

UC Irvine

UC Irvine Electronic Theses and Dissertations

Title

Going on Adventures with Binding Free Energy Calculations:\newline About Pitfalls and Flashlights, Shackles and Broken Chains

Permalink

<https://escholarship.org/uc/item/7273k33z>

Author

Baumann, Hannah M.

Publication Date

2023

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,
IRVINE

Going on Adventures with Binding Free Energy Calculations:
About Pitfalls and Flashlights, Shackles and Broken Chains

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY
in Pharmacological Sciences

by

Hannah M. Baumann

Dissertation Committee:
Professor David L. Mobley, Chair
Professor Andrej Lupták
Professor Ryan Hayes

2023

TABLE OF CONTENTS

	Page
LIST OF FIGURES	v
LIST OF TABLES	vii
ACKNOWLEDGMENTS	viii
VITA	ix
ABSTRACT OF THE DISSERTATION	xiii
1 Introduction	1
1.1 Background	1
1.2 Estimating binding affinities using molecular simulations	3
2 Challenges encountered applying equilibrium and non-equilibrium binding free energy calculations	6
2.1 Introduction	7
2.2 Theory	10
2.2.1 Alchemical free energy calculations are a physically rigorous approach to calculate binding free energies	10
2.2.2 Equilibrium free energy simulations require multiple separate intermediate states	11
2.2.3 Non-equilibrium calculations drive rapid transitions between physical end states	15
2.3 Systems	17
2.3.1 The T4 lysozyme mutant L99A is a simple test system, but still has sampling challenges	18
2.3.2 HSP90 is a challenging, pharmaceutically relevant target	19
2.4 Methods	20
2.4.1 We constructed a thermodynamic cycle to calculate the binding free energy	20
2.4.2 Simulation Details	22
2.4.3 We obtained uncertainty estimates by running independent repeats	25
2.4.4 Overlapping work distributions don't guarantee reliable free energy estimates	26

2.4.5	We compared the EQ and NEQ approaches based on convergence of computed free energy differences where possible	27
2.4.6	We assessed potential sampling limitations	28
2.5	Results	28
2.5.1	Comparing the efficiency of EQ and NEQ on the hydration free energy	29
2.5.2	Toluene binding to T4 lysozyme mutant L99A	29
2.5.3	3-Iodotoluene binding to T4 lysozyme mutant L99A	38
2.5.4	Summary of the T4 lysozyme system	46
2.5.5	Free energy calculations in the HSP90 system	46
2.5.6	Summary of sampling problems in the HSP90 system	57
2.6	Discussion and Conclusions	57
2.6.1	Challenges encountered here in ABFE will probably effect RBFEE as well	58
2.6.2	Strategies to assess sampling problems in alchemical FEC	58
2.6.3	Similarities and differences of the EQ and NEQ approaches in handling slow DOF	60
3	Broadening the scope of binding free energy calculations using a Separated Topologies approach	63
3.1	Introduction	64
3.2	Methods	69
3.2.1	The thermodynamic cycle for SepTop computes RBFEE by running two ABFE calculations in opposite directions	69
3.2.2	We developed heuristics for automatically picking suitable atoms for Boresch-style restraints	71
3.3	Systems	78
3.4	Simulation details	79
3.4.1	Preparation and parametrization of proteins and ligands	81
3.4.2	Setup of SepTop systems	83
3.4.3	Setup for calculations in the solvent	87
3.4.4	Running SepTop in GROMACS	88
3.4.5	Analysis of the results	90
3.5	Results	90
3.5.1	We first tested SepTop on the well studied TYK2 dataset as a sanity check	91
3.5.2	Scaffold hopping transformations with SepTop on ER α systems	93
3.5.3	We tested SepTop on a larger dataset, 16 ligands binding to MALT1	96
3.5.4	Testing SepTop on large scaffold hopping transformations in the BACE1 system	102
3.6	Discussion and Conclusions	108
3.6.1	SepTop performed well on the test systems in this study	109
3.6.2	Ways to test whether a target is within the domain of applicability of the BFE approach	110
3.6.3	Comparing SepTop to other alchemical binding free energy methods may not be advisable for some systems	111

3.6.4	Molecular shape and chemical similarity may be good metrics for planning SepTop calculations	112
3.6.5	Current limitations of SepTop	113
4	Impact of protein conformations on binding free energy calculations in the BACE1 system	114
4.1	Introduction	115
4.2	Simulation details	117
4.3	Results	120
4.3.1	We performed MD simulations of the ligands in three different protein structures	122
4.3.2	We used RBFЕ calculations between ligand poses to identify the preferred binding mode	126
4.3.3	RBFЕ calculations were impacted by the starting protein conformation and the presence of a buried water molecule	128
4.4	Discussion and Conclusion	131
5	Future directions	133
5.1	Analysis tool for identifying sampling problems in binding free energy calculations	133
5.2	Investigate the impact of different choices in enhanced sampling protocols in SepTop	135
5.3	Comparing the computational efficiency of different methods for binding free energy calculations, SepTop, ABFE, and standard RBFЕ	137
5.4	Developing a calculation planning tool for different binding free energy calculation approaches	139
5.5	Evaluating different choices of restraining ligands in the SepTop approach . .	140
5.6	Additional ideas for improvement of the SepTop software package	142
	Bibliography	144
	Appendix A Supporting Information: Challenges Encountered Applying Equilibrium and Non-equilibrium Binding Free Energy Calculations	160
A.1	Hydration free energy calculations	160
A.2	Restraints in binding free energy calculations	162
A.3	Toluene binding to T4 lysozyme L99A	163
A.4	3-Iodotoluene binding to T4 lysozyme L99A	169
A.5	The HSP90 system	176
	Appendix B Supporting Information: Broadening the scope of binding free energy calculations using a Separated Topologies approach	185
	Appendix C Supporting Information: Impact of protein conformations on binding free energy calculations in the BACE1 system	198

LIST OF FIGURES

	Page
2.1 Protocols for the equilibrium (EQ) and nonequilibrium (NEQ) free energy calculations	14
2.2 Toluene and 3-iodotoluene binding to T4 lysozyme mutant L99A	18
2.3 Ligand binding to HSP90	20
2.4 Thermodynamic cycle for computing binding free energies	21
2.5 The bubble-ligand in the HSP90 system	25
2.6 Decoupling of toluene in the binding site	30
2.7 Ile78 side chain reorientation upon binding of toluene to T4 lysozyme L99A .	31
2.8 NEQ work values for toluene binding to T4 lysozyme L99A	34
2.9 Correlation between reverse work values and the dihedral angle of Ile78 and Asp127 for toluene binding to T4 lysozyme L99A	36
2.10 Free energy difference of decoupling 3-iodotoluene in the binding site as a function of sampling time	39
2.11 Impact of the orientation of the Val111 side chain on the free energy difference for 3-iodotoluene binding to T4 lysozyme L99A	41
2.12 Correlation between $dH/d\lambda$ values and the dihedral angle of Val111 for 3-iodotoluene binding to T4 lysozyme L99A	44
2.13 Free energy difference of decoupling the HSP90 ligand in the binding site as a function of total simulation time	47
2.14 Trapped waters in the HSP90 system	49
2.15 Correlation between $dH/d\lambda$ values and the number of water molecules in the binding site fo HSP90	49
2.16 Correlation between NEQ work values and the number of water molecules in the binding site of HSP90	50
2.17 Binding mode flip of the weakly interacting ligand in HSP90	51
2.18 Correlation between $dH/d\lambda$ values and the RMSD of the HSP90 ligand . . .	52
2.19 Comparison of different protocols of decoupling the ligand in the binding site of HSP90	55
3.1 Thermodynamic cycle for SepTop for computing the free energy difference between two ligands in the binding site	70
3.2 Choices made in selecting suitable ligand atoms for the Boresch restraints . .	73
3.3 Choices made in selecting suitable protein atoms for the Boresch restraints .	76
3.4 Three scaffold series in the BACE1 system	80

3.5	Perturbation cycle in the TYK2 system	92
3.6	Ligand cycle in the ER α system	94
3.7	Different choices made in ER alpha preparation	95
3.8	Correlation between predicted and experimental binding free energies for 10 MALT1 inhibitors	97
3.9	Convergence issues in the MALT1 system	99
3.10	Correlation between calculated (NES Orion) and experimental binding free energies for 10 MALT1 inhibitors	103
3.11	Two poses of a ligand from the amide series binding to BACE1	105
3.12	Correlation between calculated and experimental binding free energies for 18 BACE1 inhibitors	107
4.1	Different protein conformations and inhibitors in the BACE1 system	118
4.2	Buried water molecule in the P2' pocket in the BACE1 system	122
4.3	Results from MD simulations of BACE1 ligand lig_CAT_4b	124
4.4	Rotatable bonds of BACE1 ligand lig_p2_36 during MD simulations	125
4.5	RBFE calculations between different poses of lig_p2_28 in different starting protein structures (4DJW, 3INF, 3IN4)	126
4.6	RBFE results from transformations between the four BACE1 inhibitors	129

LIST OF TABLES

	Page
2.1 Binding free energy ΔG° for the toluene/T4 lysozyme system	31

ACKNOWLEDGMENTS

First and foremost, I thank my PhD advisor David Mobley for all his support and guidance throughout the last 5 years. Not only did David always give great advice and suggestions in my research, he also always had my best interest in mind, always had my back and did everything in his power to help me get to where I wanted to get, proposing projects that I was excited about, encouraging me to present my research and helping me network with people in our field, supporting me in my job search process, and being a role model and encouragement for prioritizing the greater purposes in life.

I would also like to acknowledge additional mentors I had during graduate school. I am grateful to our collaborators at Pfizer, most importantly Eric Dybeck, Alan Mathiowetz and Christopher McClendon for all their advice and guidance, and I am very thankful for all the interest they showed in my research. I thank Christopher Bayly for all the scientific discussions we have had and for his ideas that helped advance my research. I am thankful to Daniel Kuhn for getting me excited about the field of computational chemistry, for encouraging me to come to California for grad school, and for giving me career advice.

I want to thank everyone in the MobleyLab, I'm very grateful for the community we have had, celebrating each other's successes and building each other up in difficult times. I appreciate and I am thankful for my friends who have supported me and were there for me through the highs and lows and helped me get through some tough times. I especially would like to thank Olivia Morgan, Elena Schmidt, Katharina Pfeuffer, Damaris Dorksen, Lisa Chen, Victoria Lim, Jessica Pazienza, Chris Zhang and Oanh Tran. I am grateful to my family for supporting my decision to come to the US and for being there for me every step of the way, even from afar. And I would like to thank my church family at Saddleback Irvine South for becoming a family away from home. I especially thank my Pastors, my LifeGroup, the Chen family and Olivia for giving me a fun and safe space to be myself and feel at home.

Finally, I acknowledge financial support from the National Institute of Health and from Pfizer that has made it possible for me to focus on my research.

Chapter 2 of this dissertation is reprinted with permission from the American Chemical Society regarding the published work: Baumann, H. M.; Gapsys, V.; de Groot, B.; Mobley, D. L., "Challenges Encountered Applying Equilibrium and Nonequilibrium Binding Free Energy Calculations" *J. Phys. Chem. B.*, 2021, 125, 17, 4241–4261. I acknowledge and thank my co-authors Vytautas Gapsys, Bert de Groot, and David Mobley.

VITA

Hannah M. Baumann

Education

Doctor of Philosophy in Pharmaceutical Sciences	2021
University of California, Irvine	<i>Irvine, California</i>
Pharmacist degree	2018
Johannes Gutenberg-University	<i>Mainz, Germany</i>

Professional Experience

Graduate Researcher	2018–2023
University of California, Irvine	<i>Irvine, California</i>
Pharmacist Intern	Nov 2017–April 2018
Hirsch Apotheke	<i>Mainz, Germany</i>
Computational Chemistry Intern	May 2017–Oct 2017
Merck KGaA	<i>Darmstadt, Germany</i>
Organic Chemistry Intern	Feb 2016–March 2016
Universitat Jaume I	<i>Castelló, Spain</i>

Teaching Experience

Graduate Teaching Assistant	Spring 2019
Pharmacotherapy, University of California, Irvine	<i>Irvine, California</i>

Publications

1. **Baumann, H. M.**; Mobley, D. L. “Impact of protein conformations on binding free energy calculations in the BACE1 system.” *In preparation*
2. **Baumann, H. M.**; Dybeck, E.; McClendon, C. L.; Pickard IV, F. C.; Gapsys, V.; Peérez-Benito, L.; Hahn, D. F.; Tresadern, G.; Mathiowetz, A. M.; Mobley, D. L. “Broadening the scope of binding free energy calculations using a Separated Topologies approach.” *Under review*
3. Ge, Y.; **Baumann, H. M.**; Mobley, D. L. “Absolute binding free energy calculations for buried water molecules.” *J. Chem. Theory Comput.*, **2022**, 18, 11, 6482–6499.
4. Khalak, Y. ; Tresadern, G.; Aldeghi, M.; **Baumann, H. M.**; Mobley, D. L.; de Groot, B.; Gapsys, V. “Alchemical absolute protein–ligand binding free energies for drug design.” *Chem. Sci.*, **2021**, 12, 13958-13971.
5. **Baumann, H. M.**; Gapsys, V.; de Groot, B.; Mobley, D. L. “Challenges Encountered Applying Equilibrium and Nonequilibrium Binding Free Energy Calculations.” *J. Phys. Chem. B*, **2021**, 125, 17, 4241–4261.
6. Schindler, C. E. M.; **Baumann, H. M.**; Blum, A.; Bose, D.; Buchstaller, H.-P.; Burgdorf, L.; Cappel, D.; Chekler, E.; Czodrowski, P.; Dorsch, D.; Eguida, M. K. I.; Follows, B.; Fuchß, T.; Gradler, U.; Gunera, J.; Johnson, T.; Jorand Lebrun, C.; Karra, S.; Klein, M.; Knehans, T.; Koetzner, L.; Krier, M.; Leiendecker, M.; Leuthner, B.; Li, L.; Mochalkin, I.; Musil, D.; Neagu, C.; Rippmann, F.; Schiemann, K.; Schulz, R.; Steinbrecher, T.; Tanzer, E.-M.; Unzue Lopez, A.; Viacava Follis, A.; Wegener, A.; Kuhn, D. “Large-Scale Assessment of Binding Free Energy Calculations in Active Drug Discovery Projects.” *J. Chem. Inf. Model.*, **2020**, 60, 11, 5457–5474.

Awards and Honors

Excellence in Research Award	2021
School of Pharmacy and Pharmaceutical Sciences, UC Irvine	
Penny J. Gilmer Grant for Women Graduate Students and Postdocs	2020
OpenEye Scientific Software	
Vertex Scholar of the year	2020
Vertex Pharmaceuticals	

**Penny J. Gilmer Grant for Women Graduate Students
and Postdocs**

2019

OpenEye Scientific Software

Erasmus Scholarship

2016

Johannes Gutenberg-University, Mainz

Invited Talks

1. “Broadening the scope of binding free energy calculations using a Separated Topologies approach”, Oct 2022, Vertex Day, UC Irvine
2. “Separated Topologies: Broadening the scope of binding free energy calculations”, Oct 2022, Nimbus Therapeutics (virtual)
3. “Broadening the scope and improving the accuracy of potency prediction in drug discovery”, Aug 2022, Merck KGaA Darmstadt
4. “Broadening the scope and improving the accuracy of potency prediction in drug discovery”, Aug 2022, Bayer, Berlin
5. “Broadening the scope and improving the accuracy of potency prediction in drug discovery”, July 2022, Boehringer Ingelheim, Biberach
6. “Broadening the scope and improving the accuracy of potency prediction in drug discovery”, July 2022, Janssen, Beerse
7. “Separated Topologies: Unchained from similarity”, March 2022, OpenEye CUP, Santa Fe, NM
8. “Binding free energy calculations: Sampling challenges, nonequilibrium switching, and alternate approaches”, June 2021, 2021 Virtual Workshop on Free Energy Methods in Drug Design
9. “Binding free energy calculations: Sampling challenges, nonequilibrium switching, and alternate approaches”, June 2021, OpenEye Scientific Software (virtual)
10. “Absolute binding free energy calculations using a non-equilibrium alchemical approach”, Oct 2022, Max-Plank Institute Gottingen

Contributed Presentations

1. “Broadening the scope and improving the accuracy of potency prediction in drug discovery”, talk at Gordon Research Seminar Computational Chemistry, Castelldefels, Spain, July 2022
2. “Broadening the scope and improving the accuracy of potency prediction in drug discovery”, poster at Gordon Research Conference Computational Chemistry, Castelldefels, Spain, July 2022
3. “Broadening the scope and improving the accuracy of potency prediction in drug discovery”, poster at ACS virtual meeting, August 2021
4. “Applying non-equilibrium and equilibrium free energy calculation methods on HSP90 systems”, poster at OpenEye CUP, March 2020
5. “Evaluation of alchemical non-equilibrium free energy calculations”, poster at ACS annual meeting San Diego, August 2019
6. “Evaluation of alchemical non-equilibrium free energy calculations”, poster at OpenEye CUP, March 2019
7. “Synthesis of a new cysteine protease inhibitor”, talk, University of Mainz, July 2016

ABSTRACT OF THE DISSERTATION

Going on Adventures with Binding Free Energy Calculations:
About Pitfalls and Flashlights, Shackles and Broken Chains

By

Hannah M. Baumann

Doctor of Philosophy in Pharmacological Sciences

University of California, Irvine, 2023

Professor David L. Mobley, Chair

Binding free energy calculations estimate ligand binding affinities in a physically rigorous manner. Therefore, these calculations play an important role in drug discovery, estimating ligand potency of compounds before made for experimental testing to help guide discovery efforts. Different approaches to calculating binding free energies have been developed, each with their benefits and weaknesses. Absolute Binding Free Energy (ABFE) calculations calculate the binding free energy of a single ligand to its protein. ABFE calculations can be computationally expensive as they require sampling (potentially slow) conformational changes that happen upon ligand binding. This has thus far limited widespread use of the method in the pharmaceutical industry. We investigated the performance of two different approaches to calculating absolute binding free energies, a non-equilibrium and an equilibrium approach, on two different test systems. We highlighted common pitfalls or sampling problems of these calculations and identified ways on how to shed light into and assess slow degrees of freedom that impact calculated binding free energies. Relative Binding Free Energy (RBFE) calculations are an alternative approach and calculate the difference in binding free energy between two ligands. RBFE calculations have some benefits over ABFE calculations, namely that slow conformational changes of the system upon ligand binding do not need to be sampled since one of the ligands always occupies the binding site. In common

RBFE approaches, one ligand is mutated into the other ligand, and the application of the method is restricted to the comparison of structurally related ligands. We implemented and tested an alternative approach for RBFE calculations, Separated Topologies, that allows for larger ligand transformations compared to standard RBFE methods, breaking the chains of ligand similarity restrictions in RBFE. Finally, we investigated the impact of protein conformational sampling on calculated binding free energies. We find there are often many different metastable protein conformations. Falling into the pit of one of these by failing to sample the other relevant states leads to inaccurate and biased results, highlighting the importance of adequate conformational sampling to obtaining converged calculated binding free energies.

Chapter 1

Introduction

1.1 Background

Developing a pharmaceutical drug and testing it in clinical trials is a long and very costly process, and the probability of failure is high. Each new drug costs about \$1.4-2.6 billion, from which over \$400 million are spent in preclinical studies.[85, 33] Computer-aided drug design serves to help accelerate early stage drug discovery, both in academia and industry, and to avoid labor intensive experiments of ligands with undesired properties.

Drugs are commonly molecules that bind to a protein in the body and either block or enhance its activity. A molecule can bind to a cavity of a protein by interacting with atoms in the binding site. A hypothesis, introduced by Emil Fischer in 1894[42], compares this process with a key (drug) that can close or open – meaning inhibit or enhance – a lock (receptor/enzyme). The key has to have two characteristics. First, it has to perfectly fit in the lock and be able to smoothly open it. Secondly, it has to be specific for this lock and should not be able to nonspecifically fit into any other lock. A drug needs to be similar. It is important to find a molecule that strongly binds to a protein, and at the same time

binds specifically to this protein, preventing side effects due to nonspecific binding to other proteins. In this model, a molecule is made specific and potent by giving it a shape that fits well into the cavity of the receptor and makes specific interactions with atoms in the binding site. However, the lock-and-key hypothesis misses some important concepts, such as that proteins can undergo conformational changes upon ligand binding (induced fit) and that some ligands bind to allosteric sites instead of binding to the active site.

Discovering such molecules proves extremely difficult, in part because the space of molecules which could be created is vast, so for many years computational methods have helped enhance the process and complement experiments. Physics-based methods, for instance, can model important biological processes, such as protein-ligand binding and permeation through biomembranes. These are highly complex processes and thus only approximations make these methods computationally tractable. Different techniques make different trade-offs between the accuracy and the speed. Therefore, the extent of the approximations in different methods determines the value of each method in different stages of the process. Some approaches scan huge chemical libraries, such as the Enamine REAL Space with 36 billion molecules[38], in order to explore the protein cavity and find putative binders. Fast scoring methods, such as 2D similarity searches[14] or three-dimensional tools like molecular docking[132], are used to filter this space. The output of these methods often does not correlate well with experiment. However, at this stage, even weakly-binding compounds provide a good starting point for further development making these methods valuable for hit finding. More accurate – but slower – methods can be used later in the process. In a design cycle, these initial weakly-binding compounds are improved iteratively until the key perfectly fits into its lock. Free energy methods, like those we focus on here, are more accurate and thus help guide this process. The computational expense of these methods has limited their broad application in the past. However, recent advances in computing technology, like graphical processing units, and low-cost parallel computing, enable the pharmaceutical industry to routinely and successfully apply these methods in drug discovery projects.[98, 133, 22, 74, 81, 24, 82] These

techniques prove especially valuable to test compounds before they are created experimentally, in order to avoid wasted time and effort. These more rigorous methods are less valuable when the compounds being considered are already available for assay, since in some cases the ease and cost of experiments may be less than that of computation unless synthesis is required.

1.2 Estimating binding affinities using molecular simulations

Drugs bind to receptors because of favorable interactions between the atoms of the receptor and ligand, as well as some other effects, such as hydrophobic interactions and entropic contributions. The strength of these protein-ligand interactions is measured by the binding affinity or the related quantity, the binding free energy (see Equation 2.1). The standard binding free energy is defined as the difference between the free energy of the ligand in solution at a standard concentration of 1 M and the free energy of the ligand bound to the protein. The more negative the binding free energy, the more thermodynamically stable the protein-ligand complex and the stronger the interaction. Since experimental measurements require that the ligand be synthesized first, computational estimates of binding free energies can help prioritize the synthesis towards the most promising molecules.

Molecular dynamics (MD) calculations simulate the motion of protein-ligand systems and give insight into the motions of biomolecules at atomic resolution. For instance, ligand binding and unbinding can be observed in MD simulations[135] and the binding free energy can in principle be approximated from such a simulation if multiple binding and unbinding events are observed. However, MD simulations are computationally demanding, e.g. simulating a protein-ligand complex for 100 ns can take approximately one day. Therefore, directly

simulating binding and unbinding events, which are on the millisecond time scale, becomes prohibitively expensive and is not practical in drug discovery projects. Instead, principles of statistical mechanics serve to estimate the binding affinity of the ligand in a physically rigorous manner without having to simulate binding events directly (see Equations 2.3, 2.4, 2.6, 2.7).

Different approaches to calculating binding free energies are introduced in Chapter 2 with details on the methods being explained in Section 2.2. In that work, we compared equilibrium and non-equilibrium absolute binding free energy calculations on two systems, the T4-lysozyme mutant L99A and the molecular chaperone HSP90. We identified sampling problems that led to slow convergence of the free energy estimate in calculations with both equilibrium and non-equilibrium approaches, and we discussed ways to identify sampling problems in molecular simulations.

Some of the slow degrees of freedom that can lead to sampling problems in absolute binding free energy calculations, such as slow displacement of water molecules upon ligand binding in a solvent exposed binding site, can be less severe in relative binding free energy (RBF) calculations, where one ligand always occupies the binding site. However, in common RBF approaches, ligands have to share a common scaffold, limiting the method's domain of applicability. In Chapter 3 we introduce an alternate approach for relative binding free energy calculations, called Separated Topologies, and demonstrate on four diverse and pharmaceutically relevant test systems that the accuracy of the approach is sufficiently high to rank order ligands with an accuracy comparable to traditional RBF calculations, while maintaining the additional flexibility of the Separated Topologies approach.

Finally, we investigate the impact of the starting protein conformation on calculated binding free energies in a beta-secretase test system in Chapter 4. If slow degrees of freedom and transitions between different metastable states occur on a timescale far beyond the duration of the simulation, the binding free energy estimate can seem converged. However, this apparent

convergence can be false, because other important metastable states were never visited during the simulation. Here, we started separate RBFE calculations of beta-secretase inhibitors using different crystal structures as starting structures. We found that the starting protein conformation had an impact on the free energy differences, highlighting the importance of protein conformational sampling in free energy calculations.

Chapter 2

Challenges encountered applying equilibrium and non-equilibrium binding free energy calculations

Binding free energy calculations have become increasingly valuable to drive decision making in drug discovery projects. However, among other issues, inadequate sampling can reduce accuracy, limiting the value of the technique. In this chapter we apply absolute binding free energy calculations to ligands binding to T4 lysozyme L99A and HSP90 using equilibrium and non-equilibrium approaches. We highlight sampling problems encountered in these systems, such as slow side chain rearrangements and slow changes of water placement upon ligand binding. These same types of challenges are likely to show up in other protein-ligand systems as well and we propose some strategies to diagnose and test for such problems in alchemical free energy calculations. We also explore similarities and differences in how the equilibrium and the non-equilibrium approaches handle these problems. Our results show the large amount of work still to be done to make free energy calculations robust and reliable and provide insight for future research in this area.

2.1 Introduction

In drug discovery, computational methods help navigate through a vast potential chemical space. There is often a trade-off between accuracy and speed of the methods which determines the value of each technique at different stages of the discovery process. A common goal in drug discovery is to search for molecules which bind to proteins with therapeutic potential.

In order to explore a binding site of interest and find putative binders, large chemical libraries like Enamine, with 1.36 billion molecules[38], can be scanned[52]. Fast and less accurate methods, such as ligand-based approaches, molecular docking and physicochemical property prediction, are used to filter this large space. Promising compounds can then be purchased, tested and active compounds improved iteratively.

In lead optimization, experiments may require difficult synthesis and thus be costly and slow. Therefore, during optimization, more accurate, but also more computationally expensive methods, such as free energy calculations, are used to prioritize the synthesis of compounds based on the *in silico* predictions. Although all these methods are routinely applied in drug discovery projects, improvements in their accuracy are still needed to increase their impact[115, 136, 24, 133].

The binding free energy, ΔG° , is defined as the difference between the free energy of the ligand in solution at a standard reference concentration, and the free energy of the ligand bound to the protein. This property can be calculated in a variety of ways, but two – which we consider here – are becoming relatively more common. In the so-called “absolute” approach, the ΔG° is calculated for each ligand independently. Alternatively, the “relative” approach determines the difference in binding free energy, $\Delta\Delta G^\circ$, between the binding of two ligands. In the latter case, commonly only atoms which differ between the two molecules are transformed, and the absolute binding free energy can be obtained if the binding free

energy of one of the ligands is known. This relative approach, then, is most suitable for comparing or ranking binding of related ligands.

Absolute binding free energy (ABFE) calculations have some benefits over relative ones (RBFEE)[24, 104, 4], but are thought to require more sampling because of the generally larger transformations involved compared to RBFEE. In ABFE calculations, ligands of interest can be very chemically diverse since they need not share a common scaffold. In addition, different ligands need not share a common binding mode. The method can even be used for selectivity predictions, predicting binding of a ligand to multiple proteins[4]. However, ABFE calculations are thought to require more sampling and simulations may take a long time to converge which can make them computationally expensive[24]. ABFE calculations in some systems likely require additional sampling of slow protein motions relative to RBFEE calculations, such as changes in protein conformation which occur on ligand binding (e.g. HIV protease flap opening/closing[63], kinase activation loop[29], etc.).

Alchemical free energy calculations employ an unphysical path that connects two physical end states in order to obtain free energy differences[102, 157, 141, 99, 8, 82, 117]. Let us consider for example the process of turning the interactions of a ligand off in the binding site. The initial state in this case is the fully interacting ligand bound to the protein and the final state is the decoupled ligand, not interacting with its surroundings, and a separate protein and solvent box. Here the unphysical path describes how the electrostatic and steric interactions between the ligand and the protein and surrounding solvent molecules are gradually turned off. The unphysical intermediates ensure sufficient overlap of phase space as the system is modified, allowing for calculation of accurate free energies. The Hamiltonian of the system is coupled to a parameter λ which controls the progress between the physical end states. λ can have discrete or continuous values between 0 (initial state) and 1 (final state) and the intermediate states are often referred to as λ windows. Different alchemical approaches have been developed, each having its strengths and weaknesses.

In equilibrium free energy calculations (EQ) a separate molecular dynamics (MD) simulation is performed for each of a set of discrete λ windows corresponding to different states of the system (Figure 2.1). This method often requires long simulation times since every state has to first reach equilibrium before data can be used for the analysis. If this is not achieved – and it can be difficult to ensure that it has been achieved – the approach can suffer from systematic errors[25].

In contrast, with a non-equilibrium approach (NEQ), only the physical end states (i.e. the interacting state and the noninteracting state) have to be sampled at equilibrium while the alchemical path is explored in fast switching transitions, as the ligand is switched from interacting to non-interacting (or vice versa) in a single simulation. The coupling parameter λ is continuous in this case. Multiple trials are performed, and the exponential average over the non-equilibrium work necessary to transform the ligand is related to the free energy difference of the process by the Jarzynski equality[65] (or, better if transformations in both directions are possible, the Crooks formalism[25, 26]).

Another potential benefit of the NEQ approach is that it is highly parallelizable which can reduce wall clock time when distributed computing is available. In NEQ calculations, end state simulations can also benefit from enhanced sampling techniques[93] to overcome energy barriers. Intuitively such strategies seem likely to be beneficial. However, recent literature[76] using partial replica exchange MD (an enhanced sampling technique involving exchanges between replicas with modified Hamiltonians) showed no effect of the enhanced sampling. Work in preparation from Gapsys and de Groot also suggests that in some cases, enhancing the sampling of the apo/holo states can drive the system away from the starting crystallographic pose by accumulating force field artifacts and equilibrating in a new free energy minimum which may differ from the experimental one.

While most literature applications of free energy calculations have used EQ or NEQ approaches, a third possibility is expanded ensemble simulations and related approaches, such

as λ dynamics[79]. Here, the parameter λ is a dynamic variable which is propagated throughout a single simulation. These calculations are not our focus here, but provide an interesting alternative.

In this chapter, we test the non-equilibrium approach for absolute binding free energy calculations and compare it with equilibrium free energy calculations. We highlight challenges encountered applying the two methods and explore similarities and differences in handling these issues. Since the challenges we identify are likely to show up in other protein-ligand systems as well, we propose some strategies to assess sampling problems in alchemical free energy calculations (FEC). Although we find ways to work around some of the problems, our results highlight the large amount of work still to be done to make these calculations robust and to reliably detect problems in an automated way. Most literature studies report only successes, and calculations generally seem to work, though industry reports have highlighted the need for more careful analysis of failures[136]. Here, we report on a relatively large number of “failures” with common protocols and highlight how they fail and why. We provide insights and directions for future method development and research in this area.

2.2 Theory

2.2.1 Alchemical free energy calculations are a physically rigorous approach to calculate binding free energies

The strength of protein-ligand interactions can be measured by the binding affinity or the related quantity, the binding free energy, given by:

$$\Delta G^\circ = -k_B T \ln(c^\circ K_B) = -k_B T \ln\left(c^\circ \frac{[PL]}{[P][L]}\right) \quad (2.1)$$

where k_B is the Boltzmann constant, T the temperature, K_B the binding affinity, c° the standard state concentration of 1 mol/l and $[PL]$, $[P]$, $[L]$ the concentrations of the protein-ligand complex, the protein and the ligand, respectively.

Binding free energy calculations are based on the relationship between the free energy and the configurational integral (partition function)[50, 145]. This partition function cannot be calculated explicitly for complex systems, but can be approximated by an ensemble of conformations from the Boltzmann distribution. In practice, molecular dynamics or Monte Carlo simulations are used to sample the configurational space. Several approaches serve to analyze the resulting data and obtain the change in free energy[165].

2.2.2 Equilibrium free energy simulations require multiple separate intermediate states

For the EQ approach, the Bennett Acceptance Ratio (BAR)[16] and its variant the Multistate BAR (MBAR)[140] are the most efficient estimators.

As input, BAR and MBAR take potential energy differences between adjacent alchemical states. These potential energy differences can be written (in “reduced”/dimensionless form) as Δu_{ij} where i and j denote the thermodynamic states in question. Particularly, these states correspond to different intermediate states or λ windows per the Introduction.

The potential energy differences Δu_{ij} are evaluated for samples taken from a single simulation and assess the work which would be required to move the system from one state (or one Hamiltonian) to the other. A simulation in state i can be said to have coordinates x_i , and from that simulation we evaluate $\Delta u_{ij}(x_i)$, the energy difference to alter the energy function to that corresponding to state j . Likewise, we also use snapshots from the simulation run in state j to compute $\Delta u_{ij}(x_j)$. Here, $\Delta u_{ij}(x) = u_j(x) - u_i(x)$ [165].

In this notation, the potential energy $u_i(x)$ refers to a dimensionless reduced potential function that includes the pressure-volume work; this allows easier accommodation of the equations to different choices of ensemble; here, calculations are performed in the isothermal-isobaric ensemble[140]:

$$u_i(x) = \beta(U_i(x) + p_i V(x)) \quad (2.2)$$

where x is the configuration of the system, $V(x)$ the volume and $U_i(x)$ the potential energy function.

Overall in BAR, the free energy difference is obtained by numerically solving the following equation for ΔG :

$$\sum_{i=1}^{n_i} \frac{1}{1 + e^{(M + \Delta u_{ij}(x_i) - \Delta G)}} = \sum_{j=1}^{n_j} \frac{1}{1 + e^{-(M + \Delta u_{ij}(x_j) - \Delta G)}} \quad (2.3)$$

with $M = \ln \frac{n_i}{n_j}$ and n_i and n_j being the number of samples at state i and j and $\beta = 1/k_B T$. MBAR generalizes this expression to multiple thermodynamic states rather than pairs of states as in BAR.

The sum of all free energy differences ΔG between the λ windows gives then the total free energy difference of the process.

In the case of BAR, only potential energy differences of adjacent states are computed while in the case of MBAR all states are considered and analyzed. MBAR provides a more generalized version of BAR which uses all available data to compute free energy estimates, rather than relying on information only from adjacent thermodynamic states.

It is important to mention that these estimators rely on phase space overlap between adjacent λ windows to obtain reliable free energy estimates (Figure 2.1).

Thermodynamic integration (TI) provides another common approach for EQ calculations[77]. Here, the free energy difference between two states is defined as

$$\Delta G = \int_0^1 \left\langle \frac{\delta U(\lambda)}{\delta \lambda} \right\rangle_{\lambda} d\lambda. \tag{2.4}$$

The alchemical path is sampled at a set of discrete λ values. Therefore, this integral has to be solved numerically which introduces a bias especially when there is (strong) curvature along the alchemical pathway[142].

Since sufficient sampling in all λ states is necessary to obtain reliable free energy estimates for both estimators, enhanced sampling techniques, such as Hamiltonian Replica Exchange (HREX)[160, 146, 161], are often used to improve the sampling of configuration space. Here, λ windows are run in parallel, so that the simulations can swap configurations of the system between the states. Energy barriers of the potential energy surface might be crossed more often and earlier since the energy landscape differs among the states of the alchemical pathway. The Metropolis acceptance criterion ensures that simulations in all states still converge towards the equilibrium distribution[146].

As mentioned in the introduction, all λ windows need to be simulated at equilibrium in EQ calculations, and some of these can require sampling unphysical states which might have even slower correlation times than the physical states[114, 147, 120], necessitating very long sampling times despite the use of enhanced sampling techniques.

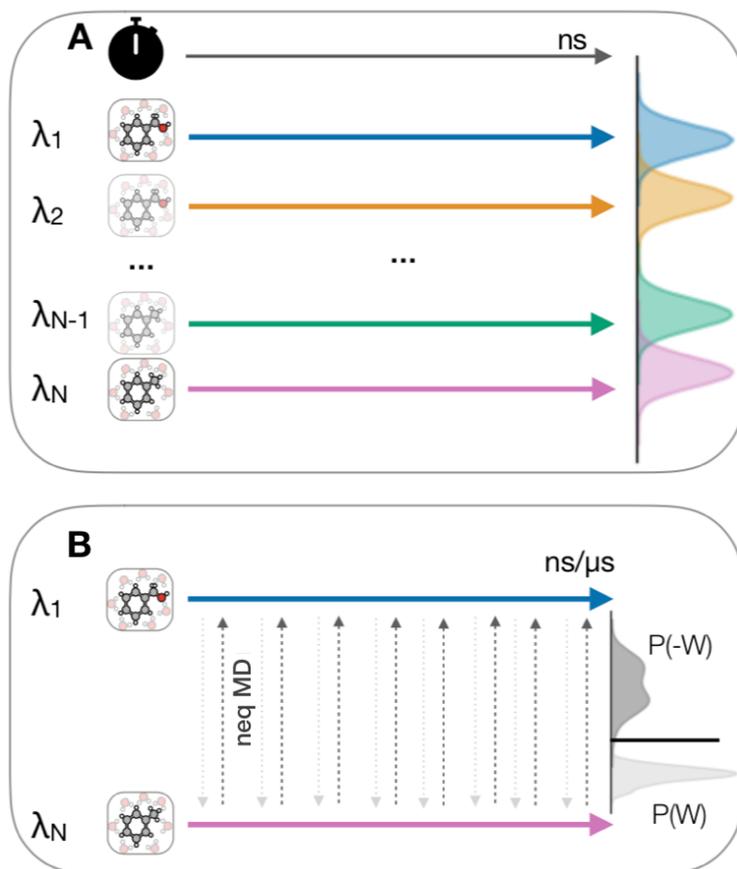


Figure 2.1: Protocols for the equilibrium (EQ) and nonequilibrium (NEQ) free energy calculations. (a) EQ sampling scheme. A separate molecular dynamics (MD) simulation (colored arrows) is performed for each of a set of discrete λ windows corresponding to different states of the system ($\lambda_1 - \lambda_N$). (b) NEQ approach, where two equilibrium simulations at the physical end states are run (λ_1 and λ_N , blue and pink arrow). The alchemical path is explored in fast switching transitions in a single simulation (grey arrows). In the bidirectional approach, many trials are performed in both directions and the non-equilibrium work distributions of the forward and reverse direction ($P(W)$ and $P(-W)$) are directly related to the free energy difference between the states (thick horizontal line between the distributions) by the Crooks fluctuation theorem. Figure was adapted from Mey et al.[99] under the CC-BY 4.0 license.

2.2.3 Non-equilibrium calculations drive rapid transitions between physical end states

In contrast, non-equilibrium FEC sample only the physical end states at equilibrium while the alchemical path is explored in fast (non-equilibrium) transitions. After generating an ensemble of equilibrium conformations in the initial and final state, MD simulations are performed which rapidly switch the ligand from interacting to non-interacting and vice versa (Figure 2.1). By changing the parameter λ , the system is forced to quickly change to the other end state without reaching equilibrium at any intermediate state. Consequently, there is dissipated work along the path and the average work exceeds the free energy difference according to the second law of thermodynamics and the maximum work theorem

$$\langle W \rangle \geq \Delta G, \tag{2.5}$$

where W is the reduced work that includes the pressure-volume work similarly as described for the potential energy above. Jarzynski showed that this inequality becomes an equality when taking the exponential average of the irreversible work done in taking the system from one state to another[66, 65]:

$$\langle e^{-\beta W} \rangle = e^{-\beta \Delta G}. \tag{2.6}$$

Since this average over the work distribution is exponential, the tails of the distribution carry the most statistical weight[2]. The samples in these tails are rare, however. This introduces a bias when the tails are not sampled adequately, meaning that this estimator frequently performs poorly.

Convergence is in general expected to be worse in the “deletion” direction than the “insertion” direction[67], unless the particular choice of restraints alters this balance. This may

seem counter-intuitive since dissipation can be larger when inserting a ghost ligand with the protein in the apo conformation than when deleting the coupled ligand in the holo state (Figure 2.8, 2.11a). While dissipation is important, convergence is partly determined by how frequent important realizations are and Jarzynski showed[67] that important realizations are particularly infrequent in the deletion direction, so that approaches like Widom insertion are practical, though not Widom deletion. Wu and Kofke showed[162] that transitions in the direction that increases the entropy will not converge with a unidirectional estimator. However, as mentioned above, restraints can alter this balance.

In contrast to the unidirectional Jarzynski equality, the Crooks Fluctuation Theorem (CFT)[25, 26, 27] takes transformations in both directions (forward and reverse) into account. This bidirectional approach was shown to converge faster than exponential averaging[128]. It relates the probability distribution of forward and reverse work values with the free energy difference

$$\frac{P_F(W)}{P_R(-W)} = e^{\beta(W-\Delta G)}. \quad (2.7)$$

If only unidirectional transformations are performed, the CFT reduces to the Jarzynski equality. The CFT can be solved for ΔG with the Bennett acceptance ratio method mentioned above (Sec. 2.2.2). Here, the summation is performed over the number of simulations in the forward (n_f) and reverse (n_r) direction:

$$\sum_{i=1}^{n_f} \frac{1}{1 + \frac{n_f}{n_r} e^{\beta(W_f - \Delta G)}} = \sum_{j=1}^{n_r} \frac{1}{1 + \frac{n_r}{n_f} e^{-\beta(W_r - \Delta G)}} \quad (2.8)$$

The work performed on the system is obtained by accumulating the energy changes as the coupling parameter is changed during the course of the switching transition[31].

$$W = \int_0^1 \frac{\delta H(\lambda)}{\delta \lambda} d\lambda. \quad (2.9)$$

Other uni- and bi-directional estimators are based on a Gaussian approximation of the probability distribution. However, these introduce significant errors even with large amounts of sampling, since the underlying work distributions are not always Gaussian, as demonstrated in the recent SAMPL6 SAMPLing challenge[128].

Non-equilibrium switching transition data must be collected from independent and uncorrelated trials in order for the average to converge to the true free energy. This is beneficial, in that switching transitions can be run in parallel and thus deploy extremely well in distributed computing environments. This can reduce the wall-clock time of these calculations relative to the EQ approach, at least when a large number of processors is available.

In order to obtain an unbiased estimate of the free energy change, the work distributions should overlap sufficiently (Figure 2.1). In the case of poor overlap, the CFT result becomes the mean of forward and reverse exponential averages individually computed by the Jarzynski equality. As mentioned above, the exponential average is highly sensitive to the tails of the distribution and therefore, depending on the amount of sampling, a biased estimate. Thus, when overlap is poor, computed free energies are likely incorrect.

2.3 Systems

We compared the performance of the EQ and NEQ methods on two different protein targets, examining two ligands binding to the T4 lysozyme mutant L99A and a ligand binding to HSP90. Each system has its own challenges which we will introduce in this section.

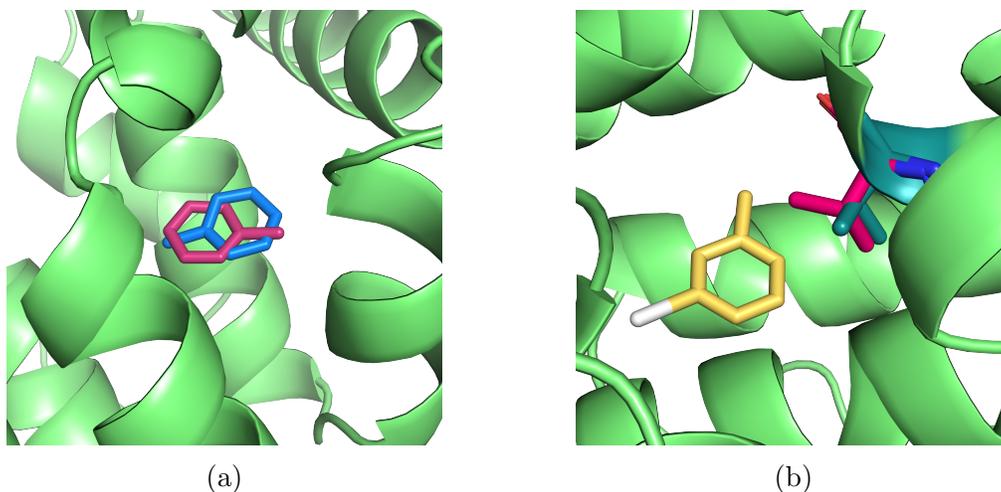


Figure 2.2: Toluene and 3-iodotoluene binding to T4 lysozyme mutant L99A. (a) Two binding poses of toluene in T4 lysozyme L99A (green). The crystallographic binding pose (pose I) is shown in blue and the alternative binding mode (pose II) is shown in magenta sticks. (b) Val111 side chain reorientation upon 3-iodotoluene binding. In the apo state, the Val111 side chain (magenta) points inside the binding pocket. It reorients upon binding (dark green sticks) to make room for 3-iodotoluene (yellow). PDB ID: 4W53

2.3.1 The T4 lysozyme mutant L99A is a simple test system, but still has sampling challenges

The T4 lysozyme mutant L99A is a simple model binding site that has been extensively studied both experimentally and computationally[39, 57, 32, 103, 102, 105, 69]. The leucine to alanine mutation leads to the formation of an apolar cavity that binds small molecules. We chose this system for several reasons. First, the dry binding site and the simple, mostly rigid ligands make the system appropriate for the comparison of methods, as sampling times are not excessively long. Secondly, since it is a well studied binding site, it is possible to compare the results with other methods. Lastly, it poses some interesting challenges for alchemical FEC. Toluene is known to bind with two different binding modes (Figure 2.2a). If the population or free energy of each pose is assessed separately, the crystallographic pose should be the more favorable.

Other ligands, such as 3-iodotoluene and *p*-xylene, induce slow side chain rearrangements

upon binding[103]. In particular, the valine 111 side chain rotates to make room for the slightly larger ligands (Figure 2.2b). Here, it will be interesting to see if this slow motion can be captured in the fast non-equilibrium transitions and whether correct binding free energies can be obtained.

2.3.2 HSP90 is a challenging, pharmaceutically relevant target

Even though simple systems, for which convergence can be achieved on reasonable timescales, are highly valuable for method comparison, we are ultimately interested in how the methods perform on pharmaceutically relevant systems. HSP90 is an anti-cancer drug target, and multiple inhibitors of the protein have been developed[7, 62]. Computationally, it is a more challenging system than the T4 lysozyme L99A mutant, though still regarded as relatively straightforward on the spectrum of potential pharmaceutical targets. The binding site is solvent exposed and, relative to T4 lysozyme L99A, the ligands are larger, more polar and flexible (Figure 2.3). While some ligands introduce larger protein conformational changes upon binding[7], these are not observed for the ligand considered in this study.

Changes in water placement and occupancy on binding can lead to sampling problems in HSP90. In the unbound state, the solvent completely fills the binding site. In the presence of the ligand, three water molecules are deeply buried in a space between the ligand and the protein (Figure 2.3). In our MD simulations where these water molecules were not included in the starting structure, we found they do not enter the space even after 40 ns (Sec. 2.5.5). We ran calculations both with and without these bound water molecules to examine their contribution to the binding free energy. In the EQ approach, we studied whether the use of HREX alleviates the sampling of this slow motion.

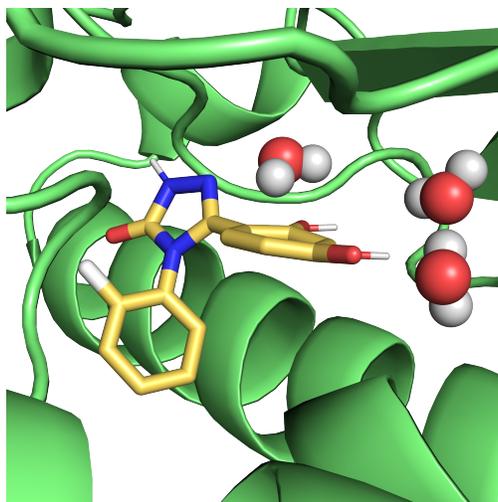


Figure 2.3: Ligand (yellow) binding to HSP90. Three buried binding site waters are represented as spheres. PDB ID: 5J64

2.4 Methods

2.4.1 We constructed a thermodynamic cycle to calculate the binding free energy

Ligand binding/unbinding events (association-dissociation) are typically too slow for direct simulation. Therefore, it is more convenient to compute binding free energies via a thermodynamic cycle. Free energy is a state function, meaning that this allows us to construct artificial (more convenient) pathways to get between the states of interest. Figure 2.4 shows the thermodynamic cycle used in this chapter. The binding free energy (panel F) is not simulated directly but obtained through summation along the cycle.

For the decoupling and recoupling of the ligand, we performed both NEQ and EQ calculations. Since these steps are most computationally challenging and expensive, we sought to investigate how NEQ performs compared to EQ.

Some of our simulations use a “bubble-ligand” that repels water to overcome water sampling problems during NEQ switching transitions. Below we show why and how it is used. To

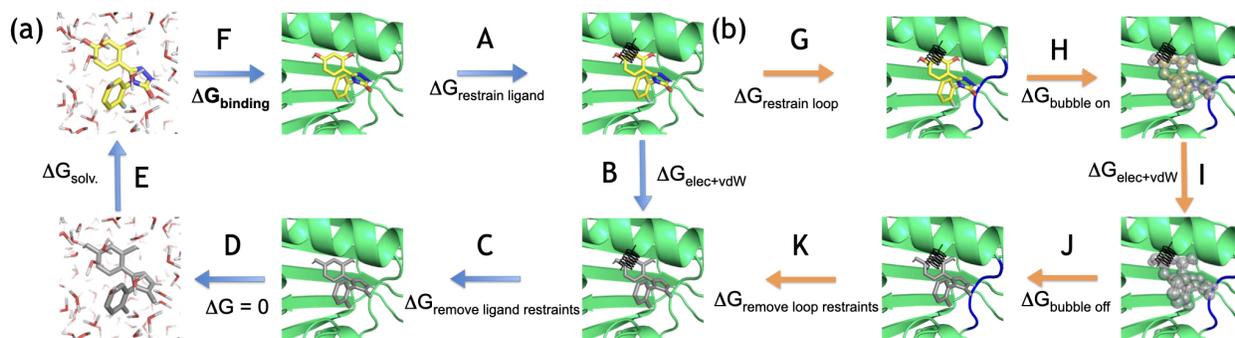


Figure 2.4: Thermodynamic cycle for computing binding free energies. The left part of the cycle (a) represents the standard cycle used in this chapter, the right part (b) was only added for calculations including a bubble-ligand (Sec. 2.5.5). The ligand is restrained using orientational (Boresch-style)[19] restraints (A) and then decoupled in the binding site, meaning that the intermolecular electrostatic and van der Waals (vdW) interactions of the ligand are turned off (B). The restraints are released analytically (C), the ligand is transferred from the complex into the solvent (D) and then coupled in the solvent (E). The next leg (F) represents the free energy difference upon ligand binding which is obtained through summation along the cycle. (b) For calculations that include a bubble-ligand to overcome insufficiencies in water sampling, a loop and three buried water molecules are restrained (G, Sec. 2.5.5), then the vdW interactions between the bubble-ligand and water are turned on (H). In the next step the ligand is decoupled in the binding site (I) and the bubble-ligand prevents water from entering the site. The bubble-ligand is then turned off again (J), the restraints on the loop and buried water molecules are released (K) which leads back to the regular cycle.

incorporate this bubble-ligand into the thermodynamic cycle (right part of Figure 2.4), we first have to restrain three buried water molecules and a protein loop close to the binding site. This is necessary because the bubble-ligand creates a vacuum in the empty binding site which some side chains of this loop try to fill resulting in conformational changes of the entire loop (Sec. 2.5.5). After restraining the loop, the bubble-ligand is turned on, so that in the next step, where the ligand is decoupled, water cannot enter the binding site. This separates the water sampling problems from the decoupling of the ligand in HSP90.

2.4.2 Simulation Details

To prepare the structures for T4 lysozyme (PDB ID: 4W53) and HSP90 (N-terminal domain, PDB ID: 5J64), first, missing heavy atoms were added with PDBfixer[37] and protonation states were assigned at pH 7 using the PDB2PQR server[35, 34]. ParmEd (<http://parmed.github.io/ParmEd>) was used to convert the pqr output to a pdb file. For the ligands, OpenEye toolkits[111] were used to add missing hydrogen atoms and to check the bond order and connectivities. For toluene and the HSP90 ligand the binding mode was extracted from the crystal structure. An alternative binding mode of toluene and the binding mode of 3-iodotoluene were obtained from the supporting information of the BLUES paper[49].

The systems of the protein-ligand complex and the ligand in solution were then generated using YANK[127]. The protein was parameterized with Amberff14SB[95] and the ligands with GAFF2[153] and AM1-BCC charges[64]. The protein-ligand complex was solvated with TIP3P water[71] in a cubic box and Na^+ and Cl^- ions were added until a salt concentration of 50 mM (T4 lysozyme) or 150 mM (HSP90) was reached. The Amber output files of the system were then converted to GROMACS structure and topology files using ParmEd. For HSP90 two protein-ligand systems were prepared, one where all water molecules from the crystal structure were removed, while in the second crystallographic waters in the binding site

were retained. The topology and coordinate files are available in the supporting information (SI).

Oriental (Boresch-style) restraints were applied by restraining 3 atoms in the protein and 3 atoms in the ligand through one distance, two angle and 3 dihedral restraints with a force constant of $20 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ for the bond and $20 \text{ kcal mol}^{-1} \text{ rad}^{-2}$ for angles and dihedrals. A table of the protein and ligand atoms used to restrain the complex can be found in SI Table SA.1. The atoms were selected during the setup of the system with YANK. YANK’s selection process picked heavy atoms that are not co-linear and ensured that the distance restraint involved a protein-ligand atom pair within 1-4 \AA . Previously, the calculated binding free energy was shown[19, 100] to be independent of the atom selection and force constants. Here, the restraint contribution (Figure 2.4 panel A) was calculated by running two equilibrium states, one of the coupled, restrained ligand and one of the coupled ligand without restraints for 10 ns each. The free energy difference was estimated using BAR and three independent replicates were run to check for convergence and to estimate the uncertainty.

All MD simulations were performed using GROMACS 2018.3[17, 1]. In the EQ approach, every λ window was energy minimized using steepest descent for 5000 steps, then equilibrated in the canonical ensemble for 10 ps at 298.15 K. Afterwards, the system was equilibrated in the isothermal-isobaric ensemble for 100 ps at a pressure of 1 bar. The production run was performed at constant NPT as well. All MD simulations were performed using the leap-frog integrator at a timestep of 2 fs. In the alchemical pathway, first the electrostatic interactions were decoupled, then the sterics, applying a soft-core potential[18]. The soft-core potential is used to avoid instabilities in intermediate lambda windows. Although new soft-core potential functions, including the smoothstep softcore potentials[83], have been developed recently, the soft-core potential from Beutler et al.[18] was used in this work as this is widely available in GROMACS. EQ and NEQ methods were treated equivalently with respect to the soft-core

potential, thus both approaches were evaluated at equivalent conditions. The intramolecular interactions of the ligand were not changed during the process. Complete settings can be found in the mdp files provided in the SI. Although the convergence rate can be different among λ windows[114, 147, 120], we ran all alchemical states for an equal amount of time for convenience. The overall convergence of the free energy difference will therefore be determined by the slowest converging window.

For the NEQ approach, end state simulations were performed in the interacting state (state A) and the decoupled state (state B) using the same general procedure (minimization and equilibration) as in the EQ approach. Then, either the first nanosecond (in the case of toluene) or the first 5 ns of the production run was discarded for equilibration and frames extracted every 40 ps (T4 lysozyme) or every 50 ps (HSP90). Fast switching transitions from state A to B (A2B) and vice versa (B2A) were performed where the $\Delta\lambda = 1/n_{steps}$ where n_{steps} is the number of steps. Since electrostatics and sterics were switched simultaneously, a soft-core potential was applied for both non-bonded interactions with the soft-core parameter $sc\text{-}\alpha=0.5$, $sc\text{-}power=1$ and $sc\text{-}\sigma=0.3$, as is typical.

Some of the simulations included a bubble-ligand which overlaps with the normal ligand but interacts only with water molecules. The bubble-ligand interacts with the water (not including the three buried water molecules) through Lennard-Jones interactions while not interacting with any other molecules in the system, and is designed (Sec. 2.5.5) to alleviate water sampling problems in the binding site at intermediate lambda values. The Lennard-Jones parameters σ and ϵ of this bubble-ligand were set to 0.33 nm and 0.41 kJ/mol for all atoms of the bubble-ligand (see topology file SI).

We also introduced position restraints to prevent the side chains of a nearby protein loop from filling the vacuum of the empty binding site created by the bubble-ligand. The restraints were applied to the side chain backbone atoms using a force constant of $1000 \text{ kJ}/(\text{mol} * \text{nm}^2)$. The three buried water molecules which do not interact with the bubble-ligand were restrained

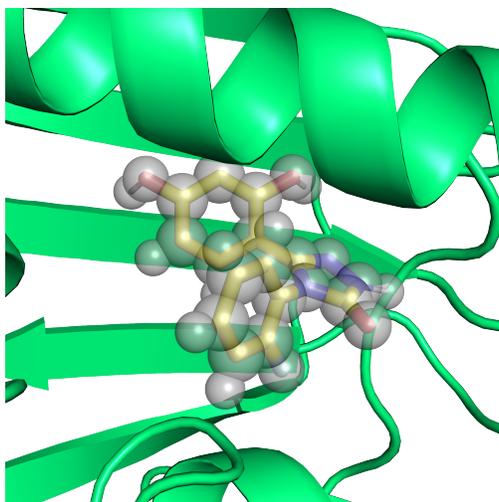


Figure 2.5: The bubble-ligand sits on top of the normal HSP90 ligand, but only interacts with water molecules in order to separate water sampling problems from the decoupling of the ligand. The bubble ligand is shown in grey spheres and the real ligand in colorized sticks.

with position restraints as well. The restraints were turned on and off using 15 λ windows with varying force constants.

2.4.3 We obtained uncertainty estimates by running independent repeats

In order to obtain reliable error estimates, we performed independent replicates, both for the EQ and the NEQ approach. Given that MD simulations are sensitive to the initial conditions, two trajectories from the same system can diverge quickly. The divergence of an ensemble of independent simulations from various starting conditions provides insight into sensitivity to initial conditions. Although in the NEQ approach the fast-switching transitions are already independent of one another (at least to the extent that the equilibrium snapshots from which they are started are assumed uncorrelated), we also ran repeats of the end state simulations. Extracting snapshots from independent shorter MD simulations rather than one larger one can ensure better sampling of phase space. Bootstrapping of NEQ work values was used for the individual replicates but we found that it underestimated the uncertainty.

For the NEQ approach, analytical error estimates can help detect lack of overlap of the work distributions. For maximum likelihood methods, like BAR, the variance asymptotically converges to the inverse of the Fisher information of the joint distribution of, in this case, forward and reverse work values.[139] Using this variance estimate, the analytical uncertainty can be vastly overestimated when overlap is poor (Sec. 2.5.5 and especially Figure A.15). We therefore only use this method as a diagnostic metric in the NEQ approach. A high uncertainly estimate indicates a potential overlap problem and that results cannot be trusted.

2.4.4 Overlapping work distributions don't guarantee reliable free energy estimates

Overlapping work distributions are a necessary but not sufficient requirement to obtain reliable free energy estimates. Consider for example a system where slow side chain rearrangements occur upon ligand binding. If this rearrangement is neither sampled in the equilibrium end state simulation nor in the fast switching transitions, the work distributions might overlap well. Computed free energies, however will not be correct, because the contribution of structural rearrangement in the binding site is not accounted for (Sec. 2.5.3, Figure 2.11). Running independent replicates helped assess these issues in this work. Phase space overlap measurements can serve as a diagnostic metric and have been examined fairly extensively for equilibrium calculations[78, 99]. Even though sufficient overlap of the potential energy distributions of neighboring λ states is crucial to obtain an unbiased estimate, it is important to remember that overlap alone does not guarantee that sampling is adequate.

2.4.5 We compared the EQ and NEQ approaches based on convergence of computed free energy differences where possible

Since the two approaches start their simulations from systems parameterized and prepared in the same way, the calculated free energy differences should converge to the same value when sampling is sufficient. In order to assess which method converges faster, we looked at how the cumulative ΔG , averaged over trials, changes over time.

A method can be considered efficient if it quickly reduces the standard deviation and bias. In the cases where convergence was achieved we therefore investigated how standard deviation and bias behave over the observed simulation time scale for the two methods. For the bias we used the converged final ΔG estimate as the true reference, although the supposedly converged free energy difference might not be correct if the most important conformations have not been sampled.

We had originally hoped to compare efficiency of EQ and NEQ approaches on all the systems considered here, but to assess efficiency, we must obtain well converged free energy estimates. However, in some of our simulations convergence is not achieved due to sampling problems we explain elsewhere (Sec. 2.5), making it impossible to compare the efficiency of the two approaches.

In all systems, we used the MBAR estimator as implemented in `alchemlyb/pymbar`[140, 36] to estimate free energy differences in the EQ approach and computed then the mean and standard deviation of three independent repeats.

For the NEQ approach we calculated free energy differences using the CFT/BAR method as implemented in `pmx`[45, 134]. Instead of calculating the mean of three independent replicates, we pooled together all work values from the repeats to estimate one free energy difference. For the uncertainty estimate, we calculated the standard deviation obtained by treating the

individual repeats separately. This analysis was performed for decoupling the ligand in the binding site as well as coupling it in the solvent.

2.4.6 We assessed potential sampling limitations

To help assess sampling problems in this work, we find it helpful to examine several properties relating to free energies, as we will further discuss in Sec. 2.5. Particularly, we compute the Pearson correlation coefficient between the work values and all side chain (ξ) dihedral angles in the protein using the SciPy package in Python, with ξ angles calculated via GROMACS. Additionally, we examine $dH/d\lambda$ values and their relationship to other degrees of freedom. Here, we do not use $dH/d\lambda$ values for free energy estimates, but they can be used to do so via TI, and these $dH/d\lambda$ values provide a view into the potential contributions of specific motions in specific trajectories to binding free energies (since they are measured for individual snapshots, allowing us to identify rearrangements/motions which lead to sudden changes in $dH/d\lambda$). We find (Sec. 2.5.3 and 2.5.5) that these are a helpful diagnostic.

2.5 Results

In this study we compare equilibrium and non-equilibrium approaches to calculating binding free energies of ligands to two protein targets, T4 lysozyme L99A and HSP90, and examine the different challenges these approaches encounter for the systems studied.

We first compare the performance of the two approaches for calculation of hydration free energies of the three ligands, which were computed as part of the thermodynamic cycle (Figure 2.4E). Those calculations converged well enough to compare the efficiency of the methods. After examining hydration free energies, we focus on protein-ligand systems.

2.5.1 Comparing the efficiency of EQ and NEQ on the hydration free energy

The EQ and NEQ approach yielded similar results for the hydration free energy of toluene, 3-iodotoluene and the HSP90 ligand. In Supporting information (SI) Figure A.1, we show how the cumulative free energy difference, the standard deviation and the bias converged for the two methods.

Using the same amount of sampling, the two approaches converged within uncertainty to the same values for toluene and 3-iodotoluene. The methods agreed to within 0.2 kcal/mol of each other fairly quickly (~ 40 ns total simulation time). In the case of the HSP90 compound, it took ~ 200 ns to agree within 0.2 kcal/mol, and EQ and NEQ free energy estimates did not converge to statistically the same result, but were very close ($\Delta\Delta G=0.14$ kcal/mol). When running longer (500 ps) but fewer NEQ switching transitions the methods agreed, but the standard deviation was slightly higher (SI Figure A.1). For these three ligands, the NEQ approach was able to reduce the standard deviation and the bias faster than the EQ approach, suggesting that it is more efficient. However, this needs investigation on a larger set of molecules.

2.5.2 Toluene binding to T4 lysozyme mutant L99A

With the force field used here, toluene binds to the cavity in this T4 lysozyme mutant with two (non-symmetric) binding modes (Figure 2.2). The crystallographic binding pose (pose I) has a more favorable binding free energy than the other pose (pose II) and interconversion between these two binding modes is very slow compared to typical simulation timescales (e.g. 100 ns or more)[49]. Therefore we ran separate calculations for each pose and applied restraints to keep the ligand from transitioning between binding modes (SI Figure A.2).

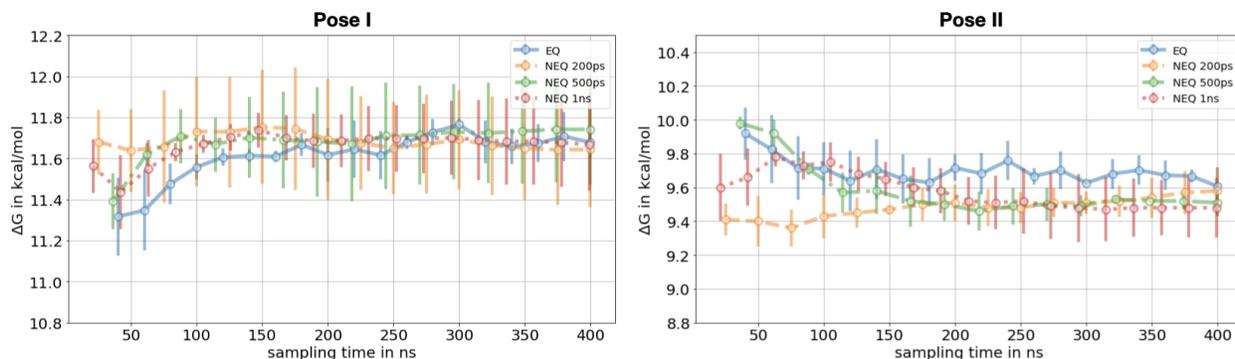


Figure 2.6: Decoupling of toluene in the binding site. The free energy difference for the EQ and NEQ approach for both poses is shown as a function of the sampling time. For the EQ approach we show the mean free energy difference across three replicates while for the NEQ approach the free energy estimate was obtained by pooling work values from three replicates and estimating one free energy difference using BAR (Sec. 2.4.5). The uncertainty estimate in both approaches is the standard deviation across three replicates. The NEQ protocols differ in the lengths of the NEQ switching transitions, which are 200 ps, 500 ps and 1 ns, respectively. The number of switching transitions differs among the protocols in such a way that the same amount of overall sampling time is used in all protocols. All protocols converged to the same free energy difference within uncertainty.

While all the protocols we report used the same amount of overall sampling time, we investigated the effect of the transition length versus number of switching transitions on the ΔG estimate in the NEQ calculations. For both binding poses, the EQ and NEQ approaches converged to a statistically equivalent value of the free energy difference (Figure 2.6). The standard deviation in pose I, however, was higher with the non-equilibrium method for all three transition lengths.

Finally, the overall binding free energy of toluene for both poses was obtained through the thermodynamic cycle (Table 2.1, results of individual legs of the cycle in SI Table SA.2). A symmetry correction of $-k_B T \ln 2 \approx -0.4$ kcal/mol was added to account for the symmetry-equivalent binding mode[100]. The crystallographic pose was correctly predicted with a more negative ΔG with both methods (Table 2.1). While the binding free energy for the non-crystallographic binding mode is not an experimental observable, when combined with the binding free energy for the crystallographic pose, it tells us the relative population of the two binding modes.

	EQ [kcal/mol]	NEQ [kcal/mol]	experimental [kcal/mol]
pose I (crystallo- graphic pose)	-4.6 ± 0.1	-4.7 ± 0.3	-5.52 ± 0.04
pose II	-3.7 ± 0.1	-3.7 ± 0.1	

Table 2.1: Binding free energy ΔG° for the toluene/T4 lysozyme system.

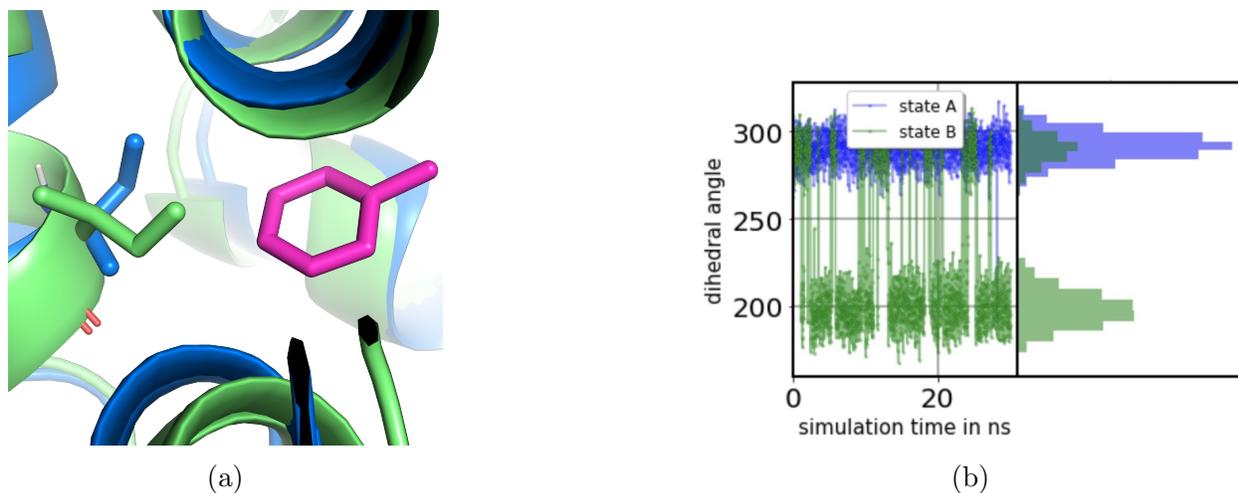


Figure 2.7: The side chain Ile78 reoriented upon binding of toluene to T4 lysozyme L99A. (a) In the bound state the Ile78 side chain occupied one rotamer state (blue sticks), while the side chain showed partial occupancy of two rotamers (blue and green sticks) in the unbound state. (b) The ξ_1 dihedral angle of Ile78 as a function of simulation time in the simulation of the interacting (state A, blue) and the noninteracting state (state B, green) and its distribution. The non-interacting state had to be sampled long enough to observe multiple transitions between the rotamers.

Isoleucine 78 reorients upon binding of toluene

The side chain of isoleucine 78 (Ile78) reorients upon binding of toluene, causing problems for both the EQ and NEQ approaches if this reorientation is not sampled adequately. While the side chain showed partial occupancy of both rotamers in the noninteracting state (state B), Ile78 had one preferred orientation in the bound state (state A, Figure 2.7).

The reorientation of the Ile78 dihedral angle occurred relatively fast on timescales of a few hundred ps – which is faster than simulation timescales and we seemed to be able to

adequately sample the motion here (Figure 2.6 and A.3). This helped us develop techniques to better assess important protein motion.

Side chain reorientation in NEQ switching transitions

In the NEQ approach, end states have to be sampled long enough to start transitions from the correct distribution of dihedral angles and/or the transitions have to be slow enough to allow for reorientation (if needed) on ligand binding, otherwise the final distribution of dihedral angles will closely match the starting distribution.

According to the Crooks fluctuation theorem, provided that transitions are started from the correct distribution of conformational states in the end states and that a sufficient number of transitions are performed, the method will converge to the correct free energy difference. This is true even if the final distribution of dihedral angles after the transitions does not match the equilibrium distribution in that end state.

As can be seen in SI Figure A.3, fast transitions (200 ps) can be long enough to allow for the side chain to reorient, however, the rotamer populations do not resemble those after the transition as observed in a long end state simulation. This does not mean that resulting values are necessarily wrong as they would be if the dominant rotamer were never sampled in the non-equilibrium switches (Sec. 2.5.3). Here, running shorter transitions, where we sampled the rotamer switching only sometimes, resulted in free energy estimates that were statistically equivalent to those obtained using longer transitions, in which the dominant rotamer was sampled most of the time, presumably because the free energy estimator gives the most weight to those transitions which discover the favorable rotamer. The standard deviation was also comparable among different NEQ protocols (Figure 2.6). This shows that sampling the side chain reorientation in some switching transitions is sufficient. The standard deviation between independent replicates for pose I is higher with the NEQ approach

than with the EQ approach for all transitions lengths (0.2-0.3 kcal/mol vs. 0.1 kcal/mol), suggesting that the protein undergoes additional conformational changes.

In this system, sampling of different rotamers of Ile78 led to a bimodal work distribution in the direction of restoring the ligand-protein interactions (the direction we call B2A, as our B state is the noninteracting state), and this was especially common for fast transitions (Figure 2.8). While a bimodal distribution of work values itself does not necessarily pose a problem, it *can* pose a problem when it results from drawing from an incorrect or biased distribution of states at the end states – such as when a slow conformational transition is inadequately sampled for one or both end states – or when it results from transitions which are too fast to sample an important conformational change.

Concretely, for example, consider a situation where a minor rotameric state for a particular sidechain ought to be slightly populated in the bound state at equilibrium and becomes dominant in the unbound state. If the actual sampling of the bound state misses that minor conformer entirely and switching to the unbound state is so rapid as to miss rotameric transitions, there is a high likelihood that resulting free energy estimates will be biased (unless sampling of reverse transformations is somehow sufficient to recover). Here, then, we find it helpful to assess dihedral sampling in the end states and assess whether switching transitions ever sample rotamers appropriate for their target end state.

If the rotamer population is not sampled correctly in the end states, the results will be inaccurate or imprecise, and a bimodal distribution can in some cases be a warning sign that sampling might be complex or slow. Restraining the Ile78 side chain resulted in a unimodal work distribution (SI Figure A.6) which further supports our findings that different rotamers of Ile78 in the B state led to a bimodal work distribution in the reverse transitions. The situation could have been worse if the noninteracting state did not already spend time sampling both rotamers and if the reorientation occurred on longer timescales. If a conformational transition is truly needed on binding then instead of having a bimodal distribution the final

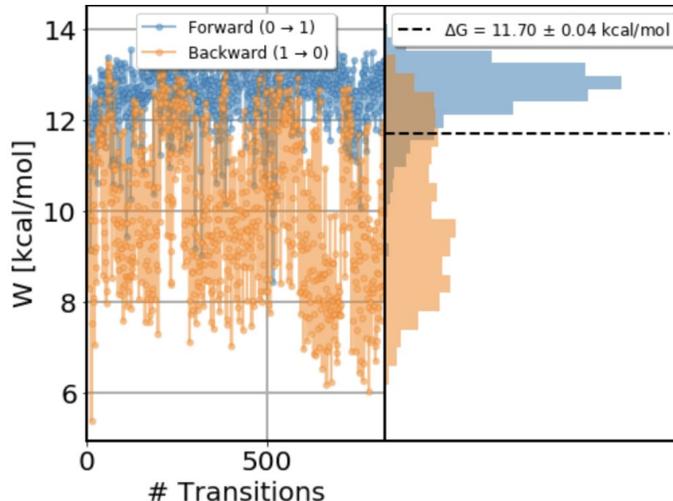


Figure 2.8: Work values for toluene binding to T4 lysozyme L99A, in forward (blue) and reverse (orange) direction. Shown are values measured for each attempted transition, as well as the distribution of the work values. The transition length for each transition was 200 ps. Data from trials in the reverse direction shows sudden changes in work values and the distribution of work values is bimodal indicating a slow degree of freedom in the system.

distribution might be unimodal, but wrong, as discussed in Sec. 2.4.4 and in Sec. 2.5.3.

There are generally two scenarios that can lead to dissipation and non-overlapping work distributions (SI Figure A.9, A.15) or bimodal work distributions (Figure 2.8): In the first scenario the system ends up in the correct end state but during NEQ switches was driven rapidly and accordingly heat was dissipated (friction). Secondly, the system does not end up in the equilibrium end state at the end of the short transition. An example for this second scenario is the Ile78 side chain orientation discussed above.

To investigate whether the Boresch-style restraints were a source for the large dissipation seen in the reverse process (orange work distribution Figure 2.8), we performed calculations applying a lower force constant on the orientational (Boresch-style) restraints (see SI Figure A.4). The average dissipated work W_d , where $W_d = W^R + \Delta G = W^F - \Delta G$ [67], was the same for both protocols (1.7 kcal/mol in the reverse direction and 0.9 kcal/mol in the forward direction).

Center of mass (COM) – COM flat-bottom harmonic distance restraints are an alternative to orientational restraints and have been applied to binding free energy calculations[137]. We performed calculations restraining the distance between the COM of the ligand and the COM of a protein side chain and compared the results of the two restraint approaches (see SI Table SA.3). The dissipated work was higher in the COM-COM restraining protocol (3.8 kcal/mol in the reverse direction and 1.4 kcal/mol in the forward direction) compared to the one using Boresch-style restraints. With a single COM-COM distance restraint, the ligand can sample the sphere with the radius of that distance restraint which presumably caused the larger dissipation. These findings suggests that the Boresch-style restraints were not a major source for the dissipation seen in Figure 8.

The Ile78 orientation correlated with the work values as detected with the Pearson correlation coefficient

In the NEQ approach, we detected slow protein motions by calculating the Pearson correlation coefficient (PCC) between the work values and all side chain dihedral angles in the protein. We calculated the dihedral angles from the set of first frames and the set of last frames of NEQ switching transitions. Then, we took the work values for the set of forward transitions and correlated it with the set of initial dihedral angles; we also did the same for the set of final dihedral angles. This gave two separate PCC for each dihedral angle (one using the dihedral angles from the first frames and one the dihedral angle of the set of last frames). We performed the calculation in both the forward and reverse direction (decoupling and coupling). When the ligand appeared in the binding site (backward direction), a high PCC was found between the ξ_1 dihedral angle of Ile78 and the reverse work values in both poses. The correlation was highest in the NEQ protocol with 200 ps/transitions using the structure of the first frame of the switching transitions (PCC = 0.5-0.6), as can be seen in Figure 2.9a. The PCC was also high using the last frame of the transitions (0.5-0.6), suggest-

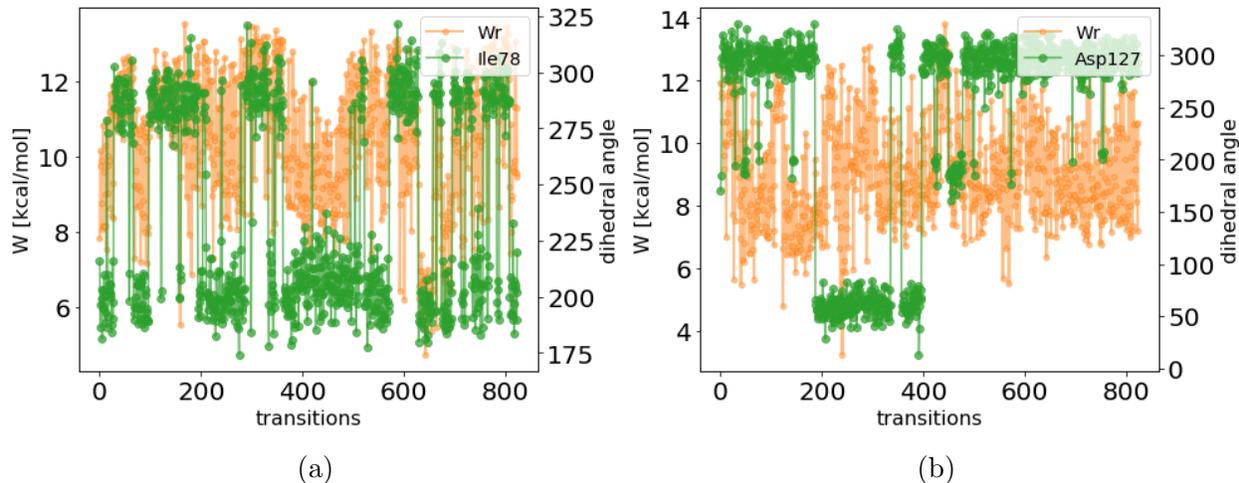


Figure 2.9: Correlation between reverse work values (W_r) and the dihedral angle of Ile78 (a) and Asp127 (b) for toluene binding to T4 lysozyme L99A. The work values (orange) and the dihedral angles (green) are plotted as a function of the transition number. The dihedral angle was calculated from the set of first frames of the NEQ transitions. (a) Work values are correlated with the dihedral angle of Ile78, which is supported by a high Pearson correlation coefficient ($PCC=0.5$). (b) A low PCC of 0.04 likely indicates no significant correlation between dihedral angle Asp127 and work values. Asp127 was included as an example of a motion that is not correlated with the free energy difference. Even though the motion is slow, it appears not to be important.

ing that it is important to sample the side chain reorientation at least in some transitions. The PCC decreases as the transition length increases, especially the correlation between work and the starting orientation (first frame of transitions)(mostly $PCC < 0.5$), since the side chain reorients more often in longer switching transitions. Figure 2.9b shows an example of a slow motion that is not correlated with work values. Asp127 is not close to the binding site and although the motion is slow, it does not seem to be important.

Calculating the correlation between work values and dihedral angles in the protein may be a useful general approach for identifying potentially problematic degrees of freedom. This approach will work best when the number of such degrees of freedom is low. If multiple side chains reorient upon binding, the correlation between the work and the orientation of each individual side chain will be reduced, making it more difficult to identify slow motions. Furthermore, the metric is prone to false positives due to accidental correlations that can

cross significance testing thresholds, as illustrated in this comic[163]. If 200 residues are predicted not to correlate with the work values, a 5% false positive rate will mean that there are 10 false correlations. Careful inspection (e.g. considering proximity to the binding site) and cross-validation with replicates can help ensure apparent correlations are not spurious.

However, it is important to note that NEQ free energy calculations do not require switching transitions to sample all relevant conformational states; what is required is that end state sampling populate the correct equilibrium distribution of states, and switching transitions must be sufficiently numerous to provide an adequate average over possible work values. Thus, the presence of work values that correlate with a particular slow conformational degree of freedom does not necessarily indicate sampling problems; rather, it means that practitioners need to check that the end states in fact sample the correct equilibrium distribution of states. If they do not, results will be inaccurate or at least imprecise.

In the equilibrium FEC, the Ile78 orientation correlates with $dH/d\lambda$ values

In the EQ approach, sampling problems can in some cases be detected by examining sudden changes in the $dH/d\lambda$ values and identifying related structural degrees of freedom. We plotted the running average of the $dH/d\lambda$ values over the simulation time for every state, picked a λ window where there was an obvious jump in the $dH/d\lambda$ value and looked at frames from the trajectory near that jump. Once any potential sampling problem – such as a motion which might be responsible for the sudden jump in $dH/d\lambda$ – was identified we correlated the motion with the $dH/d\lambda$ curves of all λ windows to assess correlations between that structural degree of freedom and $dH/d\lambda$ values.

Here, we find that sampling problems primarily occurred in the alchemical intermediate states. In the EQ approach, each state has to be sampled at equilibrium, meaning that the correct equilibrium distribution of the Ile78 dihedral angle has to be sampled at every

λ window. This can be especially challenging in intermediate states where energy barriers between the rotamers can be high. In SI Figure A.5 we show an example of the correlation between the $dH/d\lambda$ values in one λ window and the dihedral angle of Ile78. The three replicates did not sample the same rotamer distribution of the side chain indicating that sampling time was not sufficient to reach equilibrium.

This approach of correlating $dH/d\lambda$ values with structural degrees of freedom can be a useful general strategy for identifying potentially problematic degrees of freedom in equilibrium FEC. This method is comparable to examining sudden changes in work values in the NEQ approach in that it helps to narrow down a large amount of trajectory data to a specific set of frames in which a slow but important rearrangement may occur.

Here, independent replicates converged well (Figure 2.6) suggesting that the motion of the Ile78 rearrangement, although not sampled adequately in every alchemical state, was overall sampled well enough to only have a minor impact on the free energy difference.

2.5.3 3-Iodotoluene binding to T4 lysozyme mutant L99A

For 3-iodotoluene binding to T4 lysozyme (L99A), in both EQ and NEQ calculations the independent replicates did not converge to the same result in 400 ns total simulation time, indicating inadequate sampling. In Figure 2.10 we show the mean and standard deviation for the independent repeats of the EQ, and different protocols of the NEQ approach; for details on the individual repeats see SI Figures A.7 and A.8. The standard deviation σ for the EQ approach was higher than in the toluene case ($\sigma_{3\text{-iodotoluene}} = 0.4$ kcal/mol, $\sigma_{\text{toluene}} = 0.1$ kcal/mol) suggesting that sampling problems were more severe. For the NEQ protocols, which all used the same total simulation time, the ΔG estimate depended on the length of the NEQ switching transitions. Since overlap of the work distributions was poor, even with 4 ns/transition (SI Figure A.9), results are likely not to be trusted. The overall binding

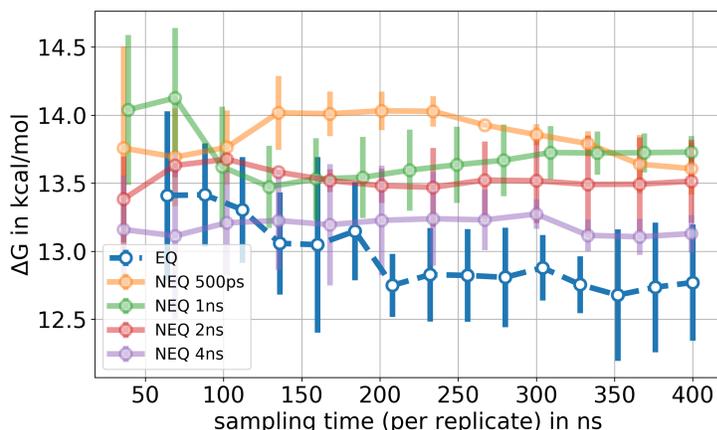


Figure 2.10: Free energy difference of decoupling 3-iodotoluene in the binding site as a function of sampling time. For the EQ approach we show the mean free energy difference across three replicates while for the NEQ approach the free energy estimate was obtained by pooling work values from three replicates and estimating one free energy difference using BAR (Sec. 2.4.5). The uncertainty estimate in both approaches is the standard deviation across three replicates. Same amount of total simulation time was used for the EQ and the four protocols of the NEQ approach. The different NEQ protocols differ in the length and the number of NEQ transitions. The standard deviation, especially for the EQ approach, was high and the ΔG estimate of the NEQ protocols depended on the transition length, suggesting insufficiencies in sampling.

free energy of 3-iodotoluene, which was obtained through the thermodynamic cycle, varies between -4.8 and -5.7 kcal/mol among different protocols, as shown in SI Table SA.4. The experimental binding affinity of 3-iodotoluene to T4 lysozyme L99A was not found in the literature. Here our focus is not on agreement with experimental values, since this is also a function of the force field and other factors; rather, we focus on whether we have obtained converged values.

Valine 111 reorients upon binding of 3-iodotoluene

Slow conformational changes in the protein prevent convergence in both methods on the timescale of our simulation. 3-Iodotoluene, being more bulky than toluene, induces a rearrangement of the valine 111 (Val111) side chain upon binding (Figure 2.2b). The timescale

for this reorientation is relatively slow (a couple of nanoseconds) and consequently results are highly sensitive to the initial protein conformation (Figure 2.11). The predicted free energy change is either too favorable or too unfavorable depending on which structure was used as a starting point. If this conformational change is not sampled properly in the free energy calculation, running simulations starting in the apo protein structure results in a predicted free energy which is too unfavorable. The contribution of protein rearrangement upon binding and potential steric clashes with the protein are not accounted for in this scenario. On the other hand, the free energy change is predicted as too favorable when simulations are started in the holo protein structure. The protein is deformed upon binding and the energetic cost of this protein strain energy is missing [103].

This side chain rearrangement is not sampled adequately in NEQ and EQ calculations

In the NEQ approach it is important to sample the correct orientation of Val111 in the end states at equilibrium and to run the transitions between the end states slowly enough to allow for a rearrangement at least in some transitions (as discussed in Sec. 2.5.2). Although the free energy estimate should be independent of the switching rate, provided that a sufficient number of work values was collected, we found that the efficiency of a protocol can depend on the switching rate. Performing the transitions at a slower rate increases the probability of observing the reorientation of a side chain, because the transition has less dissipation and follows closer the free energy gradient. Here, dissipation was large, and reducing the switching rate can be more efficient than running faster transitions (see Figure 2.10) which would require a longer total simulation time. Simulations were started in the bound protein conformation. The end state simulation of the decoupled state has therefore be run long enough so that the Val111 side chain can reorient adequately to obtain a correct equilibrium distribution of conformations. Here, the alternate rotamer (after reorientation) should be

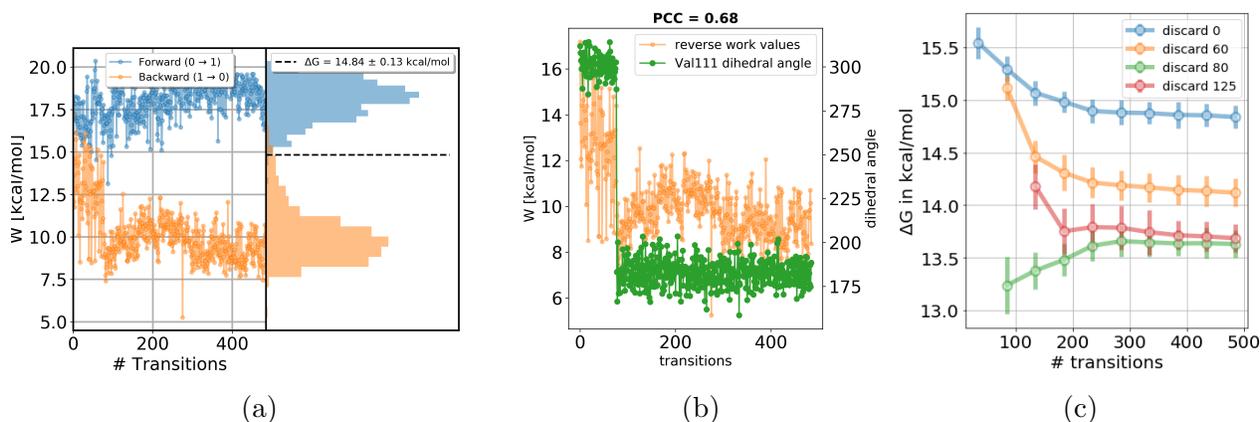


Figure 2.11: (a) Work values for 3-iodotoluene binding to T4 lysozyme L99A, in forward (blue) and reverse (orange) direction. Shown are values measured for each attempted transition, as well as the distribution of the work values. In contrast to the general protocol (Figure 2.10) where 5 ns of the end state simulation were discarded, here no data were discarded for equilibration. Data from trials in the reverse direction (coupling) shows a sudden change in work values at transition 77. This indicates sampling of a slow motion in the end state. (b) The reverse work values (orange) and the ξ_1 dihedral angle of Val111 (green) as a function of transition number. The rotation of the Val111 side chain is correlated with the NEQ work (Pearson correlation coefficient $PCC=0.68$). The sudden change in work values is likely caused by the reorientation of Val111 in the end state simulation. (c) Free energy difference as a function of transition number. The analysis was performed for discarding 0, 60, 80 or 125 transitions to study the impact of the Val111 orientation on the final ΔG estimate. The final free energy estimate is highest when no frames were discarded for equilibration, underestimating the contribution of protein strain relaxation upon unbinding. When 17 transitions start from the wrong V111 rotamer (discard 60, reorientation at transition 77), the ΔG estimate did not converge to the same result as when starting all transitions in the dominant rotamer (discard 80 and 125). This shows the impact of the orientation of the Val111 side chain on the free energy difference.

dominant in the unbound ensemble, so we discard all frames before reorientation to equilibration. This would not be necessary if the end state simulation was run long enough to capture the equilibrium distribution of conformations instead of being biased by the starting configuration of Val111.

Starting NEQ transitions from structures where Val111 had not reoriented yet led falsely to good overlap of the work distributions but a wrong estimate of the free energy difference (Figure 2.11). Only a few transitions starting from the “wrong” structure (the overpopulated minor rotamer) had a huge impact on the result since the tails of the distribution have a large statistical weight for the ΔG estimate. When as few as 17 transitions starting from the wrong Val111 rotamer were included (Figure 2.11c, discard 60), the free energy did not converge to the same result as when starting all transitions in the correct dominant rotamer (Figure 2.11c, discard 80). We attribute this to the fact that the alternate rotamer ought to be present to a vanishingly small degree at equilibrium.

Not only is it important to start the transitions with correctly populated orientations of the Val111 side chain, but the switching also has to be slow enough to allow for rearrangement towards the orientation in the other end state at least in some transitions. Note that it is also possible that, with sufficiently large end state ensembles which populate minor rotamers at exactly the right degree, such transitions during switching might not be necessary, but that was not observed here. Unfortunately, it is not obvious how the NEQ switching rate couples to the rate of side chain reorientation, making it difficult to know the required transition length *a priori*. As discussed in Sec. 2.5.2, although not all non-equilibrium switches have to sample this reorientation, it has to be sampled in some transitions to obtain overlapping work distributions and a reliable free energy estimate. Here, even with 4 ns/transition, the side chain reoriented only in 8 % of the transitions in the coupling direction (B2A) and 74 % during decoupling of the ligand (A2B), which results in too few transitions sampling the correct rotamer in the coupling direction. Since the conformational change was not sampled

sufficiently in 400 ns total simulation time, we decided that it was better to restrain the side chain.

To support our findings that the large dissipation (SI Figure A.9) was caused by a side chain rearrangement upon binding and not by the Boresch-style restraints, we performed calculations applying a lower force constant on the orientational (Boresch-style) restraints (see SI Figure A.10). The average dissipated work W_d in the forward and reverse directions was similar for the protocols using weaker and stronger restraints (3.7 vs. 3.9 kcal/mol in the reverse direction and 3.4 vs. 4.1 kcal/mol in the forward direction). These findings suggests that the Boresch-style restraints were not a major source for the dissipation in this system.

Equilibrium FEC also suffered from sampling problems due to the rearrangement of the Val111 side chain. These issues were most severe in intermediate λ windows where the Val111 side chain showed partial occupancy of two rotamers and a slow rotation between them. In order to investigate this, we looked for sudden changes in $dH/d\lambda$ values. Correlating these with the dihedral angle ξ_1 of Val111 and comparing results among three independent replicates shows that the side chain was kinetically trapped in various intermediate states (Figure 2.12). Approaches to overcome these sampling problems include the use of enhanced sampling techniques, like HREX or incorporating BLUES-like side chain moves[49, 20].

We restrained the Val111 side chain to avoid the sampling of slow DOF

In this work, we decided to work around inadequate sampling of this motion by restraining the dihedral angle of the Val111 side chain for both EQ and NEQ approaches. We restrained the side chain to its preferred orientation in the bound and in the decoupled state and switched it during the alchemical path.

With this protocol, the work distributions in the NEQ case overlapped well already using 500 ps/transition (SI Figure A.13), indicating that indeed the Val111 side chain sampling

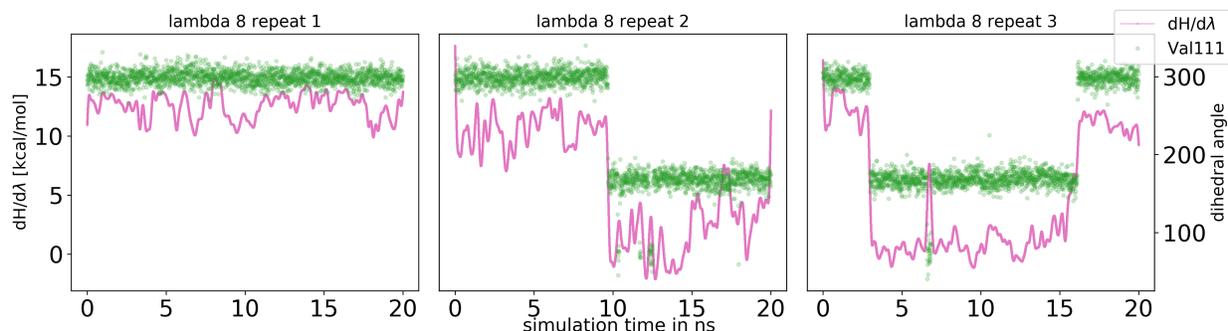


Figure 2.12: Correlation between $dH/d\lambda$ values and the dihedral angle of Val111 for 3-iodotoluene binding to T4 lysozyme L99A. A running average of the $dH/d\lambda$ values (pink, averaged across 2000 data points) and the dihedral angle of Val111 (green) is plotted as a function of simulation time. Here we only show λ 8 for the three independent replicates, additional λ windows can be found in the SI. Rotation of the Val111 side chain resulted in sudden changes of $dH/d\lambda$ showing the impact of the side chain orientation on the free energy difference. The side chain did not rotate in repeat 1, while it did in the other two repeats indicating that the correct equilibrium distribution was not sampled adequately in 20 ns simulation time.

led to poor overlap in the unrestrained calculations. However, a lot of NEQ transitions (~ 320 , SI Figure A.12) were necessary in order to converge independent replicates which was probably due to additional slow DOF in the system (perhaps Ile78 rotamer sampling). The approach of switching the dihedral during the alchemical path was challenging using the EQ protocol. The standard deviation across replicates was higher ($\text{std}=0.4$ kcal/mol, SI Figure A.11) than in the NEQ approach ($\text{std}=0.07$ kcal/mol) caused by insufficient sampling of the Val111 switching in some intermediate states.

Since we restrained the dihedral, to calculate binding free energies, we must account for the energetic costs of the dihedral restraints in the thermodynamic cycle. We did this by first turning the restraints on with the ligand interacting in the binding site, then decoupling the ligand while simultaneously switching the restraints to the orientation in the apo structure. Finally the restraints were then turned off again. The cost for restraining the side chain was very low (0.1 kcal/mol for restraining and -0.1 kcal/mol for unrestraining) since Val111 was restrained to its preferred orientation in both the holo and apo structure.

We also tested a different protocol, where we restrained the dihedral to the orientation in the holo structure both in the bound and unbound state and did not switch it upon decoupling of the ligand. This introduced new sampling problems: Other side chains in the binding site (methionine 102, valine 103, isoleucine 78) reoriented in response to the forced unfavorable orientation of Val111 in the unbound state (probably to fill the space which would otherwise be empty) making it necessary to restrain those as well. The Pearson correlation coefficient between the dihedral angles and the work values helped detect some of these DOF. This metric was however sometimes subject to noise since many dihedral angles correlated with work values. Correlations were also not always consistent across replicates because different side chain caused problems in different replicates. Restraining the dihedral angles of those side chains as well led to a low standard deviation (0.2 kcal/mol) in the NEQ approach already with short transitions (100 ps/transition).

With these additional restraints, it was more challenging to account for the restraining costs. More side chains were restrained and the costs for restraining those to an unfavorable rotameric state in the unbound state were high (~ 7.5 kcal/mol, see SI Figure A.14).

Although we are aware that the use of restraints is not a good general solution to inadequate side chain rotamer sampling, we think that it helps illustrate a potential pitfall of these approaches. Particularly, restraining the side chains gives the same outcome one might get in a shorter simulation, or a simulation where the environment is more sterically constrained (in which the side chain might not switch rotamers at all). In such cases, one might see “good overlap” in the non-equilibrium work distributions but only because the simulations missed an important motion that needs to be sampled.

2.5.4 Summary of the T4 lysozyme system

In the T4 lysozyme cases examined, convergence of the ΔG estimate was challenging due to side chain rotamer sampling. Slow side chain rearrangements, like Val111 in the 3-iodotoluene case, proved to be very challenging both in EQ and NEQ approaches. The length of NEQ transitions became very important here. Sufficient sampling of faster side chain-rearrangements upon binding, like Ile78 in the toluene system, was achieved by running longer/many NEQ transitions and long equilibration in the equilibrium FEC. In principle this could be true for slower rearrangements too, but in some cases the relevant motions can take nanoseconds to hundreds of nanoseconds for a single transition, making it less desirable to simply run longer simulations. In the NEQ approach, sampling problems were detectable from work value trends and their correlation with structural degrees of freedom in equilibrium snapshots they were derived from. Here, we restrain the dihedral angles of the side chains to work around the issues but our findings suggest that better and more robust methods are necessary to solve the problem.

2.5.5 Free energy calculations in the HSP90 system

In the HSP90 system, in both EQ and NEQ calculations, independent replicates did not converge to the same free energy estimate within a standard deviation of 0.2 kcal/mol – even with $\sim 1 \mu s$ sampling time per replicate – indicating inadequate sampling (Figure 2.13). In addition to the high standard deviation among replicates and convergence of the cumulative ΔG estimate, several additional factors suggest the presence of one or more slow degrees of freedom in the system.

In the NEQ approach, the overlap of the work distributions was very poor, even with 10 ns/transition (SI Figure A.15). Running 10 ns per transition requires considerable computational expense and removes much of the benefit of the NEQ approach. Given the same

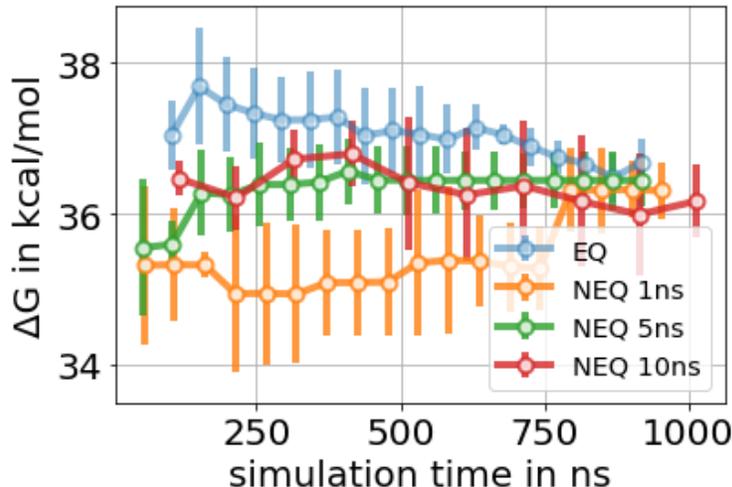


Figure 2.13: Free energy difference of decoupling the HSP90 ligand in the binding site as a function of total simulation time. For the EQ approach we show the mean free energy difference across three replicates while for the NEQ approach the free energy estimate was obtained by pooling work values from three replicates and estimating one free energy difference using BAR (Sec. 2.4.5). The uncertainty estimate in both approaches is the standard deviation across three replicates. The three NEQ protocols differ in the length and number of the switching transitions. A high standard deviation indicates inadequate sampling.

total simulation time for different NEQ protocols, the lengths of the switching transitions had a huge impact on the estimated free energy change and different protocols only converged to the same result within large uncertainty and after long simulation times. This suggests that the uncertainty of the ΔG estimates is larger when work distributions lack overlap.

On the other hand, for the EQ approach, the final ΔG estimate depended on the equilibration time (how much data were discarded to equilibration). Discarding more data (20 ns per λ window) made a difference of ~ 1 kcal/mol in all three replicates compared to not discarding any data to equilibration (SI Figure A.16). This and a low number of uncorrelated samples in some λ windows can indicate a slow degree of freedom (DOF). Similar to the T4 lysozyme system, analyzing sudden changes in $dH/d\lambda$ values helped identifying these slow DOF, as shown below.

The overall binding free energy of the HSP90 ligand, which was obtained through the thermodynamic cycle, varied between -11.0 and -11.8 kcal/mol among different protocols, as

shown in SI Table SA.5. The binding free energy of the protocols, where three buried water molecules were not present in the starting structure (Sec. 2.5.5), differs from the other protocols by ~ 2 kcal/mol (-8.6 to -9.9 kcal/mol). These protocols agreed best with the experimental binding free energy of 9.34 ± 0.07 kcal/mol, although inadequate sampling was most severe. This is possible since agreement with experimental values is not only a function of adequate sampling, but also of the force field and other factors. Therefore, we focus on whether we have obtained converged values rather than on agreement with experimental values.

Slow water sampling in the solvent exposed binding site of HSP90

The first issue we identified was slow water sampling in this solvent exposed binding site. In the unbound state, water molecules completely fill the binding site. Upon binding the ligand displaces the solvent molecules except three buried water molecules between the ligand and the protein (Figure 2.3).

In equilibrium FEC, inadequate water sampling was mainly observed in the intermediate λ windows. As the interactions of the ligand were decoupled, more water molecules entered the space between the ligand and the protein, joining the three buried water molecules (Figure 2.14). However, since the ligand was still partly interacting and blocking or slowing solvent access to this space, this motion was slow to sample. For multiple λ windows, the number of water molecules in the binding site differed among the independent replicates, and 40 ns sampling time/ λ window was not always sufficient to reach equilibrium. Jumps in $dH/d\lambda$ curves correlated with water molecules entering the buried space (Figure 2.15). This suggests that sampling the correct number of waters in the binding site has an impact on the ΔG estimate. For the fully decoupled ligand, the timescales of the water molecules filling the empty binding site were relatively fast (roughly a few hundred ps).

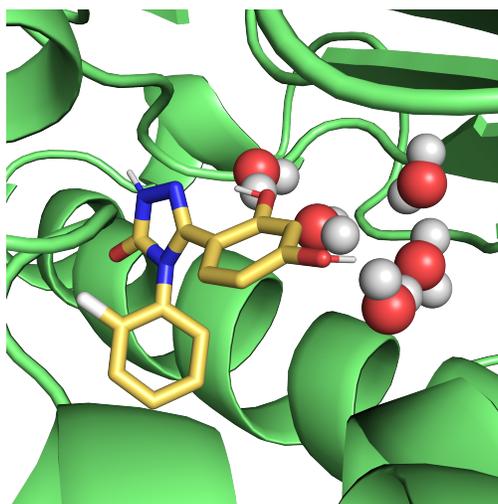


Figure 2.14: In the HSP90 system, water molecules got trapped in the binding site when the ligand appeared quickly during NEQ transitions as well as in intermediate λ windows in the EQ approach. HSP90 is shown in green, ligand in yellow sticks and water molecules are depicted as spheres.

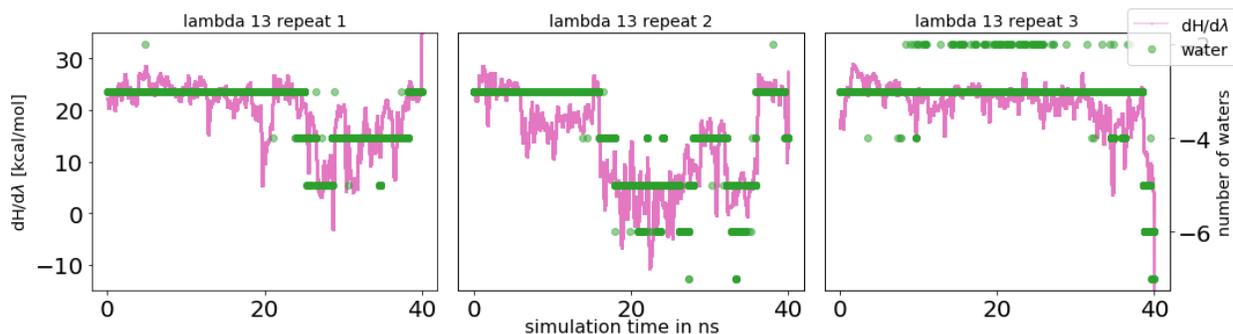


Figure 2.15: Correlation between $dH/d\lambda$ values and the number of water molecules in the binding site fo HSP90. A running average of $dH/d\lambda$ values (pink, averaged across 2000 data points) and number of water molecules (green, negative number to better show the correlation) as a function of simulation time per λ window. Here we show λ 13 across all three replicates as an example; additional data is in the SI. The entry of water molecules (more than the three that were already present in the starting structure) was a slow motion and had an impact on the potential energy of the system. Equilibrium was not reached at this particular λ window in any of the three repeats.

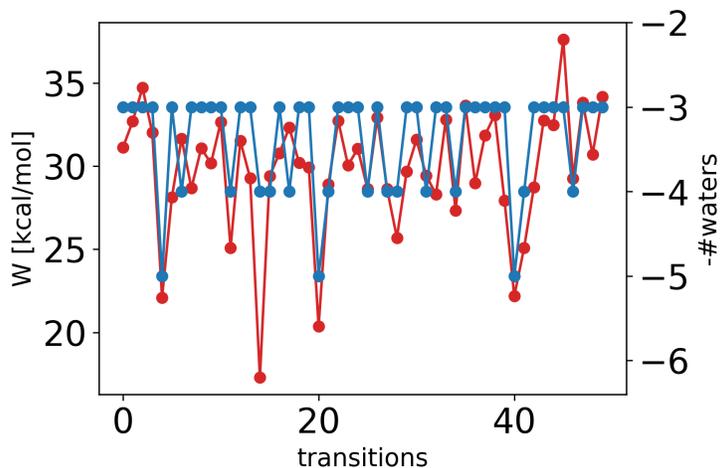


Figure 2.16: Correlation between NEQ work values (red) and the number of water molecules (blue, with negative numbers to better show the correlation) in the binding site of HSP90 as a function of transition number. Transitions – appearing the ligand in the binding site – were performed for 10 ns and the last frame of that switching transition was used for counting the number of water molecules. Most of the transitions ended up with three binding site waters, which is correct.

In the NEQ approach, water molecules got trapped in the same buried space that caused problems in the EQ approach (Figure 2.14). Particularly, when the ligand appeared in the solvent-filled binding site too quickly, water molecules were not displaced properly and got stuck between the ligand and the protein. This led to poor overlap of the probability distributions of forward and reverse work values and therefore an imprecise (and potentially biased) estimate of the free energy difference. Long transitions were necessary to properly displace the water molecules upon ligand binding. When running each transition for 10 ns, most transitions (63%) correctly ended up with the three buried water molecules in the binding site, for shorter transitions less did so (5 ns/transitions: 41% and 1 ns/transition 7%). But such transitions are computationally expensive and remove much of the benefit of the non-equilibrium approach.

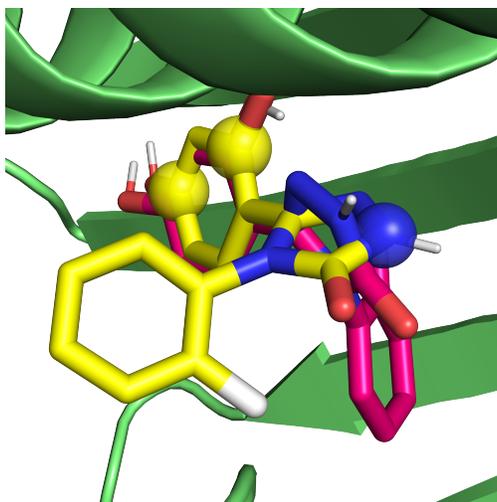


Figure 2.17: Binding mode flip of the weakly interacting ligand in HSP90. The crystallographic binding mode is shown in yellow, the flipped binding mode in pink and the protein in green. The three ligand atoms that are included in the orientational restraints are represented as spheres. The one aromatic ring that was not included in the restraints flipped into the binding pocket in the weakly and non-interacting alchemical states.

We identified other slow DOF in this system

A second sampling problem in this system was a binding mode flip that occurred when the ligand was decoupled or only weakly interacting (Figure 2.17). In this case, orientational restraints only included atoms from two of the three aromatic rings of the ligand (see spheres, Figure 2.17), hence the third ring was able to flip to a different location.

This only occurred in the weakly interacting states of the ligand, but affected the free energy difference as shown in the correlation between $dH/d\lambda$ values and the RMSD of the ligand (Figure 2.18). The binding mode flip also had an impact in the NEQ approach as judged by a high PCC (PCC=0.61) between reverse work values and the ligand RMSD (SI Figure A.17). To prevent this conformational change in the ligand, one could change the ligand atoms involved in the orientational restraints to also include the ring that flipped in the binding site in order to stabilize the binding mode better in the weakly- to non-interacting alchemical region. Another option would be the use of RMSD restraints[159].

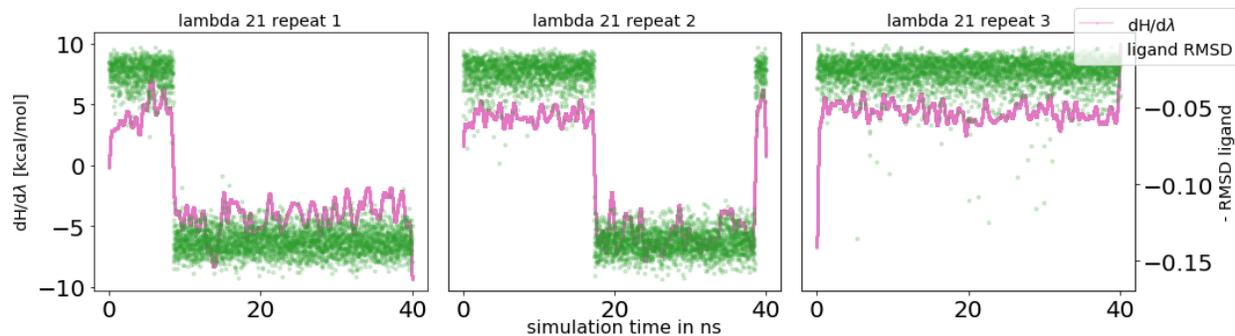


Figure 2.18: Running average of $dH/d\lambda$ values (pink, averaged across 5000 data points) and the RMSD of the HSP90 ligand (green) as a function of the simulation time. We show data for λ 21 (second last state) for three independent replicates as an example, figures of all λ windows can be found in the SI. The correlation between $dH/d\lambda$ values and the RMSD of the ligand indicates that the binding mode flip had an impact on the ΔG estimate, even though these only occurred when the ligand was weakly interacting.

Finally, minor problems were caused by side chain rearrangement of asparagine 36 (Asn36) upon binding. But since reorientation seemed not to be directly correlated with $dH/d\lambda$ values nor to the work values in the NEQ approach, the impact of this DOF on the estimate of the free energy differences is not clear. However, this Asn36 rearrangement does have a huge impact on the ΔG estimate in the “bubble approach” as described in Section 2.5.5. It is not clear whether the sampling problem here is overshadowed by more severe problems in water and binding mode sampling, or whether it is directly linked to the bubble approach.

Similar to the T4 lysozyme systems, we checked the impact of the Boresch-style restraints on the dissipation by running calculations with a weaker force constant (1/4 of the original restraint strength). SI Figure A.21 shows the comparison of these two protocols. We found that reducing the restraint strength did not have a huge impact on dissipation and the overlap of the work distributions. The average dissipated work W_d in the reverse process was 18.7 kcal/mol for the protocol with a weaker force constant and 15.2 kcal/mol for the original protocol and in the forward direction 11.5 vs. 13.9 kcal/mol, respectively.

Addressing inadequate water sampling

The water sampling problem can be separated from the decoupling of the ligand by introducing a bubble-ligand that repels the water and keeps it out of the binding site. This bubble-ligand sits on top of the normal ligand but interacts only with the water molecules (Figure 2.5). The three buried water molecules were chosen not to interact with this bubble-ligand and were restrained with position restraints. This bubble-ligand reduces the dissipation in NEQ and overcomes sampling problems in the EQ approach. Even though the water problem should be accounted for in the NEQ theorem, we found that large dissipation required long transition times or a huge number of switching transitions which became computationally expensive. Our hope was that this approach might isolate some water sampling problems. However, we found that when the ligand was decoupled, the bubble-ligand created a vacuum in the binding site which created its own problems. Side chains of a loop close to the binding site moved in to this vacuum making it necessary to restrain their backbone atoms as well.

Running NEQ with this bubble approach resulted in overlapping work distributions (already with 1 ns/transition, see SI Figure A.18) which supports our hypothesis that water sampling was causing the overlap problems. However, in both the EQ and NEQ approaches the standard deviation across independent replicates was high (0.3-0.5 kcal/mol, SI Figure A.19). This was due to slow side chain rearrangement of Asn36 (Section 2.5.5). Rotamer populations of this side chain in the unbound state differed among independent replicates which showed that sampling time was not sufficient to sample the correct Boltzmann weight of the rotamers (see SI). Both NEQ work values and $dH/d\lambda$ values in the EQ approach correlated with the orientation of this side chain (SI).

In order to obtain the binding free energy, this bubble approach has to be included in the thermodynamic cycle (Figure 2.4, right). This makes the approach not only computationally demanding, but also difficult to generalize. The protocol had to be improved iteratively,

gearing it towards this specific system. We had to optimize the van-der-Waals parameters of the bubble-ligand to sufficiently repel the water. We also had to restrain the loop that moved into the binding site to fill the vacuum created by the bubble (Sec. 2.4). For a different system it would be difficult to predict such structural changes *a priori*. Additionally, water sampling problems, although less severe, were still present when turning the bubble-ligand off (SI Figure A.20). The approach only separated the water sampling challenges from the decoupling of the ligand, but did not solve the problem directly. These factors suggest this approach is not a good general strategy but our results further support our hypothesis that water sampling caused issues in both EQ and NEQ calculations.

The overall binding free energy obtained with this bubble approach is statistically equivalent to the binding free energy obtained with the standard protocol (SI Table SA.6). The uncertainty, however, is slightly higher with the bubble approach, partly because four additional legs in the thermodynamic cycle (Figure 2.4) combine to increase the overall uncertainty.

The enhanced sampling technique HREX was not able to overcome inadequate sampling

Enhancing sampling in the EQ approach using HREX did not lead to major improvements with respect to convergence and standard deviation. As shown in Figure 2.19, the standard deviation is actually higher using HREX (0.5 kcal/mol) than without enhanced sampling (0.3 kcal/mol), as we explore below.

Both water sampling and the rearrangement of the Asn36 side chain were improved by HREX mixing of replicates. Sampling of the correct binding mode, however, seemed to be more challenging, since unfavorable (flipped) binding modes were passed between different alchemical states through replica exchange swaps (SI). This was possible because some alternate binding modes were favorable enough to be swapped to alchemical states retaining significant

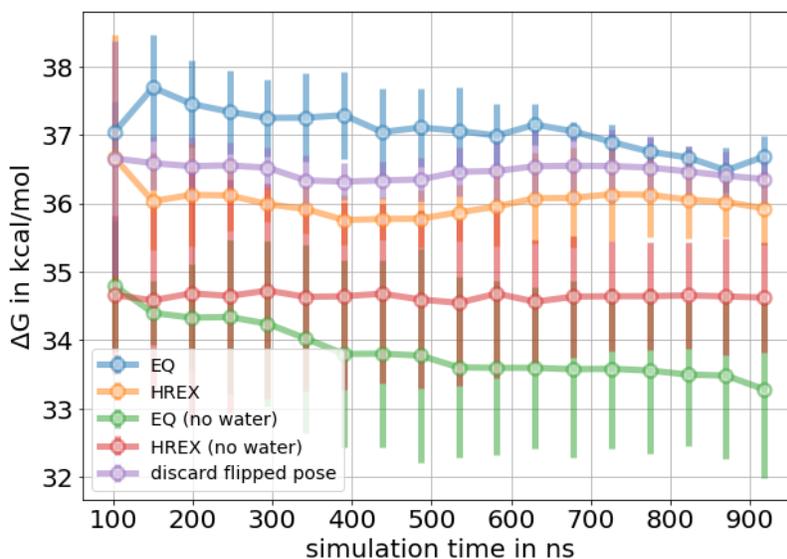


Figure 2.19: Mean free energy of the ligand binding to HSP90 as a function of total simulation time. We show the standard EQ approach (blue) and the same EQ protocol, but enhancing the sampling with HREX (orange). The purple line shows the free energy estimate of the HREX protocol after discarding all data that included the flipped binding mode. The flipped binding mode had an impact on the ΔG estimate. A high standard deviation on the HREX protocol indicates that the sampling enhancement here came with its own disadvantages. Green and red lines show simulations where the three buried water molecules were not present in the starting structure. The free energy differences from these calculations differs from the simulations where the buried waters were present (blue, orange, purple) which shows the impact of the binding site waters on the ΔG estimate.

interactions and became temporarily trapped, but not so favorable that they ought to be populated substantially at equilibrium. We reran the analysis after discarding all data that included the flipped binding mode. This changed the estimate of the free energy change by 0.4 kcal/mol and shows the impact of the binding mode flip (Figure 2.19).

In order to explore whether the NEQ approach benefits from enhanced sampling in the equilibrium end states, we performed NEQ switches using the structures from the EQ HREX run as input. As can be seen in SI Figure A.22, using enhanced sampling in the end state did not improve results here. On the contrary, some simulations appearing the ligand crashed. NEQ switches starting from the flipped binding mode, which was sampled more frequently in the HREX protocol (as discussed above), sometimes sampled a binding mode where both phenyl rings switched position (SI Figure A.23). This caused those simulation to crash as interaction of the ligand got stronger.

The presence of three buried water molecules in the binding site in the starting structure was crucial for both the EQ and NEQ approach. Indeed, not including these crystallographic waters in the starting structure had a great impact on the computed free energy difference as shown in Figure 2.19. Only when the ligand was already mostly decoupled, water molecules entered the binding site (SI). Using HREX in the EQ approach did not enhance the sampling sufficiently (Figure 2.19, red line), and other enhanced sampling methods, like GC/MC or MC/MD, are likely necessary here[131, 15] unless far longer simulations are used, which would raise computational costs far beyond what is currently standard for these calculations[133, 46, 24]. Interestingly, inserting the ligand with NEQ transitions “found” these three water molecules in exactly the same positions as in the crystal structure.

Overall, it seems likely that the use of enhanced sampling methods like HREX impacts sampling in a complex way and will affect equilibrium and nonequilibrium free energy calculations differently. While it seems certain that HREX will help dramatically in some cases, in our tests here it does not provide dramatic benefits in terms of overall sampling quality and

accuracy.

2.5.6 Summary of sampling problems in the HSP90 system

In HSP90, water sampling in a solvent exposed binding site caused major problems in alchemical FEC. For the EQ approach, sampling of the correct number of waters in alchemical intermediate states was most challenging while in the NEQ calculations water molecules got trapped in the binding site when quickly coupling the ligand. We attempted to work around the water sampling problem by introducing a bubble-ligand that keeps the water from coming into the binding site. This approach worked to some extent and helped us understand the impact of the water sampling problem better, but is not a good general protocol and more robust methods are required to solve this issue. For both methods, it was crucial to start calculations from protein structures that had three buried binding site water molecules present. Besides the inadequate water sampling, ligand binding mode flips and slow side chain rearrangement caused sampling problems as well. These severe sampling problems highlight the limitation of both approaches in ABFE calculations and reveal clear directions for further investigation.

2.6 Discussion and Conclusions

In this work we highlight challenges encountered applying equilibrium and non-equilibrium binding free energy calculations. We provide insight into how to identify sampling problems and explore similarities and differences between these two methods in handling these issues.

Alchemical calculations frequently pose sampling problems for both the EQ and NEQ approach. We observed slow side chain rearrangements (T4 lysozyme), water sampling in a solvent exposed binding site (HSP90) and flips in ligand binding mode (HSP90). These

problems are likely to show up in other protein-ligand systems and may also effect other techniques for binding free energy calculations.

2.6.1 Challenges encountered here in ABFE will probably effect RBE as well

Some of these challenges will probably affect relative binding free energy calculations (RBE) as well. For instance, slow side chain rearrangements will pose challenges whenever different ligands induce slow side chain rotations, as in the case of toluene vs. 3-iodotoluene, here, which are not that dissimilar from R-groups that might be modified in RBE calculations on a larger, drug-like ligand. Water sampling can also cause severe issues in relative binding free energy calculations when a buried water molecule is present with one ligand and displaced by another one [152, 131]. If ligands have a similar shape or displace the same water molecules, RBE will probably converge faster than the absolute protocol. Binding mode flips, on the other hand, are less likely to occur with RBE since those happen mostly in the alchemical states where the ligand is only weakly interacting in the binding site. Still, the same issue can affect R-groups of drug-like ligands.

2.6.2 Strategies to assess sampling problems in alchemical FEC

Sampling problems in these systems helped us to establish some general strategies on how to diagnose and test for problems in alchemical FEC. First, we found that in both EQ and NEQ, running independent replicates is important to identify inadequate sampling and to obtain uncertainty estimates. The analytical error – as well as bootstrapping in NEQ – underestimated the true error.

In the EQ approach, the impact of the equilibration time on the ΔG estimate and autocor-

relation analysis can give a hint to the presence of slow DOF in the system (Sec. 2.5.5).

To identify slow motions in the system that can lead to inadequate sampling, we found it helpful to examine $dH/d\lambda$ values and their relationship to other degrees of freedom. We do this by identifying sudden changes in $dH/d\lambda$ values as simulations progress. These jumps then tell us when exactly to look for sudden/slow motions by analyzing frames near those jumps. Once a slow motion is identified, the correlation between the motion and the $dH/d\lambda$ values across all alchemical states can be easily investigated. This method was successfully applied for the three ligands in this chapter and helped identify sampling problems. The approach of detecting sampling issues affecting $dH/d\lambda$ values could be automated in the future, including perhaps with machine learning, e.g. using functional mode analysis.

The overlap of forward and reverse work distributions provides a helpful further diagnostic for the NEQ approach. Having sufficient overlap of the distributions is necessary, but not sufficient for obtaining unbiased free energy estimates; as we found in Section 2.5.3, in some cases, inadequate sampling leads to a false appearance of good overlap. Slow degrees of freedom also often lead to poor overlap, which results in a high analytical error estimate. In addition, bimodal work distributions and sudden changes in work values can be a warning sign that sampling might be complex and slow. Calculating the cross-correlation between side chain dihedral angles in the protein and the non-equilibrium work values can help to diagnose problems. Instead of dihedral angles one could also choose other metrics for protein motion, such as side chain RMSD values. It is important to mention however, that analysis of correlations is prone to false positives – both for correlations between DOF and $dH/d\lambda$ curves in the EQ approach and DOF with work values in the NEQ approach. It should therefore be applied with caution and the use of replicates for cross-correlation.

In a solvent exposed binding site, counting the number of water molecules in the binding pocket can help detect changes in water placement and occupancy on binding. We also found that it is crucial to include crystallographic water molecules, especially when coming from

a high resolution structure with clear density showing ordered waters in the binding site. Here, this was important because the water molecules were buried in the binding site and the timescale for entering the buried space was very slow.

Although not exploited in this chapter, it can be a good strategy to run simulations starting in apo and holo structures if both are available. Since both calculations should converge to the same answer this can help to determine when sampling is sufficient[103, 87].

We attempted to address these sampling problems and discovered and found ways to work around some of them. We restrained side chains and introduced a bubble approach for water sampling. However, better solutions are needed to make alchemical FEC more general and robust. In the long-term, greater use of enhanced sampling techniques is likely needed, such as perhaps BLUES-like moves[49, 20] to accelerate side chain sampling.

The sampling challenges in the systems we investigated made it difficult to directly compare the efficiency of the EQ and the NEQ approaches as we had originally hoped (and as some prior studies have attempted to do [128]). For the hydration free energy of the three ligands however, calculations converged sufficiently and we were able to compare the efficiency of the EQ and NEQ approaches. Both methods performed very similar on these systems, but the NEQ approach was able to reduce the standard deviation and the bias faster than the EQ approach, suggesting that it is more efficient here, at least. However, differences were very small, and thus a larger data set is needed to investigate this more.

2.6.3 Similarities and differences of the EQ and NEQ approaches in handling slow DOF

For the decoupling of the ligands in the binding site, we concentrated on providing some understanding as to why each method behaves the way it does.

First, we compared how the two methods handle side chain rearrangements upon binding. A benefit of the NEQ approach is that the equilibrium orientation of the side chain has to be sampled only in the physical end states and not in alchemical intermediates. Crystal structures can give insight here and help assess sufficient sampling.

A weakness, however, is that when end states differ and the conformational change is not sampled during NEQ switching, dissipation can be large and a slow switching rate or numerous transitions may be required to obtain a converged free energy difference. In the case of Val111 sampling in 3-iodotoluene, we saw that this can be very limiting. Possibly separation of states approaches can help with this issue[101], e.g. having the side chain restrained during the calculations and accounting for it later. This however, creates its own problems in some cases.

An advantage for EQ, on the other hand, is that every state is sampled for a long time, allowing for rearrangements. Every state has to be sampled at equilibrium. When there are slow side chain rotations upon binding it can be difficult to reach equilibrium as well as to detect when equilibrium has been reached. Experimental structures also do not indicate the preferred side chain orientation in the unphysical alchemical region.

Water sampling in a solvent exposed binding site caused problems for both methods. For the NEQ approach, very long transitions times (5 ns - 10 ns per transitions) were required to sample changes in water placement which removed much of the benefit of the non-equilibrium approach.

In our experience, a benefit of the NEQ approach is that it can be easier to do sanity checks. Only two long simulations are performed in the physical states which is easier than having to check all trajectories along the alchemical path in EQ. The long end state simulations also provide a point of reference to check that the transition simulations are capturing the right motions. Specifically, an $A \rightarrow B$ transition needs to (at least some fraction of the time) end

up at snapshots that are representative of the same B state seen in the long equilibrium B simulation. This could be done in a systematic and automated way by ensemble comparison techniques.

In the NEQ approach, systematic errors can be introduced by starting NEQ transitions from local minima structures drawn from an incorrect or biased distribution (i.e. if the starting structures overpopulate some conformations and underpopulate others). In these cases, end state simulations can benefit from enhanced sampling techniques. In this work enhanced sampling methods were not investigated in detail in the NEQ approach, since we based this study on work of Gapsys et al. which has been successful without this[46]. In hindsight, given our results (e.g. Figure 2.9a), it is worth exploring if faster convergence can be achieved with such enhancements. We tested enhanced end state sampling in the HSP90 system, but starting transitions from structures of the EQ HREX calculations did not improve results in this case.

We learned in this study that there is no “free lunch” switching from one alchemical approach to another. If a transition is challenging, the convergence issues will remain for both EQ and NEQ approaches. An advantage of using multiple protocols might be that sampling problems can manifest differently. This may be helpful in identifying where the main sampling issues are and therefore in devising ways to alleviate them.

Chapter 3

Broadening the scope of binding free energy calculations using a Separated Topologies approach

Binding free energy calculations predict the potency of compounds to protein binding sites in a physically rigorous manner and see broad application in prioritizing the synthesis of novel drug candidates. Relative binding free energy calculations (RBFEE) have emerged as an industry standard approach to achieve highly accurate rank-order predictions of the potency of related compounds; however, this approach requires that the ligands share a common scaffold and a common binding mode, restricting the methods' domain of applicability. This is a critical limitation, since complex modifications to the ligands, especially core hopping, are very common in drug design. Absolute Binding Free Energy calculations (ABFE) are an alternate method, which can be used for ligands that are not congeneric. However, ABFE suffer from a known problem of long convergence times, due to the need to sample additional degrees of freedom within each system, such as sampling rearrangements necessary to open and close the binding site. Here, we report on an alternative method for RBFEE, called

Separated Topologies (SepTop), which overcomes the issues in both of the aforementioned methods, by enabling large scaffold changes between ligands with a convergence time comparable to traditional RBF E. Instead of only mutating atoms that vary between two ligands, this approach performs two absolute free energy calculations at the same time in opposite directions, one for each ligand. Defining the two ligands independently allows the comparison of binding of diverse ligands without the artificial constraints of identical poses or a suitable atom-atom mapping. This approach also avoids the need to sample the unbound state of the protein, making it more efficient than absolute binding free energy calculations. Here, we introduce an implementation of SepTop. We developed a general and efficient protocol for running SepTop, and we demonstrated the method on four diverse, pharmaceutically relevant systems. We report the performance of the method, as well as our practical insights into strengths, weaknesses, and challenges of applying this method in an industrial drug design setting. We find that the accuracy of the approach is sufficiently high to rank order ligands with an accuracy comparable to traditional RBF E calculations, while maintaining the additional flexibility of SepTop.

3.1 Introduction

Binding free energy calculations are a physically rigorous approach to prospectively estimate ligand potency, even before the ligand is synthesized. Although initial applications of these methods were reported decades ago[92, 158, 72], recent advances in computing technology, such as graphical processing units and low-cost parallel computing, have enabled the pharmaceutical industry to routinely and successfully apply these methods to drug discovery projects[98, 133, 22, 74, 81, 24, 82].

First, we review two common computational methods to estimate binding free energy values: relative binding free energy and absolute binding free energy. Then, we introduce a third

approach that combines the strengths of these two, giving accurate results and providing an alternative when neither absolute nor relative calculations are well suited to the problem. In this chapter we focus on alchemical approaches, which employ an unphysical path to connect two physical end states in order to obtain free energy differences.

The so-called “relative” binding free energy (RBF E) approach calculates the difference in potency between two similar ligands. During the simulations, one ligand is converted into the other by alchemical transformations of the atoms that vary between the two ligands. Common atoms from one ligand are mapped on top of those from the other ligand, resulting in either a single set of coordinates of the two ligands in the end states (single topology) or a single set of coordinates for the common substructure while representing atoms that differ between the two ligands separately (hybrid topology)[126]. This approach is less computationally demanding and has lower statistical uncertainties than absolute binding free energy calculations (ABFE). It also has advantages, *e.g.* if both ligands introduce similar conformational changes in the protein, such slow motions do not have to be sampled since the binding site is never empty. This RBF E approach is most suitable for comparing binding of related ligands and is routinely applied in drug discovery[98, 113, 30, 133, 22, 74].

The RBF E approach, however, essentially requires that the two ligands share a common scaffold which can be preserved (allowing the ligand to retain its binding mode) while modifying atoms which are not retained. This requirement for a common scaffold provides a critical challenge, especially in early-stage drug discovery where complex modifications to ligands are common. This limitation prevents these techniques from being as useful as they could be in guiding drug design. Scaffold hopping approaches[154, 169] allow for larger transformations, for example ring opening and ring size change transformations; however, the transformation size is limited, since ligands still need to share a common core. Moreover, in common practice RBF E calculations need ligands to have a shared binding pose and/or protein conformation. Additionally, RBF E calculations require atom mapping and

the construction of the “dummy atoms” must be done carefully to ensure that the energy contribution of the decoupled dummy atoms cancels out between the complex and the solvent legs of the thermodynamic cycle[43]. For example, if dummy atoms are connected to the rest of the system by more than one bond, the energy contribution does not cancel out automatically[74]. Additionally, angle and torsional terms can introduce considerable complexity if not handled with great care[43]. This concern does not apply to systems where all ligand atoms are transformed to dummy atoms, such as in ABFE.

Alternatively, the “absolute” approach computes the potency of individual ligands directly, usually through a thermodynamic cycle where a ligand is decoupled in the binding site – meaning all its non-bonded interactions are turned off – and coupled in the solvent where the interactions are turned back on[55]. Since ligands are treated individually, they do not need to share a common scaffold and can be structurally diverse. This means that ABFE could be used even in early project stages where structurally diverse ligands are common, and has been proposed to serve as a final scoring stage in virtual high throughput screening before selecting molecules for experimental testing[24, 41]. Recent studies showed that ABFE can achieve a good correlation between predicted and experimental binding free energies across different systems[3, 47, 75, 5], and even can be used to estimate binding to different proteins, allowing computation of the selectivity of ligands for a particular target[4].

However, a major limitation of the ABFE approach is that it can produce larger statistical uncertainties in the predicted potency of the ligand compared to relative approaches (see below), especially in systems where the target undergoes larger conformational changes upon ligand binding. For example, consider a protein undergoing a slow flap-closing motion upon ligand binding, such as HIV protease[63]; an ABFE calculation would need to sample the unbound state to correctly compute the true binding free energy. Such protein motions are not sampled on the typical timescale of MD simulations, resulting in inaccurate potency predictions. Slow degrees of freedom require long sampling times or the use of enhanced

sampling techniques, which can increase computational costs[75]. As a result, the ABFE method is not routinely applied in drug discovery projects. If, when the ligand is decoupled, all structures are metastable in something like the bound state one can obtain relative results from ABFE without having to sample apo-holo protein conformational transitions. However, this is not always the case, as discussed in Section 3.6

An alternate approach for RBFE, “Separated Topologies”, which we will refer to as “SepTop” throughout this paper, has the potential to combine the advantages of ABFE and standard RBFE. This protocol performs two ABFE calculations simultaneously in opposite directions, by (alchemically) inserting one ligand into the binding site, while removing the other ligand at the same time. In contrast to the standard RBFE protocol, the two ligand topologies are completely separate, making atom mapping unnecessary. Consequently, the two ligands can be structurally diverse and do not need to share a fully overlapping binding mode and/or a common scaffold, overcoming limitations of the common RBFE approach mentioned above. This protocol also never needs to sample the *apo* state of the protein as long as the protein retains its *holo* structure in the presence of both ligands, since one ligand (or a fraction of both) is always present in the binding site. Therefore, larger protein conformational changes between the bound and unbound state never need to be sampled, giving this approach a benefit in comparison to the absolute protocol. Additionally, if both ligands have the same non-zero charge, the SepTop approach conserves that charge during the transformation in the binding site while in ABFE the net charge in the binding site changes, which can lead to a wide variety of sampling and theoretical problems[124, 130, 88].

The SepTop method was introduced in 2013 in a proof-of-principle study[129]. In that study, the authors compared three RBFE methods, a single topology approach, dual topology, and SepTop, and studied the binding of two ligands to an engineered site in Cytochrome C Peroxidase. In dual topology calculations, a separate set of coordinates is used for each ligand, in contrast to the setup in single or hybrid topology approaches. Separated Topologies

can be considered a subcategory of dual topology where ligands are restrained spatially to a specific area. In the study by Rocklin *et al.* [129], dual topology referred to a different subcategory, the “linked dual topology approach” where the ligands are restrained to each other using *e.g.* distance restraints[126]. Rocklin *et al.* found that all three approaches gave comparable results when ligand reorientation was not required, while in the presence of multiple ligand binding modes, SepTop had advantages over the other RBF E approaches. In this latter case, only SepTop gave accurate results by treating individual poses separately using orientational restraints[19].

In the prior SepTop work of Rocklin *et al.*, the RBF E between binding modes was calculated and the contributions of different poses combined to obtain the overall free energy difference. This earlier study seemed to show the approach is viable, but did little to make it practical for applications or to show that it could be useful for pharmaceutically relevant systems. In addition to SepTop, there have been other recent reports of similar approaches to resolve the difficulties of ABFE calculations such as dual topology approaches[121] and the alchemical transfer method[10].

In this chapter, we reintroduce SepTop and show that it works on pharmaceutically relevant systems. We develop a prototype Python package to set up SepTop calculations in GROMACS[1] and discuss heuristics for picking atoms for orientational restraints. We test the method on several diverse, pharmaceutically relevant systems and report performance and the resulting insights into strengths, weaknesses, and challenges. We first test the approach on systems with small ligand transformations, allowing us to compare SepTop to standard RBF E and validate that it yields correct binding free energies. On more ambitious transformations, we find that SepTop performs well, even when such transformations fall outside the scope of standard RBF E methods.

3.2 Methods

3.2.1 The thermodynamic cycle for SepTop computes RBF E by running two ABFE calculations in opposite directions

The relative binding free energy between two ligands A and B, $\Delta\Delta G_{bind}$, can be obtained by transforming one ligand into the other ligand both in the solvent and in the binding site.

$$\begin{aligned}\Delta\Delta G_{bind} &= \Delta G_{bind,B} - \Delta G_{bind,A} \\ &= \Delta\Delta G_{site} - \Delta\Delta G_{solvent}\end{aligned}\tag{3.1}$$

In SepTop we obtain the relative free energy difference between two ligands in the binding site by running what is essentially two ABFE calculations at once in opposite directions. The thermodynamic path for the transformation of one ligand into the other ligand in the binding site (ΔG_{site}) is shown in Figure 3.1. To obtain the relative solvation free energy, $\Delta G_{solvent}$, we perform two absolute hydration free energy calculations if all ligands are neutral. If, on the other hand, ligands have the same non-zero charge we use a SepTop protocol in the solvent leg in order to preserve the net charge in the system (see Section 3.4.3).

Restraints are required to keep the weakly coupled and fully decoupled ligand in the binding site region and thereby reduce the phase space that needs to be sampled. In this study we apply orientational restraints, which we call “Boresch-style” restraints (after the seminal work of Boresch *et al.* which first employed these to make ABFE calculations practical).[19] In principle, however, numerous other kinds of restraints could be used for this step (affecting only the efficiency) and an assessment of the convergence of different pose restraint strategies is outside the scope of the present study. The efficiency of the approach naturally depends

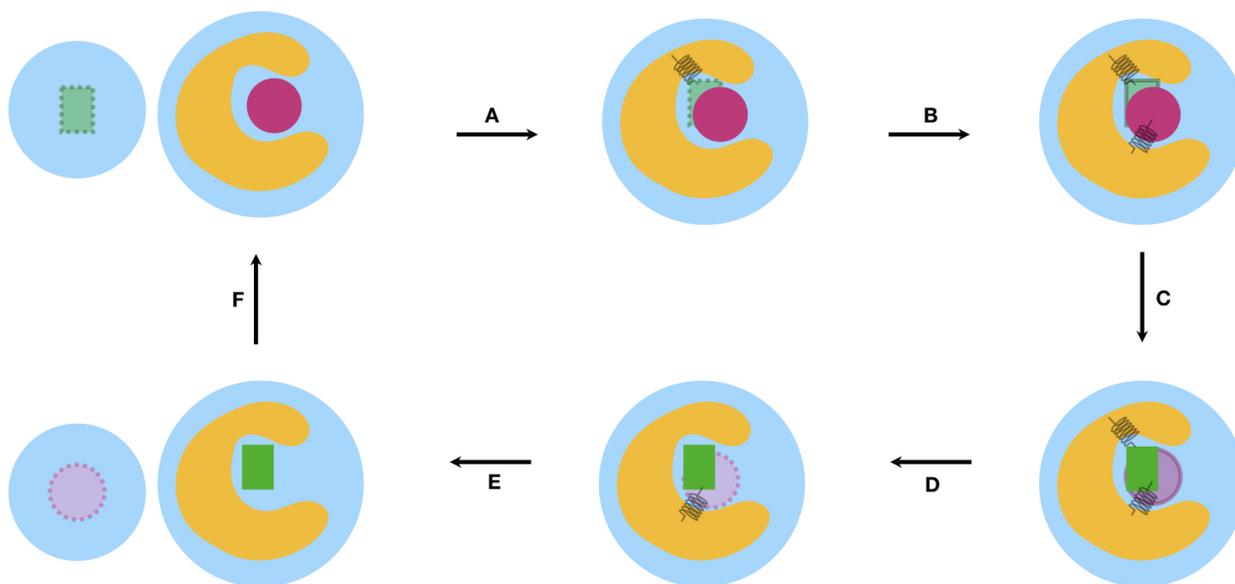


Figure 3.1: Thermodynamic cycle for SepTop for computing the free energy difference between two ligands in the binding site (F). The non-interacting dummy ligand (green) is inserted into the binding site and restrained using orientational (Boresch-style) restraints[19] (A). The van der Waals (vdW) interactions of the green ligand are turned on and the magenta ligand is restrained (B) and in the next step the electrostatic interactions of the green ligand are turned on while at the same time the electrostatics of the magenta ligand are turned off (C). Then, vdW interactions of the magenta ligand are turned off while at the same time releasing restraints on the green ligand (D). Lastly, the restraints of the now dummy magenta ligand are released analytically and the ligand transferred into the solvent (E). The free energy difference between the ligands in the solvent was obtained separately by running either two absolute hydration free energy calculations or a relative hydration free energy calculation using a SepTop approach.

on the choice of restraints, *e.g.* if two ligands share a similar shape, simulations would likely be most efficient if the shapes of the two ligands overlap well in all alchemical states. If the two ligands have a similar shape, one could restrain the shape of the first ligand to the shape of the second ligand so that in states where one of the ligands is only weakly interacting or fully decoupled, it samples the phase space of the interacting ligand it is restrained to. Such issues have not yet been carefully explored and are not the focus of the present work.

3.2.2 We developed heuristics for automatically picking suitable atoms for Boresch-style restraints

The orientational restraints used here restrain 3 atoms in the protein and 3 atoms in the ligand through 1 distance, 2 angle, and 3 dihedral restraints. Although the binding free energy should be independent of the atom selection[19, 100], the selection can impact convergence and (numerical) stability of the simulations. Therefore, we implemented a tool that selects suitable atoms for the restraints.

Multiple approaches to selecting stable atoms for Boresch-style restraints have been reported[82, 5, 21, 89], with some selection criteria being similar across implementations while other criteria differ. In these studies, equilibration simulations ranging from 1 ns[21] to 20 ns[5] were performed to help identify stationary points in the protein and ligand. Different approaches for identifying these stable structural elements were explored, such as selecting sets of atoms with the most frequent hydrogen bond and salt bridge interactions during MD[21], looking for buried residues with likely low mobility by calculating the minimal solvent-exposed surface area from the MD simulation[82] or choosing a combination of protein and ligand atoms that result in the lowest standard deviation across all six bond, angle and dihedral terms calculated across the equilibration simulation[5]. All methods have in common that only protein backbone (and C beta) atoms were considered for the restraints. Approaches varied

in the selection of the ligand atoms. Neighboring heavy atoms were considered[21], while others only considered heavy atoms within rigid scaffolds[82] to avoid restraining rotatable bonds and therefore locking-in conformations. A different approach selects ligand atoms with the farthest distance from one another[89].

Our approach builds on previous work by developing a heuristic algorithm aimed at identifying atoms that are likely to remain relatively stable as long as the ligand maintains the same binding mode, thus allowing them to effectively restrain the ligand’s motion as ligand interactions are removed. An example is shown in Figure 3.2a. Restraining ligand atoms that change their position during the simulation (Figure 3.2a I.) may lead to slow convergence while restraining stable ligand atoms (Figure 3.2a III.) can be more efficient. Similarly, convergence can be negatively affected if protein atoms involved in the restraints change their position substantially during a simulation. The algorithm is therefore designed to pick protein atoms that are likely to maintain a fairly constant position. Our aim here is to develop a restraining protocol that should work well on a large fraction of systems, however, we are aware that it is possible that no single algorithm will be ideally suited to all cases. The code for this restraining protocol is detailed below can be found in the GitHub repository `SeparatedTopologies`[11].

As one option, the equilibrated and minimized complex structure could be used to determine the Boresch restraint atoms. However, this has the drawback that it may not always be obvious which atoms will be stable from a single set of coordinates. A way to get around this is to use an entire trajectory. For example, protein-ligand complexes are often equilibrated and some data is collected prior to binding free energy calculations. Such simulations can be analyzed to help select atoms for restraints. In this work, we designed our restraints selection tool so that, if an input trajectory is provided, it is used to ensure that only atoms with relatively minimal fluctuations in their positions are considered as possible reference atoms for restraints. In particular, all protein and ligand atoms with an RMSF ≤ 0.1 Å are

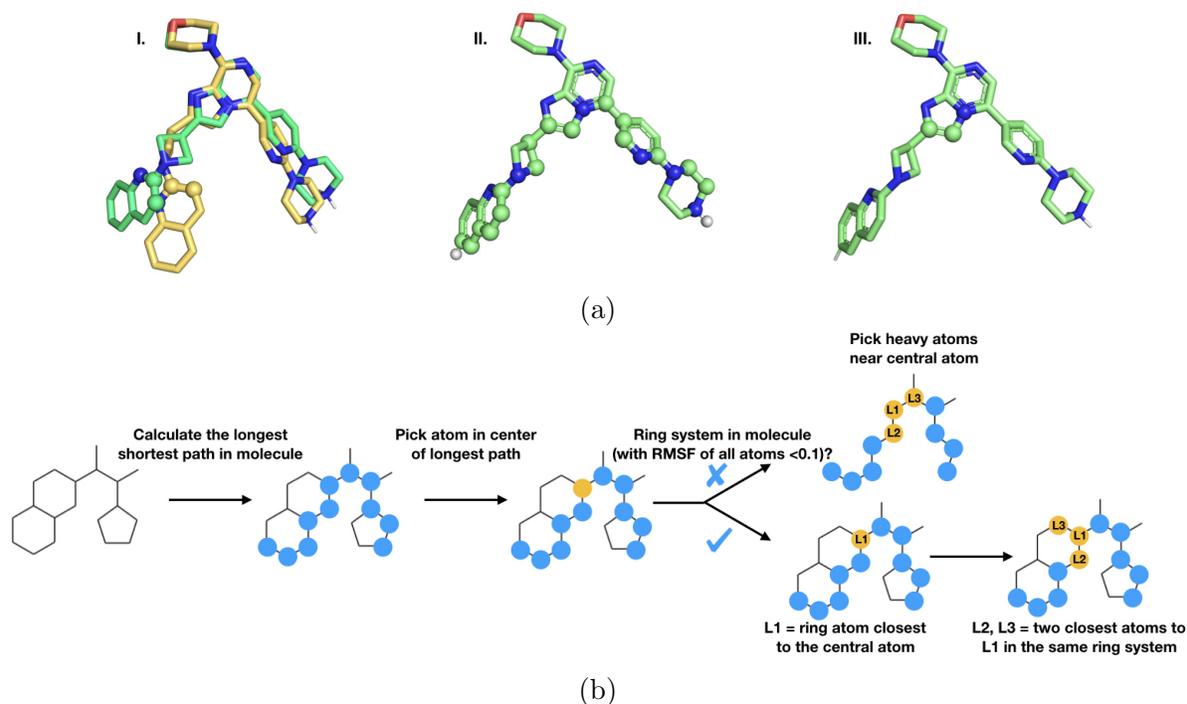


Figure 3.2: Choices made in selecting suitable ligand atoms for the restraints. (a) Selection of ligand atom for Boreesch restraints for a PDE10 ligand. (I.) Our initial version of the algorithm picked atoms from the largest ring system closest to the ligand center of mass (COM). The three selected atoms are shown in spheres. During simulations, this ring rotated from its initial structure (green) to a different binding mode (yellow). (II.) The new algorithm calculates all shortest paths between two atoms in the molecule and selects the longest path among those (shown in spheres). (III.) The three ring atoms closest to the middle of the longest path are selected for the Boreesch restraints. (b) Selection algorithm of ligand atoms based on the longest path in the molecule.

excluded and not considered for the restraints. While the ideal cutoff value might depend on the length of the input simulation, we found this cutoff at 0.1 \AA to be a practical threshold for simulations of 2 ns where frames are saved every 4 ps.

For ligand atoms, reference atoms for restraints can be chosen either automatically or by the user. The latter can be very useful if a ligand series has a structural element that is known to be stable or to be involved in key interactions in the binding site. We implemented an option for users to define their selections through a substructure search via SMARTS patterns. Automatic selection of ligand atoms on the other hand can be used when no prior information on stable ligand groups is available or if the series does not share a common

group. Here, the algorithm selects ligand atoms in a central ring system since a central ring system is likely more stable than other parts of the molecule.

More specifically, the tool computes all shortest paths between two atoms in the molecule graph, selects the longest path among those and picks the ring atoms closest to the middle of the longest path (Figure 3.2b). The algorithm then picks the center of the molecule using the longest path instead of the center of mass (COM) because we found that in one of the systems (PDE10) the later method led to the selection of atoms in an outer ring system which exhibited significant movements away from its original orientation during the simulation (Figure 3.2a I.). This suggested that if a distal ring system is used, even small local rearrangements of the ring could incorrectly appear to be substantially changing the ligand binding mode (as far as restraint calculations are concerned). In contrast, using a relatively central ring system to define restraints will ensure detection of substantial changes in ligand binding mode (such as an overall rotation or translation of the ligand in the binding site) and will certainly result in substantial changes of the relevant degrees of freedom in this case. Here, the PDE10 system helped improve the restraining protocol and is included here because it helped us develop the heuristics employed in our selection algorithm; however, it is not a focus of our study as we moved to other systems as soon as the selection algorithm was in place and did not further study PDE10.

Boresch-style restraints also require the selection of three reference protein atoms; for these, our algorithm initially considers all protein atoms, and then progressively filters out undesirable atoms (Figure 3.3). In this filtering process, the algorithm retains protein backbone atoms (as well as C-beta atoms) that are in the middle of an alpha helix or beta sheet since those are typically the most stable secondary structure elements. A trivial approach from here would be to select alpha helices only. However, this fails for proteins like galectin which do not have any helices. Therefore, we find that a more rigorous approach is to include backbone atoms in alpha helices if those are the dominant secondary structure and if not, include

both backbone atoms from helices and beta sheets. The algorithm picks atoms from central residues in those helices/sheets since outer residues can be more flexible and less stable. As mentioned above, only protein atoms with a RMSF ≤ 0.1 Å are retained. In addition to this, atoms have to be at least 10 Å and no more than 30 Å away from the ligand. The rationale for the 10 Å minimum distance is that binding site residues can undergo conformational changes upon ligand binding and therefore might be less stable as an anchor point.

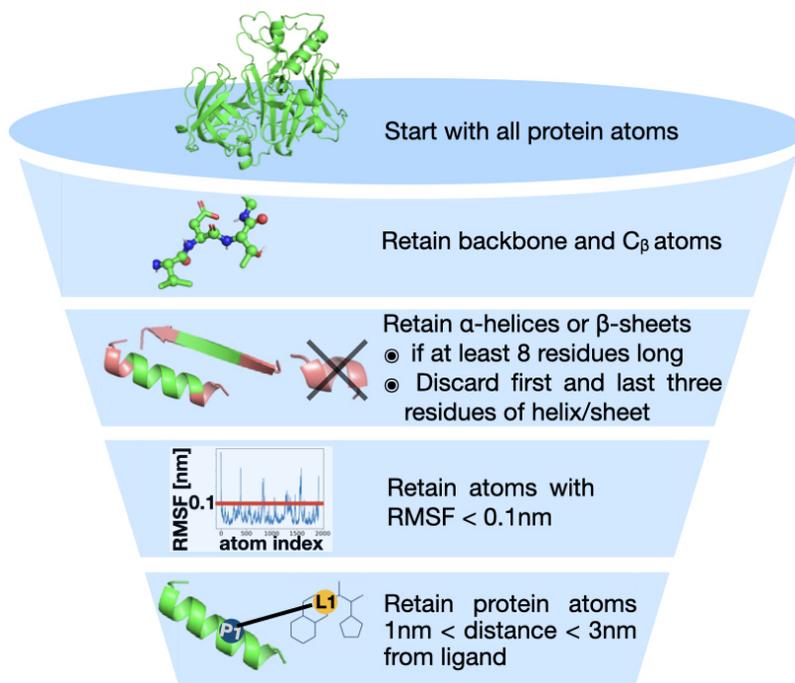
Adequate reference atoms in the protein must also satisfy several other factors. For example, if either of the two angle restraints are too close to 0° or 180°, the simulations crash due to numerical instability. To avoid this we implemented criteria to avoid near 0 or 180 degree angle restraints, more specifically that the two angle cutoffs a_{cut} are above 10 RT:

$$a_{\text{cut}}(180^\circ) = 0.5 * f_c * \left(\frac{\text{angle} - 180^\circ}{180^\circ} * \pi \right)^2 / \text{RT} \quad (3.2)$$

$$a_{\text{cut}}(0^\circ) = 0.5 * f_c * \left(\frac{\text{angle}}{180^\circ} * \pi \right)^2 / \text{RT}, \quad (3.3)$$

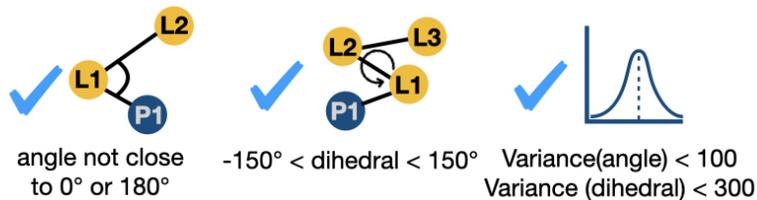
where f_c is the force constant.

In addition, atoms involved in the restraints must be sufficiently near one another, often less than half the shortest box edge away, to avoid problems due to periodic boundary conditions and the minimum-image convention. While larger separations would not in principle be a problem, restraints in certain simulation packages like GROMACS do not smoothly handle the minimum-image convention. For example, in one system (estrogen receptor alpha), a small movement of atoms involved in the restraints led to one of the restrained dihedral angles jumping between being computed “through” versus “around” the box, based on the

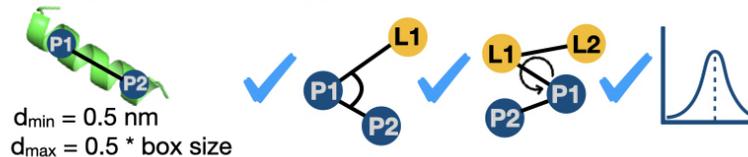


↓ Candidate protein atoms

For P1 pick an atom that passes these checks:



For P2 pick an atom that passes these checks:



For P3 pick atom that passes these checks:



Figure 3.3: Selection of suitable protein atoms for the restraints based on finding atoms that are likely to remain a constant position in MD simulations and other criteria that ensure numerical stability of the simulations.

minimum-image convention. The sudden jumps in dihedral angle due to these imaging issues then resulted in the decoupled ligand, even though restrained, leaving the binding site. Of course, this was simply an artifact of periodicity – but because of the handling of periodicity in the calculation of restraints in GROMACS, this resulted in sudden jumps in restraint energy/forces and caused problems. Thus, we adjusted our restraints selection procedure to avoid this problem.

Therefore, for the first protein atom, the algorithm takes all the protein atoms that came out of the filtering process described above and picks the first atom (P1) where the angle P1-L1-L2 between that protein atom and two of the ligand atoms is at least 10 kT from 0° or 180° (see Equation 3.2), where the dihedral angle P1-L1-L2-L3 is between -150° and 150° , and where the circular variance (as implemented in the SciPy package in Python) of that angle and dihedral angle are smaller than 100 or 300 degrees², respectively.

For the second protein atom P2 the algorithm picks an atom that is at least 0.5 nm away from the first protein atom, but no more than half the box size, and that passes the same angle (P2-P1-L1) and dihedral angle (P2-P1-L1-L2) checks as described above.

For the third atom P3, the algorithm picks the protein atom that is farthest from P1 and P2, but no more than half the box size away from them and where the dihedral angle P3-P2-P1-L1 passes the same checks as above. The same protein atoms are used for restraining both ligands if selected atoms pass the above checks in both systems. If the protein atoms selected for one protein-ligand complex are not suitable for the other protein-ligand complex, the algorithm selects different protein atoms for the second system. Ligand atoms, on the other hand, are selected independently for each ligand.

After the algorithm selects the six atoms for the restraints, the equilibrium position values for the bond, angles, and dihedrals are calculated either from a single input structure, or, if a trajectory is provided, from all provided frames and the mean value is used for the restraints.

We also found that the equilibrium length of the distance restraint has an impact on the mobility of the ligands, meaning that the chosen restraint force constant should vary with the distance restraint length chosen (if a constant level of ligand motion is the goal). In particular, the arc length s the ligand can move along around the surface of a sphere, where $s = r\theta$ and r is the distance and θ the angle P1-L1-L2, scales roughly quadratically with the distance. Therefore, we increase the force constant of that angle restraint quadratically with the distance.

3.3 Systems

We evaluated the performance of SepTop on four pharmaceutically relevant test systems.

We picked three ligands binding to tyrosine kinase 2 (TYK2) as a first test system. Those three ligands were structurally very similar and differ by small R-group changes (Figure 3.5). This allowed us to compare SepTop to standard RBF E making this system a good sanity check to ensure that the method is giving correct results. Using the same input structures and force field parameters from a previous study[48], the $\Delta\Delta G$ values of different RBF E methods should converge to the same result within statistical uncertainty.

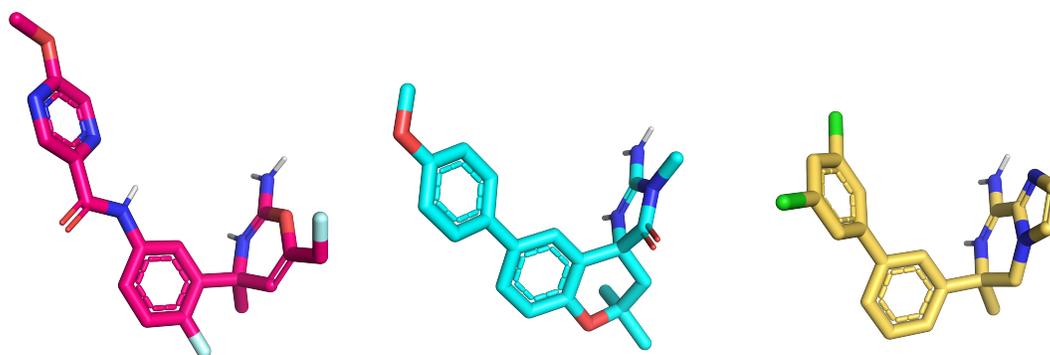
Estrogen receptor alpha ($ER\alpha$) systems have been studied by multiple groups for scaffold hopping transformation and were therefore an interesting next test system for SepTop. The transformations involve ring extensions; here, in particular, the key ring change is a transformation from a five to a six-membered ring (Figure 3.6). These scaffold hopping transformations fall outside the scope of regular RBF E, although they can be calculated using a soft-bond potential[154], additional restraints[169], or the alchemical transfer method[10]. Here, we wanted to investigate the performance of SepTop on challenging transformations like these and compare results with other methods.

We then tested the approach on a larger dataset of 16 ligands binding to MALT1 (Mucosa-associated lymphoid tissue lymphoma translocation protein 1). The ligands mostly differ by small R-group changes, however, there are a few ring formation transformations (isopropyl to cyclopropyl) and one stereo inversion transformation (see SI Figure B.1) that can pose challenges in standard RBF E if the chimeric molecule (the molecule that is composed of atoms to represent both ligands) is not created carefully.

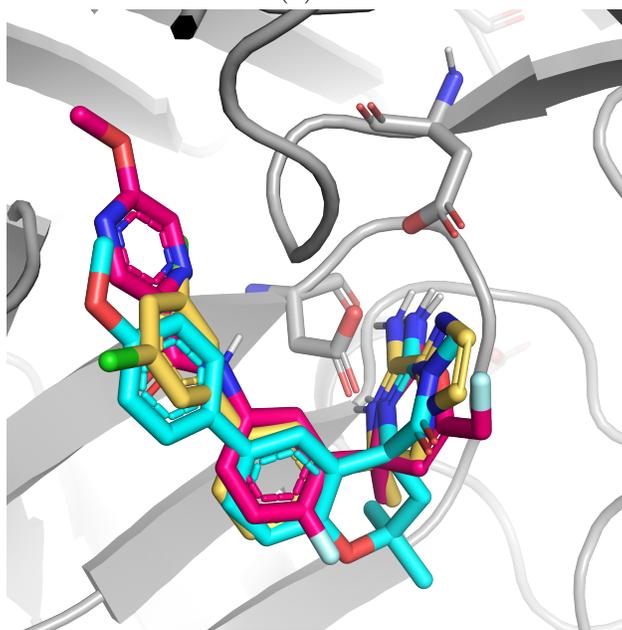
Lastly, we tested the performance of SepTop on charged ligands with diverse scaffolds binding to β -secretase 1 (BACE1). BACE1 has been used as a benchmark system for free energy calculations[156, 54]. Different scaffold series of BACE inhibitors have always been treated separately, meaning that separate RBF E maps were run for each scaffold[46, 22]. Here, we tested SepTop on ligand transformations both within different scaffold series, but also across different series. This dataset and those scaffold hopping transformations would be very challenging using standard RBF E methods making this a good test case for the domain of applicability of SepTop. We ran calculations for three different ligand series, the amide series, a spirocyclic series and a biaryl series (Figure 3.4). Different series conserve the interaction with the catalytic aspartate dyad, while the rest of the scaffolds are diverse. Especially the amide series, having a longer linker, extends into the P3 pocket, displacing some water molecules that are present in the two other series. We chose six ligands per series and performed calculations both within each series as well as across different series.

3.4 Simulation details

In the following section we describe the preparation of the proteins and ligands of the systems in this study. We then go over the setup of the systems for SepTop, as well as provide details about running and analyzing the free energy calculations.



(a)



(b)

Figure 3.4: Three scaffold series in the BACE1 system. (a) A ligand from the amide series is shown in magenta, spirocyclic series in blue, and a ligand from the biaryl series in yellow. (b) Overlay of the three scaffolds in the binding site. The catalytic aspartate dyad is shown in sticks. The three ligand scaffolds are diverse, *e.g.* the pyrazine ring of the ligand from the amide series extends into the P3 pocket of BACE1.

3.4.1 Preparation and parametrization of proteins and ligands

The proteins and ligands for the four systems in this work were prepared in order to generate input structures for the MD simulations. Topology and coordinate files of all systems can be found in the Supporting Information (SI).

For the TYK2 system, we selected three ligands from the protein-ligand Benchmark set (ligand codes: ejm 42, ejm 54, ejm 55 as defined in the Benchmark set[54]). We obtained the input coordinates and force field parameters from the protein-ligand benchmark set. Using the same input structures and parameters as in the previous study by Ge *et al.* [48] allows for a comparison of SepTop and traditional RBF E in specifically the efficiency of the free energy perturbation independent from the effect of the force field or system preparation. The ligands were parameterized with Open Force Field version 1.0.0 (codenamed “Parsley”)[123] and AM1-BCC charges[64]. The AMBER ff99sb*ILDN force field[90] was used to parameterize the protein and the TIP3P model[71] was used for the water. GROMACS[17, 1] was used to solvate the ligands and the protein-ligand systems and to add ions to reach a salt concentration of 150 mM. The output topology and coordinate files were then used to create the input files for SepTop.

For the ER α system, we ran calculations using structures prepared in three different ways. We prepared the first structure ourselves using a protocol we detail below. The other two input structures were used in the studies of Azimi *et al.*[10] and Zou *et al.*[169] and were provided by the authors. These systems were parameterized using Gaff1.8[153], Amber ff14SB[95], and TIP3P water[71]. To prepare our own structure, we began with PDB 2Q70. We performed the structure preparation on OpenEye’s Orion cloud computing platform, using their workflow (“floe”) “SPRUCE - Protein Preparation from PDB Codes” with the default parameters.[112] In that workflow hydrogen atoms were added, missing loops were built and crystallographic waters were retained. The two chains of the homodimer were

separated and chain A was used for the rest of the study. The binding mode and coordinates of the ligands were obtained from the SI of Zou *et al.* [169]. The prepared protein was aligned onto the protein provided by Zou *et al.* to be in the correct reference frame for the ligand coordinates. Orion floes “Bound Protein-Ligand MD” for the protein-ligand complex and “Solvate and Run MD” for the unbound ligands were used to solvate, parameterize, and equilibrate the systems. Ions were added to each to achieve a salt concentration of 50 mM. The GAFF1.81 force field[153] and AM1-BCC charges[64] were used for the ligands, Amber ff14SB parameters[95] for the protein, and TIP3P[71] for the water. In the Orion floes, the systems were energy minimized, equilibrated and then a production run of 2 ns was performed. The trajectory of that simulation was then used to create the input files for SepTop.

For the MALT1 system, we used PDB 7AK1 as the input structure since the ligand in that crystal structure was similar to the ligands in this study. We prepared the protein using OpenEye Spruce (Orion floe “SPRUCE - Protein Preparation from PDB Codes”). Missing loops were built and crystallographic waters were retained. The 16 ligands were then aligned onto the crystallographic ligand using OpenEye’s ShapeFit method[110, 73] as implemented in the SystemBuilder package[118]. Similarly as described above, Orion floes “Bound Protein-Ligand MD” for the protein-ligand complex and “Solvate and Run MD” for the unbound ligands were used to solvate and parameterize the systems. The Open Force Field version 1.3.1[151] and AM1-BCC charges[64] were used for the ligands, Amber ff14SB[95] for the protein, and TIP3P[71] for the water.

For the fourth system, BACE1, we chose PDB 6OD6 as input structure. We prepared the protein as described above using OpenEye Spruce. Chain A of PDB 6OD6 was then used for further simulations. The ligand SDFfiles from the amide and the biaryl series were selected from previous Janssen reports and had measured bioactivities from the same assay while ligand input files from the spirocyclic series were obtained from the protein-ligand Benchmark

set with measured bioactivities from a different assay[60]. The starting binding modes for the different scaffolds were obtained by overlaying the structures onto a crystallographic ligand using the SystemBuilder package as described above. For the amide series, PDB 6OD6 and its crystallographic ligands were used while for the spirocyclic series PDB 4JPC was used as input for the ShapeFit algorithm after aligning the PDB structure onto PDB 6OD6. PDB 3IN4, aligned onto PDB 6OD6 was used for the biaryl series. Different PDB structures were used in this step in order to perform the structural alignment onto a reference ligand that was most similar to a particular scaffold series. However, all protein-ligand complexes from different series were then prepared for MD simulations using the same protein structure, PDB 6OD6. In the bound state, the BACE1 catalytic aspartates Asp32 and Asp228 were both treated in their ionized forms. For some ligands, multiple potential binding modes had to be considered. For non-symmetric substituted phenyl rings, we performed SepTop calculations between different orientation of the R-groups and the more favorable binding mode was then used for further calculations. The solvated ligand and protein-ligand systems were then created with the Orion files as described above using the Open Force Field version 2.0.0[151] and AM1-BCC charges[64] for the ligands, Amber ff14SB[95] for the protein, and TIP3P[71] for the water. Sodium and chloride ions were added to reach a salt concentration of 150 mM.

3.4.2 Setup of SepTop systems

The solvated and parametrized systems were then further prepared for SepTop using a set of python scripts. Python scripts for performing these operations are currently housed in the GitHub repository `SeparatedTopologies`[11], which is under active development. This package contains multiple python scripts that can be used to generate all necessary input files for running SepTop in GROMACS, namely

- a coordinate file with the solvated protein-ligand complex having both ligands in the binding site,
- the topology files including the details for the orientational restraints, and
- the input topology, and coordinate files for calculations in the solvent.

The package takes as input the topology and coordinate files of the solvated protein-ligand complexes and the ligands in solution as well as a coordinate file for each of the two ligands (*e.g.* in the MOL2 or SDF format), and optionally a trajectory of an equilibrium simulation for assisting with the Boresch Restraint setup (see Section 3.2.2).

Coordinate preparation and topology file generation

This package performs a number of steps to set up each FEP transformation. First, for every transformation (or edge) it aligns the coordinates of the two protein-ligand systems using OpenEye Spruce[112]. Then it inserts the coordinates of ligand B into the coordinate file of the protein-ligand A complex, giving a coordinate file with both ligands present in the binding site. As a default, in a transformation from ligand A to ligand B the protein structure in complex with ligand A is used for the simulations, however, this can be easily adapted by the user to insert ligand A into the protein-ligand B complex.

In a next step the script creates the topology files needed to perform all transformations in the thermodynamic cycle. Two topology files are generated to describe all transformations in the binding site. This is necessary due to our multi-step alchemical pathway that first perturbs some of the vdW interactions (Figure 3.1 leg B), then the electrostatics (leg C) and finally the rest of the vdW (leg D). The first topology file describes leg B and leg C in the thermodynamic cycle and the second topology file describes the end states in leg D. The respective end states of the transformations were defined as an A and a B state in the [

`moleculetype`] section of the combined ligands in the topology files. Below we will describe the details for generating the topology files. Example output topology files are provided in the SI.

To generate the topology files, the script first inserts the topology of ligand B into the topology of the protein-ligand A complex. Then the [`atomtypes`] section is modified by adding atom types for dummy atoms that describe the non-interacting ligand, as well as “scaled” atom types, where the LJ- ϵ (see Equation 3.4) was scaled down by multiplying the LJ- ϵ by a scaling factor γ . The later atom types were then used for the enhanced sampling protocol, as described below. In addition, the atom type names for the ligand atoms were renamed to produce distinct atom type names for ligand A and ligand B. This was necessary in order to be able to exclude interactions between the two ligands in a next step by defining special non-bonded interactions between atom types. More specifically, the script adds a [`nonbond_params`] section to the topology file in which the vdW interactions between atom types of ligand A and atom types of ligand B are defined as zero.

The Coulomb interactions between the two ligands did not have to be excluded since in this thermodynamic cycle there are no end states where both ligands have electrostatic interactions turned on at the same time. As implemented in GROMACS, the Hamiltonian of an alchemical intermediate state is constructed by the linear interpolation of the Hamiltonians, rather than charges, *i.e.* $H = (1 - \lambda) * H_0 + \lambda * H_1$, where λ is the alchemical parameter and H is the Hamiltonian of the alchemical state, H_0 the Hamiltonian of end state A, and H_1 the Hamiltonian of end state B. This means that even though both ligands may have partial electrostatic interactions at the same time, the ligands will not interact with one another at any state of the alchemical transition, as long as the ligands are not interacting with each other in the end states. The vdW interactions between the two ligand, however, have to be excluded since there are end states in the alchemical path (both end states in leg C Figure 3.1) where both ligands have vdW partially or fully turned on.

Setup for ϵ -HREX

We scaled down the LJ parameters which, combined with Hamiltonian Replica Exchange (HREX)[160, 146, 161], can enhance the sampling of slow degrees of freedom, as has been shown in the implementation REST2[155]. Instead of running different replicas at different temperatures as in temperature replica exchange, all replicas are run at the same temperature while the potential energy of every replica is scaled differently. This can lead to an increase in the “effective” temperature of the system in the region where the interactions are being scaled and it was shown that this can improve the sampling efficiency compared to actual temperature replica exchange[155]. Both nonbonded and bonded parameters can be scaled down to enhance the sampling, however, here we only scale down the LJ- ϵ parameters of the ligands since we found that this was sufficient to improve sampling in most systems. In some cases we additionally scaled down the force constant of dihedral angles in the ligand to enhance slow rotamer sampling (see Section 3.5.4). We will refer to this protocol, as it modifies the LJ- ϵ and performs HREX, as ϵ -HREX throughout this work. The LJ potential, $V_{\text{LJ}}(r)$, is defined using an ϵ and a σ parameter

$$V_{\text{LJ}}(r) = 4\epsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^6 \right], \tag{3.4}$$

where r is the distance between two atoms.

The modification of the LJ parameters was implemented by adding new “scaled” ligand atom types. We multiplied the ϵ parameter of all original ligand atom types with a scaling factor γ to soften the potential. In this work we used a scaling factor of $\gamma = 0.5$, however, we found that in some cases it was necessary to further reduce the value to $\gamma = 0.01$ in some parts of the molecule in order to achieve sufficient sampling (see Section 3.5.3).

We adapted the thermodynamic cycle to incorporate this enhanced sampling method. In leg B of the thermodynamic cycle (Figure 3.1), instead of fully turning on the vdW interactions

of the green ligand, vdW interactions were only partially turned on to the modified LJ- ϵ (scaled atom type as described above). At the same time, the vdW interactions of the magenta ligand are partially turned off by transitioning from the original atom type to the scaled atom type. Both ligands have part of their vdW interactions turned on, and therefore, in the new leg D of the ϵ -HREX protocol, the vdW interactions of the magenta ligand are now fully turned off while the vdW interactions of the green ligand are fully being turned on. We compared the performances of this ϵ -HREX protocol with a protocol that did not use enhanced sampling in Section 3.5.1.

Restraint atom selection

In a next step in setting up the topology files, atoms for the Boresch-style restraints were selected according to the algorithm as described in Section 3.2.2. The restraints are defined in .itp files and added to the corresponding topology files. In this work, a force constant of 20 kcal mol⁻¹ Å⁻² was used for the bonds and 20 kcal mol⁻¹ rad⁻² for one of the angles and three dihedrals. The force constant for the other angle varied depending on the bond distance. We set it up such that for a bond distance of 5 Å a force constant of 40 kcal mol⁻¹ rad⁻² would be used for that angle restraint and with increasing distance that force constant was scaled quadratically. The contribution for releasing the orientational restraints when the ligands are non-interacting (Figure 3.1 leg A and leg E) was calculated analytically using equation 32 from Boresch *et al.* [19].

3.4.3 Setup for calculations in the solvent

For calculations in the solvent, we used two different protocols, depending on whether the ligands have a net charge. If all ligands in the dataset were neutral, we performed absolute hydration free energy calculations. If ligands were charged we performed an analogous

'separated topology' hydration free energy calculation, where the solvation of the two ligands were performed simultaneously in opposite directions. (As a test of this approach, we transformed a charged ligand into itself in solution and confirmed that the computed free energy converged to zero.) The two approaches required different scripts to generate the respective topology and coordinate files. For the absolute hydration free energy calculations the tool writes a topology file with a fully non-interacting ligand in the B state. For the charged ligands in the BACE1 dataset we ran SepTop calculations for the solvent part of the thermodynamic cycle in order to preserve the net charge throughout the simulation. We restrained the ligands to be half the box edge away from each other using a single harmonic distance restraint between the heavy atoms closest to the COM of the two ligands using a force constant of $2.4 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$. One ligand was then fully decoupled while at the same time the other ligand was fully coupled using a similar protocol as in the binding site with the difference that in the solvent, the LJ- ϵ was scaled with a scaling factor $\gamma = 0.3$ rather than $\gamma = 0.5$ as in the binding site. The system was first equilibrated for 10 ps in the NVT ensemble at a force constant of $0.0024 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ to allow the ligands to gently adjust to the distance restraint, followed by a 10 ps equilibration in the NVT ensemble at the full restraint force constant.

3.4.4 Running SepTop in GROMACS

All MD simulations were performed using GROMACS 2021.2[17, 1].

For simulations in the binding site, we used an alchemical path with a total of 20 λ windows, 8 states for leg B, 5 for leg C and 8 for leg D. Leg B and leg C were run in one step, meaning that a single topology file and therefore only two end states were used for leg B and leg C together. This combined step then used 12 λ windows, making this a total of 20 λ windows. The details of the alchemical path can be found in the free energy calculations section of the

MDP files provided in the SI.

For the unbound part of the thermodynamic cycle, we performed absolute hydration free energy calculations for all neutral ligands. Here, we used 14 λ windows, first turning off the Coulomb interactions and then the vdW interactions of the ligand. For the charged ligands in the BACE1 dataset we ran SepTop calculations (to avoid changing the formal charge of the system [130]) in the solvent using a total of 28 λ windows (see MDP files in SI for details); essentially, this involved 14 for each ligand.

All λ windows were first energy minimized using steepest descent for 5000 steps, and then equilibrated in the canonical ensemble for 10 ps at 298.15 K. A production run of 10 ns per λ window was performed in the NPT ensemble at a pressure of 1 bar. The MD simulations were performed using the stochastic dynamics integrator at a timestep of 2 fs. The soft-core potential from Beutler *et al.* [18] was applied to avoid instabilities in intermediate λ windows. Replica exchange swaps were attempted every 200 steps. Full details of simulation parameters can be found in the MDP files in the SI.

In the TYK2 system, we restrained rotatable bonds in the ligand using dihedral restraints after noticing that slow rotamer sampling led to poor convergence. We adapted the thermodynamic cycle to account for the contribution of these restraints. In step A (Figure 3.1) the non-interacting ligand B (green) was inserted from the standard state having all its rotatable bonds restrained. Then, dihedral restraints on the interacting ligand A (magenta) were turned on simultaneously with the Boresch-style restraints (step B). The next step is identical to that in the original protocol, except that it has dihedral angles on both ligands restrained. In step D the restraints on ligand B (green) are released and in step E the non-interacting ligand A with its rotatable bonds restrained was transferred to the solvent. The interactions of ligand A were then turned back on in the solvent and the dihedral restraints were released. Similarly, in the hydration free energy calculations of ligand B, first the dihedral restraints were turned on, followed by the decoupling of the ligand and then

non-interacting ligand B was inserted into the binding site in step A as described above.

Restraining rotatable bonds can lead to faster convergence in some cases, however, it can also lead to slower convergence of the free energy estimate in other cases. If *e.g.* the dihedral angle rotates freely in the solvent, releasing dihedral restraints in the solvent can require the addition of λ windows to obtain sufficient overlap of the work distributions. Therefore, our general SepTop protocol does not restrain rotatable bonds.

3.4.5 Analysis of the results

The free energy difference was obtained from simulation data using the MBAR estimator[140] as implemented in the `alchemlyb/pymbar` package[140, 36]. The first nanosecond of the 10 ns production run was discarded, as additional equilibration. In the TYK2, ER α and BACE1 systems, we ran three independent repeats to assess the convergence of the free energy estimate and to get an estimate of the uncertainty, though not in the MALT1 system to reduce compute costs and to get a better idea of typical production-level accuracy. In the MALT1 and BACE1 systems, absolute binding free energies were obtained from the $\Delta\Delta G$ values using a maximum likelihood estimator as implemented in the `arsenic` package[94] (later renamed to `cinnabar`). That same package was used to generate correlation plots and the error and correlation statistics used in this chapter.

3.5 Results

In this study, we investigated the performance of an alternate approach for RBF, SepTop, on several pharmaceutically relevant systems – specifically, TYK2, ER α , MALT1 and BACE1. Our focus here is on validating the method for such systems rather than the model system studied previously[129] and on ensuring the approach is relatively robust and accurate across

several different representative targets.

3.5.1 We first tested SepTop on the well studied TYK2 dataset as a sanity check

The TYK2 ligands that we investigated here only differ by small R-group changes which allowed us to compare SepTop with standard RBFE to ensure results are reasonable on a system where we can obtain correct binding free energies for the model with a well-established method. Using the same input structures and force field parameters as in a previous study[48], SepTop and standard RBFE using a non-equilibrium switching protocol gave similar relative binding free energies for three transformations in the TYK2 system. A low cycle closure error of 0.0 ± 0.5 kcal mol⁻¹ (SepTop ϵ -HREX Figure 3.5) suggests that $\Delta\Delta G$ estimates were well converged.

After finding that the SepTop approach worked well on this test system, we used this system as a test case to help us optimize our protocol and increase its efficiency. We found that running each λ window for 10 ns vs. 20 ns gave relative free energies which were statistically indistinguishable, so we decided to use a simulation time of 10 ns per λ window in future systems. Our initial protocol used 45 λ windows for calculations in the binding site, but we found that we could reduce the number of windows to 20 while retaining good overlap between neighboring states. We used the overlap matrix as implemented in `alchemlyb`[36] to help find a reasonable alchemical pathway and spacing with sufficient overlap.

We also investigated whether the use of enhanced sampling improved the convergence and efficiency of the simulations. We compared three protocols, one without enhanced sampling, another using HREX and a third protocol, ϵ -HREX, where we scale down the LJ- ϵ by a factor of $\gamma = 0.5$. Scaling down the LJ- ϵ can soften interactions and lower energy barriers between conformational states, and can therefore, coupled with replica exchange swaps,

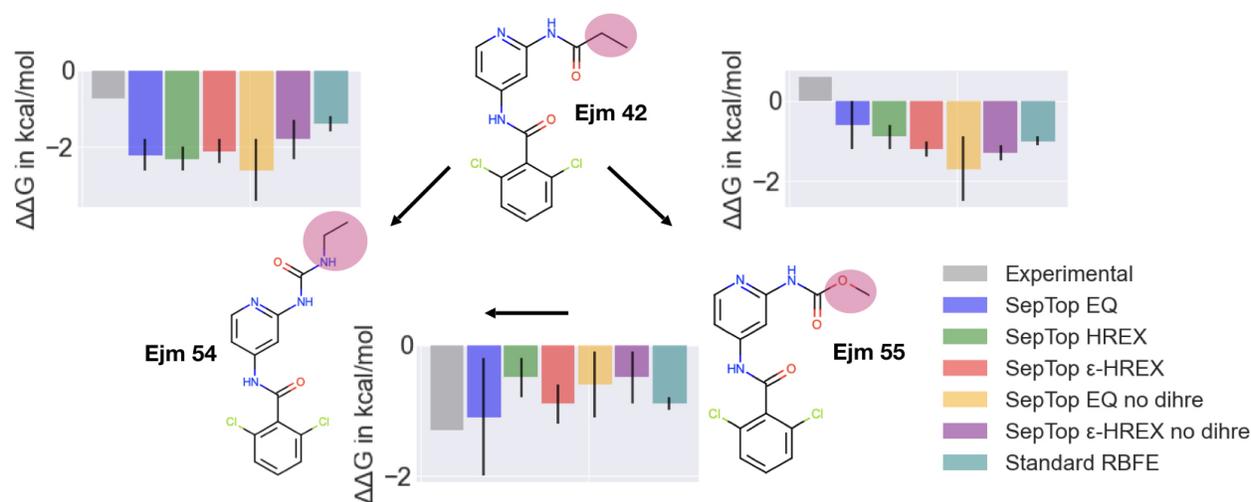


Figure 3.5: Perturbation cycle in the TYK2 system. For all three transformations, we report the experimental (grey) and calculated relative binding free energies. In the SepTop approach, five different protocols were tested, two without enhanced sampling (EQ and EQ no dihre), one using HREX, and two protocols using ϵ -HREX. In the first three protocols, rotatable bonds were restrained using dihedral restraints while in the fourth and fifth protocol (purple and yellow) no dihedral restraints were applied. The last protocol (teal) shows values obtained from the work of Ge *et al.* using standard RBF E[48]. SepTop and standard RBF E produced similar results. With SepTop, different protocols converged to the same relative free energies within uncertainty and the standard deviation across trials was lower when using enhanced sampling techniques.

improve sampling. All three protocols converged to statistically the same relative binding free energies (Figure 3.5). The standard deviation, calculated across three independent repeats, was highest when no enhanced sampling was used, while the SepTop HREX and SepTop ϵ -HREX protocols had similar values for the standard deviation. Here, we were primarily focusing on comparing results from different methods with one another rather than comparing to experimental binding affinities, because methodological improvements do not always improve agreement with experiment. In particular, agreement with experiment is not just a function of sampling, but of multiple factors, such as the force field and the choice of protonation states, counterions, etc., meaning that the best method may not necessarily agree best with experiment except when all other issues have been addressed.

Restraining all rotatable bonds in the ligands increased the efficiency of these simulations,

suggesting that sampling of different rotamers might be slow in this system. The standard deviation of the protocols that used dihedral restraints was overall lower than in the two protocols where the rotatable bond was not restrained (Figure 3.5). The thermodynamic cycle was adapted accordingly to account for the contribution of the restraints (Section 3.4). Using an enhanced sampling protocol that softens the vdW parameters and performs HREX (Figure 3.5 `SepTop ϵ -HREX no dihre`) improved the sampling of different rotamers over the standard protocol (`SepTop EQ no dihre`), shown by the smaller standard deviation across three independent repeats.

3.5.2 Scaffold hopping transformations with SepTop on ER α systems

ER α has been studied by multiple groups benchmarking different scaffold hopping approaches and therefore appeared to be a good test system for our method, giving us the possibility to compare SepTop with other methods. Here, we compared SepTop with three different RBF E methods, the alchemical transfer method (ATM)[10], as well as a RBF E method that uses auxiliary restraints for scaffold hopping transformations[169] and lastly FEP+[154]. For these three ligand transformations, different methods converged to similar relative binding free energies (Figure 3.6). Using the same input structures provided as in the auxiliary restraints paper[169], SepTop converged to the same $\Delta\Delta G$ value within uncertainty as the method using auxiliary restraints in the two edges where calculated values had been reported.

The comparison of results from different methods was challenging due to differences in system preparation in the different studies. We found that the modeling of missing loops (Figure 3.7a) and the protonation state of histidine residues (Figure 3.7b) had an impact on the predicted binding free energy. We ran calculations with the SepTop approach using an input

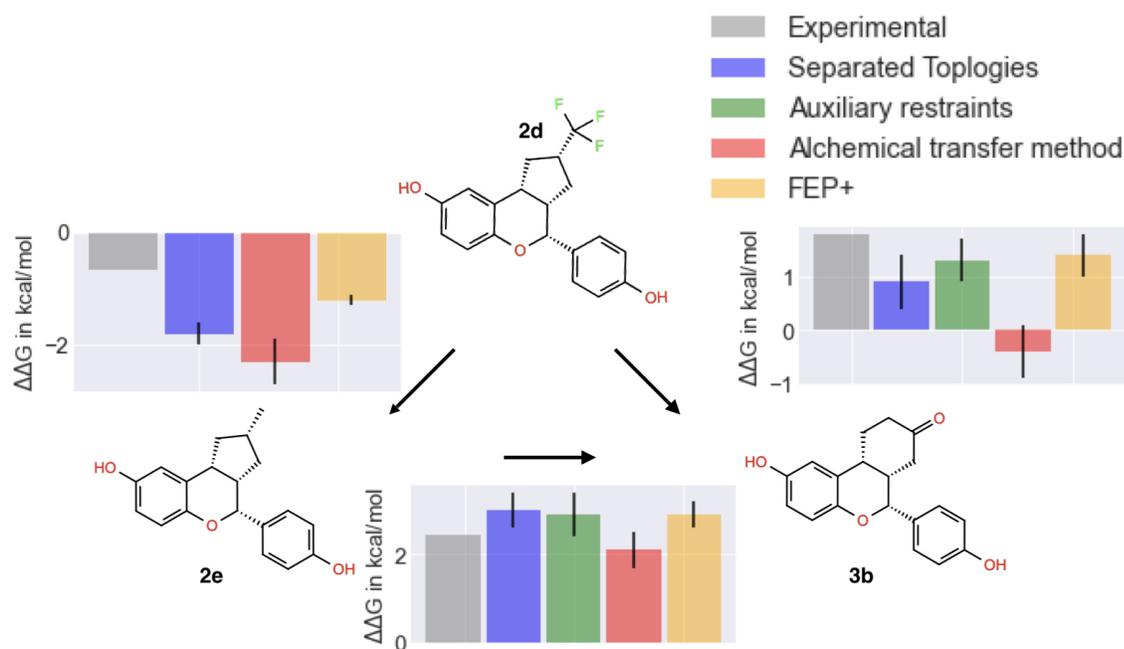


Figure 3.6: Ligand cycle in the ER α system. For all three transformations, we report the experimental and calculated relative binding free energies. The error estimate in SepTop is the standard deviation calculated across three independent repeats. Overall, different methods gave similar results, however, they did not always converge to the same $\Delta\Delta G$ values.

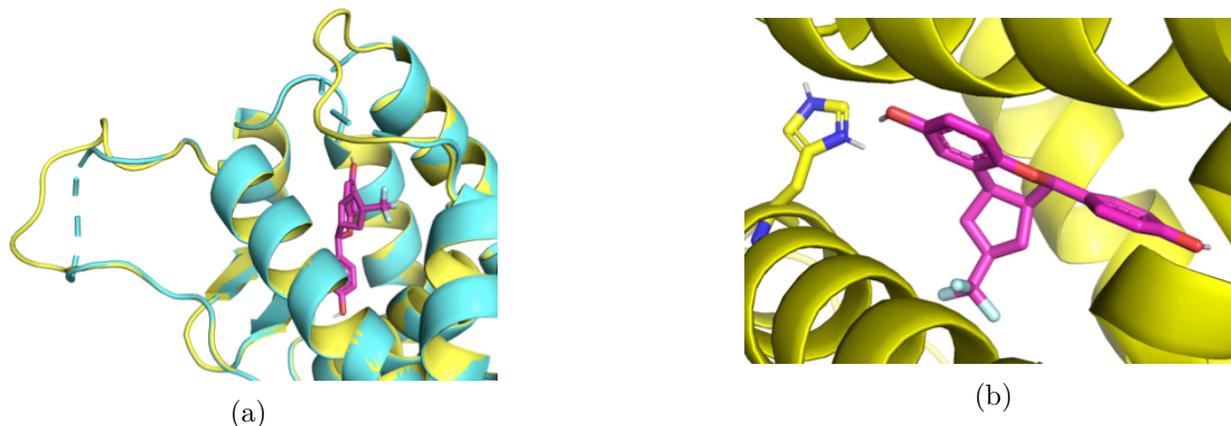


Figure 3.7: Different choices made in ER alpha preparation. (a) Two loops were not resolved in the crystal structure and missing residues were modelled in some studies[169], while in other studies the residues bordering the missing loop were capped and restraints used to keep the protein from drifting apart[10]. (b) Different software for the assignment of protein side chain protonation states resulted in differences in protonation state of some histidine residues. One of those residues is located in the binding and the selected protonation state impacted the $\Delta\Delta G$ estimate, at least in our study.

structure we prepared using OpenEye’s Spruce[112] (SI Table B.1 input `spruce`), and with prepared structures used in the studies of Azimi *et al.*[10] (input `ATM`) and Zou *et al.*[169] (input `Aux`) and obtained very different results (SI Table B.1). When using the input structure that had missing loops we had to restrain the position of backbone protein atoms to prevent the protein from drifting apart. This could have potentially impacted the results and partly explain why calculations starting from different input structures did not give equivalent results. For one of the transformations (edge 2d - 2e), we performed additional calculations to study the impact of the protonation state of histidine on the results. In one protocol the histidine in the binding site was charged (HIS220, Figure 3.7b) and in another protocol HIS220 was neutral, while in one protocol the protonation state was not reported. We found that the protonation state of that histidine in the binding site affected the calculated $\Delta\Delta G$ by $\approx 1 \text{ kcal mol}^{-1}$. Our finding also shows that depositing prepared protein structures along with a publication is critical for reproducibility of results.

3.5.3 We tested SepTop on a larger dataset, 16 ligands binding to MALT1

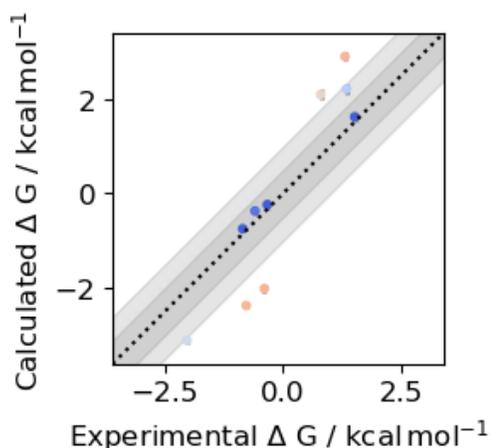
We also tested SepTop on a larger dataset, examining a series of 16 ligands binding to MALT1, and obtained good correlation and error statistics. We created a ligand map in which the 16 ligands were connected with 30 edges. To setup the ligand map, we chose one of the most potent ligands as the reference ligand and connected all other ligand to that reference ligand through edges, creating a star map. We then added additional edges between ligands to create ligand cycles. Absolute binding free energies were obtained from calculated $\Delta\Delta G$ values and experimental binding free energies using the arsenic code[94]. Since the experimental binding affinities for 6 of the compounds were outside the assay limit, we excluded those ligands from the correlation and error statistics. Correlating the calculated binding free energy of the 10 remaining ligands with the experimental values resulted in an $RMSE = 1.06$ and $R^2 = 0.85$. Additional statistics can be found in Figure 3.8a. The full set of calculated and experimental values can be found in the SI.

Assessing and addressing sampling problems in the MALT1 system

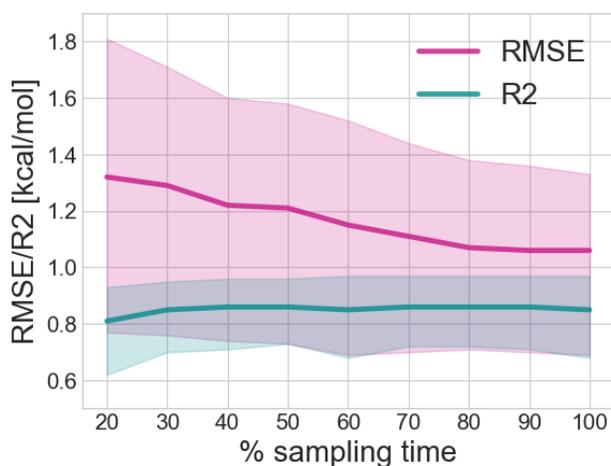
A cumulative and cycle closure analysis suggested that a sampling time of 10 ns per λ window was adequate for this system. Simulations were mostly converged after 10 ns based on both the confidence interval of the RMSE getting within 0.7 kcal/mol and the overall low cycle closure errors. We analyzed the correlation and error statistics and the cycle closure at different amounts of sampling time to assess the convergence of the free energy estimate. As seen in Figure 3.8b, the RMSE decreased with increasing simulation time until reaching a value of $RMSE \approx 1.1$ kcal/mol after around 80% sampling time, while the R^2 remained at a value around 0.85 after 30% sampling time. Since the RMSE stopped decreasing after 80% sampling time, we concluded that simulations had likely converged.

100 % sampling time
(N = 10)

RMSE:	1.06	[95%: 0.71, 1.37]
MUE:	0.86	[95%: 0.45, 1.27]
R2:	0.85	[95%: 0.67, 0.98]
rho:	0.92	[95%: 0.83, 0.98]
RAE:	0.86	[95%: 0.48, 1.88]
KTAU:	0.69	[95%: 0.27, 0.90]



(a)



(b)

Figure 3.8: (a) Correlation between predicted and experimental binding free energies for 10 MALT1 inhibitors. The plot was generated using arsenic[94]. Binding free energies centralized around zero are shown. Root mean square error (RMSE), mean unsigned error (MUE), coefficient of determination (R^2), Pearson correlation coefficient (ρ), relative absolute error (RAE), and Kendall tau (KTAU) are reported with a 95% confidence interval. A good correlation with experimental binding free energies was obtained in this system. (b) Analysis of the RMSE and the R^2 at different amounts of sampling time. The 95% confidence interval is shown as a shaded area. The RMSE remained at a value below 1.1 kcal/mol after 80% sampling time.

Assessing how well $\Delta\Delta G$ values in a closed ligand cycle sum up to zero (cycle closure) is an additional metric for convergence. If the edges in a closed ligand cycle do not sum up to zero, simulations are not converged. However, if a cycle closes to zero, this does not necessarily mean that there are no sampling problems since there might be cancellation of errors. Here, we normalized the cycle closure to obtain an approximate contribution of a single edge to the cycle closure error:

$$CC_j = \frac{\sum_{i=1}^{n_i} \Delta\Delta G}{\sqrt{n}} \quad (3.5)$$

where CC_j is the cycle closure of cycle j and n is the number of edges in cycle j . For the cycle closure analysis, we considered the entire ligand dataset, including the compounds with measured binding affinities outside the assay limit. Ligand cycles were enumerated using functions from the NetworkX package[53] which were modified by prior authors[70] to handle this problem. Most of the ligand cycles (51/87) had a cycle closure below 0.5 kcal/mol, however, for 6 ligand cycles, the cycle closure was above 1 kcal/mol, indicating sampling problems (see SI Figure B.2).

We introduced a voting system to help identify bad edges, meaning transformations that likely have not converged yet. Even this relatively small dataset of 16 ligands resulted in 87 different ligand cycles which made it difficult to identify which specific transitions were unconverged by manually looking at the cycle closure errors. Our voting system provided a more automated way to assess this. In this voting system, if a cycle had a cycle closure above a certain threshold (0.7 kcal/mol), each transformation in the cycle was assigned a penalty (+1), while each transformation in a cycle with a cycle closure below a certain threshold (0.3 kcal/mol) received a positive vote (-1). The votes for each transformation were then summed up and the edges with the most positive overall votes were investigated further to identify potential sampling problems.

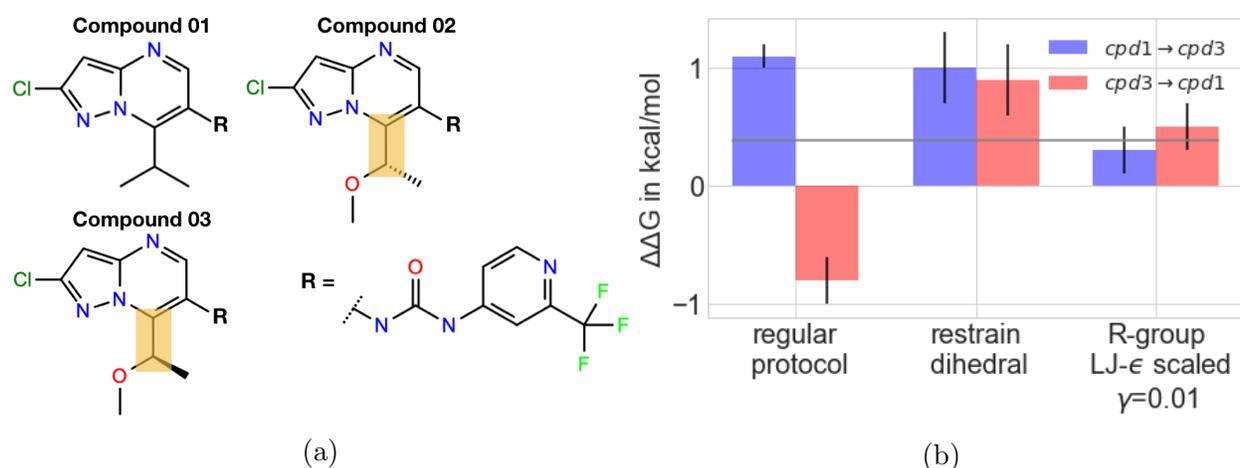


Figure 3.9: (a) 2D structure of MALT1 ligands, compounds 01-03. The yellow box highlights the rotatable bond whose rotamers were not sampled sufficiently and therefore caused sampling problems (b) $\Delta\Delta G$ values for transformations from compound 01 \rightarrow compound 03 (blue) and compound 03 \rightarrow compound 01 (red, here the negative of the $\Delta\Delta G$ is shown) using three different protocols. In the “regular protocol” the LJ- ϵ of the entire molecules was scaled down by multiplying it with a scaling factor of $\gamma = 0.5$. In the “restrain dihedral” protocol the rotatable bond shown in yellow (a) was restrained during the free energy calculation and the contribution of the restraint was accounted for by releasing the restraints afterwards. In protocol “R-group $\epsilon=0.01$ ” the LJ- ϵ of the ethyl methyl ether group was scaled down by a scaling factor of $\gamma=0.01$, allowing Hamiltonian lambda exchange to help accelerate sampling of this bond rotation. The grey line shows the experimental binding free energy. Using the regular protocol, transformation from both directions do not converge to the same relative free energy due to sampling problems, while transformations in both directions using the two other protocols converged to the same free energies as well as did transformations within the same protocol.

This voting system identified the transformation from compound 03 to compound 01 (Figure 3.9a) as the worst edge and indeed, follow up work showed that this transformation suffered from significant sampling problems. For this transformation the standard deviation across three independent repeats was low (0.2 kcal/mol). However, running the transformation in the opposite direction (compound 01 \rightarrow compound 03) gave very different results ($\Delta\Delta G = 1.1 \pm 0.1$ kcal/mol) than the original direction ($\Delta\Delta G = 0.8 \pm 0.2$ kcal/mol, before accounting for the sign flip). This difference of 1.9 kcal/mol depending on the direction indicated sampling problems in this edge.

We found that insufficient sampling of different rotamers around the bond highlighted in

yellow (Figure 3.9a) caused these sampling problems. Rotation around a rotatable bond in compound 03 (Figure 3.9a) was slow in some λ windows and depended on whether this ligand started as a dummy or fully interacting. Analyzing the free energy change caused by modifying the vdW/restraints versus the Coulomb interactions separately showed that the difference in the $\Delta\Delta G$ depending on the direction mostly happened in the λ windows modifying the Coulomb interactions. As described in Section 3.4.2, replica exchange swaps were carried out between all λ windows in legs B and C in the thermodynamic cycle (Figure 3.1) while leg D was run in a separate calculation. In the transformation going from compound 03 to compound 01, compound 03 started as a fully interacting ligand, and only a single rotamer of the R-group (Figure 3.9a) was sampled in all λ windows of leg B and leg C. Meanwhile, when going in the opposite direction, from compound 01 to compound 03, compound 03 started as a dummy ligand in leg B, sampling multiple rotamers. Here, in contrast to the transformation in the other direction, two different rotamers were sampled in the λ windows in which the Coulomb interactions were modified (leg C), benefiting from broad rotamer sampling in non-interacting and weakly interacting states through replica exchange. In this example, the rotamer distribution differed across different states in the alchemical pathway, and since for some states the equilibrium rotamer distribution was not sampled, results were incorrect.

We were able to improve the sampling of different rotamers by scaling the LJ- ϵ down by multiplying it by a scaling factor of $\gamma = 0.01$. Free energies for transitions in the two opposite directions now converged to the same result (Figure 3.9b). Only the LJ- ϵ of atoms in the ethyl methyl ether group (Figure 3.9a) were scaled down by a factor of $\gamma = 0.01$, while for the rest of the molecule a scaling factor of $\gamma = 0.5$ was used. We had tried to scale down the LJ- ϵ of the entire molecule, however, found that this led to instabilities and convergence problems. We also found that convergence could be apparently achieved by restraining the dihedral of that rotatable bond during the simulation and releasing the restraints afterward. This however then led to higher standard deviations and slow convergence when releasing

the dihedral restraint in the solvent. In the solvent, different rotamers were sampled in the interacting state which led to poor overlap of λ windows when releasing the dihedral restraint (results not shown). Overall, in this case the protocol which scaled back interactions of the mutated R-group performed best.

Comparing SepTop with a standard RBE method

We compared SepTop with standard RBE using the non-equilibrium switching (NES) method as implemented in Orion[111] (Orion file “Equilibration and Non Equilibrium Switching”) using the default 6 ns equilibration of the end states and 80 non-equilibrium switching transitions with 50 ps switching time. We used the same ligand map and force field parameters as with the SepTop protocol. We only included ligands with measured binding affinities within the assay limit in our analysis as we did above.

We found that SepTop produced better correlation and error statistics than the more standard NES approach (Figure 3.10). SepTop used more sampling time than NES Orion (for 30 transformations: Bound state: 6000 ns vs. 336 ns; Unbound state: 2240 ns vs. 336 ns) which, in addition to problems detailed below, might be part of the reason why SepTop performed better on this system. Since it would have been cost prohibitive to increase the length of the simulations on Orion to the simulation time used in the SepTop protocol, we decided not to perform a direct comparison at equal sampling time. The $\Delta\Delta G$ values calculated using the two different methods mostly agree well with one another, as can be seen in SI Figure B.3 and SI Figure B.4. The outliers in the plot indicate that for some of the transformations the two methods did not agree with each other. Two of those outliers involved transformations going from a hydrogen (Pfizer-01-05) or a methyl group (Pfizer-01-04) to a cyclopropyl (Pfizer-01-07). The overlap of forward and reverse work distributions of the nonequilibrium switching transitions was poor (SI Figure B.5), indicating insufficient sampling of important motions. A third outlier was a transformation that involved the inversion of a chiral center

which was potentially not treated correctly in the NES protocol, as discussed below. For the two outlier transformations Pfizer-01 \rightarrow compound 02 and Pfizer-01 \rightarrow compound 03, NES predicted a more dramatic change in potency than SepTop. These ligands contained the R-group that had caused sampling problems due to slow rotamer sampling in SepTop (see Section 3.5.3) which potentially also caused problems in the NES approach.

With the NES approach, some of the transformations in this set can potentially be challenging for standard RBFE calculations if the chimeric molecule is not set up carefully. Those transformations involve stereo center inversions (compound 02 \rightarrow compound 03) and ring forming/ring breaking transformations (isopropyl to cyclopropyl transformations). For the latter, chimeric molecules should be created where the entire group is included in the transformation rather than forming/breaking a ring. The chiral inversion transformation was potentially not handled correctly in the NES protocol, possibly due to a bug in the OpenEye implementation.

3.5.4 Testing SepTop on large scaffold hopping transformations in the BACE1 system

We ran RBFE calculations for three different series of BACE1 inhibitors, each based on a different scaffold, to test the performance of SepTop on large and challenging transformations. We picked six ligands per series and ran calculations within each series as well as spanning between the different series.

Since all BACE1 inhibitors in this study were positively charged, instead of running absolute hydration free energy calculations in the solvent leg of the thermodynamic cycle, we performed relative hydration free energy calculations using a similar SepTop approach as in the binding site. Running absolute hydration free energy calculations of charged ligands in the solvent would have led to a change in the net charge of the system (which is difficult to treat

			(N = 10)
RMSE:	1.89	[95%:	1.08, 2.48]
MUE:	1.50	[95%:	0.83, 2.22]
R2:	0.66	[95%:	0.38, 0.88]
rho:	0.81	[95%:	0.62, 0.95]
RAE:	1.50	[95%:	0.80, 3.48]
KTAU:	0.51	[95%:	0.09, 0.83]

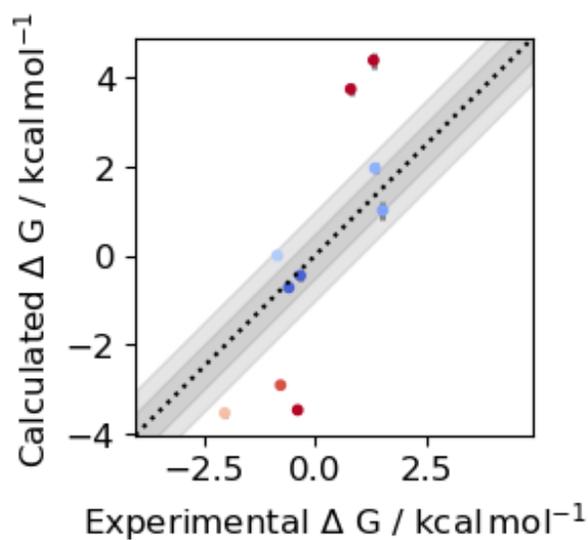


Figure 3.10: Correlation between calculated (NES Orion) and experimental binding free energies for 10 MALT1 inhibitors. Plot was generated using arsenic[94]. The correlation and error statistics are worse than the ones obtained using SepTop (Figure 3.8a). A comparison between NES and SepTop values is shown in SI Figure B.3.

for technical reasons [130]), while the net charge can be preserved using a relative approach. The ligands were restrained to be half the box edge away from each other using a single harmonic distance restraint (see Section 3.4.2). We had first attempted restraining ligands to be on top of each other using a single harmonic distance restraint, however, we found that this led to slow convergence of the free energy estimate in some cases. We therefore decided to restrain ligands to be far apart.

When multiple binding poses were plausible (as determined by overlaying the ligand onto co-crystallized ligands of different PDB structures using a maximum common substructure overlay as well as docking compounds into the site), we ran SepTop between different binding modes to identify the more favorable pose. Especially for asymmetric phenyl substituents, the orientation of the ring in the binding site was unknown and sampling different rotamers was slow such that transitions did not occur during the length of the simulation. In the amide series, different rotamers of the pyrazine ring (see Figure 3.11) affected computed relative free energies in the binding site by more than 6 kcal mol⁻¹. Running SepTop between the two poses differing in the orientation of the nitrogen atoms in the pyrazine ring, as depicted in Figure 3.11, resulted in a $\Delta\Delta G_{site} = -6.6 \pm 0.2$ kcal mol⁻¹ between the two poses in the binding site. This showed that one of the poses (Figure 3.11 (b)) was predicted to be more favorable than the other pose, possibly due to the potential to form an intramolecular hydrogen bond between one of the nitrogen atoms in the pyrazine ring and the hydrogen on the amide nitrogen, or because the alternate pose places the lone pair of the nitrogen too close to the lone pair of the amide oxygen, resulting in strong repulsion. Transitions between the two poses were only observed in a few λ windows and were very slow. The binding mode with the best docking score was not always the same pose that was predicted to be more favorable in SepTop calculations. For ligands in the biaryl series with an unknown orientation of the asymmetrically substituted phenyl ring we also ran SepTop calculations between different poses, keeping the binding mode that was predicted to be more favorable for further calculations. In the spirocyclic series we chose the same orientation of

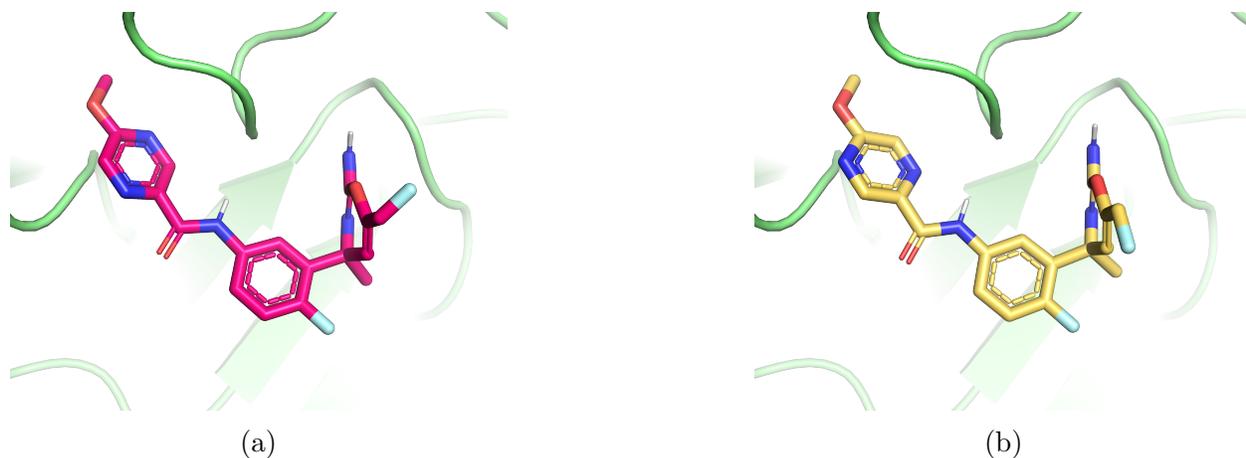


Figure 3.11: Two poses of a ligand from the amide series binding to BACE1. The orientation of the pyrazine ring in the binding site was unknown. Running SepTop between the two poses showed that the orientation of the pyrazine in the yellow pose (b) was more favorable than the magenta pose (a). The more favorable pose (b) can form an intramolecular hydrogen bond between one of the pyrazine nitrogen and the hydrogen of the amide.

the phenyl rings as given in the binding modes of the ligands in the PLBenchmark set[54]. The coordinate files of the ligands in their binding modes that were predicted to be more favorable can be found in the SI.

We adapted our alchemical protocol to try to enhance the sampling of different binding modes, however, we found that it was challenging to converge rotamer distributions at all λ windows. Here, our idea was to see whether, instead of running SepTop between different poses to identify the most favorable starting pose, we can start calculations with an unfavorable starting pose and adapt the alchemical protocol to sample the transition to the favorable binding mode. However, in the amide series, the torsion barrier around the rotatable bond between the pyrazine ring and the amide was so high, that even in the fully non-interacting state no transitions between different rotamers were observed. We adapted our enhanced sampling protocol and set the force constants to zero for all torsions that pass through the bond between the pyrazine ring and the carbonyl carbon of the amide in the non-interacting state, as well as further softening LJ-interactions of the atoms forming those torsions by scaling the LJ- ϵ by a factor of $\gamma = 0.1$. Starting SepTop from the two different

poses using this adapted protocol gave a $\Delta\Delta G_{site} = -0.6 \pm 0.3$ kcal mol⁻¹, which was much closer to zero which we would expect at sufficient sampling. The pyrazine ring now sampled different rotamers, however, sampling was not sufficient in all states along the alchemical path resulting in slow convergence. Thus, while we were eventually able to get this protocol to work, it involved considerable difficulty and manual tuning, and still exhibited signs of clear sampling problems, indicating that it would likely not work robustly for other, similar problems. Since we sought a general solution, we thus diverted our attention back towards protocols in which the preferred binding mode was an input (even if determined by an earlier SepTop calculation).

To run the SepTop calculations for this BACE1 dataset, we created ligand maps performing transformations both within each series as well as transformations spanning across different series. We ran 10 transformations for the six ligands within each series and 5 transformations between each series pair, giving a total of 45 transformations. Some transformations within a series were also scaffold hopping transformations, such as ring extensions, and we manually created the perturbation map to include both challenging transformations and transformations involving smaller R-group changes.

For this dataset, good correlation and error statistics were obtained within different scaffold series as well as for transformations across different series. Figure 3.12 shows the correlation between experimental and calculated binding free energies of the overall dataset. We show correlation both for the raw $\Delta\Delta G$ values of the 45 ligand transformations and the ΔG values of the 18 ligands. Overall, the correlation with experimental data was good with an RMSE of 1.39 kcal mol⁻¹ and an $R^2 = 0.71$. There were some outliers with larger deviation from experiment both within some of the series as well as for transformations between different series, as can be seen in the correlation of calculated and experimental $\Delta\Delta G$ values, broken up by transformations between and within different series in SI Figure B.6. We found that for this system the $\Delta\Delta G$ values from transformations between ligands

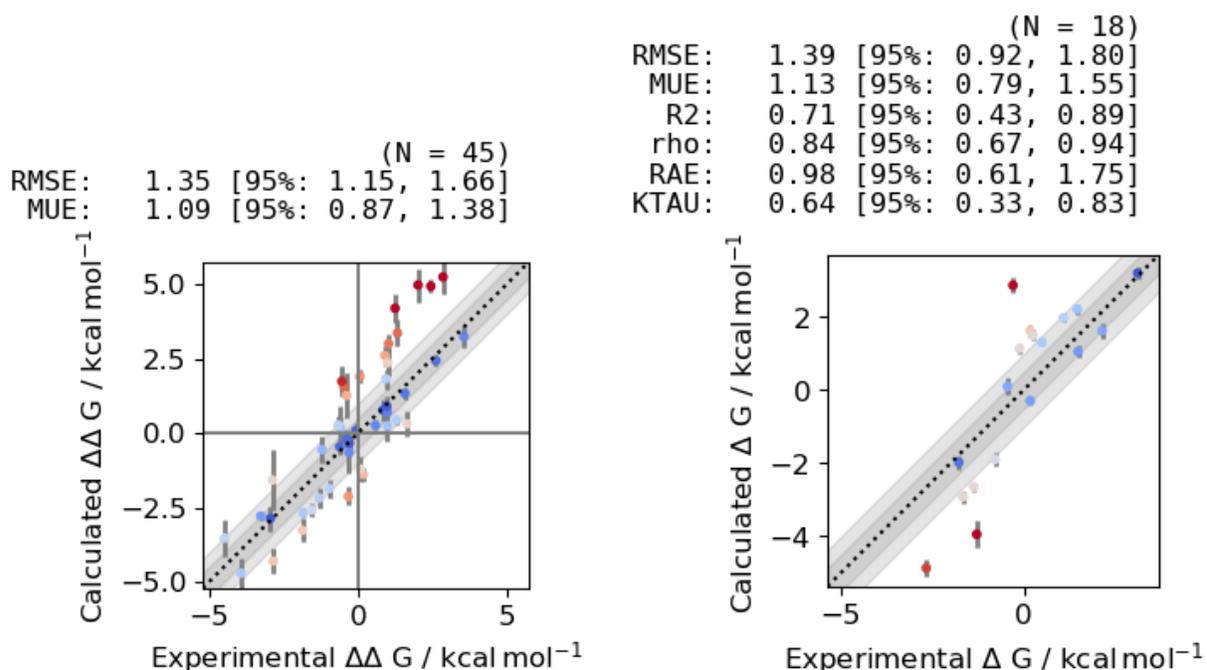


Figure 3.12: Correlation between calculated and experimental binding free energies for 18 BACE1 inhibitors. The upper plot shows the $\Delta\Delta G$ values of 45 ligand transformations and the lower plot the final calculated absolute binding free energies of the 16 ligands. Plot was generated using arsenic[94]. Calculated and experimental binding free energies correlated well, given an RMSE of $1.39 \text{ kcal mol}^{-1}$ and $R^2 = 0.71$.

within the same scaffold correlated better with experiment than transformations between ligands from different scaffolds (RMSE = 1.07 vs. RMSE = 1.78), which is unsurprising since transformations across scaffolds are clearly more challenging. However, transformations across scaffolds are impossible for standard RBEF calculations, potentially making these transformations still appealing.

We ran three independent repeats of the simulations and reported the mean and standard deviation across those. We then repeated the analysis only considering a single repeat to evaluate the impact of running simulations in triplicates on the results. Correlation between calculated and experimental ΔG values was very similar when considering only a single repeat (RMSE = 1.40 and $R^2 = 0.69$) versus three replicates (RMSE = 1.39 and $R^2 = 0.71$).

The full set of calculated and experimental values can be found in the SI.

We investigated the outliers in this system to try and identify potential sampling problems. All transformations between the amide and the biaryl series had calculated $\Delta\Delta G$ values that were too unfavorable (SI Figure B.7a). Extending the length of the simulation from 10 ns per λ window to 20 ns improved the results slightly ($RMSE_{\Delta\Delta G} = 2.23$ kcal mol⁻¹ vs. $RMSE_{\Delta\Delta G} = 1.83$ kcal mol⁻¹), suggesting that simulations had not reached convergence yet. Preliminary free energy calculations initiated from different conformations of the 10s loop (PDB ID: 4FS4 vs. 6OD6) showed that results substantially depend on the starting structure since the conformational change was not sampled adequately.

3.6 Discussion and Conclusions

In this chapter we introduce a generalized and efficient Separated Topologies approach and show that it works on pharmaceutically relevant systems. We developed a stable implementation of the approach and provide a package to setup input files for running SepTop in Gromacs[11]. For the orientational restraints applied in the SepTop approach we wrote an algorithm that picks atoms for restraints automatically based on heuristics we developed here to select relatively stable atoms in the ligand and protein (Section 3.2.2). This selection algorithm avoids several pitfalls and provides a robust and fairly general solution, at least in the systems investigated herein. We also developed a general and efficient protocol for running SepTop RBFEE with a reasonable λ schedule that worked on a variety of targets and we implemented an ϵ -HREX protocol for enhanced sampling.

3.6.1 SepTop performed well on the test systems in this study

We tested the method on several diverse, pharmaceutically relevant systems and reported performance and any resulting insights into strengths, weaknesses and challenges. We first validated the method on a system with standard RBF E results available (TYK2) and obtained comparable results to a standard RBF E method as well as achieved good convergence. We then applied the method to scaffold hopping transformations in the ER α system such as ring breaking transformations and compared results to other RBF E methods that allow scaffold hopping transformations. Here, we found that the preparation of the protein, and more specifically modeling of missing loops and the protonation state of histidine residues, had an impact on the binding free energy, highlighting both the importance of careful protein preparation as well as the challenge to identify whether differences in results between different RBF E methods originate from the methods itself or differences in the input structures. After validating the method on these smaller datasets, we tested SepTop on two larger datasets, MALT1 and BACE1, with transformations not as straightforward to run with typical RBF E. For both targets, results correlated well with experiment and good error statistics were achieved. We analyzed cycle closures as a convergence check and identified sampling problems due to slow rotamer sampling of different dihedral angles in the ligand.

Overall, SepTop performed well for large scaffold transformations in the BACE dataset. These transformations were challenging or impossible for traditional relative free energy calculations for multiple reasons. Firstly, the maximum common substructure shared by the different ligand scaffolds was very small. In addition, in transformations between the amide series and the two biaryl series, compounds from the amide series extended into the P3 pocket of BACE1 and displaced water molecules that were present in the two biaryl series. These large modifications were captured well in this dataset using SepTop.

For transformations between ligands in the same scaffold series in the BACE system, SepTop

achieved an RMSE of 1.07 kcal mol⁻¹. Two of the three series (the amide series and the spiro-cyclic series) have been studied using standard RBFE approaches[156, 46, 74]. Those studies treated the different ligand series separately and reported an RMSE of around 1 kcal mol⁻¹ on each. Thus, on this BACE system, SepTop achieved a performance comparable to prior studies, however, this is the first study that carried out transformations spanning across these distinct chemical series, possible here because our approach is more general.

3.6.2 Ways to test whether a target is within the domain of applicability of the BFE approach

In general, free energy calculations require a particular target to be within the domain of applicability of a method[133, 23, 136], but it is often difficult to know *a priori* what that domain of applicability is. Thus, when starting to work on a new system, it can make sense to first test convergence of the free energy estimate on a smaller subset to assess whether a new target is within the domain of applicability of the method. By domain of applicability we here mean that the binding mode is known and that there are no slow binding mode changes as ligands are swapped, *e.g.* slow water rearrangement or slow protein conformational changes. Convergence of the free energy estimate can be tested with different methods. For example, running calculations in replicates and looking at the standard deviation across repeats, as well as running transformations in opposite directions ($A \rightarrow B$ and $B \rightarrow A$) can help assess convergence. Additionally, a high cycle closure in a cycle of ligand transformations is another indicator that an important degree of freedom was not sampled sufficiently. In some cases more ambitious convergence tests may be beneficial, such as starting simulations from different available protein structures (*e.g.* structures that had been crystallized with different ligands bound) as well as starting simulations with different water rearrangements in the binding site. If the starting structures impact the calculated binding free energy, convergence has not been achieved yet.

If the binding mode of a ligand is unknown, we have found it to be helpful to run SepTop between different poses to determine the more favorable binding pose which we then used for further simulations (Section 3.5.4). We also found that adapting the alchemical protocol by modifying bonded and non-bonded interactions of the ligand (*e.g.* scaling down the LJ- ϵ (ϵ -HREX), and scaling down dihedral force constants) may not be sufficient to sample the correct binding mode sufficiently when starting simulations from the wrong pose (Section 3.5.4). Based on our results so far, we strongly advise against trying to engage in ligand binding pose prediction *while* doing free energy calculations, but instead selecting one binding pose *prior* to the calculation (*e.g.* by running SepTop between binding poses to determine the more favorable pose).

3.6.3 Comparing SepTop to other alchemical binding free energy methods may not be advisable for some systems

If ligands in a dataset are structurally very diverse and fall outside the scope of standard RBFEE approaches, in addition to RBFEE methods like SepTop, binding free energies can be calculated using standard absolute binding free energy calculations. For some targets, however, ABFE calculations are expected to be extremely difficult or nearly impossible without extraordinary long simulations or new algorithmic developments. For example, a protein may undergo a substantial and slow conformational change on ligand binding (*e.g.* HIV-1 protease, which has a large flap motion on inhibitor binding), or a protein may undergo a change in protonation state on ligand binding (*e.g.* BACE1, here). In such cases, a normal ABFE calculation which simply removes the ligand from the binding site would leave the protein in an metastable “unbound” state unless simulations are extremely long (in the case of conformational change) or special algorithms are employed (in the case of protonation state changes). In more detail, the two catalytic aspartates in BACE1 were both ionised in the protein-ligand bound state, but that is likely to change to a single

ionised and one neutral Asp in the apo state. Thus, we deliberately avoid testing ABFE calculations on BACE1 as we wish to avoid problems caused by an incorrect protonation state for the unbound state. Likewise, we also do not compare to standard RBFE calculations for the scaffold-hopping transformations considered here, as these are simply disallowed by standard RBFE calculations, making comparison impossible. In general, we expect SepTop to be computationally more efficient than ABFE and slightly less efficient than standard RBFE (on transformations where RBFE is applicable).

3.6.4 Molecular shape and chemical similarity may be good metrics for planning SepTop calculations

RBFE calculations, including with SepTop, can be planned more effectively using some measure of ligand “similarity” to assess which transformations will be easy and which will be hard. Here, we have not investigated planning and similarity metrics, reserving this for future work. However, it seems likely that SepTop will benefit from different similarity metrics than standard RBFE, since SepTop replaces one ligand with another in the binding site rather than mutating one ligand into the other. In this work we did not explore the impact of the design of the transformation map on the efficiency of the calculations, but we propose that molecular shape and chemical similarity may be good metrics to use for planning SepTop calculation networks, in contrast to the maximum common substructure and 2D graph similarity metrics that are often used in standard RBFE approaches[91, 23]. There is a better chance that ligands introduce similar conformational changes in the protein and that they displace similar water molecules (and therefore converge faster), if the two ligands have a high molecular shape and chemical similarity.

3.6.5 Current limitations of SepTop

In theory, transformations performed using the SepTop approach are not restricted by structural similarity of the ligands and ligand binding modes, however, there might be limits to the domain of applicability of the method. We expect slow convergence and biased results if the two ligands bind to different conformations of the protein and if transformations between those conformations happen on timescales longer than the simulation run time. Slow convergence can also be expected if one ligand displaces buried water molecules that are not displaced in the presence of the other ligand, especially if the entry/exit of that water molecule is not sampled throughout the simulation. In addition, results are expected to be incorrect if the ligands bind to different protonation states of the protein or if one or more ligands bind covalently, and results may be slow to converge if the ligands bind in disparate regions of the binding site.

Overall, we found that SepTop performed well on pharmaceutically relevant test systems which had been previously studied, and then we applied it to a larger number of compounds two different targets (MALT1 and BACE1) involving both transformations which are possible for typical RBFE calculations, and those which are challenging or impractical. Some of these challenging transformations involved scaffold-hopping transformations. These tests also involved larger numbers of ligands. We found that SepTop performed as well as standard RBFE calculations for transformations within a given congeneric series, but avoided complexities of atom mapping and required minimal human intervention to set up the calculations. For scaffold-hopping transformations, accuracy was predictably somewhat lower, likely because these transformations are dramatically more difficult to sample, but these preliminary results are nevertheless encouraging. Overall, this suggests that SepTop may be a general and broadly useful approach for relative binding free energy calculations that expands their domain of applicability.

Chapter 4

Impact of protein conformations on binding free energy calculations in the BACE1 system

In binding free energy calculations, simulations must sample all relevant conformations of the system in order to obtain unbiased results. For instance, different ligands can bind to different metastable states of a protein, and if these protein conformational changes are not sampled in relative binding free energy calculations, their contribution to binding is not accounted for and thus calculated binding free energies are inaccurate. In this work, we investigate the impact of different beta-secretase 1 (BACE1) protein conformations obtained from X-ray crystallography on the binding of BACE1 inhibitors. We highlight how these conformational changes are not adequately sampled in typical molecular dynamics simulations. Furthermore, we show that insufficient sampling of relevant conformations induces substantial error in relative binding free energy calculations, as judged by a variation in calculated relative binding free energies up to 2 kcal/mol depending on the starting protein conformation. These results emphasize the importance of protein conformational sampling and pose this

BACE1 system as a challenge case for further method development in the area of enhanced protein conformational sampling, either in combination with binding calculations or as an endpoint correction.

4.1 Introduction

Free energy calculations are a quantitative and rigorous approach to calculating the free energy difference between different thermodynamic states. For example, binding free energy calculations determine the free energy difference between an unbound ligand and protein and the protein-ligand complex, allowing calculation of the binding free energy. The calculations find widespread use in the pharmaceutical industry, *e.g.* in prospectively estimating ligand potency during lead optimization to help guide discovery efforts.[6, 9, 133, 22, 74, 81] The free energy difference between two states is related to the ratio of the configurational partition functions of the two states. In order to approximate the partition functions, all relevant conformations of the system need to be sampled. Hence, the accuracy of the calculated change in free energy depends on how well the phase space of the system is sampled.

In binding free energy calculations, inadequate sampling of relevant protein conformations can lead to inaccurate, biased results.[168, 74] For instance, ligand binding may require the protein to undergo a structural rearrangement. In absolute binding free energy calculations, conformational changes going from the apo (unbound) structure to the holo (bound) structure, need to be sampled[75, 13, 40]. Similarly, in relative binding free energy (RBF) calculations, if two ligands bind to distinct protein conformations, the transition between those conformations needs to be sampled. However, the protein can be kinetically trapped in a metastable conformation in molecular dynamics (MD) simulations, and it has been shown that inadequate sampling of these slow protein conformational changes can lead to biased results[101, 87, 56, 122]. For example, two binding free energy calculations that are started

from two different metastable states of the protein should converge to the same free energy difference (if the same underlying model is used), however, it has been shown that starting calculations from different protein conformations can lead to different calculated binding free energies.[87]

Beta-secretase, or BACE1, is a protein with known conformational flexibility, as judged by X-ray crystallography[164, 59, 116, 138] and MD simulations[74, 164, 51], and different ligands sometimes bind to different metastable states of the protein, meaning conformations that are different in a significant way and that interconvert slowly. BACE1 is an aspartyl protease that cleaves amyloid precursor protein (APP), generating amyloid beta; and a wide range of inhibitors have been developed with the goal of targeting Alzheimer’s disease by reducing amyloid beta peptide production[167, 58, 143, 61, 97, 68, 107, 80]. Among other conformational changes, the flap region and the 10s loop in BACE1 undergo slow rearrangements upon ligand binding[164, 74] and the protein conformations obtained using X-ray crystallography can differ even when structurally similar ligands are bound (Figure 4.1a).

Previous binding free energy studies on structurally similar BACE1 inhibitors used different crystal structures as input structures for free energy calculations[156, 74, 46], however, the impact of the choice of the input crystal structure has not yet been investigated. In those previous studies, calculations of one set of BACE1 ligands have been carried out using a crystal structure where the 10s loop occupies a closed conformation (“CAT set” studied using PDB ID 4DJW, see Figure 4.1)[156, 46], while another set of ligands that were structurally very similar to the first set have been studied in a crystal structure with an open 10s loop conformation (“P2 set” studied using PDB ID 3IN4)[74, 46]. However, the two ligand sets employed in these studies differed and the same ligands have not yet been run in the two different crystal structures, making it impossible to know how much impact the choice of protein structure had on calculated binding free energies.

In this work we investigate the stability of BACE1 ligands in different metastable states of the protein as well as the impact of the starting crystal structure on relative binding free energies. We performed MD simulations of ligands from the two sets mentioned above starting with two different crystal structures and evaluated the stability both of the protein structures and the ligands. We then performed relative binding free energy calculations between the ligand pairs using a Separated Topologies approach[129, 12], performing the ligand transformations using different crystal structures as starting coordinates to investigate the impact of protein conformations on calculated relative binding free energies. This work highlights the importance of the sampling of different protein conformations and presents this BACE1 system as a challenge case for further method development in the area of enhanced protein conformational sampling.

4.2 Simulation details

The three crystal structures (PDB IDs 3IN4, 3INF and 4DJW) and the four ligands were prepared, giving 12 combinations, in order to generate the input for MD simulations. The GROMACS topology and coordinate files of all protein-ligand systems can be found in the Supporting Information (SI).

The two protein structures 3IN4 and 4DJW were prepared on OpenEye’s Orion cloud computing platform, using their workflow (“floe”) “SPRUCE - Protein Preparation from PDB Codes” with the default parameters.[112] The crystal structure 3INF was prepared using the OpenEye Spruce toolkit. In order to build missing loops, the gap penalty was set to -5 after the default sequence alignment had failed and had resulted in the terminal region being misaligned. In both protocols, crystallographic water molecules were retained, hydrogen atoms were added at pH 7 and missing loops were built.

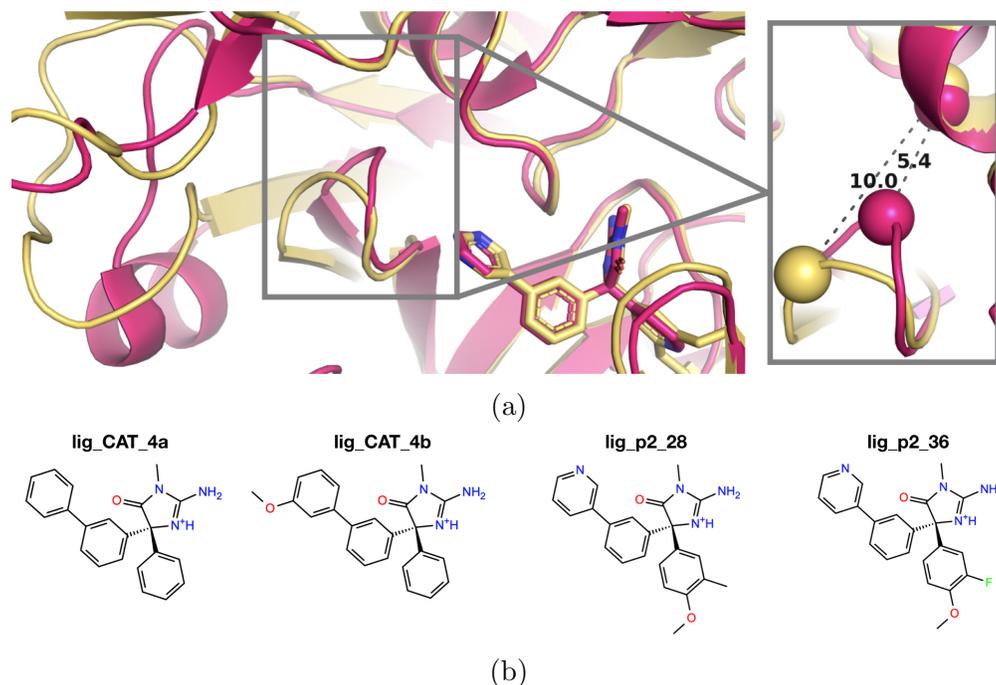


Figure 4.1: Different protein conformations and inhibitors in the BACE1 system. (a) Binding site view of PDB ID 3IN4 (yellow) and PDB ID 4DJW (magenta). The two crystal structures differ in the conformation of the 10s loop which is in a closed (magenta) versus open (yellow) conformation. The structures also differ in a part that is adjacent to the 10s loop (to the left) where the 3IN4 structure has a large loop while the 4DJW structure a helix and a loop. The grey box to the right shows a zoomed in view of the 10s loop. The C α of a serine (S71 in PDB 4DJW, S72 in PDB 3IN4) and the carbon of an alanine (A396/A397) are depicted as spheres and were used in distance measurements to determine whether the 10s loop was open vs. closed. (b) 2D depiction of the four BACE1 inhibitors investigated in this study. The first two ligands are from the CAT series, the other two ligands from the P2 set.

We selected two ligands from each of the two ligand sets and obtained `.sdf` files for the ligands from the SI of those publications[156, 74]. We will be referring to the two sets as the CAT set[28, 156] and the P2 set[96, 74] throughout this chapter, according to their naming scheme in the Protein-Ligand-Benchmark set.[54] The ligands were aligned onto the crystallographic ligand in PDB ID 4DJW using OpenEye’s ShapeFit method[110, 73] as implemented in the SystemBuilder package.[118] For the two ligands in the P2 set, multiple binding modes, more specifically, different orientations of the nitrogen in the pyridine ring as well as of the R-group in non-symmetrically substituted phenyl rings, had to be considered and prepared. The systems were solvated, parameterized and equilibrated using Orion files “Bound Protein-Ligand MD” for the protein-ligand complex and “Solvate and Run MD” for the unbound ligands. Sodium and chloride atoms were added to reach a salt concentration of 150 mM. The Open Force Field version 2.0.0[151] and AM1-BCC charges[64] were used to parameterize the ligands, Amber ff14SB[95] was used for the protein, and TIP3P[71] for the water.

The prepared systems were then used to run MD simulations and binding free energy calculations in GROMACS 2021.2.[17, 1]

For each of the four ligands, 500 ns long MD simulations were run in both crystal structures. The systems were first energy minimized using steepest descent for 5000 steps, and then equilibrated for 10 ps in the canonical ensemble at 298.15 K. Then, a production run of 500 ns was performed in the NPT ensemble using the stochastic dynamics integrator at a timestep of 2 fs and a pressure of 1 bar. Full details of simulation parameters can be found in the MDP files in the SI. MD simulations were run in triplicates.

For relative binding free energy calculations between the ligands in the two different crystal structures, a Separated Topologies approach (SepTop) was used. This approach essentially performs two absolute binding free energy calculations at the same time in opposite directions, inserting one ligand while removing another. Details of the method can be found in

other publications[129, 12]. The input files for SepTop were generated using the SepTop tools `SeparatedTopologies`[11]. Relative binding free energy calculations between different binding modes of the same ligand were performed to determine the more favorable binding mode. Rotatable bonds were restrained using a force constant of $10 \text{ kcal mol}^{-1} \text{ rad}^{-2}$. Once the more favorable binding mode was confirmed, relative binding free energy calculations between the four ligands were carried out in all three crystal structures. For calculations in the binding site, 20 λ windows were used while in the solvent leg, 28 λ windows were used. The details of the simulation parameters and the alchemical path can be found in the MDP files in the SI.

Free energy differences were calculated using the MBAR estimator[140] as implemented in the `alchemlyb` interface to the `pybar` package[140, 36]. The first nanosecond of a 10 ns production run (per λ window) was discarded as additional equilibration. Three independent repeats of the simulations were run and the mean and standard deviation across the repeats was reported.

4.3 Results

In this study, we investigated the impact of the starting protein conformation on the calculations of relative binding free energies in the BACE1 system. We first performed MD simulations of four BACE1 inhibitors (Figure 4.1b) using three different crystal structures (total of 12 combinations) and then ran RBFEE calculations between those ligands starting from the different protein conformations. The difference in the protein conformations between the structures (PDB ID 4DJW and PDB ID 3IN4) is shown in Figure 4.1a. Specifically, in the crystal structures 3IN4 (and 3INF, not shown), the 10s loop (grey box, Figure 4.1a) is crystallized in an open conformation (yellow) while the 4DJW crystal structure has a closed 10s loop conformation (magenta) and the residues in the vicinity of that 10s loop

(bottom left, Figure 4.1a) form a loop (yellow), vs. a helix (magenta).

In previous BACE1 RBFE studies, calculations of the P2 ligand set were performed starting from an open 10s loop conformation[74, 46] (PDB 3IN4) while for ligands of the CAT set a closed 10s starting structure was used[156, 46] (PDB 4DJW). As shown in Figure 4.1b, the ligands from the two sets are structurally very similar. For some of the ligands from the two sets the respective metastable protein conformations from the starting crystal structures were known to be the dominant bound state, as judged by crystal structures that were co-crystallized with ligands from those sets. For other ligands, however, the dominant bound state was not known. We selected two ligands from each of the two datasets. Those four ligands are structurally very similar, as can be seen in Figure 4.1b. However, the crystal structures that are available for two of the ligands (PDB 3INF and 4DJV) show that these chemically similar ligands can bind to different metastable states of the protein.

After initial calculations we noticed that there was an additional difference between the two selected crystal structures; the presence or absence of a buried water molecule in the P2' pocket (4.2). We found that in MD simulations where the water was not present in the starting structure (3IN4), the water was not able to enter that site. Therefore, we decided to run additional calculations with another open 10s structure, 3INF, that had a crystallographic water molecule at the same location as in the closed 10s loop structure (4DJW), however, had the same open 10s loop conformations as the 3IN4 structure. The water molecule was absent in the 3IN4 structure, presumably because the bulkier diethyl phenyl ring of the co-crystallized ligand displaces the water molecule. Here, we wanted to investigate whether computed binding free energies are impacted by the presence of the water molecule when using the two open 10s loop protein conformations.

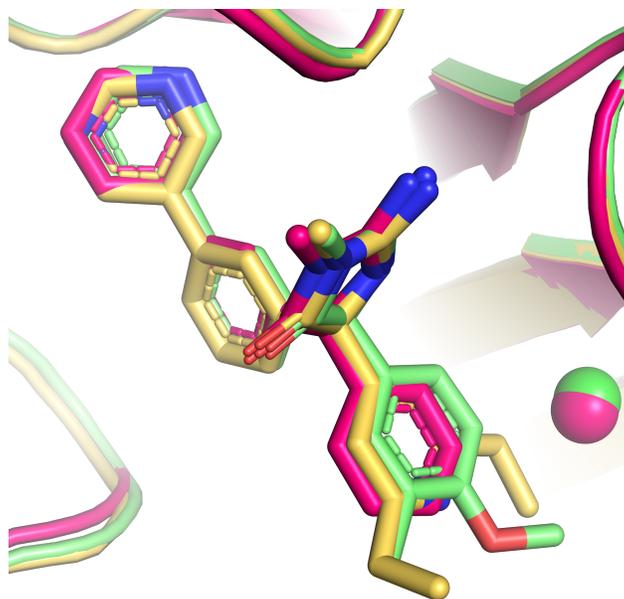


Figure 4.2: Buried water molecule in the P2' pocket in the BACE1 system. The co-crystallized ligands are depicted as sticks, the buried water molecules as spheres and a part of the protein as cartoon. PDB structures 4DJW (magenta) and 3INF (green) have a buried water molecule present in the P2' pocket which is absent in the 3IN4 (yellow) structure, presumably because the ethyl group displaces the water upon binding.

4.3.1 We performed MD simulations of the ligands in three different protein structures

The conformational transition going from the closed 10s loop conformation to the open 10s loop conformation and vice versa was not sampled adequately in 500 ns long MD simulations. If an MD simulation is started from the open 10s loop protein conformation for a ligand whose dominant bound state is the closed 10s loop conformation, we would expect to observe the loop transition from an open to a closed conformation during the MD simulation, and vice versa for a ligand binding to the open 10s loop conformation. For all four ligands, we ran 500 ns of MD simulations in triplicate both starting from the two open 10s loop structures and the closed 10s loop structure, giving 12*3 separate MD simulations. We measured the distance between the C α of a serine (S71 in PDB 4DJW, S72 in PDB 3IN4 and 3INF) and the carbon of an alanine (A396/A397) as a metric for whether the 10s loop is open ($d =$

1.0 nm, as measured in PDB ID 3IN4) or closed ($d = 0.54$ nm, as measured in PDB ID 4DJW) (see grey box Figure 4.1a). As an additional metric, we calculated the root mean square deviation (RMSD) of the backbone atoms of seven loop residues (residues 68 - 74 in PDB 3IN4/3INF and 67 - 73 in PDB 4DJW) with respect to an average structure, calculated from one of the trajectories from a calculations started with 3IN4 and 4DJW. In Figure 4.3 we show the results for one of the ligands, additional results of the other ligands can be found in the SI folder. In the MD simulations that started in the open 10s loop conformation, the loop stayed open throughout the simulation time. On the other hand, when MD simulations were started from the closed 10s loop structure, the loop slightly opened, but did not fully open up to the open 10s conformation.

In addition to the stability of the protein conformation, we analyzed the binding mode of the ligands and found that it was mostly stable in these long MD simulations. To analyze ligand binding modes, we calculated the ligand RMSD (SI Figure C.1) as well as two torsion angles in the ligand (Figure 4.4). A low RMSD of 0.5-1.5 Å shows that the ligand binding mode was relatively stable in the MD simulations. Sudden increases in ligand RMSD resulted from rotations of the aromatic ring in the P3 pocket (RB2 Figure 4.4 b). For two of the ligands (`lig_p2.28` and `lig_p2.36`) we performed two separate MD simulations in two different binding modes, where the nitrogen in the pyridine ring and the substituent of the phenyl ring in the P2' pocket pointed in versus out. During the MD simulation, the pyridine ring transitioned multiple times between different binding modes and the two binding modes were similarly populated (see histograms rotatable bond RB1 Figure 4.4 a). The phenyl ring on the other hand was not able to rotate, but remained in the respective starting conformation (rotatable bond RB2 Figure 4.4 b). With sufficient sampling we would expect the rotamer distribution to be the same, independent of the starting rotamer.

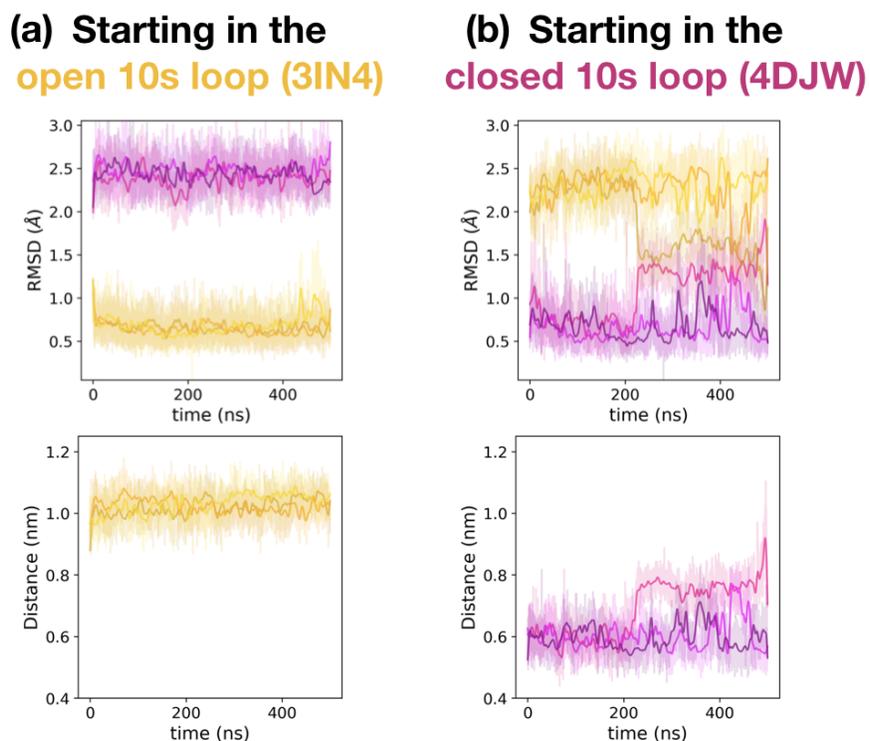


Figure 4.3: MD simulations of BACE1 ligand `lig_CAT_4b`. Top shows RMSD of the loop conformation relative to reference structures; bottom shows P3 pocket size. Simulations were initiated from protein structure 3IN4 (see (a), open 10s loop) or 4DJW (see (b), closed 10s loop). In the upper plots we show the RMSD of the 10s loop protein backbone with respect to an average open 10s loop structure (yellow) and an average closed 10s loop structure (magenta) as a function of simulation time. The bottom plots show as metric of the size of the P3 pocket the distance between $C\alpha$ of a serine (S71 in PDB 4DJW, S72 in PDB 3IN4) and the carbon of an alanine (A396/A397) as a function of simulation time. All plots show data for all frames in a lighter color, and a running average (averaged over 100 frame) in a darker color for the triplicate MD runs. (a) Starting MD simulations with the open 10s loop structure, the protein remained in the open 10s loop state (distance $d \approx 1 \text{ nm}$, lower RMSD to open 10s loop structure (yellow) than to closed 10s loop structure (magenta)). (b) Starting MD simulations with the closed 10s loop structure, the 10s loop opened up slightly, but not fully, after around 200 ns in one of the MD simulations, as can be seen by an increase in RMSD to the closed 10s loop structure (magenta, upper right plot) and a decrease in RMSD to the open 10s loop structure (yellow, upper right plot) as well as in increase in the distance (lower right plot).

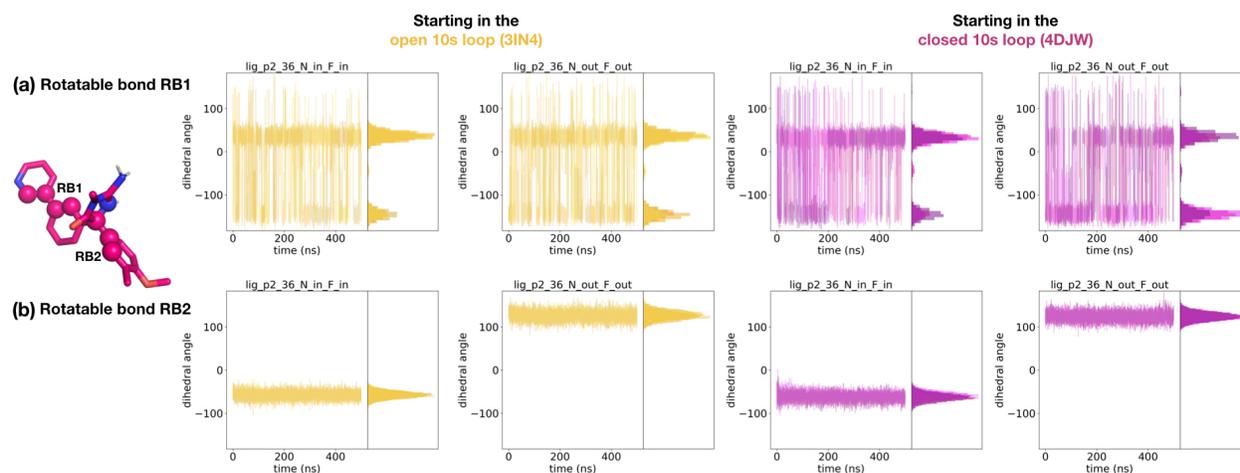


Figure 4.4: Rotatable bonds of BACE1 ligand `lig_p2_36` during MD simulations. Simulations were initiated from protein structure 3IN4 (left, yellow) or 4DJW (right, magenta). The dihedral angle of two rotatable bonds in the ligand are shown as a function of simulation time as well as a histogram. Separate plots are shown for initializing the MD simulations in two different binding modes (`N_in_F_in` and `N_out_F_out`) in both protein structures (3IN4, yellow and 4DJW, magenta). In both protein structures, independently of the starting binding mode, the pyridine ring rotated multiple times (a, upper plots), in contrast to the phenyl ring that stayed stuck in the orientation the simulations were started in (b, lower plots). This means that sampling was not sufficient since the rotamer distribution should not depend on the starting structure.

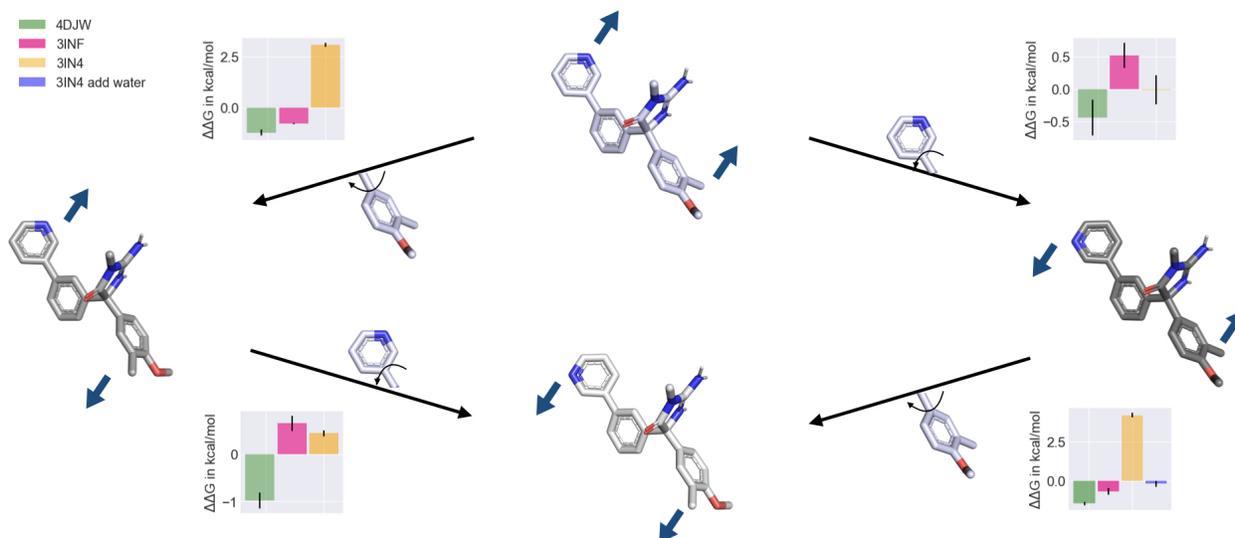


Figure 4.5: RBFE calculations between different poses of `lig_p2_28` were performed using different crystal structures (4DJW, 3INF, 3IN4). Four different poses were considered, namely two different orientations of the pyridine ring and two different orientations of the substituted phenyl ring. Calculated relative binding free energies for each transformation are shown as bar graphs. For one of the transformations (lower right) an additional calculation was run in the 3IN4 structure after manually adding a buried water molecule. Starting calculations from different crystal structures and different water placements resulted in different relative binding free energies and different preferred ligand binding modes, though if conformational sampling captured the relevant protein and water motions, no difference would be expected.

4.3.2 We used RBFE calculations between ligand poses to identify the preferred binding mode

Since the MD simulations were not able to clearly reveal the preferred binding mode of the two ligands `lig_p2_28` and `lig_p2_36`, we ran RBFE calculations between different potential binding modes of those ligands to identify the preferred binding mode. Two different rotamers of the pyridine ring as well as two orientations of the asymmetrically substituted phenyl ring were considered, giving four different binding modes. We used SepTop to calculate the free energy change of transforming the ligand in binding mode A to its copy in binding mode B and restrained rotatable bonds to prevent transitions between different binding modes.

The predicted preferred binding mode of the ligands differed depending on the starting structure (Figure 4.5). In the open 10s loop structure 3IN4, both the nitrogen and the fluoro/methyl group were more favorable in an “in” conformation, which had been the predicted orientation in a previous study as well[74]. However, in the other open 10s loop structure 3INF, which has a buried water molecule present in the P2’ pocket (Figure 4.2), the binding mode where the methyl/fluoro group was pointing to an “out” orientation was more favorable. Thus, in the absence of the buried water molecule, the R-group pointed into the region vacated by the water molecule. In the crystal structure 3INF, ligand `lig_p2_28` is co-crystallized in the presence of the buried water molecule with the methyl group pointing outwards. Since this is the only crystal structure that was co-crystallized with ligand `lig_p2_28`, this suggests that this ligand binds to BACE1 in the presence of the buried water molecule. This also suggests that the water molecule plays an important role in ligand binding.

As an additional test for the impact of the water molecule, we took the 3IN4 structure, where the water molecule was not present, and added the water molecule manually in the same location as in the 3INF structure. We then repeated the RBFEE calculations between two of the binding poses (Figure 4.5 lower right). After adding the water molecule, the calculated relative binding free energies between the two poses agreed with the free energies calculated from the 3INF structure, suggesting that indeed the water molecule had led to the discrepancies between the two open 10s loop structures.

We also found that in the closed 10s loop conformation, both the pyridine nitrogen and the fluoro/methyl group preferred an “out” conformation. This protein structure had the same water molecule present in the crystal structure as the 3INF structure. Even though the protein conformation of the two structures in the 10s loop differed, the preferred binding mode in the P2’ region was the same.

In summary, the preferred binding mode of the pyridine ring (that points towards the 10s

loop) depended on the protein conformation of the 10s loop. When the protein has an open 10s loop structure, the pyridine ring appears to prefer to bind with the nitrogen pointing inwards, while when the protein has a closed 10s loop structure the pyridine nitrogen pointed outwards. The preferred binding mode of the substituted phenyl ring in the P2' pocket depended on the presence or absence of a buried water molecule where if the water was present, the R-group pointed outwards, while in the absence of the water it pointed inwards.

4.3.3 RBE calculations were impacted by the starting protein conformation and the presence of a buried water molecule

After determining the starting binding modes for the ligands in the three crystal structures, we calculated the RBE for the six pairs of ligands in different crystal structures to investigate how the choice of starting protein conformation impacts calculated relative binding free energies. We found that the RBE results were impacted by the starting crystal structure used in the simulations. We performed six RBE transformations using the SepTop approach and ran the same transformations both in the two open 10s loop and in the closed 10s loop structure, giving a total of 18 transformations (Figure 4.6).

The presence of a buried water molecule had an impact on the $\Delta\Delta G$ values, as can be seen in the variation in the results between the two open 10s loop structures, where a buried water was present in the 3INF crystal structure which was absent in the 3IN4 structure. For three of the six transformations, the results differed by ≈ 2 kcal/mol depending on the starting structure, while for the other three transformation there was no significant difference between the two sets of results (Figure 4.6 magenta and yellow). The three transformations where the buried water molecule impacted the results involved ligand `lig_p2_28`. The methyl group of that ligand preferred to point away from the water in the structure where the water was present and towards where the water would be in the structure without the water molecule.

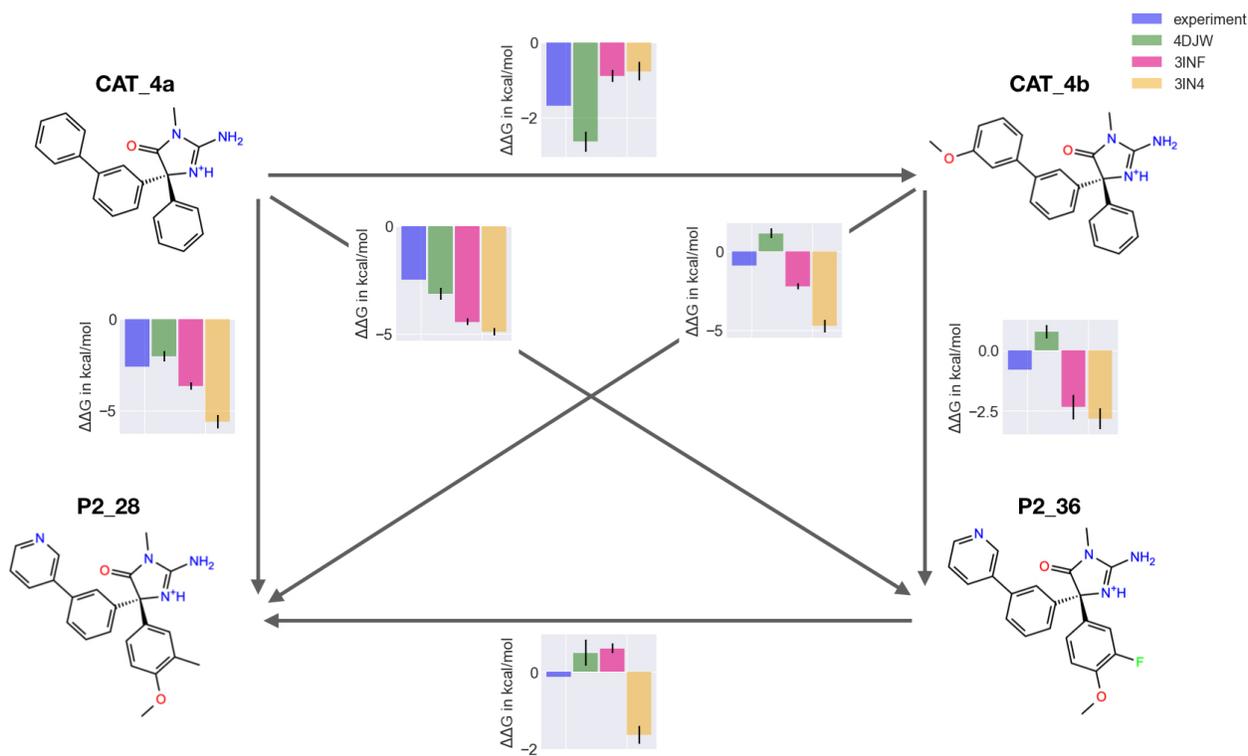


Figure 4.6: RBFEE results from transformations between the four BACE1 inhibitors. Shown are experimental and calculated $\Delta\Delta G$ values (bar graphs) for each transformation. For the calculated values, results from simulations started using different crystal structures (4DJW, 3INF, 3IN4) were reported. The starting crystal structure had an impact on the $\Delta\Delta G$ values as judged by a variation in results between calculations that were started from different protein conformations of the 10s loop (4DJW, green vs. 3INF, magenta). Simulations that started from a similar open 10s loop conformation, but different numbers of buried waters (3INF, magenta and 3IN4, yellow), also gave different calculated relative binding free energies. Both the starting protein conformation and the presence of water molecules impacted calculated relative binding free energies even though with sufficient sampling, no difference would be expected.

In addition to the buried water molecule, the initial protein conformation (obtained from the crystal structures) also impacted the calculations as judged by a variation in the results between calculations started in the open 10s loop structure (Figure 4.6 3INF, magenta) and the closed 10s loop structure (4DJW, green). Results differed up to 3.5 kcal/mol (transformation `lig_CAT_4b` \rightarrow `lig_p2_28` $\Delta\Delta G=1.2$ kcal/mol (4DJW) vs. $\Delta\Delta G=-2.3$ kcal/mol (3INF)), though if protein conformational sampling captured the relevant motions, no difference would be expected. For one of the transformations (`lig_p2_36` \rightarrow `lig_p2_28`) there was no statistically significant difference between results from the two different loop conformations. These two ligands only differ in the substituent of the phenyl ring in the P2' pocket, while they both have the pyridine ring in the P3 pocket, close to the 10s loop. Therefore, any potential contributions due to the different 10s loop conformation may cancel out between these ligands since it seems to be the pyridine ring which plays the largest role in mediating the effects of loop conformation.

As an additional convergence check in calculations started from the 3INF crystal structure, we performed calculations for all edges in opposite directions, meaning going from ligand A to ligand B and from ligand B to ligand A. Results are shown in SI Figure C.2. Calculations in both directions should converge to the same change in free energy with sufficient sampling, however we found that for three of the six edges calculations did not converge to the same relative binding free energy (after accounting for the sign flip). These transformations involved ligand `CAT_4b` which is the only ligand with a methoxy phenyl group while the other ligands either have an unsubstituted phenyl or pyridine ring in this position. We found that the larger methoxy group displaced a water molecule in the P3 pocket which likely resulted in the convergence problems observed here.

4.4 Discussion and Conclusion

In this chapter we showed how inadequate sampling of conformational changes induces substantial error (as judged by variation in the results), in RBFЕ calculations for BACE1 inhibitors. We believe these results may be important both because of their implications for binding free energy calculations generally, and for any future binding free energy studies on BACE1, where the effect of the selected protein conformation had not been previously examined.

Sampling of relevant protein conformations can be challenging both in MD and in binding free energy calculations. First, the “correct” protein conformation that a ligand binds to may not be known. In BACE1, structurally similar ligands bind to either an open 10s or a closed 10s loop conformation and for some ligands it is unknown which conformation they bind to, yet the chosen conformation can substantially impact calculated binding free energies. Even for co-crystallized ligands it cannot be ruled out that they may also bind to a protein conformation that is different from the one in that crystal, *e.g.* under different crystallization conditions[164], or the preferred conformation in solution might be different from that adopted in a crystal.

Furthermore, even if the preferred protein conformation is known, it can be challenging to sample and account for different protein conformations in the presence of different ligands in free energy calculations. In BACE1, the timescale for the conformational change in the 10s loop is too slow to be sampled during RBFЕ; we found the relevant conformational changes not occurring on the 500 ns timescale here. However, we saw that the starting protein conformation had a substantial impact on the results of RBFЕ calculations. Starting calculations from different crystal structures gave different results and without accounting for the conformational change it is impossible to know which calculated relative binding free energy is correct.

These results also highlight the ongoing need for better methods to enhance the sampling of protein motions, either in combination with binding calculations or as an endpoint correction. We therefore pose this BACE1 system as a challenge case for methods development in this area and provide input files so that researchers can test their methods on this system.

Chapter 5

Future directions

5.1 Analysis tool for identifying sampling problems in binding free energy calculations

As described in Chapters 2, 3 and 4, identifying slow degrees of freedom in free energy calculations helps determine the confidence in calculated binding free energies. Assessing the quality of the results is especially crucial when applying binding free energy calculations in a prospective study where calculated values are used to drive decisions on which compounds to synthesize next. In Chapter 2 we identified some metrics that helped assess the presence of sampling problems in simulations and that helped determine slow degrees of freedom in the system. More specifically, we identified slow degrees of freedom by identifying correlations between a thermodynamic property, such as the non-equilibrium work values, and a degree of freedom in the system, such as side chain dihedral angles (see Chapter 2 Figure 2.9). However, the analysis was not fully automated and required inspecting and visualizing simulation trajectories.

Therefore, it would be valuable to have a software package that the user can run as a post-simulation analysis after every binding free energy calculation. This package should perform multiple checks and flag potential problematic slow degrees of freedom in the system. Some examples of degrees of freedom that should likely always be monitored are:

- rotatable bonds of the ligand
- protein side chain rotamers, especially around the binding site
- the number of water molecules in the binding site
- the RMSD of the ligand and/or of parts of the protein

Such a tool could then calculate the Pearson correlation coefficient between the time series of a thermodynamic property (*e.g.* the $dH/d\lambda$ in equilibrium FEC or the non-equilibrium work values in non-equilibrium FEC) and the degrees of freedom listed above. Such analysis was performed in Chapter 2 and in recent work by Zhang et. al.[166]. A high (absolute) Pearson correlation coefficient (close to 1 or -1) signifies that the slow degree of freedom is important, meaning that it impacts the binding free energy. However, careful statistical analysis may also be necessary to avoid accidentally flagging spurious correlations (*e.g.* if significance testing is used with a p-value of 5% and 20 different degrees of freedom are tested, there's a reasonable chance a spurious correlation might clear the significance threshold.)

Additionally, in RBFEC calculations the tool should routinely calculate cycle closure errors around closed cycles in the ligand perturbation graph, flagging any ligand cycles with a high cycle closure and identifying edges that likely contributed substantially to the high cycle closure error, such as with the voting system introduced in Chapter 3.5.3.

5.2 Investigate the impact of different choices in enhanced sampling protocols in SepTop

In the SepTop approach that is described in Chapter 3, we applied an enhanced sampling method, ϵ -HREX, to soften interactions of the ligands to improve conformational sampling. We found that our default scaling factor of $\gamma = 0.5$ was not always able to sufficiently sample different rotamers of the ligand (Figure 3.9) and that in some cases it was necessary to further scale down the LJ- ϵ of parts of the ligand to improve sampling. In the MALT1 system, scaling the LJ- ϵ of the R-group down by a scaling factor of $\gamma = 0.01$ led to good convergence (Figure 3.9). In that transformation, the problematic torsion was identified after careful manual inspection in a post-simulation analysis and then simulations were rerun with the modified enhanced sampling protocol. However, this required rerunning the simulations after the problem had been identified.

Therefore, a tool that automatically determines which parts of the ligand need to be “hit harder” with the enhanced sampling protocol, meaning the atoms where the LJ- ϵ needs to be scaled down further, would be valuable. As a potential way to identify such ligand atoms, a short simulation of the ligand in vacuum can be performed and analyzed prior to the simulations for free energy calculation. Analyzing the dihedral angles of the ligand from such a simulation might give insight into potential slow rotamer transitions. A different method to identifying potential slow rotamer transitions is analyzing the volume around the ligand in the binding site. If a specific R-group of the ligand would clash with protein atoms (or other ligand atoms) if it were to rotate, it is clear that such a rotation would be difficult to sample in an MD simulation. In such case (steric hindrance), scaling down the LJ interactions along the alchemical pathway might enhance rotamer transitions.

In some systems, scaling down only the LJ- ϵ in the enhanced sampling protocol fails to sample different rotamer states sufficiently. For instance, a recent study by Lee et. al.

reported that the torsion angle energy terms can provide a significant sampling obstacle.[84] If a system experiences slow rotamer sampling due to a high force constant of the torsion angles that go through the rotatable bond and therefore a high barrier between rotamers, scaling down the force constants of the dihedral angles can lead to faster sampling of the equilibrium rotamer distribution. For some ligands in the BACE1 test system, scaling down the force constants of certain dihedrals to zero resulted in improved rotamer sampling of a pyrazine ring (Section 3.5.4). In the BACE1 system, the torsion terms that needed to be included in the enhanced sampling protocol were identified after running the free energy calculations and after determining the slow degree of freedom that led to slow convergence of the free energy estimate.

An approach to help identify the torsion terms that need to be scaled down prior to running the binding free energy simulations could be to perform a short vacuum simulation of the ligand, this time in the non-interacting state. If a rotatable bond in the ligand does not rotate in the non-interacting state, scaling down the LJ- ϵ would not lead to better sampling, but scaling down the force constant of the torsion angles can improve sampling.

Additionally, different ligand rotamers and therefore groups of the ligands that should be included in the enhanced sampling protocol could be identified using docking. Docked poses could be clustered based on different R-group rotamers and if multiple rotamers of an R-group were identified through docking, these R-groups should be included in the enhanced sampling protocol by either scaling the LJ- ϵ down further or scaling down dihedral force constants.

However, in both cases, scaling down the LJ- ϵ of a part of the molecule and scaling down dihedral force constants, it is unclear whether a general protocol can be found that works on a variety of systems. For instance, we found that scaling down the LJ- ϵ of a ligand R-group by a factor of $\gamma = 0.01$ led to good convergence in the MALT1 system (Figure 3.9, however, the same protocol did not work well when applying it to scaling down the LJ- ϵ of a pyridine

ring of a BACE1 ligand.

Finally, when talking about enhanced sampling protocols it is important to consider that softening interactions of the ligand (i.e. increasing the effective temperature of the system), increases the phase space that needs to be sampled and the thermodynamic length, meaning the distance between alchemical states. This can reduce the overlap of neighboring λ windows.[166] Since the acceptance rate of replica exchange swaps depends on the thermodynamic length, it is important to not only monitor the degree of freedom that the user wants to enhance (*e.g.* rotamer sampling), but also to monitor the effect of changes in the protocol on the HREX efficiency when determining a good enhanced sampling protocol. There are several metrics that can give insight into the HREX efficiency, such as the mean first passage time of a trajectory and the number of times a trajectory traverses through lambda space.

5.3 Comparing the computational efficiency of different methods for binding free energy calculations, SepTop, ABFE, and standard RBF

In Chapter 3 we speculate that the SepTop approach is computationally more efficient than ABFE calculations. In SepTop, the unbound state of the protein does not need to be sampled meaning that slow conformational changes upon ligand binding, either of the protein or of water molecules in a solvent exposed binding site, do not need to be sampled, at least as long as both ligands bind to the same metastable state of the protein. However, we have not yet performed ABFE and SepTop calculations on the same system to investigate whether SepTop is indeed more efficient than ABFE in general. Similarly, it would be interesting to investigate and compare the computational efficiency of the two RBF approaches, the standard RBF approach using a hybrid topology or the SepTop approach.

To investigate and compare the computational efficiency of ABFE, SepTop and standard RBFE approaches, the three methods first have to be applied to a variety of test systems, calculating the binding free energy for each of the systems using the three methods, since efficiency will certainly be a function of the system studied (e.g. as noted above, ABFE will be more challenging or even impractical when the protein must undergo a large conformational change or a change of protonation state on ligand binding). It is important to pick systems with varying degree of difficulty for the individual methods. For example, for the standard RBFE approach, this could mean including transformations that involve both small and large perturbations, while for the ABFE approach including systems where the protein remains in the same metastable state or undergoes large conformational changes upon ligand binding. Picking multiple systems with varying difficulty would ensure that each method is tested on systems where the method performs well and systems that involve challenges for the method. Note that all three methods need to be run on every test system and that a challenging system for one method (e.g. scaffold hopping transformations for standard RBFE) may not be challenging for another method (SepTop or ABFE). It may make sense to pick systems with prior information on the performance of different alchemical approaches, such as the protein-ligand benchmark set[54] or the BRD4 dataset[3]. Input parameters (e.g. the force field and starting coordinates) as well as simulation parameters should be held constant across different methods, to the extent that is possible.

After performing the calculations using the different approaches, the convergence of the free energy estimate as a function of simulation time (or number of force evaluations) can be analyzed to give insight into the computational efficiency of the approaches. An example of a comparison of the computational efficiency of different methods has been conducted in the SAMPL6 sampling challenge[128].

5.4 Developing a calculation planning tool for different binding free energy calculation approaches

Since binding free energy calculations are computationally demanding, investing simulation time in RBFE ligand transformations that quickly converge helps making the best use of computational resources. For standard RBFE calculations, ligand perturbation maps are typically created such that chemically similar compounds are connected, assuming those transformations to be most efficient.[91, 119]

If there is a significant difference between the computational efficiency of standard RBFE, SepTop and ABFE approaches, and if that difference depends on the systems (*e.g.* small versus large ligand transformation), it would be beneficial to always use the approach that is most efficient in a specific case. For example, if standard RBFE calculation are more efficient than SepTop for transformations that involve small perturbations (*e.g.* small R-group changes), a standard RBFE approach should be used for such a transformation. However, if for transformations that involve large perturbations, SepTop is computationally more efficient than standard RBFE, SepTop should be used. Without having performed a thorough comparison of the computational efficiency, lets assume that standard RBFE are computationally more efficient in cases of small R-group changes while SepTop is more efficient for larger transformations.

In such a case, a calculation planning tool can cluster ligands based on ligand similarity. Standard RBFE calculations are then performed between ligands within the same cluster, and the SepTop approach applied to transformations between ligands from different clusters, connecting the clusters with each other.

Additionally, planning efficient SepTop ligand networks requires a metric to assess which transformations are “easy” versus “difficult”, and such a metric has not yet been investi-

gated. Similarity metrics that are used to create efficient ligand maps for standard RBFE calculations[91] are often based on maximum common substructure similarity between ligands, which may not be a good metric for the SepTop approach (or, indeed, even for standard RBFE; efficiency has not yet been carefully studied). Instead, the shape and electrostatic similarity between ligands might be a better way to planning SepTop perturbation networks, as discussed in Section 3.6.4.

Furthermore, the equilibrated protein-ligand complexes that may be performed prior to the RBFE calculations provide additional “similarity” insight, not only into ligand similarity, but also similarity of protein conformation (*e.g.* side chain dihedral angles in the binding site), water network in the binding site etc. Harvesting information from such equilibration trajectories can potentially lead to more efficient perturbation maps. Imagine, *e.g.*, a protein where different ligands bind to different side chain rotamers of an amino acid in the binding site, and the transition between the rotamers is slow. In such case, creating a ligand map that clusters the ligands based on the rotamer they bind to, may be more efficient than a ligand map where the slow rotamer transition needs to be sampled in every ligand transformation. Therefore, it would be interesting to develop a tool that incorporates such information into the planning of ligand perturbation maps.

5.5 Evaluating different choices of restraining ligands in the SepTop approach

In the SepTop approach, the intramolecular interactions of ligands are decoupled and therefore ligands need to be restrained to the binding site to reduce the phase space that needs to be sampled making calculations tractable. In Chapter 3 we applied Boresch-style restraints[19] in the SepTop approach. These Boresch-style restraints are convenient for multiple reasons,

namely the contribution of removing the restraints in the decoupled state can be accounted for analytically, and secondly because the standard state dependence is incorporated and accounted for conveniently in the same analytical expression that calculates the restraining contribution (see Equation 32 Boresch et al.[19]). An additional benefit of the Boresch-style restraints is that the restraints are agnostic to how similar the ligands are, in contrast to the linked dual topology approach discussed below.

Even though Boresch-style restraints are convenient, they allow significant flexibility and movement of the ligand, and further restriction of degrees of freedom may increase the efficiency of the SepTop method. The linked dual topology approach is similar to SepTop with the difference that instead of restraining the ligands to the protein using Boresch-style restraints, the ligands are restrained to each other using *e.g.* harmonic restraints[126]. There are often no simple analytical solutions for calculating the free energy cost of restraining the decoupled ligand, meaning that the contribution has to be accounted for by running simulations which increases computational time, making this approach potentially more expensive.

Instead, one could combine both Boresch-style restraints and restraints that link the two ligands together, combining the convenience of the analytical correction of the Boresch-style restraints with the decreased flexibility of the ligands due to additional linked restraints. Here, the Boresch-restraints are applied to the decoupled ligand, then the Boresch-restraints are switched to the linked restraints such that during the alchemical transformation of the ligands, the ligands are restrained to each other. Instead of restraining ligands to each other using harmonic distance restraints, an alternative could be to restrain the ligands to each other based on the shape and electrostatic overlay of the two ligands. The downside, however, of restraining ligands to each other is that the restraints and potentially the computational efficiency of the SepTop approach then depend on ligand similarity while using restraints between the individual ligands and the protein should yield results relatively independent of

ligand similarity.

5.6 Additional ideas for improvement of the SepTop software package

The SepTop tools `SeparatedTopologies`[11] are under active development and require further improvement to make them more robust. For instance, when calculating suitable atoms for the Boresch-style restraints for BACE1 ligands that contained a macrocycle, ligand atoms in the large macrocycle ring were picked. Those atoms were involved in rotatable bonds which could restrain ligand conformational sampling and also could potentially lead to slow convergence when applying the restraints. In that case the macrocycle atoms were picked because the algorithm picked atoms in a central ring system and the macrocycle was counted as a ring system. Instead of picking atoms from any ring system, the algorithm could instead take a list of SMILES strings of “acceptable” structural elements, *e.g.* aromatic ring.

Lastly, initiating simulations at different stages in the thermodynamic cycle from different input starting coordinates might increase the efficiency of the method. In SepTop, simulations in the binding site are performed in two sets of simulations due to our multi-step alchemical pathway (see thermodynamic cycle Figure 3.1). Replica exchange swaps are performed between simulations in leg B and leg C while leg D is run in separate calculations. Currently, a single set of input coordinates is used for all λ windows in both sets of calculations in the binding site. The input coordinates of protein-ligand A complex (in an ligand A to ligand B transformation) are used to initiate simulations in all λ windows. Instead, one could use two different starting coordinates for the two separate steps in the thermodynamic cycle, taking coordinates from an equilibrated protein-ligand A simulation as input for step 1 and coordinates from an equilibrated protein-ligand B simulation as input for step 2. Especially

if the two ligands bind to different protein conformations or water rearrangements in the binding site, this could reduce equilibration times for the individual λ windows. However, care must be taken in this approach to avoid overly biasing results by the choice of starting coordinates.

Overall, we think that binding free energy calculation methods play an important role in the drug discovery process and that SepTop may be a general and broadly useful approach for RBFEE calculations that expands their domain of applicability. In the future, large scale benchmarking and application of different binding free energy calculation methods will help identify which methods will be most efficient and reliable in a specific system. This will then enable the practitioner to pick the method that is best suited for a specific system and to only apply the methods in a drug discovery project when they will add value.

Bibliography

- [1] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1-2:19–25, Sept. 2015.
- [2] M. Aldeghi, B. L. de Groot, and V. Gapsys. Accurate Calculation of Free Energy Changes upon Amino Acid Mutation. In T. Sikosek, editor, *Computational Methods in Protein Evolution*, volume 1851, pages 19–47. Springer New York, New York, NY, 2019. Series Title: Methods in Molecular Biology.
- [3] M. Aldeghi, A. Heifetz, M. J. Bodkin, S. Knapp, and P. C. Biggin. Accurate calculation of the absolute free energy of binding for drug molecules. *Chemical Science*, 7(1):207–218, 2016.
- [4] M. Aldeghi, A. Heifetz, M. J. Bodkin, S. Knapp, and P. C. Biggin. Predictions of Ligand Selectivity from Absolute Binding Free Energy Calculations. *Journal of the American Chemical Society*, 139(2):946–957, Jan. 2017.
- [5] I. Alibay, A. Magarkar, D. Seeliger, and P. C. Biggin. Evaluating the use of absolute binding free energy in the fragment optimisation process. *Communications Chemistry*, 5(1):105, Sept. 2022.
- [6] B. K. Allen, M. M. Kulkarni, B. Chamberlain, T. Dwight, C. Koh, R. Samant, F. Jernigan, J. Rice, D. Tan, S. Li, K. Marino, H. Huang, E. Chiswick, B. Tesar, S. Sparks, Z. Lin, T. D. McGee, I. Kolossváry, C. Lin, S. Shechter, H. Soutter, C. Bastos, M. Taimi, S. Lai, A. Petrin, T. Kane, S. Swann, H. Gardner, C. Winter, and W. Sherman. Design of a systemic small molecule clinical STING agonist using physics-based simulations and artificial intelligence. preprint, Cancer Biology, May 2022.
- [7] M. Amaral, D. B. Kokh, J. Bomke, A. Wegener, H. P. Buchstaller, H. M. Eggenweiler, P. Matias, C. Sirrenberg, R. C. Wade, and M. Frech. Protein conformational flexibility modulates kinetics and thermodynamics of drug binding. *Nature Communications*, 8(1):2276, Dec. 2017.
- [8] K. A. Armacost, S. Riniker, and Z. Cournia. Exploring Novel Directions in Free Energy Calculations. *Journal of Chemical Information and Modeling*, 60(11):5283–5286, Nov. 2020.

- [9] E. Awoonor-Williams, C. Dickson, P. Furet, A. Golosov, and V. Hornak. Molecular Insights into the Binding Interaction of the First Class of Potent Small Molecules Disrupting the YAP-TEAD Protein-Protein Interaction. preprint, Chemistry, Mar. 2023.
- [10] S. Azimi, S. Khuttan, J. Z. Wu, R. K. Pal, and E. Gallicchio. Relative Binding Free Energy Calculations for Ligands with Diverse Scaffolds with the Alchemical Transfer Method. *Journal of Chemical Information and Modeling*, 62(2):309–323, Jan. 2022.
- [11] H. Baumann and D. L. Mobley. MobleyLab/SeparatedTopologies: SepTop v1, Mar. 2023.
- [12] H. M. Baumann, E. Dybeck, C. L. McClendon, F. C. Pickard IV, V. Gapsys, L. Pérez-Benito, D. F. Hahn, G. Tresadern, A. M. Mathiowetz, and D. L. Mobley. Broadening the scope of binding free energy calculations using a Separated Topologies approach. *chemarxiv*, DOI: 10.26434/chemrxiv-2023-9678k(DOI: 10.26434/chemrxiv-2023-9678k), Mar. 2023.
- [13] H. M. Baumann, V. Gapsys, B. L. De Groot, and D. L. Mobley. Challenges Encountered Applying Equilibrium and Nonequilibrium Binding Free Energy Calculations. *The Journal of Physical Chemistry B*, 125(17):4241–4261, May 2021.
- [14] L. Bellmann, P. Penner, and M. Rarey. Topological Similarity Search in Large Combinatorial Fragment Spaces. *Journal of Chemical Information and Modeling*, 61(1):238–251, Jan. 2021.
- [15] I. Y. Ben-Shalom, C. Lin, T. Kurtzman, R. C. Walker, and M. K. Gilson. Simulating Water Exchange to Buried Binding Sites. *Journal of Chemical Theory and Computation*, 15(4):2684–2691, Apr. 2019.
- [16] C. H. Bennett. Efficient estimation of free energy differences from Monte Carlo data. *Journal of Computational Physics*, 22(2):245–268, Oct. 1976.
- [17] H. Berendsen, D. van der Spoel, and R. van Drunen. GROMACS: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications*, 91(1-3):43–56, Sept. 1995.
- [18] T. C. Beutler, A. E. Mark, R. C. van Schaik, P. R. Gerber, and W. F. van Gunsteren. Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations. *Chemical Physics Letters*, 222(6):529–539, June 1994.
- [19] S. Boresch, F. Tettinger, M. Leitgeb, and M. Karplus. Absolute Binding Free Energies: A Quantitative Approach for Their Calculation. *The Journal of Physical Chemistry B*, 107(35):9535–9551, Sept. 2003.
- [20] K. H. Burley, S. C. Gill, N. M. Lim, and D. L. Mobley. Enhancing Side Chain Rotamer Sampling Using Nonequilibrium Candidate Monte Carlo. *Journal of Chemical Theory and Computation*, 15(3):1848–1862, Mar. 2019.

- [21] W. Chen, D. Cui, R. Abel, R. A. Friesner, and L. Wang. Accurate calculation of absolute protein-ligand binding free energies. preprint, Chemistry, Apr. 2022.
- [22] M. Ciordia, L. Pérez-Benito, F. Delgado, A. A. Trabanco, and G. Tresadern. Application of Free Energy Perturbation for the Design of BACE1 Inhibitors. *Journal of Chemical Information and Modeling*, 56(9):1856–1871, Sept. 2016.
- [23] Z. Cournia, B. Allen, and W. Sherman. Relative Binding Free Energy Calculations in Drug Discovery: Recent Advances and Practical Considerations. *Journal of Chemical Information and Modeling*, 57(12):2911–2937, Dec. 2017.
- [24] Z. Cournia, B. K. Allen, T. Beuming, D. A. Pearlman, B. K. Radak, and W. Sherman. Rigorous Free Energy Simulations in Virtual Screening. *Journal of Chemical Information and Modeling*, 60(9):4153–4169, Sept. 2020.
- [25] G. E. Crooks. Nonequilibrium Measurements of Free Energy Differences for Microscopically Reversible Markovian Systems. *Journal of Statistical Physics*, 90(5/6):1481–1487, 1998.
- [26] G. E. Crooks. Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. *Physical Review E*, 60(3):2721–2726, Sept. 1999.
- [27] G. E. Crooks. Path-ensemble averages in systems driven far from equilibrium. *Physical Review E*, 61(3):2361–2366, Mar. 2000.
- [28] J. N. Cumming, E. M. Smith, L. Wang, J. Misiaszek, J. Durkin, J. Pan, U. Iserloh, Y. Wu, Z. Zhu, C. Strickland, J. Voigt, X. Chen, M. E. Kennedy, R. Kuvelkar, L. A. Hyde, K. Cox, L. Favreau, M. F. Czarniecki, W. J. Greenlee, B. A. McKittrick, E. M. Parker, and A. W. Stamford. Structure based design of iminohydantoin BACE1 inhibitors: Identification of an orally available, centrally active BACE1 inhibitor. *Bioorganic & Medicinal Chemistry Letters*, 22(7):2444–2449, Apr. 2012.
- [29] M. D’Abramo, N. Besker, G. Chillemi, and A. Grottesi. Modeling conformational transitions in kinases by molecular dynamics simulations: achievements, difficulties, and open challenges. *Frontiers in Genetics*, 5, May 2014.
- [30] F. Deflorian, L. Perez-Benito, E. B. Lenselink, M. Congreve, H. W. T. van Vlijmen, J. S. Mason, C. d. Graaf, and G. Tresadern. Accurate Prediction of GPCR Ligand Binding Affinity with Free Energy Perturbation. *Journal of Chemical Information and Modeling*, 60(11):5563–5579, Nov. 2020.
- [31] C. Dellago and G. Hummer. Computing Equilibrium Free Energies Using Non-Equilibrium Molecular Dynamics. *Entropy*, 16(1):41–61, Dec. 2013.
- [32] Y. Deng and B. Roux. Calculation of Standard Binding Free Energies: Aromatic Molecules in the T4 Lysozyme L99A Mutant. *Journal of Chemical Theory and Computation*, 2(5):1255–1273, Sept. 2006.

- [33] J. A. DiMasi, H. G. Grabowski, and R. W. Hansen. Innovation in the pharmaceutical industry: New estimates of R&D costs. *Journal of Health Economics*, 47:20–33, May 2016.
- [34] T. J. Dolinsky, P. Czodrowski, H. Li, J. E. Nielsen, J. H. Jensen, G. Klebe, and N. A. Baker. PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Research*, 35(Web Server):W522–W525, May 2007.
- [35] T. J. Dolinsky, J. E. Nielsen, J. A. McCammon, and N. A. Baker. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Research*, 32(Web Server):W665–W667, July 2004.
- [36] D. Dotson, O. Beckstein, D. Wille, I. Kenney, shuail, trje3733, H. Lee, V. Lim, B. Allen, and M. S. Barhaghi. alchemistry/alchemlyb: 0.3.1, Jan. 2020.
- [37] Eastman. Eastman P (2019) PDBFixer. <https://github.com/pandegroup/pdbfixer>.
- [38] Enamine. Enamine REAL compounds: <https://enamine.net/compound-collections/real-compounds> Accessed 04/18/2023, 2023.
- [39] A. Eriksson, W. Baase, X. Zhang, D. Heinz, M. Blaber, E. Baldwin, and B. Matthews. Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. *Science*, 255(5041):178, Jan. 1992.
- [40] M. Fajer, K. Borrelli, R. Abel, and L. Wang. Quantitatively Accounting for Protein Reorganization in Computer-Aided Drug Design. *Journal of Chemical Theory and Computation*, page acs.jctc.3c00009, May 2023.
- [41] M. Feng, G. Heinzemann, and M. K. Gilson. Absolute binding free energy calculations improve enrichment of actives in virtual compound screening. *Scientific Reports*, 12(1):13640, Aug. 2022.
- [42] E. Fischer. Einfluss der Configuration auf die Wirkung der Enzyme. *Berichte der deutschen chemischen Gesellschaft*, 27(3):2985–2993, Oct. 1894.
- [43] M. Fleck, M. Wieder, and S. Boresch. Dummy Atoms in Alchemical Free Energy Calculations. *Journal of Chemical Theory and Computation*, 17(7):4403–4419, July 2021.
- [44] K. Fuchino, Y. Mitsuoka, M. Masui, N. Kurose, S. Yoshida, K. Komano, T. Yamamoto, M. Ogawa, C. Unemura, M. Hosono, H. Ito, G. Sakaguchi, S. Ando, S. Ohnishi, Y. Kido, T. Fukushima, H. Miyajima, S. Hiroyama, K. Koyabu, D. Dhuyvetter, H. Borghys, H. J. M. Gijzen, Y. Yamano, Y. Iso, and K.-i. Kusakabe. Rational Design of Novel 1,3-Oxazine Based Beta-Secretase (BACE1) Inhibitors: Incorporation of a Double Bond To Reduce P-gp Efflux Leading to Robust Abeta Reduction in the Brain. *Journal of Medicinal Chemistry*, 61(12):5122–5137, June 2018.

- [45] V. Gapsys, S. Michielssens, D. Seeliger, and B. L. de Groot. pmx: Automated protein structure and topology generation for alchemical perturbations. *Journal of Computational Chemistry*, 36(5):348–354, Feb. 2015.
- [46] V. Gapsys, L. Pérez-Benito, M. Aldeghi, D. Seeliger, H. van Vlijmen, G. Tresadern, and B. L. de Groot. Large scale relative protein ligand binding affinities using non-equilibrium alchemy. *Chemical Science*, 11(4):1140–1152, 2020.
- [47] V. Gapsys, A. Yildirim, M. Aldeghi, Y. Khalak, D. van der Spoel, and B. L. de Groot. Accurate absolute free energies for ligand–protein binding based on non-equilibrium approaches. *Communications Chemistry*, 4(1):61, Dec. 2021.
- [48] Y. Ge, D. F. Hahn, and D. L. Mobley. A Benchmark of Electrostatic Method Performance in Relative Binding Free Energy Calculations. *Journal of Chemical Information and Modeling*, 61(3):1048–1052, Mar. 2021.
- [49] S. C. Gill, N. M. Lim, P. B. Grinaway, A. S. Rustenburg, J. Fass, G. A. Ross, J. D. Chodera, and D. L. Mobley. Binding Modes of Ligands Using Enhanced Sampling (BLUES): Rapid Decorrelation of Ligand Binding Modes via Nonequilibrium Candidate Monte Carlo. *The Journal of Physical Chemistry B*, 122(21):5579–5598, May 2018.
- [50] M. Gilson, J. Given, B. Bush, and J. McCammon. The statistical-thermodynamic basis for computation of binding affinities: a critical review. *Biophysical Journal*, 72(3):1047–1069, Mar. 1997.
- [51] A. A. Gorfe and A. Caflisch. Functional Plasticity in the Substrate Binding Site of Beta-Secretase. *Structure*, 13(10):1487–1498, Oct. 2005.
- [52] C. Gorgulla, A. Boeszoermyenyi, Z.-F. Wang, P. D. Fischer, P. W. Coote, K. M. Padmanabha Das, Y. S. Malets, D. S. Radchenko, Y. S. Moroz, D. A. Scott, K. Fackeldey, M. Hoffmann, I. Iavniuk, G. Wagner, and H. Arthanari. An open-source drug discovery platform enables ultra-large virtual screens. *Nature*, 580(7805):663–668, Apr. 2020.
- [53] A. A. Hagberg, D. A. Schult, and P. J. Swart. Exploring network structure, dynamics, and function using NetworkX. In G. Varoquaux, T. Vaught, and J. Millman, editors, *Proceedings of the 7th python in science conference*, pages 11 – 15, Pasadena, CA USA, 2008.
- [54] D. F. Hahn and J. Wagner. openforcefield/protein-ligand-benchmark: 0.2.0 Addition of new targets, Nov. 2021.
- [55] D. Hamelberg and J. A. McCammon. Standard Free Energy of Releasing a Localized Water Molecule from the Binding Pockets of Proteins: Double-Decoupling Method. *Journal of the American Chemical Society*, 126(24):7683–7689, June 2004.
- [56] G. Heinzelmann, N. M. Henriksen, and M. K. Gilson. Attach-Pull-Release Calculations of Ligand Binding and Conformational Changes on the First BRD4 Bromodomain. *Journal of Chemical Theory and Computation*, 13(7):3260–3275, July 2017.

- [57] J. Hermans and L. Wang. Inclusion of Loss of Translational and Rotational Freedom in Theoretical Estimates of Free Energies of Binding. Application to a Complex of Benzene and Mutant T4 Lysozyme. *Journal of the American Chemical Society*, 119(11):2707–2714, Mar. 1997.
- [58] L. Hong, G. Koelsch, X. Lin, S. Wu, S. Terzyan, A. K. Ghosh, X. C. Zhang, and J. Tang. Structure of the Protease Domain of Memapsin 2 (Beta-Secretase) Complexed with Inhibitor. *Science*, 290(5489):150–153, Oct. 2000.
- [59] L. Hong and J. Tang. Flap Position of Free Memapsin 2 (Beta-Secretase), a Model for Flap Opening in Aspartic Protease Catalysis. *Biochemistry*, 43(16):4689–4695, Apr. 2004.
- [60] K. W. Hunt, A. W. Cook, R. J. Watts, C. T. Clark, G. Vigers, D. Smith, A. T. Metcalf, I. W. Gunawardana, M. Burkard, A. A. Cox, M. K. Geck Do, D. Dutcher, A. A. Thomas, S. Rana, N. C. Kallan, R. K. DeLisle, J. P. Rizzi, K. Regal, D. Sammond, R. Groneberg, M. Siu, H. Purkey, J. P. Lyssikatos, A. Marlow, X. Liu, and T. P. Tang. Spirocyclic β -Site Amyloid Precursor Protein Cleaving Enzyme 1 (BACE1) Inhibitors: From Hit to Lowering of Cerebrospinal Fluid (CSF) Amyloid β in a Higher Species. *Journal of Medicinal Chemistry*, 56(8):3379–3403, Apr. 2013.
- [61] I. Hussain, J. Hawkins, D. Harrison, C. Hille, G. Wayne, L. Cutler, T. Buck, D. Walter, E. Demont, C. Howes, A. Naylor, P. Jeffrey, M. I. Gonzalez, C. Dingwall, A. Michel, S. Redshaw, and J. B. Davis. Oral administration of a potent and selective non-peptidic BACE-1 inhibitor decreases beta-cleavage of amyloid precursor protein and amyloid-beta production in vivo. *Journal of Neurochemistry*, 100(3):802–809, Feb. 2007.
- [62] J. R. Huth, C. Park, A. M. Petros, A. R. Kunzer, M. D. Wendt, X. Wang, C. L. Lynch, J. C. Mack, K. M. Swift, R. A. Judge, J. Chen, P. L. Richardson, S. Jin, S. K. Tahir, E. D. Matayoshi, S. A. Dorwin, U. S. Lador, J. M. Severin, K. A. Walter, D. M. Bartley, S. W. Fesik, S. W. Elmore, and P. J. Hajduk. Discovery and Design of Novel HSP90 Inhibitors Using Multiple Fragment-based Design Strategies. *Chemical Biology & Drug Design*, 70(1):1–12, July 2007.
- [63] R. Ishima, D. I. Freedberg, Y.-X. Wang, J. M. Louis, and D. A. Torchia. Flap opening and dimer-interface flexibility in the free and inhibitor-bound HIV protease, and their implications for function. *Structure*, 7(9):1047–S12, Sept. 1999.
- [64] A. Jakalian, D. B. Jack, and C. I. Bayly. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *Journal of Computational Chemistry*, 23(16):1623–1641, Oct. 2002.
- [65] C. Jarzynski. Equilibrium free-energy differences from nonequilibrium measurements: A master-equation approach. *Physical Review E*, 56(5):5018–5035, Nov. 1997.
- [66] C. Jarzynski. Nonequilibrium Equality for Free Energy Differences. *Physical Review Letters*, 78(14):2690–2693, Apr. 1997.

- [67] C. Jarzynski. Rare events and the convergence of exponentially averaged work values. *Physical Review E*, 73(4):046105, Apr. 2006.
- [68] F. Jeppsson, S. Eketjäll, J. Janson, S. Karlström, S. Gustavsson, L.-L. Olsson, A.-C. Radesäter, B. Ploeger, G. Cebers, K. Kolmodin, B.-M. Swahn, S. Von Berg, T. Bueters, and J. Fälting. Discovery of AZD3839, a Potent and Selective BACE1 Inhibitor Clinical Candidate for the Treatment of Alzheimer Disease. *Journal of Biological Chemistry*, 287(49):41245–41257, Nov. 2012.
- [69] W. Jiang, J. Thirman, S. Jo, and B. Roux. Reduced Free Energy Perturbation/Hamiltonian Replica Exchange Molecular Dynamics Method with Unbiased Alchemical Thermodynamic Axis. *The Journal of Physical Chemistry B*, 122(41):9435–9442, Oct. 2018.
- [70] J. Jordan. <https://gist.github.com/joe-jordan/6548029>. 10/20/2022.
- [71] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, 79(2):926–935, July 1983.
- [72] W. L. Jorgensen and C. Ravimohan. Monte Carlo simulation of differences in free energies of hydration. *The Journal of Chemical Physics*, 83(6):3050–3054, Sept. 1985.
- [73] B. P. Kelley, S. P. Brown, G. L. Warren, and S. W. Muchmore. POSIT: Flexible Shape-Guided Docking For Pose Prediction. *Journal of Chemical Information and Modeling*, 55(8):1771–1780, Aug. 2015.
- [74] H. Keränen, L. Pérez-Benito, M. Ciordia, F. Delgado, T. B. Steinbrecher, D. Oehlich, H. W. T. van Vlijmen, A. A. Trabanco, and G. Tresadern. Acylguanidine Beta Secretase 1 Inhibitors: A Combined Experimental and Free Energy Perturbation Study. *Journal of Chemical Theory and Computation*, 13(3):1439–1453, Mar. 2017.
- [75] Y. Khalak, G. Tresadern, M. Aldeghi, H. M. Baumann, D. L. Mobley, B. L. de Groot, and V. Gapsys. Alchemical absolute protein–ligand binding free energies for drug design. *Chemical Science*, 12(41):13958–13971, 2021.
- [76] Y. Khalak, G. Tresadern, B. L. de Groot, and V. Gapsys. Non-equilibrium approach for binding free energies in cyclodextrins in SAMPL7: force fields and software. *Journal of Computer-Aided Molecular Design*, 35(1):49–61, Jan. 2021.
- [77] J. G. Kirkwood. Statistical Mechanics of Fluid Mixtures. *The Journal of Chemical Physics*, 3(5):300–313, May 1935.
- [78] P. V. Klimovich, M. R. Shirts, and D. L. Mobley. Guidelines for the analysis of free energy calculations. *Journal of Computer-Aided Molecular Design*, 29(5):397–411, May 2015.
- [79] J. L. Knight and C. L. Brooks. Lambda-Dynamics free energy simulation methods. *Journal of Computational Chemistry*, 30(11):1692–1700, Aug. 2009.

- [80] Y. Koriyama, A. Hori, H. Ito, S. Yonezawa, Y. Baba, N. Tanimoto, T. Ueno, S. Yamamoto, T. Yamamoto, N. Asada, K. Morimoto, S. Einaru, K. Sakai, T. Kanazu, A. Matsuda, Y. Yamaguchi, T. Oguma, M. Timmers, L. Tritsmans, K.-i. Kusakabe, A. Kato, and G. Sakaguchi. Discovery of Atabecestat (JNJ-54861911): A Thiazine-Based Beta-Amyloid Precursor Protein Cleaving Enzyme 1 Inhibitor Advanced to the Phase 2b/3 EARLY Clinical Trial. *Journal of Medicinal Chemistry*, 64(4):1873–1888, Feb. 2021.
- [81] B. Kuhn, M. Tichý, L. Wang, S. Robinson, R. E. Martin, A. Kuglstatter, J. Benz, M. Giroud, T. Schirmeister, R. Abel, F. Diederich, and J. Hert. Prospective Evaluation of Free Energy Calculations for the Prioritization of Cathepsin L Inhibitors. *Journal of Medicinal Chemistry*, 60(6):2485–2497, Mar. 2017.
- [82] T.-S. Lee, B. K. Allen, T. J. Giese, Z. Guo, P. Li, C. Lin, T. D. McGee, D. A. Pearlman, B. K. Radak, Y. Tao, H.-C. Tsai, H. Xu, W. Sherman, and D. M. York. Alchemical Binding Free Energy Calculations in AMBER20: Advances and Best Practices for Drug Discovery. *Journal of Chemical Information and Modeling*, 60(11):5595–5623, Nov. 2020.
- [83] T.-S. Lee, Z. Lin, B. K. Allen, C. Lin, B. K. Radak, Y. Tao, H.-C. Tsai, W. Sherman, and D. M. York. Improved Alchemical Free Energy Calculations with Optimized Smoothstep Softcore Potentials. *Journal of Chemical Theory and Computation*, 16(9):5512–5525, Sept. 2020.
- [84] T.-S. Lee, H.-C. Tsai, A. Ganguly, and D. M. York. ACES: Optimized Alchemically Enhanced Sampling. *Journal of Chemical Theory and Computation*, page acs.jctc.2c00697, Jan. 2023.
- [85] S. P. Leelananda and S. Lindert. Computational methods in drug discovery. *Beilstein Journal of Organic Chemistry*, 12:2694–2718, Dec. 2016.
- [86] J. Liang, V. Tsui, A. Van Abbema, L. Bao, K. Barrett, M. Beresini, L. Berezhkovskiy, W. S. Blair, C. Chang, J. Driscoll, C. Eigenbrot, N. Ghilardi, P. Gibbons, J. Halladay, A. Johnson, P. B. Kohli, Y. Lai, M. Liimatta, P. Mantik, K. Menghrajani, J. Murray, A. Sambrone, Y. Xiao, S. Shia, Y. Shin, J. Smith, S. Sohn, M. Stanley, M. Ultsch, B. Zhang, L. C. Wu, and S. Magnuson. Lead identification of novel and selective TYK2 inhibitors. *European Journal of Medicinal Chemistry*, 67:175–187, Sept. 2013.
- [87] N. M. Lim, L. Wang, R. Abel, and D. L. Mobley. Sensitivity in Binding Free Energies Due to Protein Reorganization. *Journal of Chemical Theory and Computation*, 12(9):4620–4631, Sept. 2016.
- [88] Y.-L. Lin, A. Aleksandrov, T. Simonson, and B. Roux. An Overview of Electrostatic Free Energy Computations for Solutions and Proteins. *Journal of Chemical Theory and Computation*, 10(7):2690–2709, July 2014.

- [89] Z. Lin, J. Zou, S. Liu, C. Peng, Z. Li, X. Wan, D. Fang, J. Yin, G. Gobbo, Y. Chen, J. Ma, S. Wen, P. Zhang, and M. Yang. A Cloud Computing Platform for Scalable Relative and Absolute Binding Free Energy Predictions: New Opportunities and Challenges for Drug Discovery. *Journal of Chemical Information and Modeling*, 61(6):2720–2732, June 2021.
- [90] K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror, and D. E. Shaw. Improved side-chain torsion potentials for the Amber ff99SB protein force field: Improved Protein Side-Chain Potentials. *Proteins: Structure, Function, and Bioinformatics*, 78(8):1950–1958, June 2010.
- [91] S. Liu, Y. Wu, T. Lin, R. Abel, J. P. Redmann, C. M. Summa, V. R. Jaber, N. M. Lim, and D. L. Mobley. Lead optimization mapper: automating free energy calculations for lead optimization. *Journal of Computer-Aided Molecular Design*, 27(9):755–770, Sept. 2013.
- [92] T. P. Lybrand, J. A. McCammon, and G. Wipff. Theoretical calculation of relative binding affinity in host-guest systems. *Proceedings of the National Academy of Sciences*, 83(4):833–835, Feb. 1986.
- [93] M. Macchiagodena, M. Pagliai, M. Karrenbrock, G. Guarnieri, F. Iannone, and P. Procacci. Virtual Double-System Single-Box: A Nonequilibrium Alchemical Technique for Absolute Binding Free Energy Calculations: Application to Ligands of the SARS-CoV-2 Main Protease. *Journal of Chemical Theory and Computation*, 16(11):7160–7172, Nov. 2020.
- [94] H. B. Macdonald, D. F. Hahn, M. Henry, J. Chodera, D. Dotson, W. Glass, and I. Pulido. openforcefield/openff-arsenic: v0.2.1, Feb. 2022.
- [95] J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, and C. Simmerling. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation*, 11(8):3696–3713, Aug. 2015.
- [96] M. S. Malamas, J. Erdei, I. Gunawan, J. Turner, Y. Hu, E. Wagner, K. Fan, R. Chopra, A. Olland, J. Bard, S. Jacobsen, R. L. Magolda, M. Pangalos, and A. J. Robichaud. Design and Synthesis of 5,5'-Disubstituted Aminohydantoins as Potent and Selective Human Beta-Secretase (BACE1) Inhibitors. *Journal of Medicinal Chemistry*, 53(3):1146–1158, Feb. 2010.
- [97] P. C. May, R. A. Dean, S. L. Lowe, F. Martenyi, S. M. Sheehan, L. N. Boggs, S. A. Monk, B. M. Mathes, D. J. Mergott, B. M. Watson, S. L. Stout, D. E. Timm, E. Smith LaBell, C. R. Gonzales, M. Nakano, S. S. Jhee, M. Yen, L. Ereshefsky, T. D. Lindstrom, D. O. Calligaro, P. J. Cocke, D. Greg Hall, S. Friedrich, M. Citron, and J. E. Audia. Robust Central Reduction of Amyloid-Beta in Humans with an Orally Available, Non-Peptidic Beta-Secretase Inhibitor. *The Journal of Neuroscience*, 31(46):16507–16516, Nov. 2011.

- [98] K. Meier, J. P. Bluck, and C. D. Christ. Use of Free Energy Methods in the Drug Discovery Industry. In *Free Energy Methods in Drug Discovery: Current State and Future Directions*, volume 1397 of *ACS Symposium Series*, pages 39–66. American Chemical Society, Nov. 2021. Section: 2.
- [99] A. S. J. S. Mey, B. Allen, H. E. B. Macdonald, J. D. Chodera, M. Kuhn, J. Michel, D. L. Mobley, L. N. Naden, S. Prasad, A. Rizzi, J. Scheen, M. R. Shirts, G. Tresadern, and H. Xu. Best Practices for Alchemical Free Energy Calculations. *arXiv:2008.03067 [q-bio, stat]*, Aug. 2020. arXiv: 2008.03067.
- [100] D. L. Mobley, J. D. Chodera, and K. A. Dill. On the use of orientational restraints and symmetry corrections in alchemical free energy calculations. *The Journal of Chemical Physics*, 125(8):084902, Aug. 2006.
- [101] D. L. Mobley, J. D. Chodera, and K. A. Dill. Confine-and-Release Method: Obtaining Correct Binding Free Energies in the Presence of Protein Conformational Change. *Journal of Chemical Theory and Computation*, 3(4):1231–1235, July 2007.
- [102] D. L. Mobley and M. K. Gilson. Predicting Binding Free Energies: Frontiers and Benchmarks. *Annual Review of Biophysics*, 46(1):531–558, May 2017.
- [103] D. L. Mobley, A. P. Graves, J. D. Chodera, A. C. McReynolds, B. K. Shoichet, and K. A. Dill. Predicting Absolute Ligand Binding Free Energies to a Simple Model Site. *Journal of Molecular Biology*, 371(4):1118–1134, Aug. 2007.
- [104] D. L. Mobley and P. V. Klimovich. Perspective: Alchemical free energy calculations for drug discovery. *The Journal of Chemical Physics*, 137(23):230901, Dec. 2012.
- [105] D. L. Mobley and D. Slochower. Mobleylab/Benchmarksets: Version 1.2, Aug. 2017.
- [106] K. Nakahara, K. Fuchino, K. Komano, N. Asada, G. Tadano, T. Hasegawa, T. Yamamoto, Y. Sako, M. Ogawa, C. Unemura, M. Hosono, H. Ito, G. Sakaguchi, S. Ando, S. Ohnishi, Y. Kido, T. Fukushima, D. Dhuyvetter, H. Borghys, H. J. M. Gijzen, Y. Yamano, Y. Iso, and K.-i. Kusakabe. Discovery of Potent and Centrally Active 6-Substituted 5-Fluoro-1,3-dihydro-oxazine Beta-Secretase (BACE1) Inhibitors via Active Conformation Stabilization. *Journal of Medicinal Chemistry*, 61(13):5525–5546, July 2018.
- [107] U. Neumann, H. Rueeger, R. Machauer, S. J. Veenstra, R. M. Lueoend, M. Tintelnot-Blomley, G. Laue, K. Beltz, B. Vogg, P. Schmid, W. Frieauff, D. R. Shimshek, M. Staufenbiel, and L. H. Jacobson. A novel BACE inhibitor NB-360 shows a superior pharmacological profile and robust reduction of amyloid-beta and neuroinflammation in APP transgenic mice. *Molecular Neurodegeneration*, 10(1):44, Dec. 2015.
- [108] D. Oehlich, A. Peschiulli, G. Tresadern, M. Van Gool, J. A. Vega, A. I. De Lucas, S. A. Alonso de Diego, H. Prokopcova, N. Austin, S. Van Brandt, M. Surkyn, M. De Cleyn, A. Vos, F. J. R. Rombouts, G. Macdonald, D. Moechars, H. J. M. Gijzen, and A. A. Trabanco. Evaluation of a Series of Beta-Secretase 1 Inhibitors Containing Novel

- Heteroaryl-Fused-Piperazine Amidine Warheads. *ACS Medicinal Chemistry Letters*, 10(8):1159–1165, Aug. 2019.
- [109] D. Oehlrich, H. Prokopcova, and H. J. Gijzen. The evolution of amidine-based brain penetrant BACE1 inhibitors. *Bioorganic & Medicinal Chemistry Letters*, 24(9):2033–2045, May 2014.
- [110] OpenEye. OEDOCKING 4.1.1.0: OpenEye Scientific Software, Inc., Santa Fe, NM. <http://www.eyesopen.com>. 5/23/2022.
- [111] OpenEye. OpenEye Toolkits 2021 OpenEye Scientific Software, Santa Fe, NM. <http://www.eyesopen.com>. 5/23/22.
- [112] OpenEye. Spruce 1.4.0.0: OpenEye Scientific Software, Santa Fe, NM. <http://www.eyesopen.com>. 5/23/22.
- [113] A. Ozen, E. Perola, N. Brooijmans, and J. Kim. Prospective Application of Free Energy Methods in Drug Discovery Programs. In *Free Energy Methods in Drug Discovery: Current State and Future Directions*, volume 1397 of *ACS Symposium Series*, pages 127–141. American Chemical Society, Nov. 2021. Section: 5.
- [114] R. K. Pal and E. Gallicchio. Perturbation potentials to overcome order/disorder transitions in alchemical binding free energy calculations. *The Journal of Chemical Physics*, 151(12):124116, Sept. 2019.
- [115] C. D. Parks, Z. Gaieb, M. Chiu, H. Yang, C. Shao, W. P. Walters, J. M. Jansen, G. McGaughey, R. A. Lewis, S. D. Bembenek, M. K. Ameriks, T. Mirzadegan, S. K. Burley, R. E. Amaro, and M. K. Gilson. D3R grand challenge 4: blind prediction of protein–ligand poses, affinity rankings, and relative binding free energies. *Journal of Computer-Aided Molecular Design*, 34(2):99–119, Feb. 2020.
- [116] S. Patel, L. Vuillard, A. Cleasby, C. W. Murray, and J. Yon. Apo and Inhibitor Complex Structures of BACE (Beta-secretase). *Journal of Molecular Biology*, 343(2):407–416, Oct. 2004.
- [117] J. W. Perthold, D. Petrov, and C. Oostenbrink. Toward Automated Free Energy Calculation with Accelerated Enveloping Distribution Sampling (A-EDS). *Journal of Chemical Information and Modeling*, 60(11):5395–5406, Nov. 2020.
- [118] M. Pitman. pitmanme/system_builder: System Builder v1.0.0 (v1.0.0). Zenodo., June 2022.
- [119] M. Pitman, D. F. Hahn, G. Tresadern, and D. L. Mobley. To Design Scalable Free Energy Perturbation Networks, Optimal Is Not Enough. *Journal of Chemical Information and Modeling*, 63(6):1776–1793, Mar. 2023.
- [120] P. Procacci. Solvation free energies *via* alchemical simulations: let’s get honest about sampling, once more. *Physical Chemistry Chemical Physics*, 21(25):13826–13834, 2019.

- [121] P. Procacci. Relative Binding Free Energy between Chemically Distant Compounds Using a Bidirectional Nonequilibrium Approach. *Journal of Chemical Theory and Computation*, 18(6):4014–4026, June 2022.
- [122] L. Pérez-Benito, H. Keränen, H. Van Vlijmen, and G. Tresadern. Predicting Binding Free Energies of PDE2 Inhibitors. The Difficulties of Protein Conformation. *Scientific Reports*, 8(1):4883, Mar. 2018.
- [123] Y. Qiu, D. G. A. Smith, S. Boothroyd, H. Jang, D. F. Hahn, J. Wagner, C. C. Bannan, T. Gokey, V. T. Lim, C. D. Stern, A. Rizzi, B. Tjanaka, G. Tresadern, X. Lucas, M. R. Shirts, M. K. Gilson, J. D. Chodera, C. I. Bayly, D. L. Mobley, and L.-P. Wang. Development and Benchmarking of Open Force Field v1.0.0—the Parsley Small-Molecule Force Field. *Journal of Chemical Theory and Computation*, 17(10):6262–6280, Oct. 2021.
- [124] M. M. Reif and P. H. Hünenberger. Computation of methodology-independent single-ion solvation properties from molecular simulations. III. Correction terms for the solvation free energies, enthalpies, entropies, heat capacities, volumes, compressibilities, and expansivities of solvated ions. *The Journal of Chemical Physics*, 134(14):144103, Apr. 2011.
- [125] T. I. Richardson, J. A. Dodge, G. L. Durst, L. A. Pfeifer, J. Shah, Y. Wang, J. D. Durbin, V. Krishnan, and B. H. Norman. Benzopyrans as selective estrogen receptor Beta agonists (SERBAs). Part 3: Synthesis of cyclopentanone and cyclohexanone intermediates for C-ring modification. *Bioorganic & Medicinal Chemistry Letters*, 17(17):4824–4828, Sept. 2007.
- [126] B. Ries, S. Rieder, C. Rhiner, P. H. Hünenberger, and S. Riniker. RestraintMaker: a graph-based approach to select distance restraints in free-energy calculations with dual topology. *Journal of Computer-Aided Molecular Design*, Mar. 2022.
- [127] A. Rizzi, J. Chodera, L. Naden, K. Beauchamp, P. Grinaway, B. Rustenburg, S. Albanese, and S. Saladi. Choderalab/Yank: 0.23.0 Multi-Analysis, July 2018.
- [128] A. Rizzi, T. Jensen, D. R. Slochower, M. Aldeghi, V. Gapsys, D. Ntekoumes, S. Bosisio, M. Papadourakis, N. M. Henriksen, B. L. de Groot, Z. Cournia, A. Dickson, J. Michel, M. K. Gilson, M. R. Shirts, D. L. Mobley, and J. D. Chodera. The SAMPL6 SAMPLing challenge: assessing the reliability and efficiency of binding free energy calculations. *Journal of Computer-Aided Molecular Design*, 34(5):601–633, May 2020.
- [129] G. J. Rocklin, D. L. Mobley, and K. A. Dill. Separated topologies—A method for relative binding free energy calculations using orientational restraints. *The Journal of Chemical Physics*, 138(8):085104, Feb. 2013.
- [130] G. J. Rocklin, D. L. Mobley, K. A. Dill, and P. H. Hünenberger. Calculating the binding free energies of charged species based on explicit-solvent simulations employing lattice-sum methods: An accurate correction scheme for electrostatic finite-size effects. *The Journal of Chemical Physics*, 139(18):184103, Nov. 2013.

- [131] G. A. Ross, E. Russell, Y. Deng, C. Lu, E. D. Harder, R. Abel, and L. Wang. Enhancing Water Sampling in Free Energy Calculations with Grand Canonical Monte Carlo. *Journal of Chemical Theory and Computation*, 16(10):6061–6076, Oct. 2020.
- [132] A. A. Sadybekov, A. V. Sadybekov, Y. Liu, C. Iliopoulos-Tsoutsouvas, X.-P. Huang, J. Pickett, B. Houser, N. Patel, N. K. Tran, F. Tong, N. Zvonok, M. K. Jain, O. Savych, D. S. Radchenko, S. P. Nikas, N. A. Petasis, Y. S. Moroz, B. L. Roth, A. Makriyannis, and V. Katritch. Synthron-based ligand discovery in virtual libraries of over 11 billion compounds. *Nature*, 601(7893):452–459, Jan. 2022.
- [133] C. E. M. Schindler, H. Baumann, A. Blum, D. Böse, H.-P. Buchstaller, L. Burgdorf, D. Cappel, E. Chekler, P. Czodrowski, D. Dorsch, M. K. I. Eguida, B. Follows, T. Fuchß, U. Grädler, J. Gunera, T. Johnson, C. Jorand Lebrun, S. Karra, M. Klein, T. Knehans, L. Koetzner, M. Krier, M. Leiendecker, B. Leuthner, L. Li, I. Mochalkin, D. Musil, C. Neagu, F. Rippmann, K. Schiemann, R. Schulz, T. Steinbrecher, E.-M. Tanzer, A. Unzue Lopez, A. Viacava Follis, A. Wegener, and D. Kuhn. Large-Scale Assessment of Binding Free Energy Calculations in Active Drug Discovery Projects. *Journal of Chemical Information and Modeling*, 60(11):5457–5474, Nov. 2020.
- [134] D. Seeliger and B. L. de Groot. Protein Thermostability Calculations Using Alchemical Free Energy Simulations. *Biophysical Journal*, 98(10):2309–2316, May 2010.
- [135] Y. Shan, E. T. Kim, M. P. Eastwood, R. O. Dror, M. A. Seeliger, and D. E. Shaw. How Does a Drug Molecule Find Its Target Binding Site? *Journal of the American Chemical Society*, 133(24):9181–9183, June 2011.
- [136] B. Sherborne, V. Shanmugasundaram, A. C. Cheng, C. D. Christ, R. L. DesJarlais, J. S. Duca, R. A. Lewis, D. A. Loughney, E. S. Manas, G. B. McGaughey, C. E. Peishoff, and H. van Vlijmen. Collaborating to improve the use of free-energy and other quantitative methods in drug discovery. *Journal of Computer-Aided Molecular Design*, 30(12):1139–1141, Dec. 2016.
- [137] Y. Shi, M. L. Laury, Z. Wang, and J. W. Ponder. AMOEBA binding free energies for the SAMPL7 TrimerTrip host–guest challenge. *Journal of Computer-Aided Molecular Design*, 35(1):79–93, Jan. 2021.
- [138] H. Shimizu, A. Tosaki, K. Kaneko, T. Hisano, T. Sakurai, and N. Nukina. Crystal Structure of an Active Form of BACE1, an Enzyme Responsible for Amyloid Beta Protein Production. *Molecular and Cellular Biology*, 28(11):3663–3671, June 2008.
- [139] M. R. Shirts, E. Bair, G. Hooker, and V. S. Pande. Equilibrium Free Energies from Nonequilibrium Measurements Using Maximum-Likelihood Methods. *Physical Review Letters*, 91(14):140601, Oct. 2003.
- [140] M. R. Shirts and J. D. Chodera. Statistically optimal analysis of samples from multiple equilibrium states. *The Journal of Chemical Physics*, 129(12):124105, Sept. 2008.

- [141] M. R. Shirts, D. L. Mobley, and J. D. Chodera. Chapter 4 Alchemical Free Energy Calculations: Ready for Prime Time? In *Annual Reports in Computational Chemistry*, volume 3, pages 41–59. Elsevier, 2007.
- [142] M. R. Shirts and V. S. Pande. Comparison of efficiency and bias of free energies computed by exponential averaging, the Bennett acceptance ratio, and thermodynamic integration. *The Journal of Chemical Physics*, 122(14):144107, Apr. 2005.
- [143] D. Shuto, S. Kasai, T. Kimura, P. Liu, K. Hidaka, T. Hamada, S. Shibakawa, Y. Hayashi, C. Hattori, B. Szabo, S. Ishiura, and Y. Kiso. KMI-008, a novel Beta-Secretase inhibitor containing a hydroxymethylcarbonyl isostere as a transition-State mimic: design and synthesis of substrate-based octapeptides. *Bioorganic & Medicinal Chemistry Letters*, 13(24):4273–4276, Dec. 2003.
- [144] A. W. Stamford, J. D. Scott, S. W. Li, S. Babu, D. Tadesse, R. Hunter, Y. Wu, J. Misiaszek, J. N. Cumming, E. J. Gilbert, C. Huang, B. A. McKittrick, L. Hong, T. Guo, Z. Zhu, C. Strickland, P. Orth, J. H. Voigt, M. E. Kennedy, X. Chen, R. Kuvelkar, R. Hodgson, L. A. Hyde, K. Cox, L. Favreau, E. M. Parker, and W. J. Greenlee. Discovery of an Orally Available, Brain Penetrant BACE1 Inhibitor That Affords Robust CNS A β Reduction. *ACS Medicinal Chemistry Letters*, 3(11):897–902, Nov. 2012.
- [145] T. Steinbrecher and A. Labahn. Towards Accurate Free Energy Calculations in Ligand Protein-Binding Studies. *Current Medicinal Chemistry*, 17(8):767–785, Mar. 2010.
- [146] Y. Sugita, A. Kitao, and Y. Okamoto. Multidimensional replica-exchange method for free-energy calculations. *The Journal of Chemical Physics*, 113(15):6042–6051, Oct. 2000.
- [147] Z. X. Sun, X. H. Wang, and J. Z. H. Zhang. BAR-based optimum adaptive sampling regime for variance minimization in alchemical transformation. *Physical Chemistry Chemical Physics*, 19(23):15005–15020, 2017.
- [148] A. A. Trabanco-Suarez, F. Delgado-Jimenez, J. A. Vega Ramiro, G. J. Tresadern, H. J. M. Gijzen, and D. Oehlich. Preparation of 5,6-dihydro-imidazo[1,2-a]pyrazin-8-ylamine derivatives as beta-secretase inhibitors. *WO2012085038 A1*, 2012.
- [149] A. A. Trabanco-Suarez, F. J. R. Rombouts, G. J. Tresadern, M. L. M. Van Gool, G. J. Macdonald, C. Martinez Lamenca, and H. J. M. Gijzen. Preparation of 5,6-dihydro-2H-[1,4]oxazin-3-yl-amine derivatives useful as therapeutic inhibitors of Beta-secretase (BACE). *WO2011154431 A1*, 2011.
- [150] A. A. Trabanco-Suarez, G. J. Tresadern, and F. Delgado-Jimenez. 4,7-Dihydropyrazolo[1,5-a]pyrazin-6-yl-amine derivatives as Beta-secretase inhibitors and their preparation and use for the treatment of diseases. *WO2012038438 A1*, 2012.
- [151] J. Wagner, M. Thompson, D. Dotson, Hyejang, SimonBoothroyd, and J. Rodríguez-Guerra. openforcefield/openff-forcefields: Version 2.0.0 "Sage", Aug. 2021.

- [152] J. Wahl and M. Smieško. Assessing the Predictive Power of Relative Binding Free Energy Calculations for Test Cases Involving Displacement of Binding Site Water Molecules. *Journal of Chemical Information and Modeling*, 59(2):754–765, Feb. 2019.
- [153] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case. Development and testing of a general amber force field. *Journal of Computational Chemistry*, 25(9):1157–1174, July 2004.
- [154] L. Wang, Y. Deng, Y. Wu, B. Kim, D. N. LeBard, D. Wandschneider, M. Beachy, R. A. Friesner, and R. Abel. Accurate Modeling of Scaffold Hopping Transformations in Drug Discovery. *Journal of Chemical Theory and Computation*, 13(1):42–54, Jan. 2017.
- [155] L. Wang, R. A. Friesner, and B. J. Berne. Replica Exchange with Solute Scaling: A More Efficient Version of Replica Exchange with Solute Tempering (REST2). *The Journal of Physical Chemistry B*, 115(30):9431–9438, Aug. 2011.
- [156] L. Wang, Y. Wu, Y. Deng, B. Kim, L. Pierce, G. Krilov, D. Lupyan, S. Robinson, M. K. Dahlgren, J. Greenwood, D. L. Romero, C. Masse, J. L. Knight, T. Steinbrecher, T. Beuming, W. Damm, E. Harder, W. Sherman, M. Brewer, R. Wester, M. Murcko, L. Frye, R. Farid, T. Lin, D. L. Mobley, W. L. Jorgensen, B. J. Berne, R. A. Friesner, and R. Abel. Accurate and Reliable Prediction of Relative Ligand Binding Potency in Prospective Drug Discovery by Way of a Modern Free-Energy Calculation Protocol and Force Field. *Journal of the American Chemical Society*, 137(7):2695–2703, Feb. 2015.
- [157] B. J. Williams-Noonan, E. Yuriev, and D. K. Chalmers. Free Energy Methods in Drug Design: Prospects of “Alchemical Perturbation” in Medicinal Chemistry: Miniperspective. *Journal of Medicinal Chemistry*, 61(3):638–649, Feb. 2018.
- [158] C. F. Wong and J. A. McCammon. Dynamics and design of enzymes and inhibitors. *Journal of the American Chemical Society*, 108(13):3830–3832, June 1986.
- [159] H.-J. Woo and B. Roux. Calculation of absolute protein-ligand binding free energy from computer simulations. *Proceedings of the National Academy of Sciences*, 102(19):6825–6830, May 2005.
- [160] C. J. Woods, J. W. Essex, and M. A. King. The Development of Replica-Exchange-Based Free-Energy Methods. *The Journal of Physical Chemistry B*, 107(49):13703–13710, Dec. 2003.
- [161] C. J. Woods, J. W. Essex, and M. A. King. Enhanced Configurational Sampling in Binding Free-Energy Calculations. *The Journal of Physical Chemistry B*, 107(49):13711–13718, Dec. 2003.
- [162] D. Wu and D. A. Kofke. Phase-space overlap measures. I. Fail-safe bias detection in free energies calculated by molecular simulation. *The Journal of Chemical Physics*, 123(5):054103, Aug. 2005.

- [163] xkcd. Jelly beans comic, <https://xkcd.com/882/>.
- [164] Y. Xu, M.-j. Li, H. Greenblatt, W. Chen, A. Paz, O. Dym, Y. Peleg, T. Chen, X. Shen, J. He, H. Jiang, I. Silman, and J. L. Sussman. Flexibility of the flap in the active site of BACE1 as revealed by crystal structures and molecular dynamics simulations. *Acta Crystallographica Section D Biological Crystallography*, 68(1):13–25, Jan. 2012.
- [165] F. M. Ytreberg, R. H. Swendsen, and D. M. Zuckerman. Comparison of free energy methods for molecular systems. *The Journal of Chemical Physics*, 125(18):184114, Nov. 2006.
- [166] I. Zhang, D. A. Rufa, I. Pulido, M. M. Henry, L. E. Rosen, K. Hauser, S. Singh, and J. D. Chodera. Identifying and overcoming the sampling challenges in relative binding free energy calculations of a model protein:protein complex. preprint, Biophysics, Mar. 2023.
- [167] J. Zhao, X. Liu, W. Xia, Y. Zhang, and C. Wang. Targeting Amyloidogenic Processing of APP in Alzheimer’s Disease. *Frontiers in Molecular Neuroscience*, 13:137, Aug. 2020.
- [168] K. Zhu, C. Li, K. Y. Wu, C. Mohr, X. Li, and B. Lanman. Modeling receptor flexibility in the structure-based design of KRASG12C inhibitors. *Journal of Computer-Aided Molecular Design*, 36(8):591–604, Aug. 2022.
- [169] J. Zou, Z. Li, S. Liu, C. Peng, D. Fang, X. Wan, Z. Lin, T.-S. Lee, D. P. Raleigh, M. Yang, and C. Simmerling. Scaffold Hopping Transformations Using Auxiliary Restraints for Calculating Accurate Relative Binding Free Energies. *Journal of Chemical Theory and Computation*, 17(6):3710–3726, June 2021.

Appendix A

Supporting Information: Challenges Encountered Applying Equilibrium and Non-equilibrium Binding Free Energy Calculations

A.1 Hydration free energy calculations

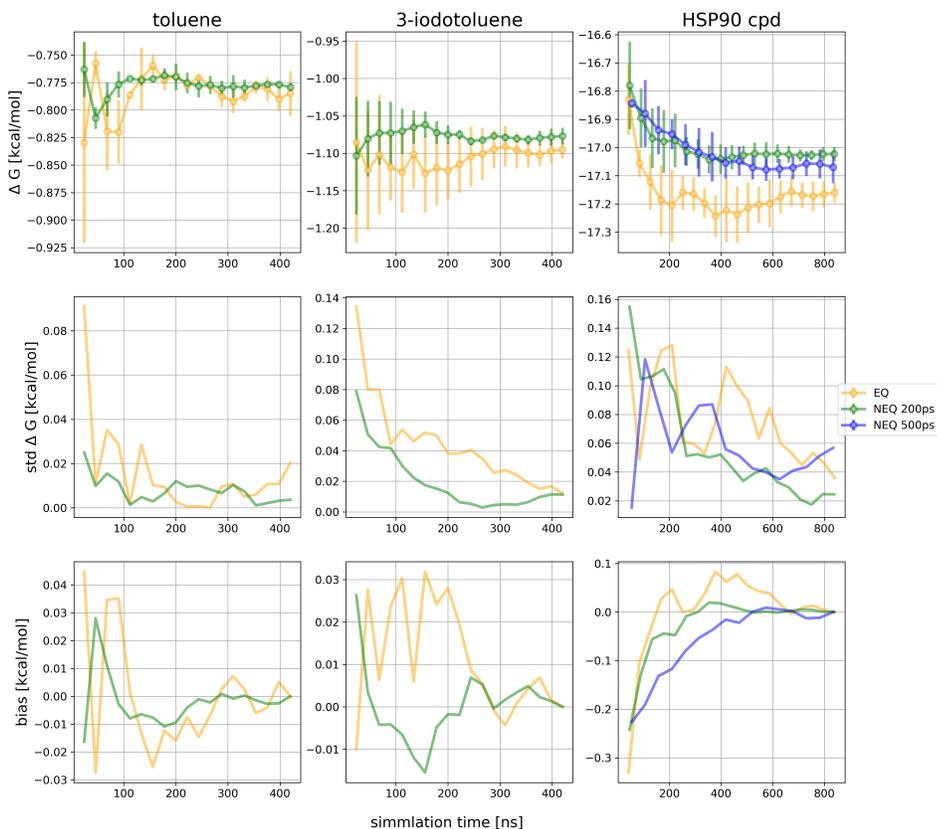


Figure A.1: Hydration free energy, standard deviation (std) and bias as a function of total simulation time for toluene, 3-iodotoluene and the HSP90 compound. For the EQ approach (yellow) we show the mean free energy difference across three replicates in the top row while for the NEQ approach (green and blue) the free energy estimate was obtained by pooling work values from three replicates and estimating one free energy difference using BAR (Sec. 4.5). The uncertainty estimate in both approaches is the standard deviation across three replicates. Second and third row represents the standard deviation and bias. For the bias we assumed that the final value of the mean free energy is the theoretical hydration free energy of the compounds, meaning that it may be underestimated if simulations did not converge. Standard deviation and bias were reduced faster with the NEQ approach, suggesting that it is slightly more efficient in this test set. For the HSP90 compound the two methods did not converge to the same mean free energy within uncertainty. We tested a different NEQ protocol with longer, but fewer transitions (blue, 500 ps/transition). EQ and NEQ approaches converged to statistically the same result using this protocol.

A.2 Restraints in binding free energy calculations

	P1	P2	P3	L1	L2	L3
toluene Pose I	Cys54 CA	Ala98 O	Ala99 O	C3	C5	C4
toluene Pose II	Cys54 CA	Ala98 O	Ala99 O	C3	C5	C4
3-iodotoluene	Cys54 CA	Ala98 O	Ala99 O	C1	C6	C3
HSP90 compound	Tyr201 CD2	Pro164 C	Gly82 O	N2	C1	C3

Table A.1: Protein (P1, P2, P3) and ligand atoms (L1, L2, L3) involved in Boresch-style restraints. One distance (P3-L1), two angles (P2-P3-L1 and P3-L1-L2) and three dihedrals (P1-P2-P3-L1, P2-P3-L1-L2 and P3-L1-L2-L3) were restrained to their value in an equilibrated reference structure.

A.3 Toluene binding to T4 lysozyme L99A

	$\Delta G_{restraints\ on}$ [kcal/mol]	$\Delta G_{decouple\ ligand}$ [kcal/mol]	$\Delta G_{restraints\ off}$ [kcal/mol]	$\Delta G_{solvate\ ligand}$ [kcal/mol]	ΔG^0 [kcal/mol]
Pose I EQ	1.96 ± 0.08	11.68 ± 0.08	-8.615	-0.79 ± 0.02	-4.2 ± 0.1
Pose I NEQ 200 ps		11.6 ± 0.3		-0.779 ± 0.004	-4.2 ± 0.3
Pose I NEQ 500 ps		11.7 ± 0.3			-4.3 ± 0.3
Pose I NEQ 1 ns		11.7 ± 0.2			-4.3 ± 0.2
Pose II EQ	1.3 ± 0.1	9.6 ± 0.1	-6.774	-0.79 ± 0.02	-3.3 ± 0.1
Pose II NEQ 200 ps		9.6 ± 0.1		-0.779 ± 0.004	-3.3 ± 0.1
Pose II NEQ 500 ps		9.51 ± 0.08			-3.3 ± 0.1
Pose II NEQ 1 ns		9.5 ± 0.2			-3.2 ± 0.2

Table A.2: Binding free energy ΔG^0 for the toluene/T4 lysozyme system. We show the free energy difference of each individual leg of the thermodynamic cycle as well as the binding free energy calculated by summation along the cycle. Reported uncertainties for decoupling the ligand in the binding site ($\Delta G_{decouple\ ligand}$) and in the solvent ($\Delta G_{solvate\ ligand}$) are the standard deviation of three independent replicates. $\Delta G_{restraints\ off}$ was calculated analytically, using the formula in Boresch *et al.*[19].

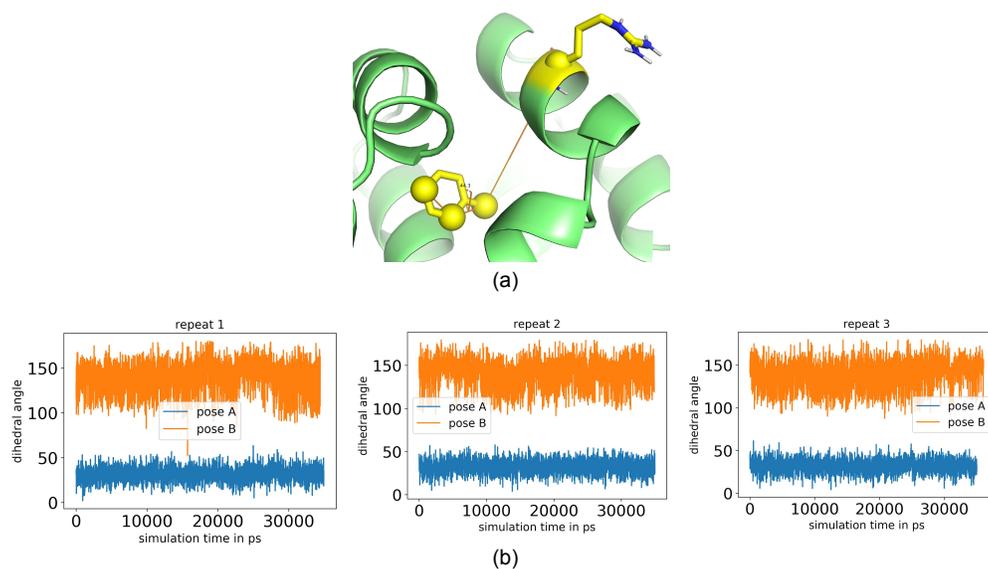


Figure A.2: Two binding modes of toluene in T4 lysozyme L99A. (a) The dihedral angle used to differentiate the two bindings modes is defined by the C1, C3 and C5 atoms of toluene and the alpha carbon of arginine 119, depicted as spheres. (b) The crystallographic binding mode (pose I) is shown in blue and the alternative mode (pose II) in orange. Orientational restraints kept toluene from transitioning between binding modes. The two binding modes were restrained with different force constants. Pose I was restrained with a force constant of $20 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ for the bond and $20 \text{ kcal mol}^{-1} \text{ rad}^{-2}$ for angles and dihedrals, while the alternative binding mode was restrained with $10 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ and $10 \text{ kcal mol}^{-1} \text{ rad}^{-2}$.

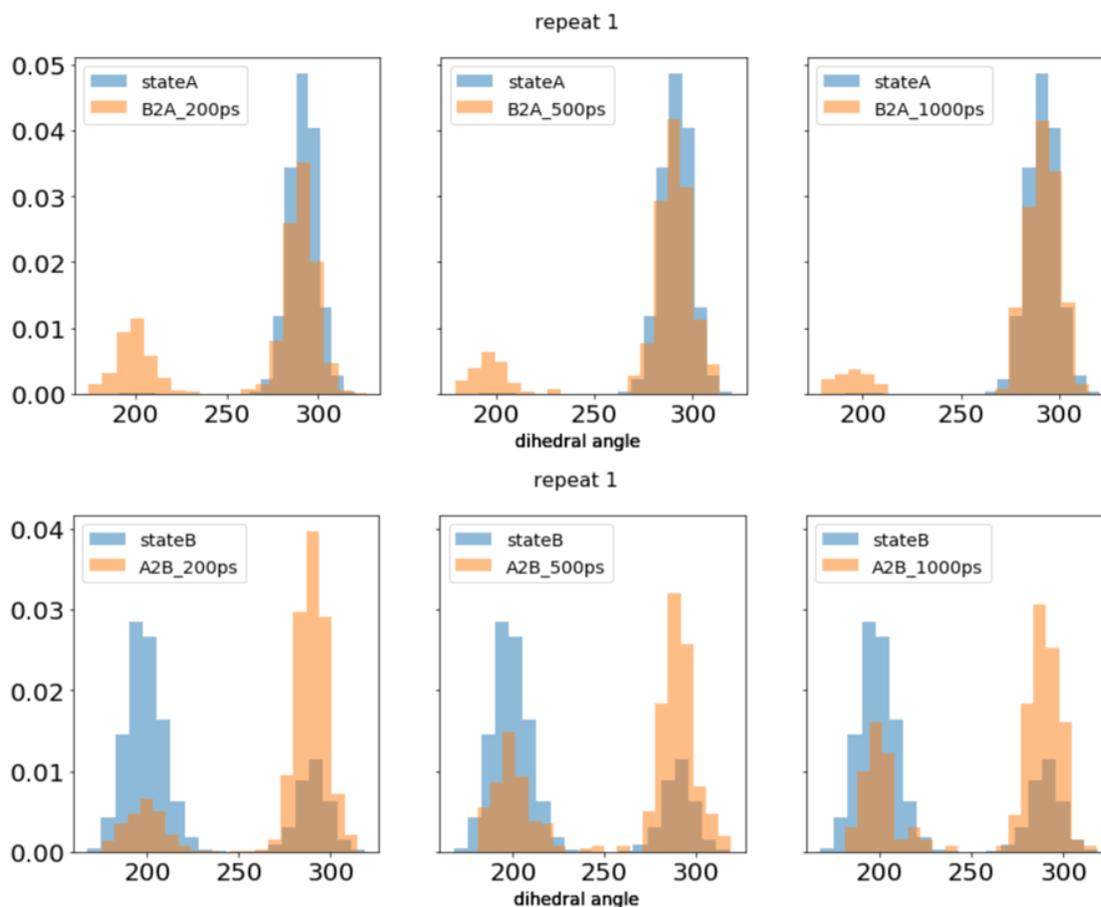


Figure A.3: The distribution of the dihedral angle ξ_1 of Ile78 for toluene binding to T4 lysozyme L99A. Shown are values in the end states (blue) and the last frame of the switching transitions towards the end state (orange; A2B means going from state A to state B) for different transition lengths. Especially for the decoupling transitions (bottom, A2B: interacting to non-interacting), the transition length had an impact on how many transitions reoriented to the other Ile78 rotamer. This figure shows only one repeat in the crystallographic pose, all repeats and the two poses have very similar distributions and can be found in the SI.

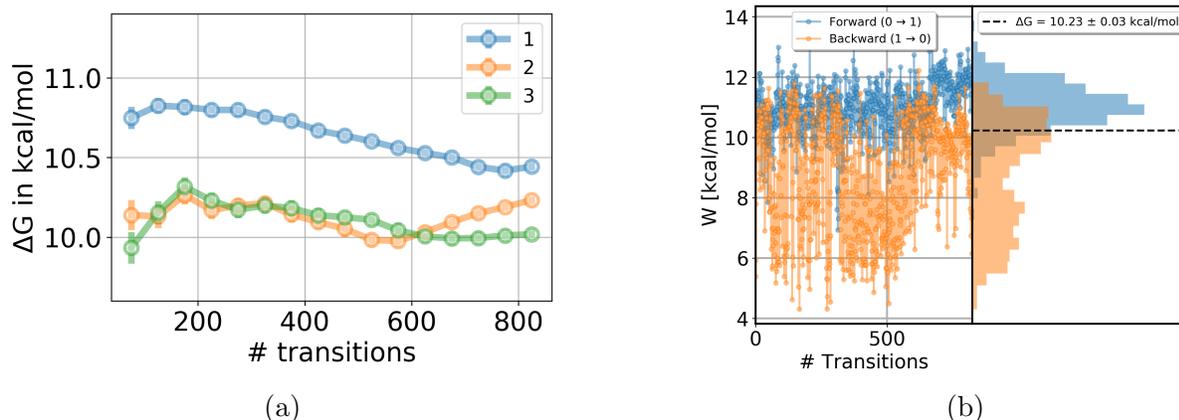


Figure A.4: Protocol using weaker Boresch-style restraints ($5 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ and $5 \text{ kcal mol}^{-1} \text{ rad}^{-2}$). (a) Free energy difference for decoupling toluene in the binding site with the NEQ protocol using weaker Boresch-style restraints. The cumulative free energy difference is plotted against the number of transitions. The standard deviation across three independent replicates is 0.2 kcal/mol which is slightly lower than in the protocol with higher force constant on the restraints (0.3 kcal/mol). (b) Work values for toluene binding to T4 lysozyme L99A. Shown are values measured for each attempted transition, in forward (blue) and reverse (orange) direction, as well as the distribution of the work values. The reverse work distribution is similar to the one in the protocol using stronger restraints and the mean dissipation was independent of the restraint strength and in both protocols 1.7 kcal/mol in the reverse direction and 0.9 kcal/mol in the forward direction.

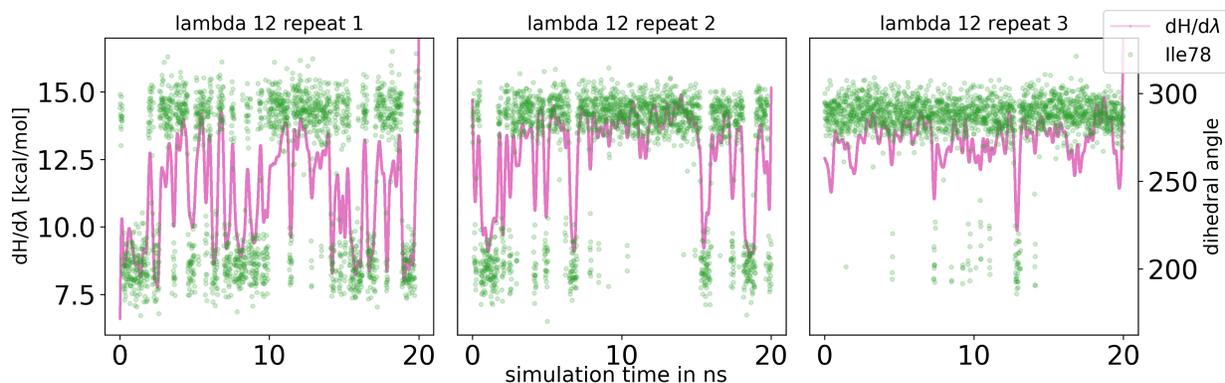


Figure A.5: Correlation between $dH/d\lambda$ values and the dihedral angle of Ile78 for toluene binding to T4 lysozyme L99A. A running average of the $dH/d\lambda$ values (pink, averaged across 2000 data points) and the dihedral angle of Ile78 (green) is plotted as a function of simulation time. Here we only show λ 12 for the three independent replicates, additional λ windows can be found in the SI. Sudden changes in $dH/d\lambda$ correlated with rotation of the Ile78 side chain.

		$\Delta G_{COM \text{ on, Boresch off}}$ [kcal/mol]	$\Delta G_{decouple \text{ ligand}}$ [kcal/mol]	$\Delta G_{COM \text{ off, Boresch on}}$ [kcal/mol]	ΔG_{total} [kcal/mol]
Only Boresch restraints	NEQ		11.6 ± 0.3		11.6 ± 0.3
COM-COM distance restraints	EQ	-1.87 ± 0.03	5.9 ± 0.1	8.38 ± 0.04	12.4 ± 0.1
	NEQ		5.9 ± 0.3		12.4 ± 0.3

Table A.3: Comparison of the free energy difference for decoupling toluene in T4 lysozyme L99A using Boresch-style restraints and center-of-mass (COM) - COM flat-bottom harmonic distance restraints. For the COM-COM restraint protocol we first turned the Boresch-style restraints off while turning on a distance restraint between the COM of the ligand and the COM of the side chain alanine 99. Then we decoupled the ligand and afterwards switched the COM-COM restraint back to the Boresch-style restraints in the noninteracting state. Reported uncertainties are the standard deviation of three independent replicates. In the NEQ approach, the dissipated work was higher in the COM-COM restraint protocol (3.8 kcal/mol in the reverse direction and 1.4 in the forward direction) than in the protocol using Boresch-style restraints. With a single COM-COM distance restraint, the ligand can sample the sphere around the radius of that distance restraint which presumably caused the larger dissipation.

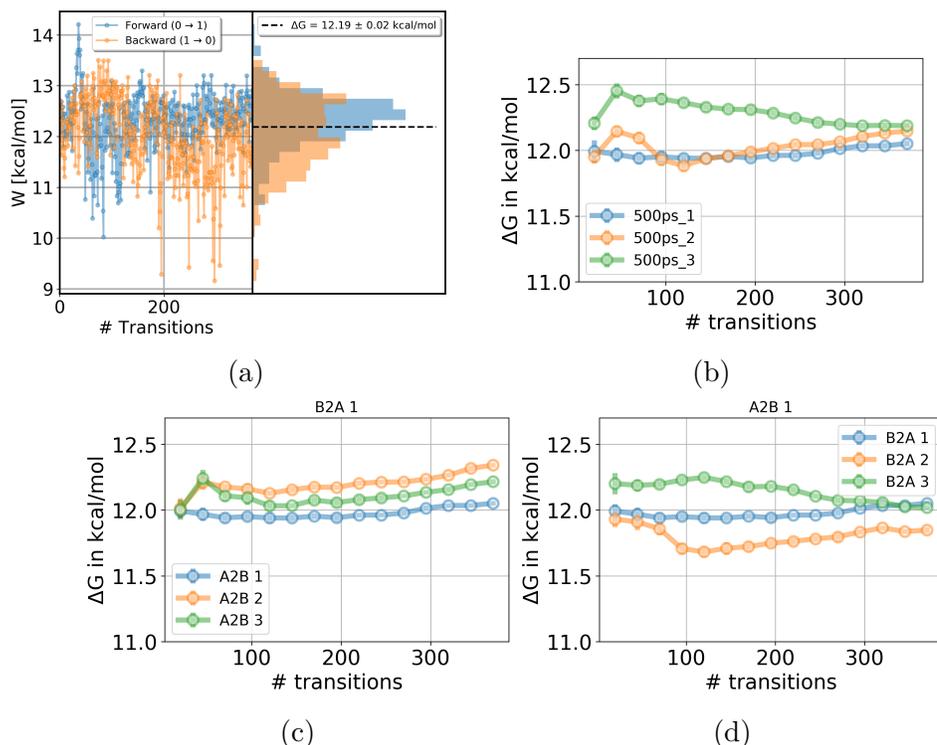


Figure A.6: Restraining the Ile78 side chain in the NEQ protocol to further investigate its impact. (a) Work values for toluene binding to T4 lysozyme L99A while restraining the Ile78 side chain. Shown are values measured for each attempted transition, in forward (blue) and reverse (orange) direction, as well as the distribution of the work values. In contrast to the unrestrained protocol, the reverse work distribution (orange) is unimodal, which suggests that sampling of two Ile78 rotamers caused the bimodal work distributions in the unrestrained simulations. (b) Free energy difference for decoupling toluene in the binding site with the NEQ protocol using dihedral restraints. The cumulative free energy difference is plotted against the number of transitions. Three independent replicates seem to converge to the same free energy difference and the standard deviation was lower in this protocol compared to the unrestrained calculations (0.06 vs. 0.3 kcal/mol). (c) and (d) Analyzing the set of forward and reverse data of replicates separately. Three replicates give three sets of transitions in the forward direction and reverse direction which are independent from one another. We took a fixed set of transitions from one replicate for one direction (in (c) the set of transitions from the first replicate was used for the reverse direction (B2A 1)) and calculated the free energy difference using data from the other direction of all three replicates (in (c) A2B 1 in blue uses transitions in the forward direction from replicate one). The same was performed for the other direction in (d). The free energy difference of the second replicate (orange) is higher in the forward direction (c) and lower in the reverse direction (d). This shows that the apparent convergence (and low standard deviation) seen in (b) is likely a coincidence.

A.4 3-Iodotoluene binding to T4 lysozyme L99A

	$\Delta G_{\text{restraints on}}$ [kcal/mol]	$\Delta G_{\text{decouple ligand}}$ [kcal/mol]	$\Delta G_{\text{restraints off}}$ [kcal/mol]	$\Delta G_{\text{solvate ligand}}$ [kcal/mol]	ΔG^0 [kcal/mol]
EQ	1.7 ± 0.1	12.8 ± 0.2	-8.638	-1.10 ± 0.01	-4.8 ± 0.2
NEQ 500 ps		13.6 ± 0.2		-5.6 ± 0.2	
NEQ 1 ns		13.7 ± 0.1		-5.7 ± 0.1	
NEQ 2 ns		13.5 ± 0.3		-5.5 ± 0.3	
NEQ 4 ns		13.1 ± 0.1		-5.1 ± 0.1	
restrain Val111 EQ		13.0 ± 0.4		-1.10 ± 0.01	-5.0 ± 0.4
restrain Val111 NEQ		12.77 ± 0.07		-1.08 ± 0.01	-4.8 ± 0.1

Table A.4: Binding free energy ΔG^0 for the 3-iodotoluene/T4 lysozyme system. We show the free energy difference of each individual leg of the thermodynamic cycle as well as the binding free energy calculated by summation along the cycle. Reported uncertainties are the standard deviation of three independent replicates. $\Delta G_{\text{restraints off}}$ was calculated analytically, using the formula in Boresch *et al.*[19]. Costs for restraining and releasing the restraints on Val111 were 0.1 and -0.1 kcal/mol.

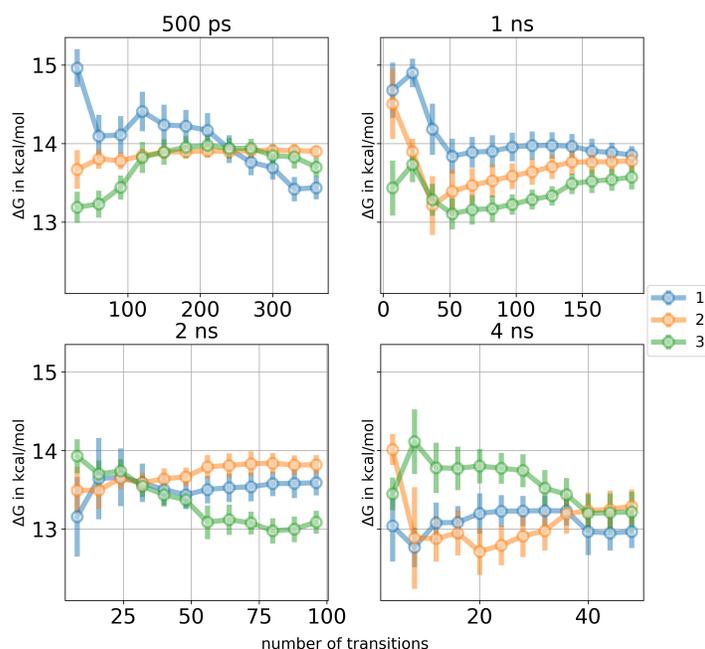


Figure A.7: Free energy difference for decoupling 3-iodotoluene in the binding site using non-equilibrium FEC. The cumulative free energy difference is plotted as a function of the number of transitions. We tested four different protocols that differ in the length of the non-equilibrium transition and the number of transitions. Three independent replicates (blue, orange, green) were run for each protocol. Uncertainties were estimated via bootstrapping. The different protocols did not converge to the same free energy difference in the simulation time.

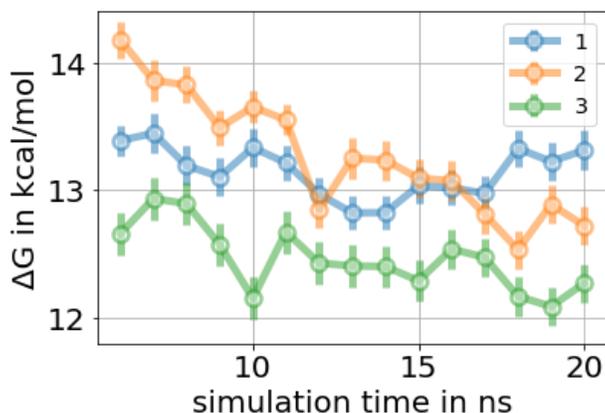


Figure A.8: Free energy difference for decoupling 3-iodotoluene in the binding site using equilibrium FEC. The cumulative free energy difference of three independent replicates is plotted as a function of the simulation time per λ window. An analytical uncertainty estimate was used. Independent replicates did not converge to the same free energy difference within uncertainty in 20 ns per λ window simulation time.

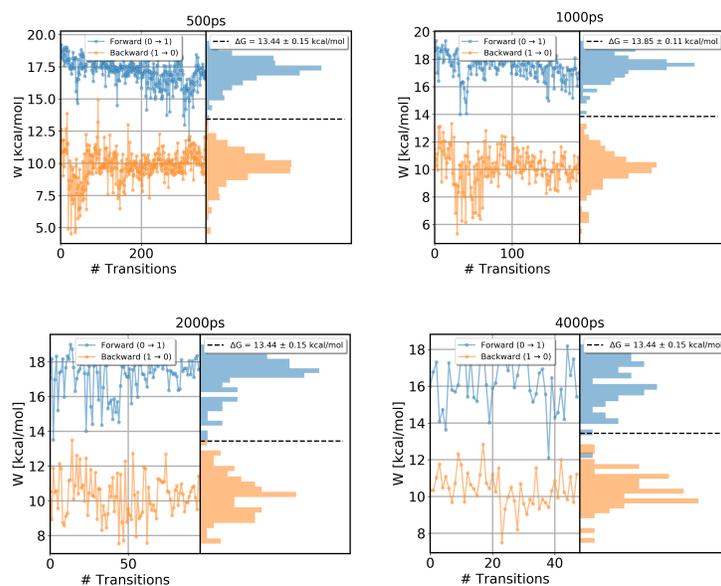


Figure A.9: Work values for 3-iodotoluene binding to T4 lysozyme L99A, in forward (blue) and reverse (orange) direction. Shown are values measured for each attempted transition, as well as the distribution of the work values. The four plots show protocols that differ in the length of each NEQ switching transitions. The overlap of the work distributions in all protocols was poor, however improved slightly with transition length.

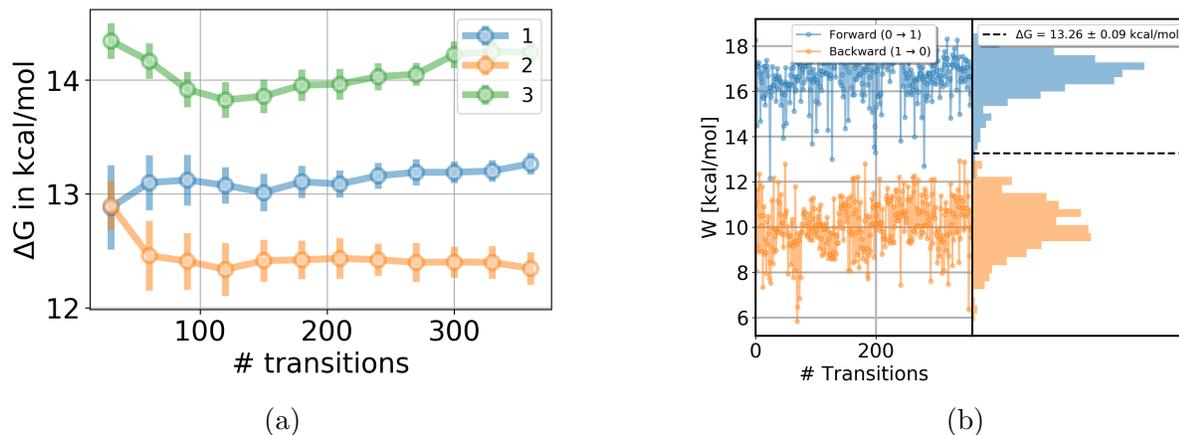


Figure A.10: Weaker Boresch-style restraints ($5 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ and $5 \text{ kcal mol}^{-1} \text{ rad}^{-2}$). (a) Free energy difference for decoupling 3-iodotoluene in the binding site with the NEQ protocol using weaker Boresch-style restraints. The cumulative free energy difference is plotted against the number of transitions. The standard deviation across three independent replicates was 0.8 kcal/mol which is higher than in the protocol with stronger force constant on the restraints (0.2 kcal/mol). (b) Work values for 3-iodotoluene binding to T4 lysozyme L99A. Shown are values for replicate 1 measured for each attempted transition, in forward (blue) and reverse (orange) direction, as well as the distribution of the work values. The reverse work distribution is similar to the one in the protocol using stronger restraints and the mean dissipation was similar in both protocols, in the reverse direction 3.7 kcal/mol for the protocol with a weaker force constant and 3.9 kcal/mol for the other protocol and in the forward direction, 3.4 vs. 4.1 kcal/mol , respectively.

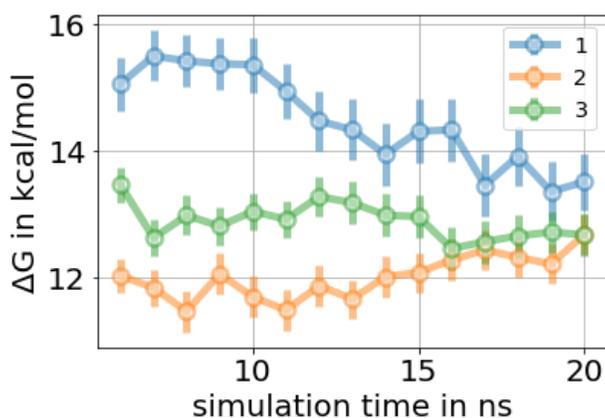


Figure A.11: Decoupling 3-iodotoluene in the binding site while switching the Val111 side chain using the EQ approach. We show the free energy difference of three replicates as a function of simulation time.

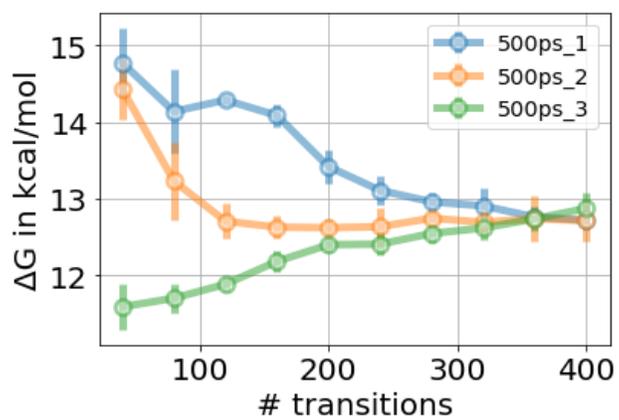


Figure A.12: Decoupling 3-iodotoluene in the binding site while switching the Val111 side chain using the NEQ approach. We show the free energy difference of three replicates as a function of number of transitions. The ΔG estimate converged approximately after 320 transitions.

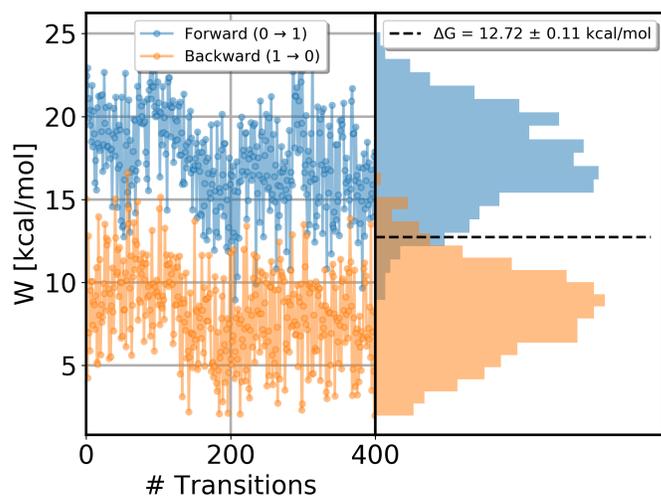


Figure A.13: Work values and work distributions for decoupling 3-iodotoluene in the binding site of T4 lysozyme L99A while switching the Val111 side chain. Work values of the forward direction (decoupling, blue) and reverse direction (coupling, orange) are plotted as a function of transition number and the distribution of work values in the vertical plot. Transitions were run for 500 ps each. Forward and reverse work distributions overlapped well when the side chain was switched during the alchemical path, compared to the unrestrained protocol (Figure A.9, 500 ps).

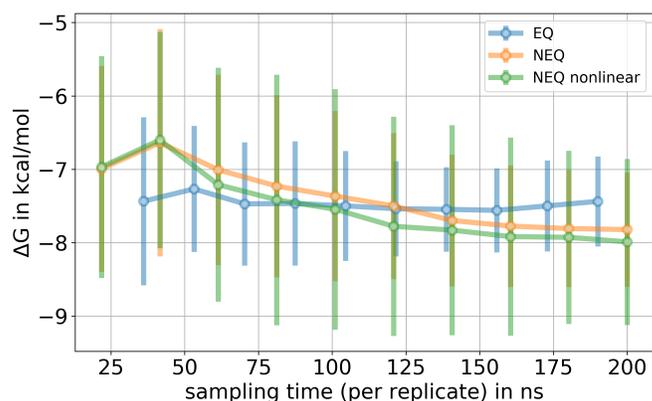


Figure A.14: Free energy difference for releasing the dihedral restraints in the decoupled state of the 3-iodotoluene/ T4 lysozyme system in the protocol where multiple dihedrals were restrained. The alchemical path was nonlinear (blue and green curve) or linear (orange curve). For the EQ approach we show the mean free energy difference across six replicates while for the NEQ approach the free energy estimate was obtained by pooling work values from six replicates and estimating one free energy difference using BAR (Sec. 4.5). The uncertainty estimate in both approaches is the standard deviation across six replicates. All protocols converged to the same free energy difference within uncertainty. In the NEQ approach, a nonlinear alchemical path, where more sampling time can be spend in the more challenging part of the transformation (lower force constants) did not perform better than the protocol using a linear alchemical path. The standard deviation was high (0.6-1.1 kcal/mol) indicating sampling problems in this protocol where multiple additional side chains were restrained. Input topology files for the nonlinear NEQ protocol are provided in the SI files.

A.5 The HSP90 system

	$\Delta G_{restraints\ on}$ [kcal/mol]	$\Delta G_{decouple\ ligand}$ [kcal/mol]	$\Delta G_{restraints\ off}$ [kcal/mol]	$\Delta G_{solvate\ ligand}$ [kcal/mol]	ΔG^0 [kcal/mol]
EQ	1.26 ± 0.08	36.7 ± 0.3	-9.003	-17.16 ± 0.04	-11.8 ± 0.3
HREX		35.9 ± 0.5			-11.0 ± 0.5
NEQ 1 ns		36.3 ± 0.4		-17.02 ± 0.02	-11.5 ± 0.4
NEQ 5 ns		36.2 ± 0.3			-11.4 ± 0.3
NEQ 10 ns		36.2 ± 0.5			-11.4 ± 0.5
EQ no water	1.4 ± 0.04	33.3 ± 1.3	-8.924	-17.16 ± 0.04	-8.6 ± 1.3
HREX no water		34.6 ± 0.8			-9.9 ± 0.8

Table A.5: Binding free energy ΔG° for the HSP90 system. We show the free energy difference of each individual leg of the thermodynamic cycle as well as the binding free energy calculated by summation along the cycle. Reported uncertainties are the standard deviation of three independent replicates. $\Delta G_{restraints\ off}$ was calculated analytically[19]. The last two rows represent protocols, where all crystallographic water molecules were removed from the input structure, including three buried binding site waters.

	EQ	NEQ
$\Delta G_{restraints\ on}$ [kcal/mol]	1.26 \pm 0.04	
$\Delta G_{restrain\ loop}$ [kcal/mol]	1.8 \pm 0.2	
$\Delta G_{bubble\ on}$ [kcal/mol]	-6.979 \pm 0.006	
$\Delta G_{decouple\ ligand}$ [kcal/mol]	56.4 \pm 0.5	56.2 \pm 0.3
$\Delta G_{bubble\ off}$ [kcal/mol]	-10.6 \pm 0.2	
$\Delta G_{restraints\ loop\ off}$ [kcal/mol]	-3.1 \pm 0.5	
$\Delta G_{restraints\ off}$ [kcal/mol]	-9.003	
$\Delta G_{solvate\ ligand}$ [kcal/mol]	-17.16 \pm 0.04	-17.02 \pm 0.02
ΔG^0 [kcal/mol]	-12.7 \pm 0.8	-12.7 \pm 0.6

Table A.6: Binding free energy ΔG° for the HSP90 system using the bubble approach. We show the free energy difference of each individual leg of the thermodynamic cycle as well as the binding free energy calculated by summation along the cycle. Reported uncertainties are the standard deviation of three independent replicates. $\Delta G_{restraints\ off}$ was calculated analytically[19].

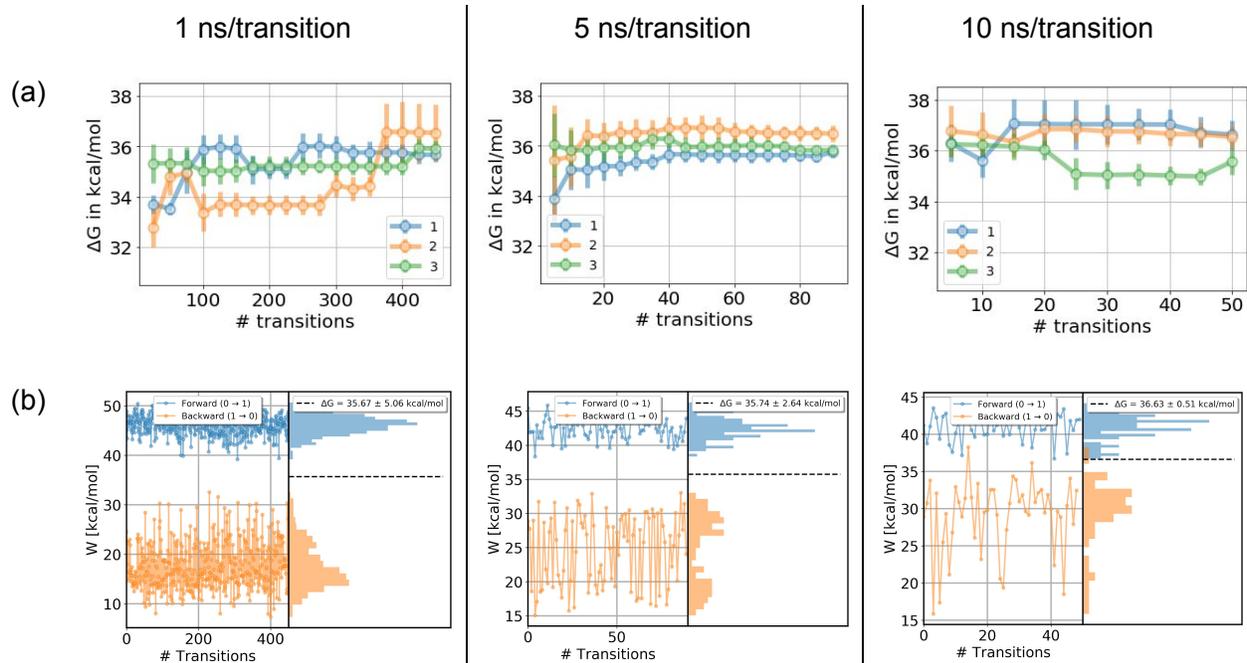


Figure A.15: (a) Free energy difference of three independent replicates as a function of number of NEQ transitions in the HSP90 system. Three protocols using 1 ns, 5 ns or 10 ns per switching transition were used in the NEQ approach and uncertainty estimates obtained via bootstrapping. (b) Work values of forward (blue) and reverse (orange) transitions as a function of transition number and the distribution of work values. Shown is one example (replicate 1) per protocol, additional plots are provided in the SI. The uncertainty is the analytical error (see section 4.3). At poor overlap the uncertainty is high (5 kcal/mol for 1 ns/transition replicate 1), indicating that the estimate is likely incorrect.

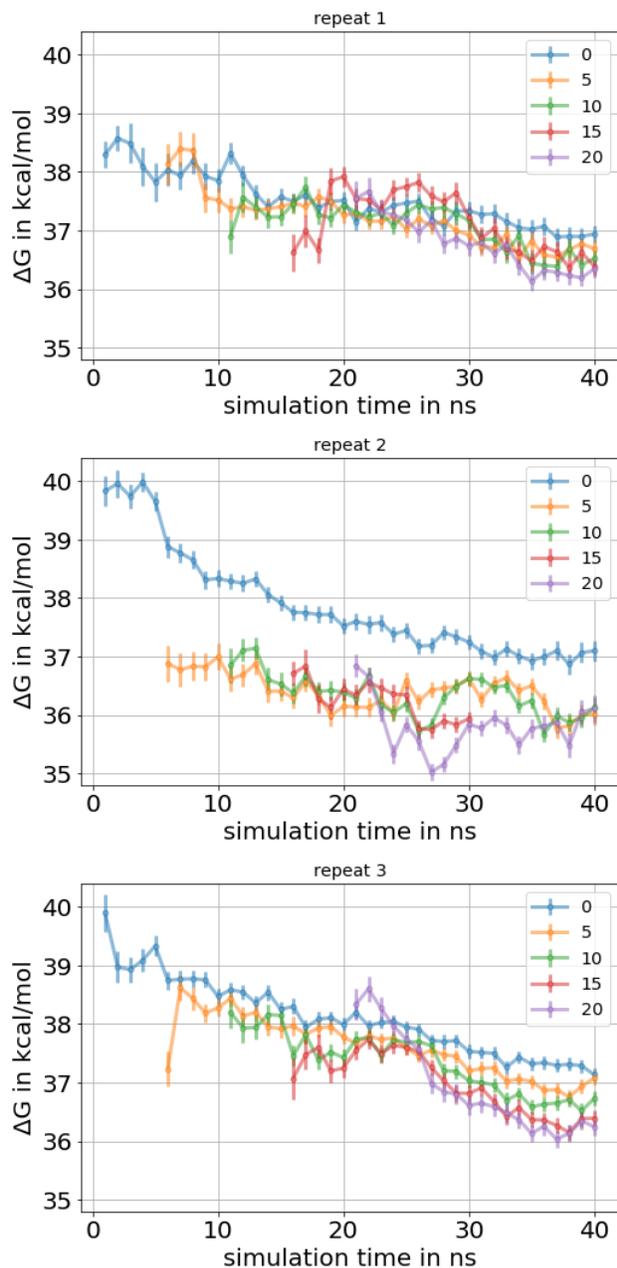


Figure A.16: Impact of the equilibration time on the free energy difference in the HSP90 system. In a separate plot for every replicate, we show the cumulative free energy difference as a function of simulation time/ λ window. Different amount of data were discarded for equilibration, e.g. 0 means all data from the production run were used to estimate the free energy difference and 20 means 20 ns/ λ window were discarded for equilibration. In all three repeats the final free energy difference depends on the amount of data discarded for equilibration indicating a slow DOF in the system.

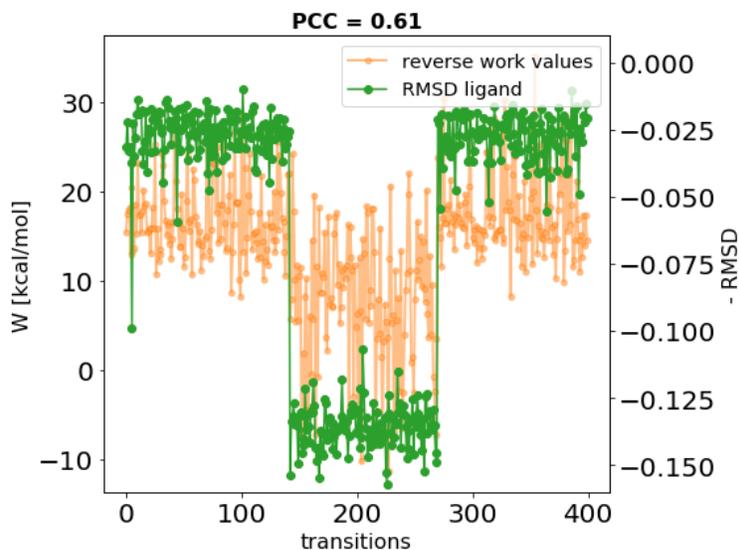


Figure A.17: Correlation between reverse work values (W_r) and the ligand RMSD in the HSP90 system. The work values (orange) and the RMSD (green) are plotted as a function of the transition number. The RMSD was calculated from the set of first frames of the NEQ transitions. Work values correlated with the ligand RMSD, which is supported by a high Pearson correlation coefficient ($PCC=0.61$). The binding mode flip had an impact on the work values.

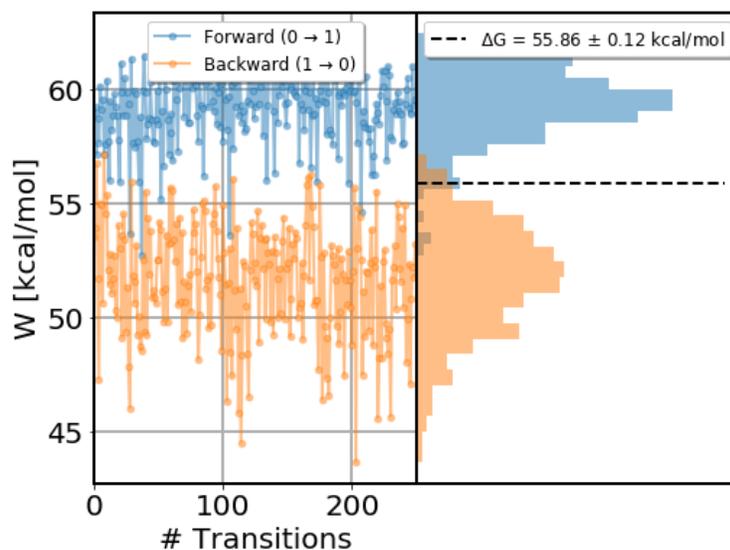


Figure A.18: Work values for the ligand binding to HSP90, in forward (blue) and reverse (orange) direction in the presence of the bubble-ligand. Shown are values measured for each attempted transition, as well as the distribution of the work values. Transitions were run for 1 ns each. Forward and reverse work distributions overlapped well in the presence of the bubble-ligand, supporting that the poor overlap was caused by inadequate water sampling.

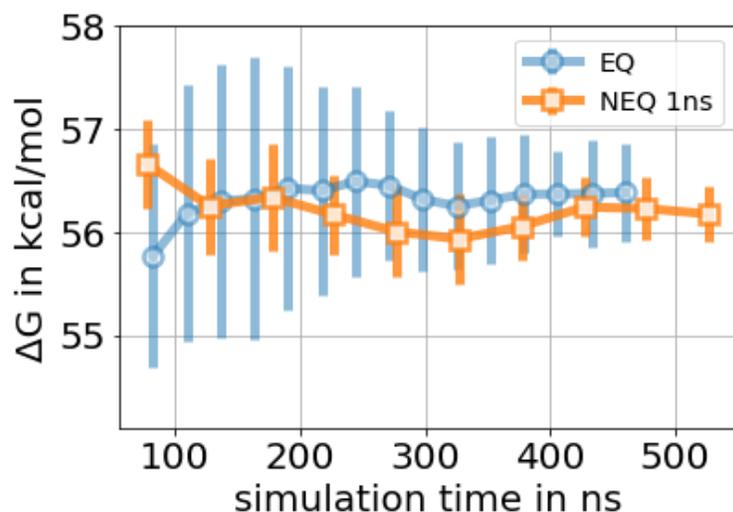


Figure A.19: Free energy difference for decoupling the HSP90 ligand in the binding site in the presence of the bubble-ligand. For the EQ approach we show the mean free energy difference across three replicates while for the NEQ approach the free energy estimate was obtained by pooling work values from three replicates and estimating one free energy difference using BAR (Sec. 4.5). The uncertainty estimate in both approaches is the standard deviation across three replicates.

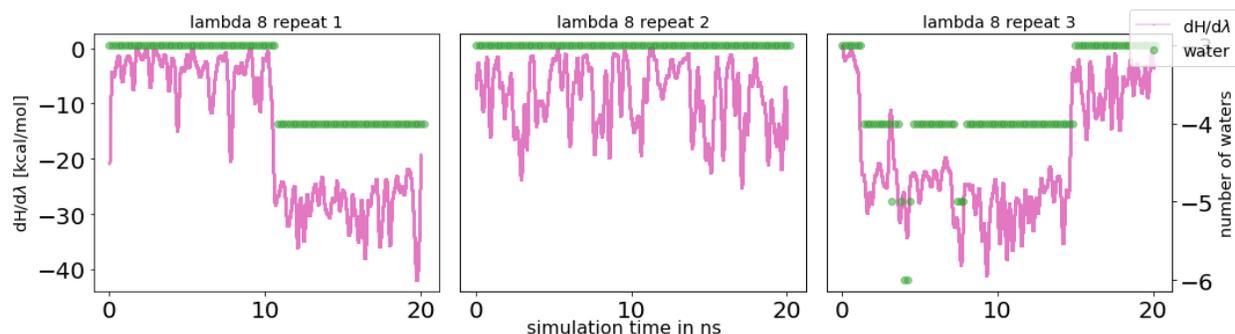


Figure A.20: Water sampling problems when turning the interactions of the bubble-ligand off. Correlation between $dH/d\lambda$ values and number of water molecules in the binding site for the ligand binding to HSP90. A running average of the $dH/d\lambda$ values (pink, averaged across 2000 data points) and the number of water molecules (green) is plotted as a function of simulation time. Here we only show λ 8 for the three independent replicates, additional λ windows can be found in the SI. Sudden changes in $dH/d\lambda$ correlated with the entry/escape of water molecules into the space between protein and ligand. Water sampling was inadequate which shows that introducing the bubble-ligand only separates the water sampling from the decoupling of the ligand, but does not solve the problem.

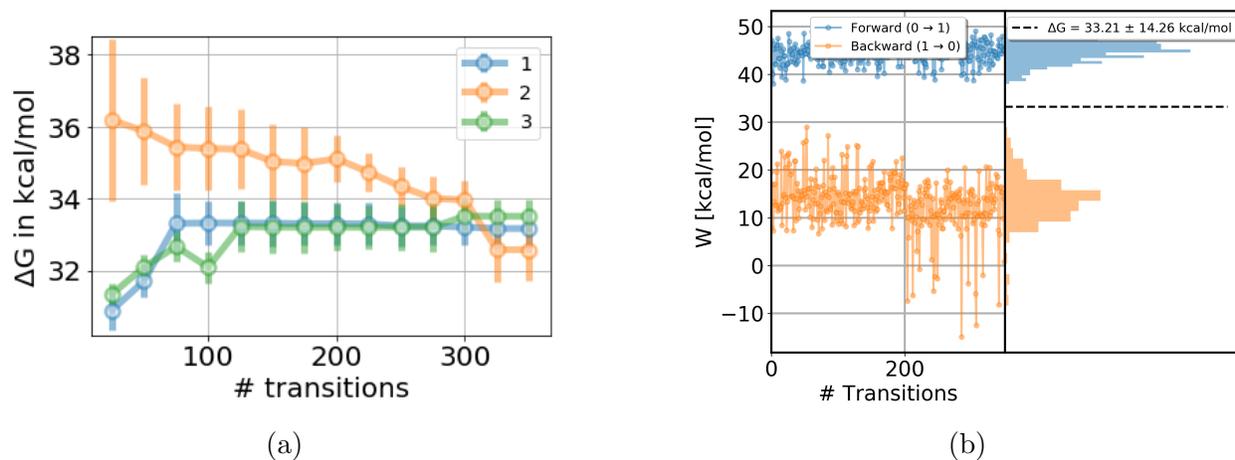


Figure A.21: Weaker Boresch-style restraints ($5 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ and $5 \text{ kcal mol}^{-1} \text{ rad}^{-2}$). (a) Free energy difference for decoupling the HSP90 ligand in the binding site with the NEQ protocol using weaker Boresch-style restraints. The cumulative free energy difference is plotted against the number of transitions. Transitions were run for 1 ns. The standard deviation across three independent replicates was 0.4 kcal/mol which is higher than in the protocol with stronger force constant on the restraints (0.2 kcal/mol). (b) Work values for the ligand binding to HSP90. Shown are values for replicate 1 measured for each attempted transition, in forward (blue) and reverse (orange) direction, as well as the distribution of the work values. The reverse work distribution is similar to the one in the protocol using stronger restraints and the mean dissipation was similar in both protocols, in the reverse direction 18.7 kcal/mol for the protocol with a weaker force constant and 15.2 kcal/mol for the original protocol, and in the forward direction 11.5 vs. 13.9 kcal/mol .

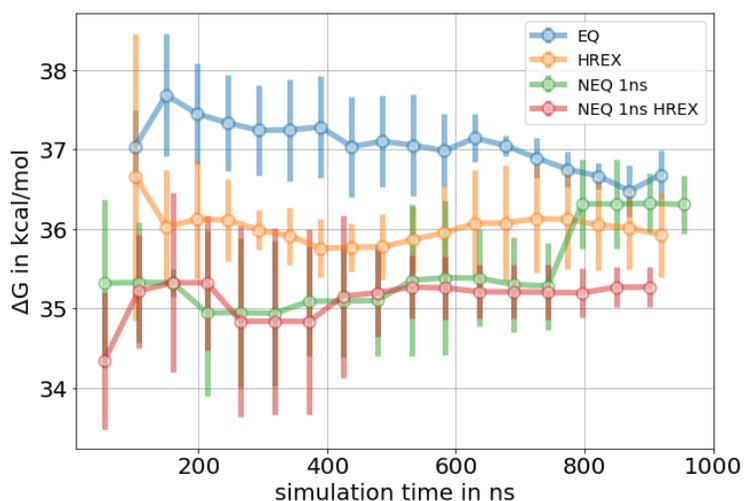


Figure A.22: Free energy difference for decoupling the HSP90 ligand in the binding site. We show EQ and NEQ protocols with and without the use of HREX as an enhanced sampling technique. In the NEQ HREX approach, end state structures were used from the EQ HREX simulation, and no enhanced sampling was performed in the NEQ switches. For the EQ approach we show the mean free energy difference across three replicates while for the NEQ approach the free energy estimate was obtained by pooling work values from three replicates and estimating one free energy difference using BAR (Sec. 4.5). The uncertainty estimate in both approaches is the standard deviation across three replicates. Using structures from HREX simulations as input for the NEQ switches did not improve results in this case. On the contrary, additional sampling problems were introduced (see Figure A.23).

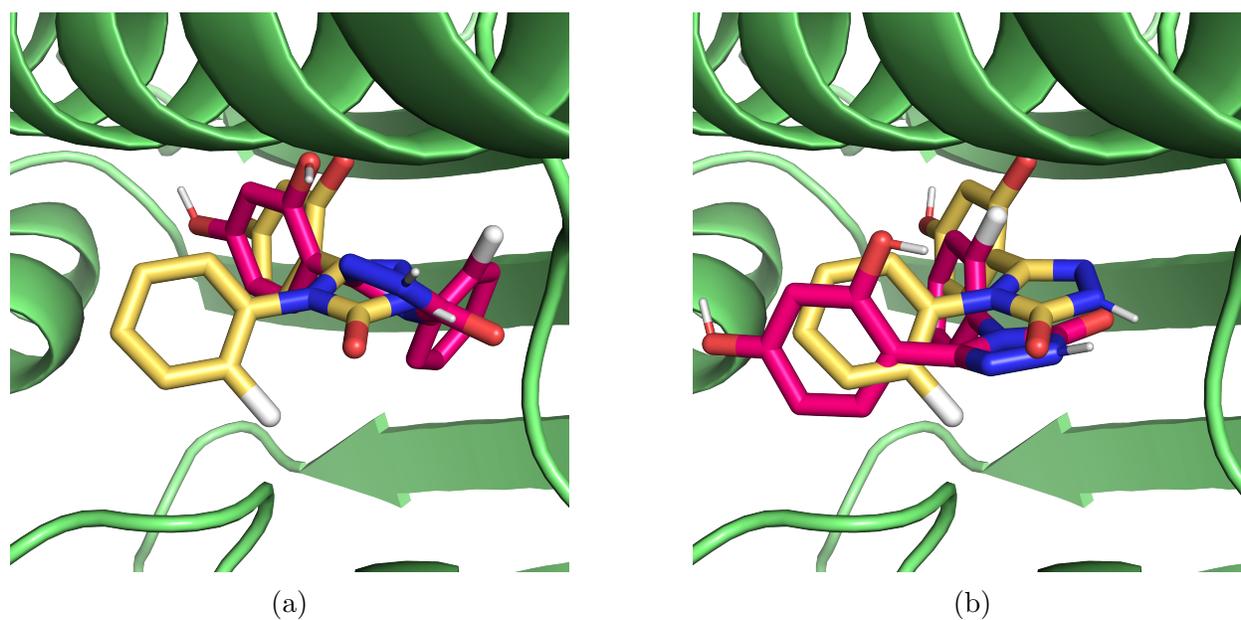


Figure A.23: Flipped binding mode of the ligand binding to HSP90. The crystallographic binding pose is shown in yellow sticks, the flipped binding mode in magenta sticks. (a) In the non-interacting state the fluorophenyl ring was flipped in some frames. (b) Some NEQ switching transitions starting from the flipped binding mode in (a) crashed when the two phenyl rings flipped their position in the binding site.

Appendix B

Supporting Information: Broadening the scope of binding free energy calculations using a Separated Topologies approach

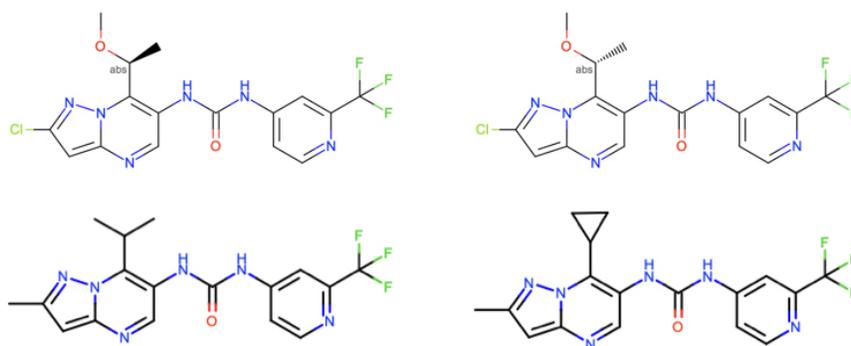


Figure B.1: 2D structures of four MALT1 ligands. Transformations here involve a chiral inversion (top) as well as the closing of a ring going from isopropyl to cyclopropyl (bottom).

	Edge 2d - 2e	Edge 2d - 3b	Edge 2e - 3b
SepTop (input Spruce)	0.0 ± 0.3	0.4 ± 0.3	0.3 ± 0.3
SepTop (input Aux)	-1.8 ± 0.2	0.9 ± 0.5	3.0 ± 0.4
SepTop (input ATM)	-3.7 ± 0.3	-3.5 ± 0.2	0.3 ± 0.2
ATM[10]	-2.3 ± 0.4	-0.4 ± 0.5	2.1 ± 0.4
Aux[169]		1.3 ± 0.4	2.9 ± 0.5

Table B.1: Relative binding free energies in the ER α system using different input structures. All values are given in kcal/mol. Predictions differed by over 3 kcal/mol depending on the input structure used in the calculations with the most significant difference being in Edge 2d-2e where different input structures led to a difference of 3.7 kcal/mol.

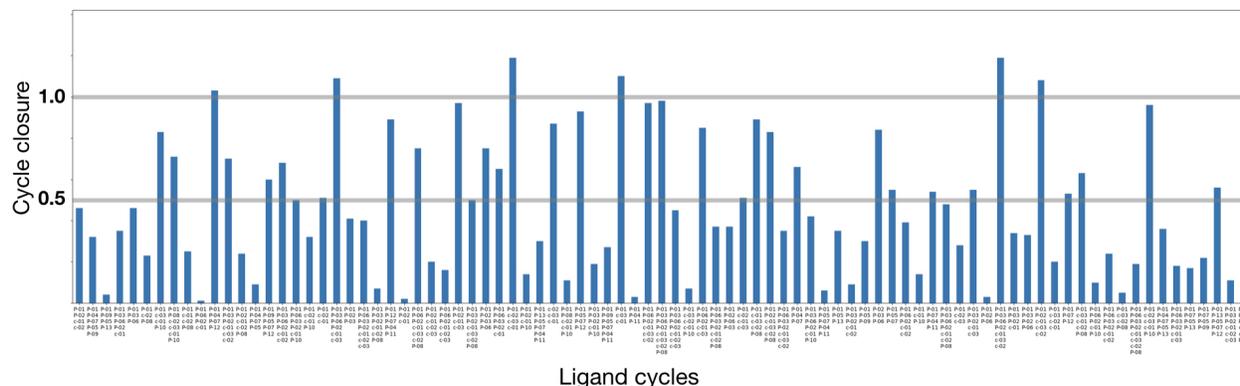


Figure B.2: Cycle closure for all ligand cycles in the MALT1 system. The majority of ligand cycles have a cycle closure below 0.5 kcal/mol. Six ligand cycles have a cycle closure greater than 1 kcal/mol, indicating sampling problems.

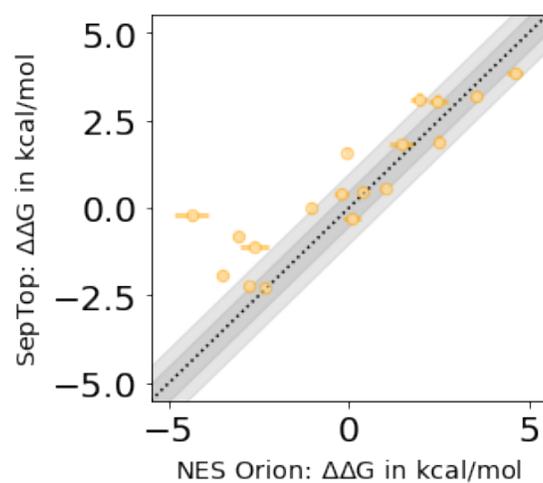


Figure B.3: Correlation between SepTop and NES Orion results for the MALT1 system. Shown are $\Delta\Delta G$ values for all edges. For most edges, both methods gave similar results, however, there are some outliers where the two methods do not agree with one another.

Edge	$\Delta\Delta G$ Exp	$\Delta\Delta G$ SepTop	$\Delta\Delta G$ NES
Pfizer-01-01 → Pfizer-01-02	0.26	0.00 ± 0.16	-1.03 ± 0.10
Pfizer-01-01 → Pfizer-01-03	0.52	0.41 ± 0.15	-0.21 ± 0.19
Pfizer-01-02 → Pfizer-01-03	0.26	-0.30 ± 0.16	0.09 ± 0.24
Pfizer-01-01 → Pfizer-01-06	2.19	3.10 ± 0.17	1.94 ± 0.21
Pfizer-01-02 → Pfizer-01-06	1.93	3.05 ± 0.17	2.46 ± 0.27
Pfizer-01-03 → Pfizer-01-06	1.68	1.90 ± 0.16	2.49 ± 0.15
Pfizer-01-01 → compound-01	0.07	-2.24 ± 0.14	-2.78 ± 0.12
Pfizer-01-01 → compound-03	0.46	-0.81 ± 0.14	-3.08 ± 0.16
Pfizer-01-01 → compound-02	-1.18	-1.92 ± 0.14	-3.53 ± 0.14
Pfizer-01-01 → Pfizer-01-04	1.65	3.18 ± 0.14	3.53 ± 0.16
Pfizer-01-01 → Pfizer-01-05	2.16	3.86 ± 0.14	4.61 ± 0.22
Pfizer-01-01 → Pfizer-01-07	2.37	1.81 ± 0.14	1.47 ± 0.36
compound-02 → compound-03	1.64	1.59 ± 0.12	-0.04 ± 0.07
Pfizer-01-05 → Pfizer-01-07	0.21	-1.09 ± 0.13	-2.63 ± 0.40
Pfizer-01-04 → Pfizer-01-07	0.72	-0.22 ± 0.12	-4.37 ± 0.45
compound-02 → compound-01	1.25	0.56 ± 0.12	1.01 ± 0.13
compound-03 → compound-01	-0.38	0.47 ± 0.12	0.38 ± 0.16
Pfizer-01-02 → compound-01	-0.19	-2.28 ± 0.14	-2.35 ± 0.08

Figure B.4: Comparing SepTop and NES Orion results in the MALT1 system.



Figure B.5: Nonequilibrium work values for the transformation between Pfizer-01-05 and Pfizer-01-07 in the MALT1 system. Forward work values are shown in red and work values from the reverse direction in blue. The distributions of forward and reverse work values do not overlap well for this transformation.

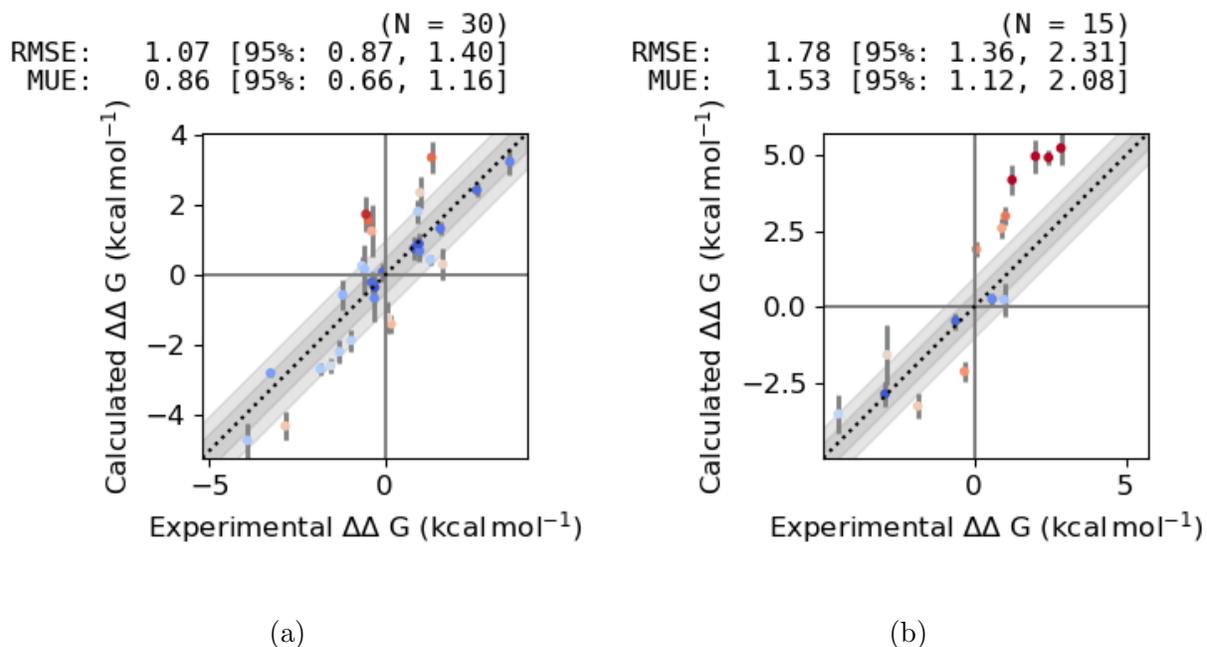
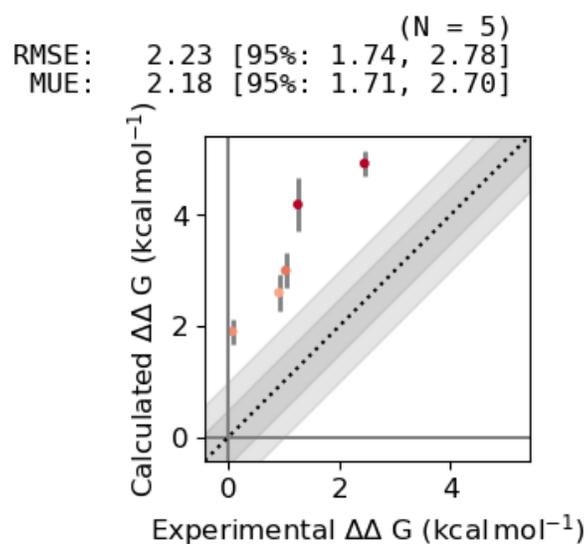


Figure B.6: Correlation between calculated and experimental relative binding free energies for transformations in the BACE1 system. Shown are results from transformations run within the three ligand scaffold series (a) and across different scaffolds (b). $\Delta\Delta G$ values from transformations between ligands within the same scaffold (RMSE=1.07) correlate better with experiment than transformations between ligands of different scaffolds (RMSE=1.78).



(a) BACE amide biaryl

Figure B.7: Correlation between calculated and experimental relative binding free energies for transformations between the amide series and the biaryl series in the BACE1 system. For all five transformations the free energy change was calculated to be more unfavorable as to compared to experiment.

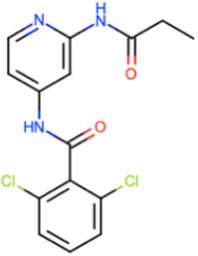
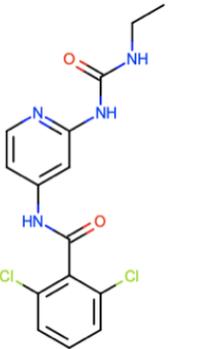
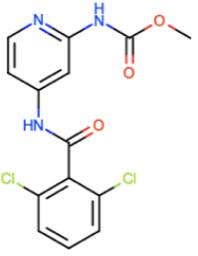
name	2D structure	Exp. IC50
lig_ejm_42		64 nM ³
lig_ejm_54		18 nM ³
lig_ejm_55		170 nM ³

Table B.2: Ligand structures and experimental binding affinities[86] for the TYK2 system.

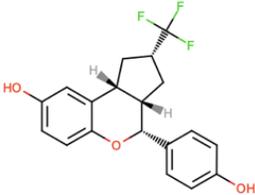
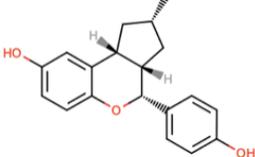
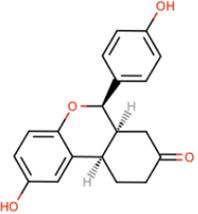
name	2D structure	Exp. IC ₅₀
2d		12.4 +/- 4.9 nM ⁴
2e		3.3 +/- 2.4 nM ⁴
3b		410 nM ⁴

Table B.3: Ligand structures and experimental binding affinities[125] for the ER α system.

name	2D structure	Exp. IC50
2714550-01-1 (compound 1)		407.4587 nM
1832576-04-1 (compound 2)		49.257 nM
1832577-09-9 (compound 3)		778.82434 nM
Pfizer-01-01		359.95273 nM
Pfizer-01-02		557.5762 nM
Pfizer-01-03		863.8893 nM
Pfizer-01-04		5810.657 nM
Pfizer-01-05		13802.098 nM

Table B.4: Ligand structures and experimental binding affinities for the MALT1 system.

Pfizer-01-06	<p>Pfizer-01-06</p>	14603.069 nM
Pfizer-01-07	<p>Pfizer-01-07</p>	19565.105 nM
Pfizer-01-08	<p>Pfizer-01-08</p>	>20000.0 nM
Pfizer-01-09	<p>Pfizer-01-09</p>	>20000.0 nM
Pfizer-01-10	<p>Pfizer-01-10</p>	>20000.0 nM
Pfizer-01-11	<p>Pfizer-01-11</p>	>20000.0 nM
Pfizer-01-12	<p>Pfizer-01-12</p>	>20000.0 nM
Pfizer-01-13	<p>Pfizer-01-13</p>	>20000.0 nM

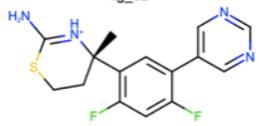
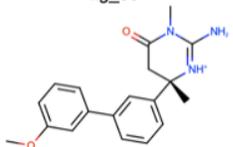
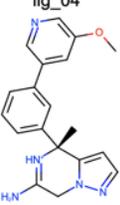
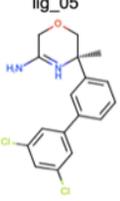
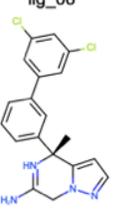
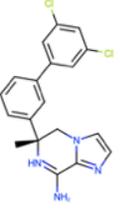
name	2D structure	Exp. IC50
lig_02		240 nM ⁵
lig_03		269 nM ⁶
lig_04		110 nM ⁷
lig_05		1122 nM ⁸
lig_06		148 nM ⁷
lig_07		2138 nM ⁹

Table B.5: Ligand structures and experimental binding affinities[109, 144, 150, 149, 148] for the biaryl ligands in the BACE1 system.

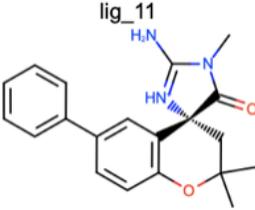
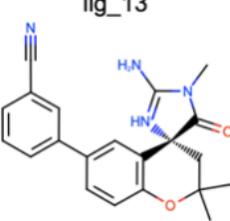
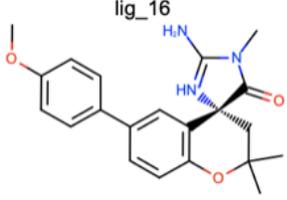
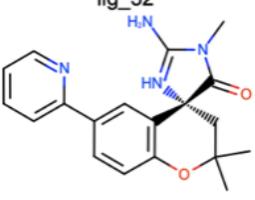
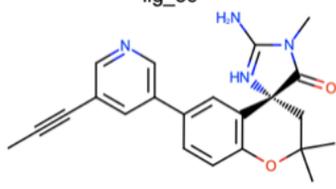
lig_1a		36830 nM ¹⁰
lig_11		455 nM ¹⁰
lig_13		94 nM ¹⁰
lig_16		7197 nM ¹⁰
lig_32		2363 nM ¹⁰
lig_36		10 nM ¹⁰

Table B.6: Ligand structures and experimental binding affinities[60] for the spirocyclic ligands in the BACE1 system.

lig_41	<p>lig_41</p>	240 nM ¹¹
lig_45	<p>lig_45</p>	48 nM ¹¹
lig_67	<p>lig_67</p>	18 nM ¹²
lig_69	<p>lig_69</p>	20 nM ¹²
lig_74	<p>lig_74</p>	2 nM ¹³
lig_81	<p>lig_81</p>	11 nM ¹³

Table B.7: Ligand structures and experimental binding affinities[44, 106, 108] for the amide ligands in the BACE1 system.

Appendix C

Supporting Information: Impact of protein conformations on binding free energy calculations in the BACE1 system

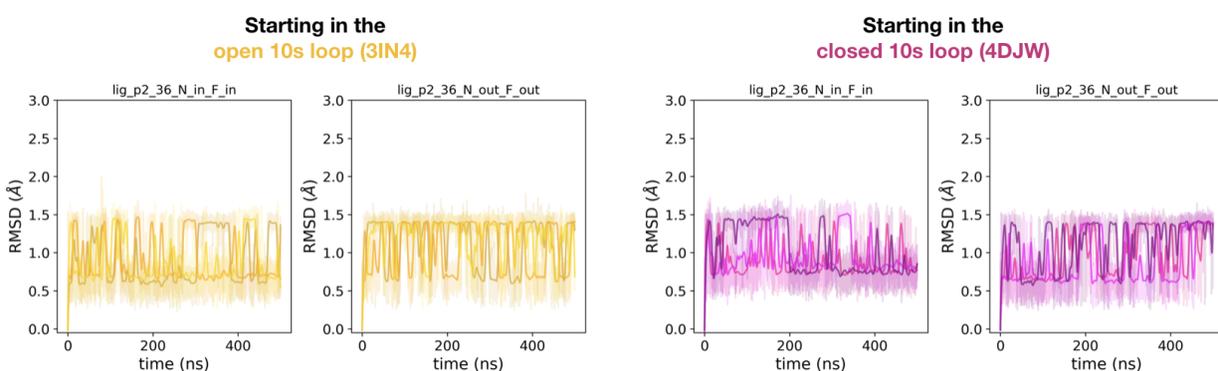


Figure C.1: RMSD of BACE1 ligand 1ig_p2_36 during MD simulations. Simulations were initiated from protein structure 3IN4 (left, yellow) or 4DJW (right, magenta). The RMSD of the ligand is shown as a function of simulation time. Separate plots are shown for initializing the MD simulations in two different binding modes (`N_in_F_in` and `N_out_F_out`) in both protein structures (3IN4, yellow and 4DJW, magenta). Data from all simulation frames are shown in lighter color, and a running average (averaged over 100 frames) in a darker color for the triplicate MD runs. In both protein structures, independently of the starting binding mode, the ligand RMSD is below 2\AA . Changes in ligand RMSD correlate with rotations around the RB1 (Figure 4.4), meaning that besides rotamer sampling, the ligand binding poses remain fairly constant during the MD simulation.

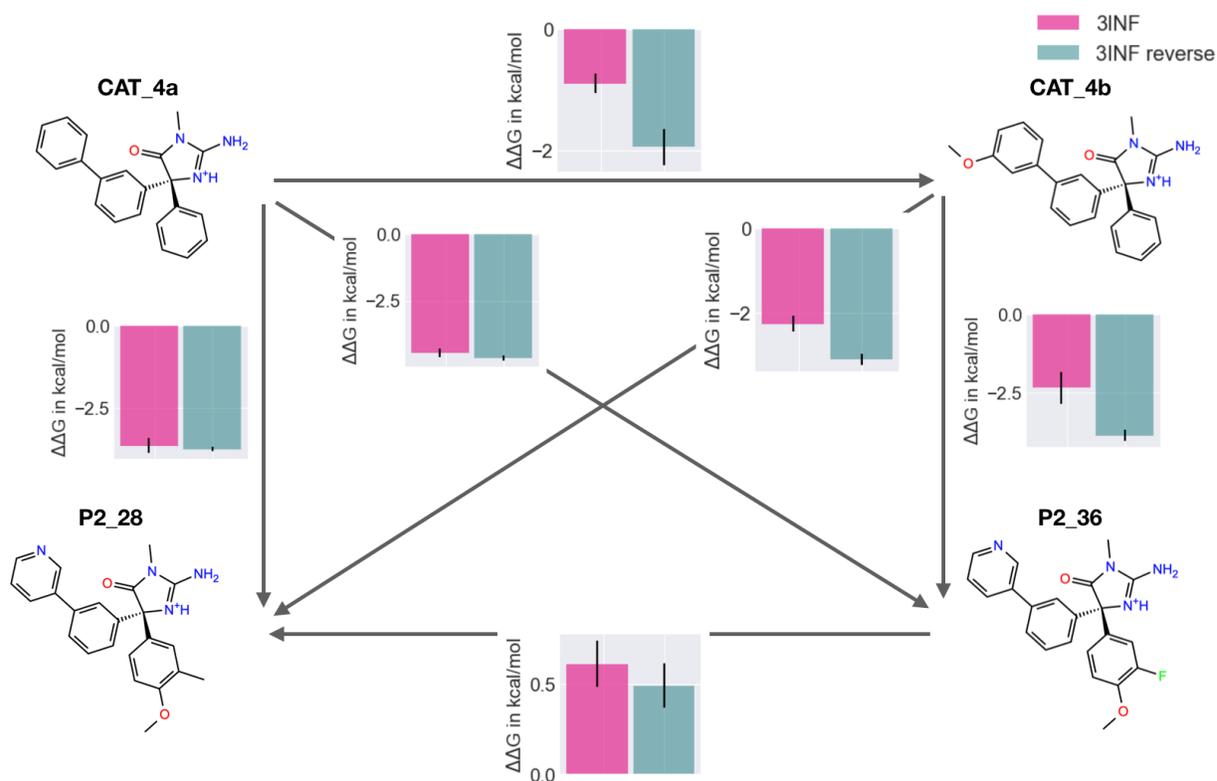


Figure C.2: RBFEE results from transformations between the four BACE1 inhibitors using crystal structure 3INF. Shown are calculated $\Delta\Delta G$ values (bar graphs) for each transformation, performed in both direction. For the reverse direction (3INF reverse) the negative of the relative free energy is shown, accounting for the sign flip when changing the direction of a transformation. For three of the six edges, statistically equivalent results for simulations from opposite directions were obtained. For the three edges that involve ligand CAT_4b, results differed depending on the direction of the transformation. Since we would expect to obtain the same $\Delta\Delta G$ independent of the direction (after accounting for the sign flip), this indicates sampling problems that are likely caused by insufficient sampling of water replacement by the methoxy group.