanian, G.; Ramsundar, B.; Pande, V.; Denny, R. A. *J. Chem. Inf. Model.* **2016**, *56* (10), 1936–1949.

(173) Kocak, A.; Erol, I.; Yildiz, M.; Can, H. *J. Mol. Graph. Model.* **2016**, *70*, 226–235.

(174) Gueto-Tettay, C.; Zuchniarz, J.; Fortich-Seca, Y.; Gueto-Tettay, L. R.; Drosos-Ramirez, J. C. *J. Mol. Graph. Model.* **2016**, *70*, 181–195.

(175) Di Pietro, O.; Juarez-Jimenez, J.; Muñoz-Torrero, D.; Laughton, C. A.; Javier Luque, F. *PLoS One* **2017**, *12* (5), 1–22.

(176) Manousiouthakis, V. I.; Deem, M. W. *J. Chem. Phys.* **1999**, *110* (6), 2753.

(177) Stachel, S. J.; Coburn, C. A.; Steele, T. G.; Crouthamel, M. C.; Pietrak, B. L.; Lai, M. T.; Holloway, M. K.; Munshi, S. K.; Graham, S. L.; Vacca, J. P. *Bioorganic Med. Chem. Lett.* **2006**, *16* (3), 641–644.

(178) Jacobson, M. P.; Friesner, R. A.; Xiang, Z.; Honig, B. *J. Mol. Biol.* **2002**, *320* (3), 597–608.

(179) Jacobson, M. P.; Pincus, D. L.; Rapp, C. S.; Day, T. J. F.; Honig, B.; Shaw, D. E.; Friesner, R. A. *Proteins Struct. Funct. Genet.* **2004**, *55* (2), 351–367.

(180) p Schrodinger Release 2016-1: Prime,Schrodinger, LLC.

(181) Wasmuth, E. V; Lima, C. D. *Nucleic Acids Res.* **2016**, *45* (November 2016), 1–12.

(182) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. *J. Mol. Biol.* **1990**, *215* (3), 403–410.

(183) Søndergaard, C. R.; Olsson, M. H. M.; Rostkowski, M.; Jensen, J. H. *J. Chem. Theory Comput.* **2011**, *7* (7), 2284–2295.

(184) Onufriev, A.; Bashford, D.; Case, D. A. *J. Phys. Chem. B* **2000**, *104* (15), 3712–3720.

(185) Onufriev, A.; Bashford, D.; Case, D. A. *Proteins Struct. Funct. Genet.* **2004**, *55* (2), 383–394.

(186) Williams, T.; Kelley, C. 2013.

(187) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graph. Model.* **1996**, *14*, 33–38.

(188) Baron, R.; McCammon, J. A. *Biochemistry* **2007**, *46* (37), 10629–10642.

(189) Karplus, M.; Gelin, B. R.; McCammon, J. A. *Biophys. J.* **1980**, *32* (1), 603–618.

(190) Wüthrich, K.; Wagner, G. *Trends Biochem. Sci.* **1984**, *9* (4), 152–154.

(191) Long, D.; Mu, Y.; Yang, D. *PLoS One* **2009**, *4* (6), 1–10.

(192) Hu, Y.; Liu, H. *J. Phys. Chem. A* **2014**, *118* (39), 9272–9279.

(193) Wang, Y.; Papaleo, E.; Lindorff-Larsen, K. *Elife* **2016**, *5* (AUGUST), 1–35.

(194) Swegat, W.; Schlitter, J.; Krüger, P.; Wollmer, A. *Biophys. J.* **2003**, *84* (3), 1493–1506.

(195) Sonoda, M. T.; Martínez, L.; Webb, P.; Skaf, M. S.; Polikarpov, I. *Mol. Endocrinol.* **2008**, *22* (7), 1565–1578.

(196) Bortolato, A.; Deflorian, F.; Weiss, D. R.; Mason, J. S. *J. Chem. Inf. Model.* **2015**, *55* (9), 1857–1866.

(197) Guo, D.; Pan, A. C.; Dror, R. O.; Mocking, T.; Liu, R.; Heitman, L. H.; Shaw, D. E.; IJzerman, A. P. *Mol. Pharmacol.* **2016**, *89* (5), 485–491.

(198) Dickson, A.; Lotz, S. D. *Biophys. J.* **2017**, *112* (4), 620–629.

(199) Schiffer, J. M.; Feher, V. A.; Malmstrom, R. D.; Sida, R.; Amaro, R. E. *Biophys. J.* **2016**, *111* (8), 1631–1640.

(200) Mulder, F. A.; Skrynnikov, N. R.; Hon, B.; Dahlquist, F. W.; Kay, L. E. *J. Am. Chem. Soc.* **2001**, *123* (5), 967–975.

(201) Ucisik, M. N.; Zheng, Z.; Faver, J. C.; Merz, K. M. *J. Chem. Theory Comput.* **2014**, *10* (3), 1314–1325.

(202) Kitahara, R.; Yoshimura, Y.; Xue, M.; Kameda, T.; Mulder, F. A. A. *Sci. Rep.* **2016**, *6*, 1–12.

(203) Mann, G.; Hermans, J. *J. Mol. Biol.* **2000**, *302* (4), 979–989.

(204) Eriksson, A. E.; Baase, W. A.; Zhang, X. J.; Heinz, D. W.; Blaber, M.; Baldwin, E. P.; Williams, B. W. . *Science (80-.).* **1992**, *255* (5041), 178–183.

(205) Korzhnev, D. M.; Orekhov, V. Y.; Dahlquist, F. W.; Kay, L. E. *J. Biomol. NMR* **2003**, *26* (1), 39–48.

(206) Mulder, F. A. A.; Mittermaier, A.; Hon, B.; Dahlquist, F. W.; Kay, L. E. *Nat. Struct. Biol.* **2001**, *8* (11), 932–935.

(207) Bouvignies, G.; Vallurupalli, P.; Hansen, D. F.; Correia, B. E.; Lange, O.; Bah, A.; Vernon, R. M.; Dahlquist, F. W.; Baker, D.; Kay, L. E. *Nature* **2011**, *477* (7362), 111–117.

(208) Wang, K.; Chodera, J. D.; Yang, Y.; Shirts, M. R. *J. Comput. Aided. Mol. Des.* **2013**.

(209) Feher, V. a; Baldwin, E. P.; Dahlquist, F. W. *Nat. Struct. Biol.* **1996**, *3* (6), 516–521.

(210) Merski, M.; Fischer, M.; Balius, T. E.; Eidam, O.; Shoichet, B. K. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112* (16), 5039–5044.

(211) Lopez, C. J.; Yang, Z.; Altenbach, C.; Hubbell, W. L. *Proc. Natl. Acad. Sci.* **2013**, *110* (46), E4306–E4315.

(212) Baldwin, E.; Baase, W. A.; Zhang, X.; Feher, V.; Matthews, B. W. *J Mol Biol* **1998**, *277*, 467–485.

(213) Vallurupalli, P.; Hansen, D. F.; Lundstrom, P.; Kay, L. E. *J Biomol NMR* **2009**, *45*, 45–55.

(214) Morton, A.; Baase, W. A.; Matthews, B. W. *Biochem* **1995**, *34* (27), 8564–8575.

(215) Lerch, M. T.; López, C. J.; Yang, Z.; Kreitman, M. J.; Horwitz, J.; Hubbell, W. L. *Proc. Natl. Acad. Sci.* **2015**, *112* (19), E2437–E2446.

(216) Quillin, M. L.; Breyer, W. A.; Griswold, I. J.; Matthews, B. W. *J Mol Biol* **2000**, *302*, 955–977.

(217) Deng, Y.; Roux, B. *J. Chem. Theory Comput.* **2006**, *2* (5), 1255–1273.

(218) Fuglestad, B.; Gasper, P. M.; Tonelli, M.; McCammon, J. A.; Markwick, P. R. L.; Komives, E. A. *Biophys. J.* **2012**, *103* (1), 79–88.

(219) Han, B.; Liu, Y.; Ginzinger, S. W.; Wishart, D. S. *J. Biomol. NMR* **2011**, *50* (1), 43–57.

(220) Eriksson, A. E.; Baase, W. A.; Wozniak, J. A.; Matthews, B. W. *Nature* **1992**, *355* (6358), 371–373.

(221) Nicholson, H.; Anderson, D. E.; Dao Pin, S.; Matthews, B. W. *Biochemistry* **1991**, *30* (41), 9816–9828.

(222) Vallurupalli, P.; Chakrabarti, N.; Pomes, R.; Kay, L. *Chem. Sci.* **2016**, *3*, 3602–3613.

(223) Xu, J.; Baase, W. A.; Baldwin, E.; Matthews, B. W. *Protein Sci.* **1998**, *7* (1), 158–177.

(224) McIntosh, L. P.; Griffey, R. H.; Muchmore, D. C.; Nielson, C. P.; Redfield, A. G.; Dahlquist, F. W. *Proc. Natl. Acad. Sci.* **1987**, *84* (5), 1244–1248.

(225) Xue, M.; Kitahara, R.; Yoshimura, Y.; Mulder, F. A. A. *Biochem. Biophys. Res. Commun.* **2016**, *478* (3), 1185–1188.

(226) Baase, W. A.; Liu, L.; Tronrud, D. E.; Matthews, B. W. *Protein Science*. 2010, pp 631–641.

(227) Morton, A.; Matthews, B. W. *Biochemistry* **1995**, *34* (27), 8576–8588.

(228) Vallurupalli, P.; Chakrabarti, N.; Pomès, R.; Kay, L. E. *Chem. Sci.* **2016**, *7* (6), 3602–3613.

(229) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. *Proc. Natl. Acad. Sci.* **2001**, *98* (18), 10037–10041.

(230) Dolinsky, T. J.; Nielsen, J. E.; McCammon, J. A.; Baker, N. A. *Nucleic Acids Res.* **2004**, *32* (WEB SERVER ISS.), 1–3.

(231) Miao, Y.; Feher, V. A.; McCammon, J. A. *J. Chem. Theory Comput.* **2015**, *11* (8), 3584–3595.

(232) Durrant, J. D.; Votapka, L.; Sörensen, J.; Amaro, R. E. *J. Chem. Theory Comput.* **2014**, *10* (11), 5047–5056.

(233) Wang, Y.; Martins, J. M.; Lindorff-Larsen, K. *Chem. Sci.* **2017**.

(234) Mulder, F. A. A.; Mittermaier, A.; Hon, B.; Dahlquist, F. W.; Kay, L. E. *Nat. Struct. Biol.* **2001**, *8* (11), 932–935.

(235) Wray, J. W. J. W.; Baase, W. A. W. A.; Lindstrom, J. D. J. D.; Weaver, L. H. L. H.; Poteete, A. R. A. R.; Matthews, B. W. B. W. *J Mol Biol* **1999**, *292* (5), 1111–1120.

(236) Liu, L.; Baase, W. A.; Matthews, B. W. *J. Mol. Biol.* **2009**, *385* (2), 595–605.

(237) Akhterov, M. V.; Choi, Y.; Olsen, T. J.; Sims, P. C.; Iftikhar, M.; Gul, O. T.; Corso, B. L.; Weiss, G. A.; Collins, P. G. *ACS Chem. Biol.* **2015**, *10* (6), 1495–1501.

(238) Eriksson, A. E.; Baase, W. A.; Matthews, B. W. *J. Mol. Bio.* **1993**, *229*, 747–769.

(239) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins: Structure, Function and Genetics*. 2006, pp 712–725.

(240) Mulder, F. A. A.; Skrynnikov, N. R.; Hon, B.; Dahlquist, F. W.; Kay, L. E. *J. Am. Chem. Soc.* **2001**, *123* (5), 967–975.

(241) Skrynnikov, N. R.; Dahlquist, F. W.; Kay, L. E. *J. Am. Chem. Soc.* **2002**, *124* (41),

12352–12360.

(242) Sitkoff, D.; Sharp, K. A.; Honig, B. *J. Phys. Chem.* **1994**, *98* (7), 1978–1988.