

UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Probing the activation of histone deacetylase 3 using computer simulations /

Permalink

<https://escholarship.org/uc/item/7g38x2cx>

Author

Arrar, Mehrnoosh

Publication Date

2014

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Probing the activation of histone deacetylase 3 using computer simulations

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

in

Chemistry

by

Mehrnoosh Arrar

Committee in charge:

Professor James Andrew McCammon, Chair
Professor Michael Gilson
Professor Judy Kim
Professor Emmanuel Theodorakis
Professor John Weare

2014

Copyright
Mehrnoosh Arrar, 2014
All rights reserved.

The dissertation of Mehrnoosh Arrar is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2014

EPIGRAPH

The truth is that we live out our lives putting off all that can be put off; perhaps we all know deep down that we are immortal and that sooner or later all men will do and know all things.

—Jorge Luis Borges

TABLE OF CONTENTS

Signature Page	iii
Epigraph	iv
Table of Contents	v
List of Figures	vii
List of Tables	viii
Acknowledgements	ix
Vita	x
Abstract of the Dissertation	xi
Chapter 1	General introduction	1
	1.1 Computer simulations of biomolecules	1
	1.2 Background and significance of systems of study	2
	1.3 The sampling problem	4
Chapter 2	Structural insight into the separate roles of IP4 and DAD in activation of histone deacetylase 3	5
	2.1 Introduction	5
	2.2 Methods	6
	2.3 Results	8
	2.4 Discussion	17
	2.5 Conclusions	20
	2.6 Supporting Information	20
	2.7 Acknowledgements	23
Chapter 3	Inactivating mutation in Histone deacetylase 3 stabilizes its active conformation.	25
	3.1 Introduction	25
	3.2 Methods	27
	3.3 Results	28
	3.3.1 Global stability of apo HDAC3 ^{R265P} resembles that of activated HDAC3 ^{WT}	28
	3.3.2 Geometry of apo HDAC3 ^{R265P} active site channel resembles that of activated HDAC3 ^{WT}	31
	3.4 Discussion	33
	3.5 Acknowledgements	35

Chapter 4	Structure-based virtual screen for novel inhibitors of histone deacetylase 3.	36
	4.1 Introduction	36
	4.2 Methods	38
	4.3 Results and Discussion	39
	4.4 Conclusions	47
	4.5 Acknowledgements	48
Chapter 5	w-REXAMD: A Hamiltonian replica exchange approach to improve free energy calculations for systems with kinetically-trapped conformations	49
	5.1 Introduction	49
	5.2 Methods	52
	5.3 Results and discussion	54
	5.3.1 Conformational Sampling	54
	5.3.2 Alchemical Free Energy Calculations	56
	5.4 Conclusions	59
	5.5 Supporting Information	61
	5.5.1 Accelerated MD Approach	61
	5.5.2 Hamiltonian Replica Exchange	61
	5.5.3 Thermodynamic Integration	63
	5.6 Acknowledgements	65
Chapter 6	General conclusions	66
Bibliography	68

LIST OF FIGURES

Figure 2.1:	Superimposed snapshots from HDAC3 simulations	9
Figure 2.2:	Change in by-residue fluctuations in HDAC3 when bound to IP4 and DAD	11
Figure 2.3:	Principal component analysis of apo and complexed HDAC3	14
Figure 2.4:	Impact of IP4 binding on protein-protein interface stability	15
Figure 2.5:	Δ RMSF of DAD backbone	16
Figure 2.6:	Surface representation of stable channel leading to active site	17
Figure 2.7:	Influence of DAD and IP4 binding on dynamics of key active site residues	18
Figure 2.8:	RMSD of HDAC3 backbone	21
Figure 2.9:	Cluster representatives	22
Figure 2.10:	RMSD of DAD backbone	23
Figure 3.1:	Cartoon representation of active HDAC3 complex	26
Figure 3.2:	Principal component analysis of HDAC3	29
Figure 3.3:	Analysis of mutant and wild-type HDAC3 changes in by-residue RMSF	30
Figure 3.4:	Geometry of active site channel	31
Figure 3.5:	Population analysis of catalytic Tyr298	32
Figure 4.1:	Active site channel stabilized by DAD and IP4	38
Figure 4.2:	Summary of virtual screen workflow	40
Figure 4.3:	Distribution of Robust Initial Enhancement metric for receptor models	41
Figure 4.4:	Projection of receptor models in HDAC3 principal component space	42
Figure 4.5:	Assessment of positive controls	43
Figure 4.6:	Summary of virtual screen results	45
Figure 4.7:	Alternative druggable sites	46
Figure 5.1:	Conformational sampling of butane	55
Figure 5.2:	Conformational sampling of cyclohexane	57
Figure 5.3:	Thermodynamic cycle of alchemical identity transformation of bromocyclohexane	58
Figure 5.4:	Free energy of identity transformation for bromocyclohexane	60
Figure 5.5:	Convergence of chair conformations in unaccelerated state	64

LIST OF TABLES

Table 2.1:	.418pt_____	22
Table 2.2:	Parametrization of bonds with Zn center	23
Table 2.3:	Parametrization of angles with Zn center	24
Table 2.4:	Charges modified near Zn center	24
Table 5.1:	.418pt_____	62
Table 5.2:	Summary table of mixing among replicas	65

ACKNOWLEDGEMENTS

First I must thank my parents, who taught me the resilience that got me this far. So many friends and family members have helped me along the way but I would like to especially thank Farnoz and Sara for their hilarious group messages; Mehrdod and Wendy for making me an aunt; and Matt and Andrew for becoming my family here in San Diego. I owe a special thank you to Adrián, who randomized my trajectory when I was just a freshman; and to Dario, Marce and Lula, for showing me how a family does science. Thank you to Santi, for taking care of me when I was in a new city. And to Leo, for reluctantly agreeing to drive to Tucumán that one winter in Argentina.

I owe a debt of gratitude to my Ph. D advisor, Andy, who has shared and taught me so much while somehow still letting me find the excitement and responsibility of independent research. I also thank all those in the group who have collaborated with me: Joseph Kaus, César Augusto F. de Oliveira, Mikolai Fajer, William Sinko, Levi Pierce, and Rigney Turnham, as well as the rest of the group for always keeping it a fun place to be. And of course, thanks to Patti Craft for making all administrative things a piece of cake.

Finally, I would also like to thank the other members of my doctoral committee, Professors Judy Kim, John Weare, Emmanuel Theodorakis, and Michael Gilson, for their careful guidance at committee meetings.

Chapter 2 is a near reprint of the material as it appears in Protein Science 2013, Volume 22, Pages 83-92. The dissertation author was the primary contributor and first author of this work.

Chapter 3 is a minimally modified version of a manuscript in Protein Science 2013, Volume 22, Pages 1306-1312. The dissertation author was the primary contributor and first author of this work.

Chapter 4 is a minimally modified version of a manuscript being prepared for submission. The dissertation author was the primary contributor and first author of this work.

Chapter 5 is a near reprint of the material as it appears in J. Chem. Theory Comput. 2013, Volume 9, Pages 18-23. The dissertation author was the primary contributor and first author of this work.

VITA

2010	B. S. in Chemistry <i>Summa cum laude</i> , University of Florida
2010-2011	Graduate Teaching Assistant, University of California, San Diego
2012	M. S. in Chemistry, University of California, San Diego
2014	Ph. D. in Chemistry, University of California, San Diego

PUBLICATIONS

Arrar, M., F. de Oliveira, C.A., McCammon, J.A. “Inactivating mutation of Histone Deacetylase 3 stabilizes its active conformation”, *Protein Science* **2013**, 22(10), pp 1306-1312.

Arrar, M., F. de Oliveira, C.A., Fajer, M., Sinko, W., McCammon, J. A. “w-REXAMD: A Hamiltonian replica exchange approach to improve free energy calculations for systems with kinetically-trapped conformations”, *Journal of Chemical Theory and Computation* **2013**, 9(1), pp 18-23.

Arrar, M., Turnham, R., Pierce, L., F. de Oliveira, C.A., McCammon, J. A. “Structural insight into the separate roles of inositol tetrphosphate and deacetylase-activating domain in activation of histone deacetylase 3”, *Protein Science* **2013**, 22(1), pp 83-92.

Boechi, L., **Arrar, M.**, Marti, M., Olson, J., Estrin, D., Roitberg, A. “Hydrophobic effect drives oxygen uptake in myoglobin via histidine E7”, *Journal of Biological Chemistry* **2013**, 288(9), pp 6754-6762.

Capece, L., **Arrar, M.**, Roitberg, A., Yeh, S., Marti, M., Estrin, D. “Substrate stereospecificity in tryptophan dioxygenase and indoleamine 2,3-dioxygenase”, *Proteins* **2010**, 78(14), pp 2961-72.

ABSTRACT OF THE DISSERTATION

Probing the activation of histone deacetylase 3 using computer simulations

by

Mehrnoosh Arrar

Doctor of Philosophy in Chemistry

University of California, San Diego, 2014

Professor James Andrew McCammon, Chair

In the past several decades, exciting research has elucidated that the structural dynamics of proteins have a profound impact on their activities. Molecular dynamics (MD) simulations have become a powerful means of exploring the ensembles of structures accessible to proteins, in order to compute thermodynamic properties that allow us to understand the forces that drive the binding of a drug to its receptor, or even the formation of macromolecular assemblages. The research presented in this dissertation utilizes MD simulations, among other computational techniques, introduced in Chapter 1, to further our understanding of the activation mechanism of histone deacetylase 3 (HDAC3), a promising target for epigenetic cancer therapy. The enzymatic activity of HDAC3 depends on its binding to other proteins and a small molecule, inositol

tetraphosphate (IP4). In Chapter 2, we show how the binding of HDAC3 to a deacetylase activating domain and to IP4 restrain the dynamics of HDAC3, suggesting a single active HDAC3 conformation. In Chapter 3, we show that the dynamics of HDAC3 can be similarly stabilized by mutating a single residue near the IP4-binding site, shedding light on the long-range allosteric networks that must be conserved for the deacetylase activity of HDAC3. In Chapter 4, we show that the restrained conformational ensembles of HDAC3 have predictive power in identifying known HDAC inhibitors among decoy drug-like molecules, and perform a structure-based virtual screen for novel active site and allosteric inhibitors. One of the challenges in using MD simulations to address these types of questions for systems is the extent to which an MD trajectory can indeed represent the conformational ensemble of the system. In Chapter 5, we present a method that aims to enhance the sampling of relevant conformations, particularly for a problematic system with kinetically "trapped" conformations.

Chapter 1

General introduction

1.1 Computer simulations of biomolecules

Over the past few decades, exciting advances in experimental and computational methodologies have brought the molecular machinery of the cell to the frontiers not only of the natural sciences, but also of mathematics, engineering, and medicine. Ever since the 1950s, pieces and patterns in the scientific community began to fuel an itch. An itch to define the amorphous magical machinery behind genetics, metabolism, catalysis. Suddenly the creative sketches of cylinders and coils in the biology textbooks were not enough. Phenomenological models for allostery and cooperativity postulated that changes in the conformations of proteins could underly the experimentally observed phenomena[72]. Not only did it become clear that a protein's structure was important for its function, but suddenly multiple conformations had to be considered to understand its activity. When Perutz and co-workers resolved the first three-dimensional structure of hemoglobin in 1960[76], eager chemists were also developing theoretical models to describe the energy of a protein as a function of its three dimensional configuration, models based on the chemical nature of interactions between the atoms[56].

With admirable creativity, scientists began to use the principles of molecular dynamics (MD), which at the time were being used to study models of liquids and gases, to propagate the atomic coordinates of proteins in time[69]. Through these early molecular dynamics simulations, it became evident that the static structures resolved from x-ray crystallography were averaged over the many configurations accessible within thermal

fluctuations of the protein. With many years of improvement on the initial models, or "force fields", used to compute the forces on protein atoms, as a function of their cartesian coordinates, this computational approach to study the dynamics of biological molecules has become a field of its own, even divided with nearly religious beliefs about which charge model is most accurate, and which compilation of MD code is most efficient when distributed among many central or graphics processing units[54, 55]. During the preparation of this dissertation, the Royal Swedish Academy of Sciences awarded the Nobel prize in Chemistry to Martin Karplus, Michael Levitt, and Arieh Warshel, "for the development of multiscale models for complex chemical systems", making this an exciting time indeed to be a computational chemist.

MD simulations have emerged as a reliable methodology to support and even challenge data obtained with experimental techniques in biophysics, and together the fields have allowed for considerable leaps in our understanding of the importance of protein structural *dynamics* in biological and chemical processes. The atomistic representation of protein structures and the ability to propagate these structures over nano- to microsecond timescales have perhaps only further emphasized and drawn attention to the rich complexity underlying the phenomena of allostery and cooperativity in proteins and other macromolecular assemblies[85, 61].

1.2 Background and significance of systems of study

The research presented in this dissertation in large part advances our understanding of the structural dynamics underlying the activation, and inhibition, of an enzyme, Histone deacetylase 3 (HDAC3). HDAC3 participates in many biochemical pathways, but perhaps most well-studied is its role in regulating gene transcription through the removal of acetyl groups from acetyllysine amino acid residues on the flexible N-termini, or "tails" of histone core proteins[71]. Through the removal of acetyl groups, HDACs effectively restore the positive charge to the Lys residues, which is thought to promote the interaction between positively charged histone tails and negatively charged DNA wrapped around neighboring histone core proteins, thus compacting the local chromatin structure and repressing transcription of the associated genes[10].

Acetylation is a post-translational modification that modulates the behavior of many proteins, for example the assembly/disassembly of tubulin, and the list of non-histone HDAC substrates, both cytoplasmic and nuclear, is growing[71, 47]. Inhibitors of HDACs are currently used in the treatment of acute T-cell lymphoma, and inhibitors with targeted specificity for a particular HDAC isoform have been shown to be promising candidates for neurological diseases, like Friedrich's Ataxia and Huntington's disease, as well as in prompting HIV-infected cells out of latency[44, 35].

From a biophysical perspective, this system presents an interesting challenge to our understanding of structure-function relationships, as the enzymatic activity of HDAC3 depends upon its assembly into a larger multi-protein nuclear receptor complex[73]. The first structure of the deacetylase was only recently resolved, using X-ray crystallography, where HDAC3 is bound to the necessary portion of the nuclear receptor complex (called the deacetylase activating domain, or DAD)[99]. In this same structure a small molecule, inositol tetrphosphate (IP4), was also resolved at the protein-protein interface, and was discovered to play a critical, yet mechanistically ambiguous, role in activating HDAC3 as well.

Chapter 2 dissects the effects, both local and long-range, that IP4 and DAD each have on the structural dynamics of HDAC3, through the use of molecular dynamics simulations of free, and IP4- or DAD-bound HDAC3 complexes. We identify stabilizing effects unique to either IP4 and DAD that we propose to be key in maintaining the conformation of a channel leading to the active site. In Chapter 3, we further probe the allosteric networks in HDAC3 by investigating the effect of an *in silico* single point mutation at the IP4 binding region. We identify this site, i.e. the arginine residue at position 265, to be an allosteric "hot spot", which, when substituted with a proline residue, can restrain the entire dynamics of the deacetylase backbone, similar to the effect that IP4 binding has on HDAC3. Despite its similarly restrained conformation, the R265P mutant has been shown to have no deacetylase activity, suggesting that the activation of HDAC3 is more fine-tuned than previously expected. We propose that the subtle differences between the dynamics of HDAC3 when bound to IP4 (or with the R265P mutation), and those of the wild-type HDAC3:IP4:DAD complex are indeed key in the activation of the enzyme.

In Chapter 4, we apply the insight gained from MD simulations in a drug discovery effort for novel inhibitors of HDAC3. Recently, the incorporation of an ensemble of receptor conformations in conventional docking protocols has been successful in identifying novel inhibitors in virtual screens[29]. Here, we expand upon this idea by further filtering ensembles of receptor conformations by first evaluating their predictive performance in ranking known inhibitors out of a larger subset of drug-like decoy compounds.

1.3 The sampling problem

Although MD simulations are routinely used to study the dynamic properties of biomolecules, there are two key challenges in the field. One significant challenge is the accuracy of the atomistic model used to represent the system of interest[8]. The model, which includes how the bonded and non-bonded interactions are accounted for in the potential energy of the system that then determines the forces which propagate the system in time, as well as the description of charge distributions on atoms, is always being tested and improved to more accurately reproduce various thermodynamic properties of proteins, solvent, ions, nucleic acids, and metallic complexes.

The other challenge encountered with MD simulations is the assumption that a given trajectory of the system over time is representative of the ensemble of states in phase space that the system may occupy, enabling us to calculate the ensemble-averaged thermodynamic properties as time averages of a molecular dynamics trajectory. The problem, however, is that often a starting configuration for a single trajectory is separated from another, perhaps biologically relevant, conformation by a barrier that is larger than the thermal fluctuations of the system. In such cases, the system is kinetically trapped, and a conventional MD simulation is no longer a sufficient tool for determining accurate thermodynamic ensemble averages.

In Chapter 5 we address this sampling problem in MD simulations by introducing a new enhanced sampling technique, which is a combination of previous techniques developed in our group. The method, w-REXAMD, is applied to the calculation of the free energy change in the self-transformation of small system with known kinetically trapped states, bromo-cyclohexane.

Chapter 2

Structural insight into the separate roles of IP4 and DAD in activation of histone deacetylase 3

2.1 Introduction

Histone deacetylases (HDACs) play an important role in the regulation of gene expression by modifying local chromatin structure[10]. HDACs directly oppose the activity of acetyl transferases by deacetylating acetyllysine residues on specific histone tails. Hypoacetylation of histone tails leads to compacted chromatin structure, limiting accessibility of DNA to necessary transcription factors, and thus repressing transcription of certain genes[81].

Abnormal gene expression is often associated with cancer cells, and epigenetic routes to cancer therapy are of growing interest. Eleven of the known human HDACs (classes I, II, and IV) have known important roles in cancer cell biology[102] and neurodegenerative diseases[49, 26, 108]. Some HDAC inhibitors are already utilized in cancer therapies and targeted inhibition of HDACs may be key to less toxic treatments. In this study we focus specifically on the activation mechanism of the class I deacetylase HDAC3. Inhibition of HDAC3 in particular has been shown to disrupt HIV-1 latency and reduce cancer cell proliferation[3, 44, 89].

The catalytic activity of HDACs generally depends on their recruitment to large multi-protein complexes[64, 73]. Structural data on active HDAC complexes is still sparse, and only recently Watson et al. reported the first crystal structure of HDAC3[99]. Unlike the other class I deacetylases, HDAC3 is recruited to the NCoR/SMRT nuclear receptor complex. Both NCoR and SMRT have critical N-terminal deacetylase-activating domains (DAD), which is necessary for deacetylase activity[100, 107, 38]. The SMRT DAD domain was also resolved in the crystal structure of HDAC3[99]. In the crystal structure, an inositol (1,4,5,6) tetraphosphate (IP4) molecule was found at the interface of HDAC3 and DAD, adjacent to the mouth of a 12 Å tunnel that leads to the zinc center of the active site[99]. Mutations that disrupt the interactions between HDAC3 and IP4 abolished both deacetylase activity and HDAC3:DAD complex formation, consistent with the necessity for DAD-binding in the activation of HDAC3[99]. Interestingly, a DAD mutation K449A which was inoffensive to HDAC3:DAD complex formation also abolished deacetylase activity[20]. Because K449 is close to the IP4 binding site, the K449A mutation may inactivate the deacetylase via mediation of IP4-HDAC3 interactions, which would suggest that IP4 plays an additional role in the activation of HDAC3, aside from stabilizing the HDAC3:DAD complex.

Here we examine the structural ensemble of HDAC3, specifically: (i) how does this ensemble of HDAC3 structures change between the active HDAC3:DAD:IP4 complex and the inactive apo HDAC3 state? (ii) Does the presence of either IP4 or DAD alone have a significant impact on the dynamics of HDAC3? and (iii) What effect does IP4 have on the HDAC3:DAD complex?

Our approach is to use molecular dynamics simulations to explore the structural ensembles of apo HDAC3 as well as the HDAC3:DAD:IP4, HDAC3:DAD, and HDAC3:IP4 complexes. In the Results section that follows, we will address each of above questions. We will then discuss our findings about the HDAC3 conformational ensembles in terms of its histone deacetylase activity.

2.2 Methods

1. System Preparation

Zinc metal center. Instead of assigning a formal charge of +2 to the zinc metal center, charges were assigned by performing quantum mechanical calculations on a simplified model of the active site. The active site model included the side chains of the zinc-coordinated ASP and HIS residues capped with methyl groups. To maintain the tetrahedral geometry and simulate the apo state of the protein we added a water molecule where an acetate ion was found in the crystal structure. A geometry optimization was performed using Gaussian (g03)[33], with B3LYP and the 6-31G(d) basis set. Electrostatic potential fits were also obtained from gaussian, using the optimized geometry, and Hartree Fock level of theory with the 6-31G* basis set. The Restrained Electrostatic Potential (RESP) module within the Antechamber program in AmberTools[14] was used to assign charges to the atoms using the Gaussian output.

The force constants for bond, angle, and torsions for the atoms bonded to zinc were taken from the literature or by analogy to similar active sites[77]. Equilibrium angles and bond distances involving zinc were taken from the crystal structure[99]. Inositol (1,4,5,6) tetrphosphate parameters. Charges for the small molecule InsP4 were obtained using the same procedure described for the zinc metal center. The Amber ff99SB force field was used (CT, OS, P, O2 atom types) for all parameters except for the C-O-P-O- dihedral which was treated like the C-O-P-OH dihedral by analogy. Equilibrium bonds and angles were adjusted slightly to agree with the crystal structure.

Side chain protonation states. PROPKA 3.1[7, 74] was run on the web-server (pdb 4A69). Only two histidine residues had particularly low pKas in HDAC3 and were modified to be charged HIP residues. Other histidines were designated as either HIE or HID (with epsilon or delta nitrogen protonated, respectively) by evaluating the local environments of each side chain in the crystal structure.

2. Molecular dynamics details

Solvation and equilibration. The systems were built using the AMBER ff99SB force field, with the ildn modification for ILE, LEU, ASP, and ASN residues[65]. Each system was solvated in a 10.0 Å TIP3P octahedron, and potassium ions were added to neutralize the charge in each case: HDAC (9), HDAC:InsP4 (17), HDAC:InsP4:DAD (13). Additional potassium and chloride ions were added to simulate physiological salt concentration (0.15 M). The updated Lennard Jones parameters for ions were used[52].

Systems were heated to 300K at constant volume (NVT) with restraints on the protein that were gradually reduced from 200 to 0 kcal/mol/Å² over a period of 150ps. The langevin thermostat was used with a collision frequency of 1.0 ps⁻¹. The SHAKE algorithm was used to constrain bonds to non-polar hydrogens, and a 1.0 fs time-step was used during dynamics[82]. An 8.0 Å cutoff was used for non-bonded interactions. 100ps equilibration runs were done in the NPT ensemble, using isotropic pressure scaling and a pressure relaxation time of 2.0 ps.

Production simulations. For each system, three independent simulations were started from different snapshots from the equilibration simulations (taken at arbitrary intervals after density of water box was equilibrated) with randomized velocities. All production runs were done in the same NVT conditions described in 'Solvation and equilibration', with the exception of a 2.0 fs time-step. All simulations were done using the PMEMD module within the Amber11 simulation package[14]. Production runs were performed on GPUs using Amber11 pmemd.CUDA, each for 100ns.

2.3 Results

Dynamic regions in apo HDAC3 are stabilized by the presence of DAD and IP4

Without DAD and IP4, HDAC3 has intrinsic flexibility There is currently only one crystal structure of HDAC3 available in the protein data bank (pdb 4A69), in which the deacetylase is in complex with DAD and IP4. We simulated the apo state of HDAC3 by removing both the DAD domain and IP4 from the crystal structure. In these simulations we observed a highly dynamic deacetylase backbone (Fig 2.1, panel A and SI Fig 2.8, panel A), particularly near the active site loops. In the ternary complex, on the other hand, we found the HDAC3 backbone maintained its crystal structure conformation (Fig 2.1, panel D and SI Fig 2.8, panel D). The only region with notable flexibility in the ternary complex was the random coil region (residues 340 to 348).

The increased fluctuations in the apo structure were localized to the active site loops and a nearby helix (Helix 6). In Fig 2.2, panel A we show the change in root mean square fluctuations (RMSF) per residue for the ternary and apo states. Negative

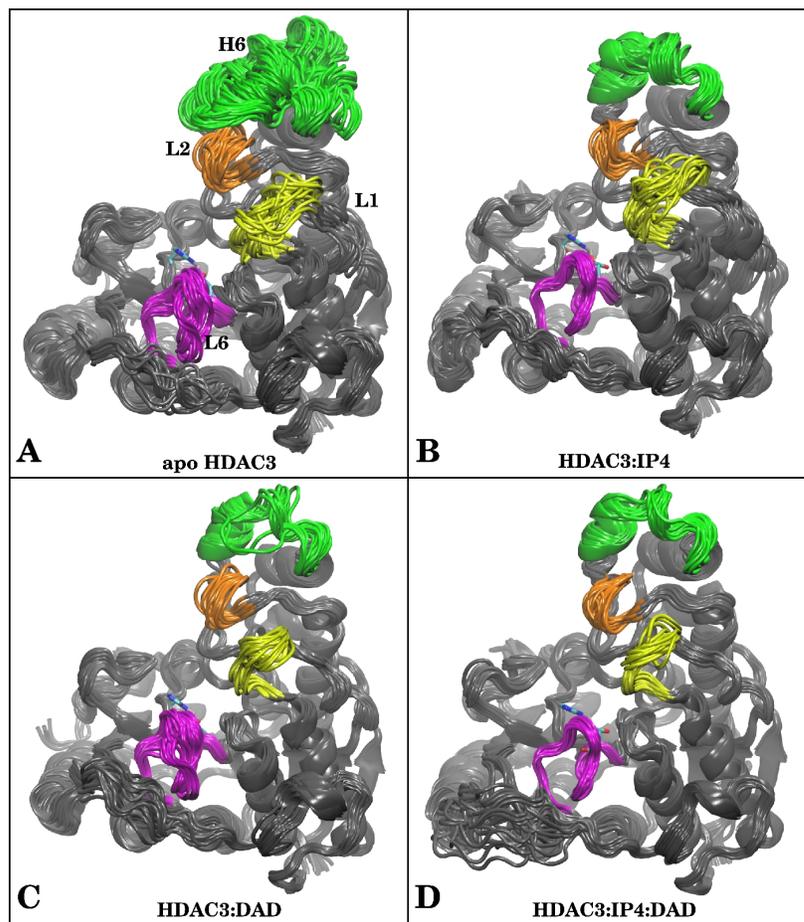


Figure 2.1: Superimposed snapshots from MD simulations for (A) apo HDAC3, (B) HDAC3:IP4, (C) HDAC3:DAD, and (D) HDAC3:IP4:DAD. Helix 6 (green), Loop1 (yellow), Loop2 (orange), and Loop 6 (magenta) are highlighted to show differences in dynamics. If present in the simulation, DAD and IP4 are shown as purple ribbons or red ball-and-stick representations, for reference

Δ RMSF values indicate flexible regions of apo HDAC3 that are stabilized in the ternary complex. The Δ RMSF values are also projected onto the protein structure, using a Blue (negative Δ RMSF)-Gray-Red (positive Δ RMSF) color scale. The regions stabilized in the presence of both co-repressors not surprisingly include Loops 1 and 6, which form part of the IP4-binding site at the protein-protein interface. Perhaps less expected was the flexibility of Helix 6 (residues 78 to 88) near Loop 2, which was most flexible and even unfolds in the apo HDAC3 simulation, but does not directly interact with either DAD or IP4. These results suggest an overall decreased flexibility of specific regions of the deacetylase backbone upon co-repressor binding.

Presence of IP4 alone reduces overall HDAC3 backbone flexibility

In an attempt to separate the contributions of DAD and IP4 to the observed stability of the deacetylase backbone in the ternary complex, we analyzed the dynamics of the intermediate HDAC3:IP4 and HDAC3:DAD complexes. In Fig 2.1, panels B and C, we show superimposed snapshots from these simulations. While the presence of the DAD domain resulted in less fluctuations in the deacetylase backbone than in the apo state (Fig 2.2, panel C), we were surprised to find that the presence of IP4 alone greatly reduced the flexibility of the deacetylase backbone (Fig 2.2, panel B). The regions stabilized by either DAD or IP4 were the same regions stabilized in the ternary complex, although their respective contributions seem to be complementary. The Δ RMSF data show that IP4 has a slightly greater stabilizing effect on Helix 6, Loop 6, and the C-terminal helix (Fig 2.2, panel B) than does the DAD domain (Fig 2.2, panel C). The DAD domain, on the other hand, has a more notable stabilizing effect on Loop 1 (residues 19 to 25). Neither the binding of DAD nor IP4 appears to affect the random coil at residues 340 to 348 substantially, but the binding of both co-repressors does seem to cause an increase in flexibility in this region. We should highlight the fact that the stabilized regions in the presence of IP4 were not necessarily near the IP4 binding site, suggesting the importance of allosteric regulation in HDAC3 activity.

Clustering of the apo HDAC3 structural ensemble suggests population-shift upon DAD and IP4-binding

Given the observed variations in the flexibility of the deacetylase backbone in

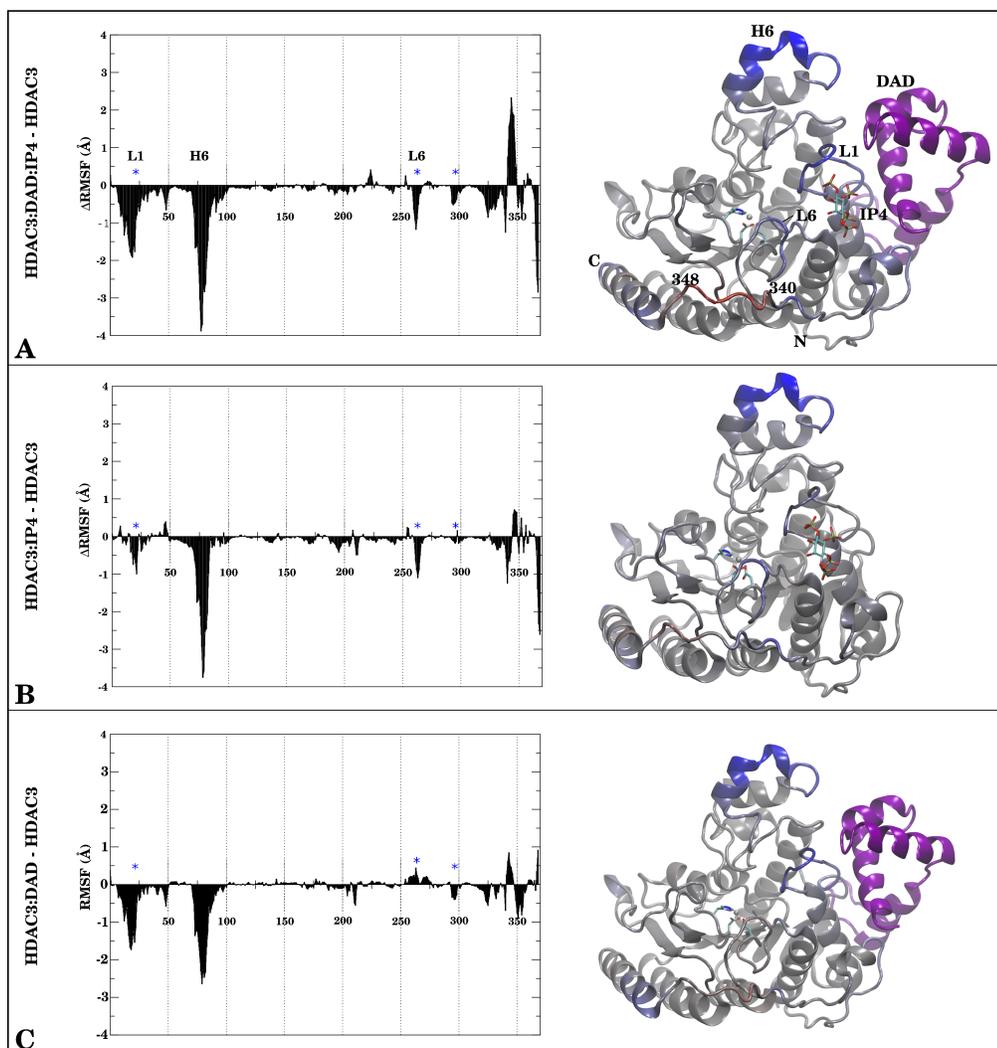


Figure 2.2: The Δ RMSF (Å) per residue in HDAC3 between (A) HDAC3:DAD:IP4, (B) HDAC3:IP4, (C) HDAC3:DAD and the apo state is shown on the left side of each panel. The same values are projected onto a structure of the protein for reference, using a Blue (Negative) \rightarrow Gray (Zero) \rightarrow Red (Positive) scale. A range of -3.00 to +3.00 Å was used to make the figure. DAD and IP4, when present in the simulations are shown for reference in purple ribbons and stick representations, respectively.

the different co-repressor-bound and apo states, we sought to answer the question: are the same HDAC3 conformations being stabilized in the different co-repressor-bound HDAC3 ensembles? We first clustered the structural ensemble of the apo state, using an RMSD metric and the hierarchical clustering algorithm (see SI and Shao et al. for details[84]). The main characteristics distinguishing each of the four clusters were the orientation and helicity of Helix 6, as well as the conformations of Loops 1 and 6. To compare the co-repressor-bound states to these reference clusters, we then measured the RMSD of the HDAC3 backbone from the ternary, IP4-bound, and DAD-bound HDAC3 ensembles against each of the four apo cluster representatives. Nearly all of the HDAC3 structures from the ternary simulations ($\approx 90\%$) are within 1.4 Å of the Cluster2 representative structure (see Tab I), which comprises only $\approx 15\%$ of the apo structural ensemble. In fact, none of the structures from the ternary simulation resemble (using a 1.4 Å cutoff) any of the other apo cluster representatives aside from Cluster2. The presence of DAD alone resulted in $\approx 70\%$ of the structures resembling the Cluster2 representative, and $\approx 10\%$ resembling the Cluster0 representative. The presence of IP4 alone surprisingly resulted in $\approx 90\%$ of the structural ensemble to lie within 1.4 Å of the Cluster2 representative as well. It appears, then, that the limited flexibility (described above) in each of the co-repressor-bound deacetylase states actually shifts the population towards a single conformation, which was only minimally populated in the apo HDAC3 ensemble.

DAD and IP4 shift HDAC3 population by restraining motions along the first two principal components in apo HDAC3 dynamics

To further characterize the dynamics of apo HDAC3 and the impact of DAD and IP4 on these dynamics, we used principal component analysis (PCA) to identify the most dominant modes in the dynamics. This analysis provides information about the low frequency motions of the dynamics, which may complement the conformational analysis above. We computed the first two principal components (PC 1 and PC2), which comprised just over 50% of the total variance, from the apo HDAC3 simulations (using only the C α atoms). Movies that show PC1 and PC2 motions can be found in Supplementary Information. We then projected the HDAC3 dynamics from the apo, HDAC3:IP4,

HDAC3:DAD, and HDAC3:DAD:IP4 simulations onto the PC space, which describes motions of the loops contracting around the active site. In Fig 2.3 we show the projections of the HDAC3 dynamics from all four simulations, along with the projection of the crystal structure conformation for reference. The PCA analysis shows that the ternary complex is notably restrained along PC1 and PC2 (Fig 2.3, panel D), which was not surprising, given the low root mean square deviation from the crystal structure conformation (SI Fig 2.8, panel D). With the sole addition of IP4 to HDAC3, the motions characterized by both PC1 and PC2 were already restrained considerably (Fig 2.3, panel B). The addition of the DAD domain to HDAC3 also notably reduces the motions along both principal components. The PCA results identify two dominant modes in the apo HDAC3 dynamics that are each limited in the presence of DAD and IP4, which we suspect results in the observed population-shift effect upon the binding of both IP4 and DAD to HDAC3.

IP4 stabilizes the protein-protein interface as well as the deacetylase backbone

Given the presence of a number of basic residues at the protein-protein interface where IP4 binds, we expected a major contribution of IP4 to be to stabilize the HDAC3:DAD complex in this region, as was also suggested previously by Watson et al¹². Upon closer comparison of the HDAC3:DAD complexes with and without the small molecule, we found that IP4 does more than simply stabilize the protein-protein interface near the binding site. The entire DAD backbone was quite flexible when bound to HDAC3 without IP4 (see Fig 2.4, panel A and SI Fig 2.10). Surprisingly, the stabilizing effect of IP4-binding affected the DAD backbone nearly uniformly, regardless of proximity to the IP4-binding site (Fig 2.5).

Changes in active site loop dynamics are expected to affect substrate recognition

Accessibility to the active site is determined by loop conformations The remarkable stability of the active site loops in the ternary structure helps to maintain a very clear tunnel to the zinc metal center (see Fig 2.6). In its stable conformations the tun-

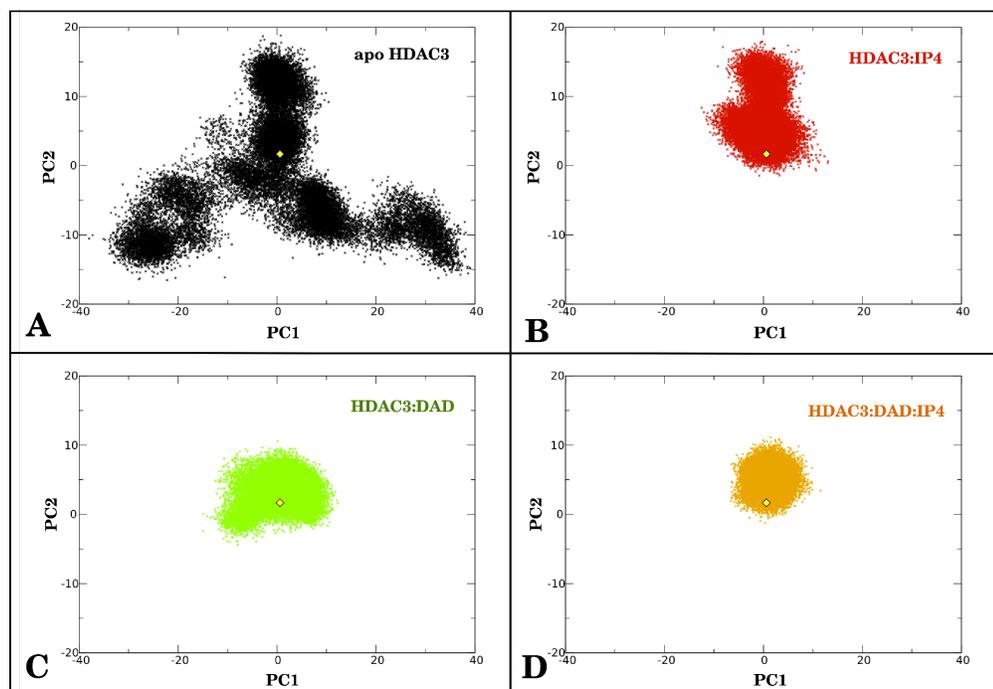


Figure 2.3: Principal component analysis of HDAC3. First two principal components were calculated from apo HDAC3 C α dynamics. Structural ensembles of HDAC3 C α atoms from (A) apo HDAC3, (B) HDAC3:IP4, (C), HDAC3:DAD, and (D) HDAC3:DAD:IP4 simulations were projected onto these two principal components. A yellow diamond in each graph marks the crystal structure conformation of HDAC3 for reference.

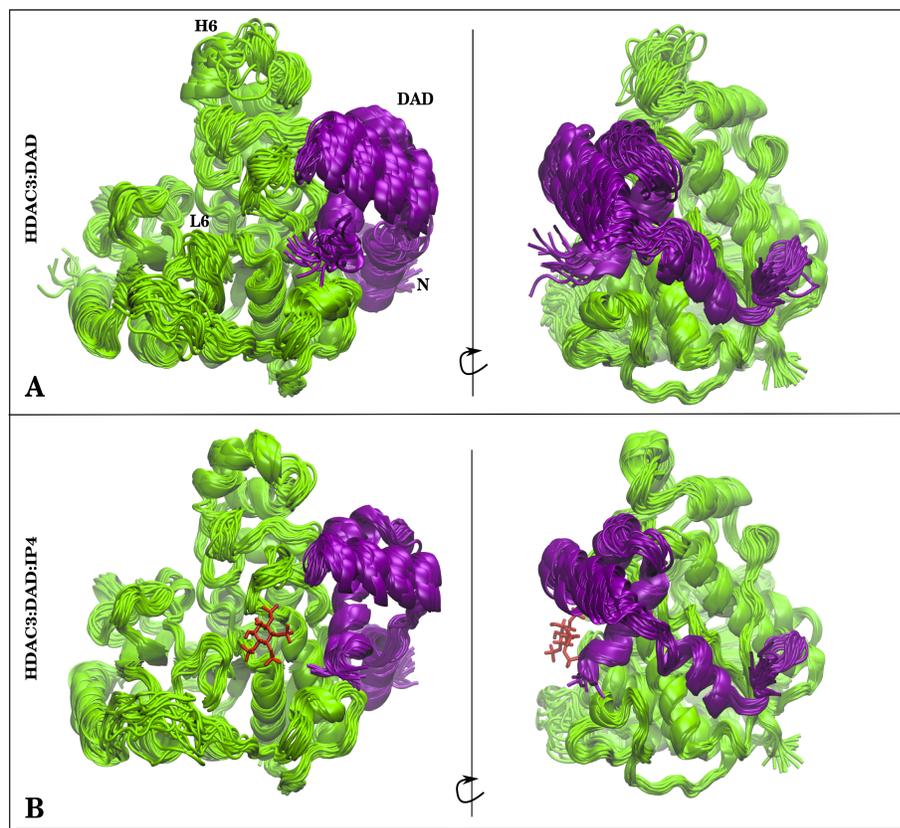


Figure 2.4: Structures from (A) HDAC3:DAD and (B) HDAC3:DAD:IP4 simulations are presented to show difference in stabilities of the protein-protein complex due to the presence of IP4. HDAC3 (green) is shown with the DAD domain (purple). If IP4 was present in the simulation, a single snapshot of IP4 is shown as red sticks for reference.

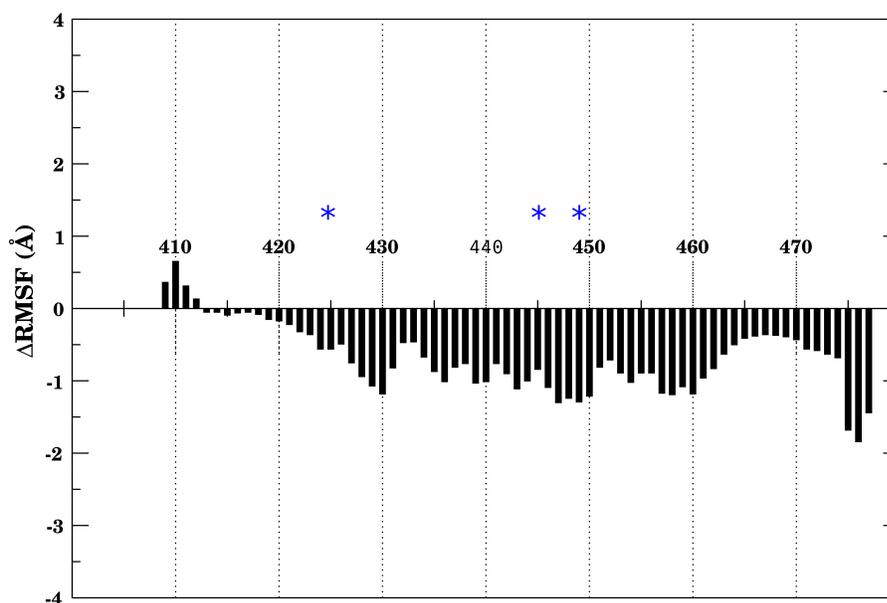


Figure 2.5: Δ RMSF of DAD backbone (HDAC3:DAD:IP4 minus HDAC3:DAD). Negative values indicate stabilization of backbone due to the presence of IP4. Blue asterisks indicate residues near IP4 binding site.

nel is lined with hydrophobic residues, with the exception of Tyr298. In the alternative loop conformations, this access to the active site is distorted into a wider opened, yet unblocked, conformation. The tunnel to the active site is likely to play an important role in substrate recognition. In the crystal structure of the ternary complex, an acetate ion was resolved in the active site along with a methionine side chain from an adjacent DAD domain in the crystal lattice. The authors suggested the resolved structure mimicked the presence of the substrate in the active site[99]. It is important to note that we observed the same tunnel without any ligands in the active site, given the presence of both DAD and IP4 bound to HDAC3.

Binding of co-repressors to HDAC3 shifts Tyr298 side chain from a solvent-exposed to an inward conformation The amino acid residues near the zinc center (D170, H172, D259, H134, H135, Q255) maintained the same side-chain conformations, regardless of the observed loop motions (see Fig 2.7). One notable exception, however, was Tyr298, located at the rim of the active site. We found that Tyr298 sampled two rotameric states, either solvent-exposed or poised inward toward the active site (as observed in the crystal structure)[99]. Fig 2.7 panel E shows the populations of the two

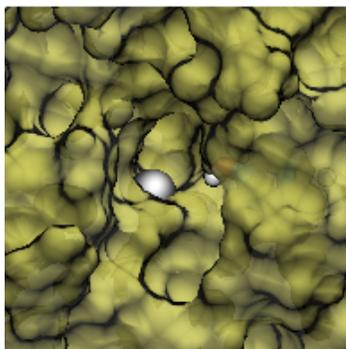


Figure 2.6: Surface representation of stable tunnel observed in ternary complex simulation. Zinc metal center is shown as a silver sphere. Tyr298 side-chain lining the tunnel is shown as sticks.

rotameric states for the ternary (HDAC3:DAD:IP4), DAD-bound, IP4-bound, and apo deacetylase states. Binding of each of the two co-repressors to HDAC3 biases the inward conformation of Tyr298, and in the ternary complex the Tyr298 side chain is exclusively inward (throughout the triplicate 100ns simulations). This population shift of the Tyr298 side chain may be necessary for substrate recognition, as it also forms part of the tunnel to the active site (Fig 2.6).

2.4 Discussion

Apo HDAC3 has flexible regions near the active site HDAC3 has recently been shown to depend on IP4 in addition to the DAD domain for its histone deacetylase activity[99]. The goal of this paper was to characterize any changes in the dynamics of HDAC3 due to the presence of its co-repressors. We observed from simulations of the apo state of HDAC3 that it samples a range of conformations, which we clustered into four groups. In the active ternary complex we found that the structural ensemble of HDAC3 consisted predominantly of just one of these four conformations. Interestingly, the presence of IP4 or DAD alone could also effect a similar and substantial population shift towards the stable conformation found in the ternary complex. Differences between RMSFs by residue suggest that DAD and IP4 stabilize flexible regions of apo HDAC3 differentially (Fig 2.2). Most of the stabilized regions not surprisingly are near the protein-protein interface where IP4 binds. Some significantly stabilized regions,

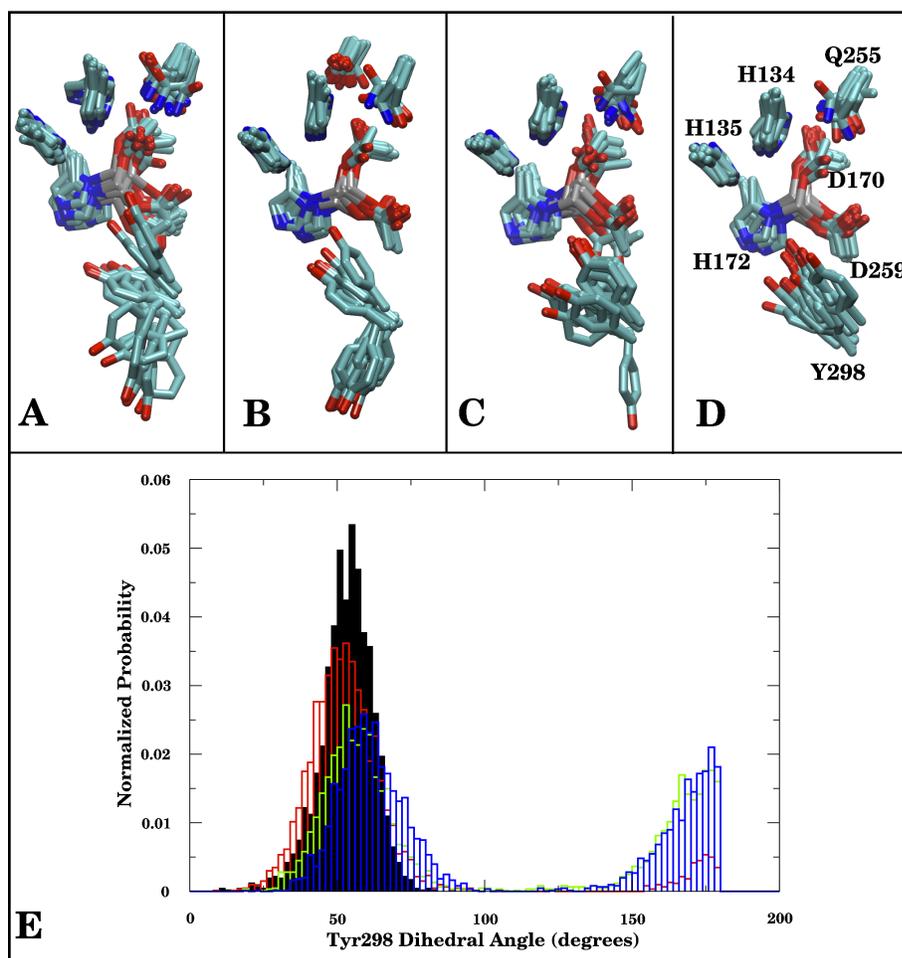


Figure 2.7: Superimposed snapshots from MD simulations show displacements of active site residues for (A) HDAC3, (B) HDAC3:IP4, (C) HDAC3:DAD, (D) HDAC3:DAD:IP4. (E) Tyr298 (N-C α -C β -C γ) dihedral angle probability distributions shown for HDAC3:DAD:IP4 (solid, black), HDAC3:DAD (red), HDAC3:IP4 (green), and HDAC3 (blue). (F) Cartoon representation of HDAC3:IP4:DAD complex, showing active site residues as sticks; the horizontal plane above the active site marks the top-down perspective of panels A-D.

however, e.g. Helix 6, or destabilized regions, such as the random coil (residues 340 to 348), in the ternary complex are not at all close to the DAD interface or the IP4 binding site, indicating allosteric effects of the co-repressors on HDAC3. Given that HDAC3 has no histone deacetylase activity in its flexible apo state, it would be tempting to suggest that the stabilized conformation is indeed the active conformation. This idea is reasonable since the observed conformational shifts, perhaps sufficiently described by the first two principal components, comprise motions that involve a contractile motion around the active site (see SI). The well-formed tunnel to the active site is distorted as a result of changes in the loop conformations. It is quite possible that this stable access route to the zinc center is important for the recognition of acetyllysines on histone tails. The observed shift in Tyr298 (which lines the tunnel) rotameric populations is also relevant to substrate recognition. In the crystal structure, Tyr298 was poised inward, hydrogen bonding with an acetate ion resolved in the crystal structure. The corresponding Tyr306 residue in HDAC8 has already been shown to be critical for both substrate recognition and in the reactivity of the deacetylase[28, 103, 1].

Making the connection from structure to activity If the stabilized conformation of HDAC3 can indeed be thought of as the active conformation, we must also acknowledge that both DAD and IP4 independently induce a population-shift towards this very structure, albeit to a lesser extent than in the presence of both co-repressors. It would seem not unreasonable then that HDAC3 have considerable deacetylase activity in the presence of just one or the other co-repressor. Experimental mutagenesis experiments, however, seem to suggest otherwise. Mutations disrupting the HDAC3:DAD complex also inactivate HDAC3[20], confirming that the binding to DAD is necessary for activity. Conversely, mutations of a DAD residue (K449A) near the IP4 binding site did not perturb the protein-protein association but still inactivated HDAC316, suggesting IP4 is also necessary for histone deacetylase activity. Recent inactivating mutations to HDAC3 Loop 1 and Loop 6 residues, that specifically interact with IP4 but not DAD, on the other hand, still disrupted complex formation with DAD; suggesting an interdependence between IP4 binding, DAD binding, and HDAC3 deacetylase activity[99]. The observation that HDAC3 is active in the presence of both co-repressors may suggest (i)

that the subtle change in dynamics from one co-repressor bound to both co-repressors bound is actually critical to deacetylase activity, and (ii) that the inability to bind one of the co-repressors, e.g. IP4, reduces the affinity for the second co-repressor, i.e. DAD, thereby increasing the equilibrium concentration of apo HDAC3. Future binding studies or free energy calculations would be needed to better evaluate the interdependence of DAD and IP4, but from our results it is clear that IP4 stabilizes the HDAC3:DAD complex as well as key regions (e.g. Loop 6) of the deacetylase backbone. It is also important to recall that DAD, in addition to stabilizing other regions (in particular Loop 1) of the deacetylase backbone, is critical in localizing HDAC3 near histones.

2.5 Conclusions

From a thorough analysis of multiple independent 100ns MD simulations, we discussed the conformational ensemble of HDAC3, both in the apo state, and bound to its two co-repressors necessary for activity: the DAD of the SMRT/NCOR nuclear complex and IP4. We showed that IP4 stabilizes the entire HDAC3:DAD protein-protein interface, well beyond its binding region. We further studied the independent effects of IP4 and DAD binding to HDAC3 and found that each is capable of stabilizing key regions of the deacetylase backbone. Together, IP4 and DAD restrain the conformation of HDAC3 to that found in the crystal structure, with a clear tunnel-like pathway to the zinc center. Based on a discussion of our results and available mutagenesis data, it is likely that activation of HDAC3 relies on a delicate interdependence between DAD and IP4, which only together can induce a sufficient population-shift and activate HDAC3.

2.6 Supporting Information

Clustering Details for apo HDAC3 structural ensemble. Clustering was done within Ptraj within the AmberTools package. Hierarchical, average linkage, and centroid linkage algorithms were tested. Since each algorithm requires the number of clusters to be chosen by the user, we also tested various cluster sizes (3, 4, 5) and the different algorithms on a subset of apo HDAC3 trajectories. Visual inspection of clusters as well

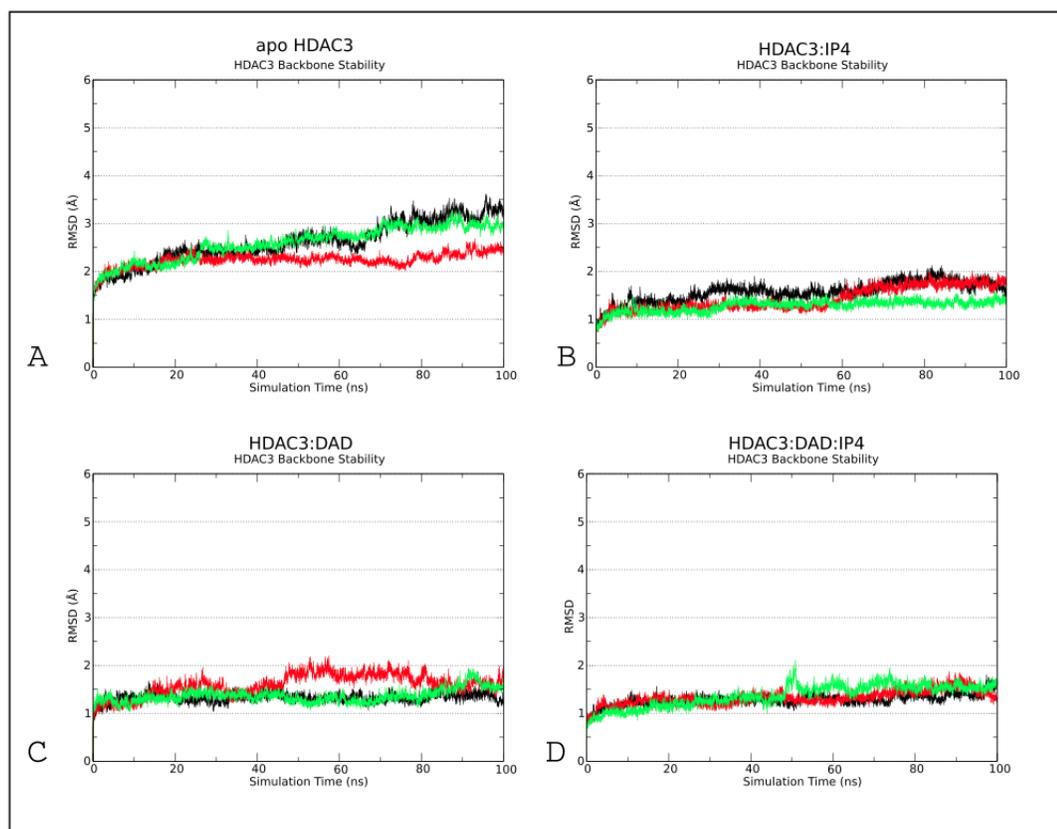


Figure 2.8: Root mean square deviations (\AA) from initial frames for MD simulations of (A) apo HDAC3, (B) HDAC3:IP4, (C) HDAC3:DAD, and (D) HDAC3:DAD:IP4. Initial frames were different snapshots from pre-production MD (once density of water box equilibrated). Three colors represent independent trials. For each trajectory, all protein backbone atoms were first aligned to its initial coordinates, and RMSD of HDAC3 backbone is plotted during the course of 100ns.

as consideration of three different clustering metrics (DBI, pSF, and SSR/SST) were used to choose the algorithm and number of clusters, as suggested in Ref [84].

Table 2.1: Details of clustering analysis and clustering metrics for apo HDAC3 structural ensemble.

Desired Clusters	4
Clustering Algorithm	Hierarchical
Distance Metric	RMS
Davies-Bouldin Index	2.82
Pseudo F-statistic	35311
Sum of squares regression / Total sum of squares	0.92

Cross comparison of structural ensembles

Using a RMSD cutoff of 1.4 Å between the apo structures and each of the cluster representatives, we were able to assign the same cluster occupancies for the apo trajectory as the clustering algorithm (+/- 5%). We then used the same cutoff to classify resemblance between structures from the different ensembles and the apo cluster representatives.

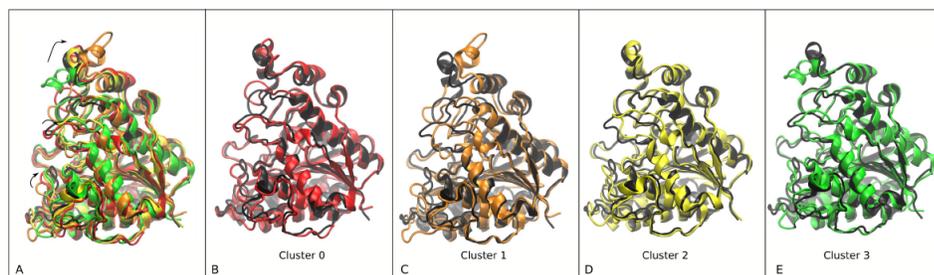


Figure 2.9: Representatives from clustering analysis of apo HDAC3 simulations. (A) All four cluster representatives are superimposed on the crystal structure of HDAC3. Arrows indicate distinguishing motifs among the clusters; (B) Cluster 0 representative in red; (C) Cluster 1 representative in orange; (D) Cluster 2 representative in yellow; (E) Cluster 3 representative in green.

Assignment of protonation states. Neutral histidine residues were simulated with either the N_ε (HIE: 16, 21, 32, 37, 55, 160, 172, 192, 233, 338, or N_δ (HID: 26, 61, 134) imidazole nitrogen protonated depending on local environments. PropKa was used to estimate the pKa, in addition to local environments in the crystal structure, in order to

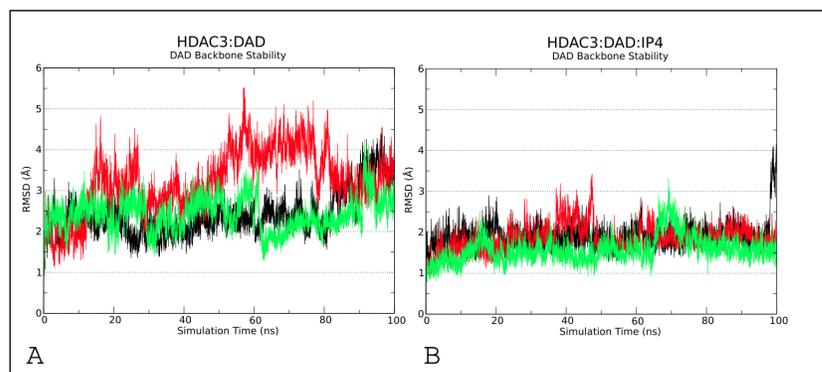


Figure 2.10: RMSD (\AA) of DAD backbone atoms during MD simulations of (A) HDAC3:DAD, and (B) HDAC3:DAD:IP4 complexes. For each trajectory the entire protein backbone was aligned to its initial coordinates and then the RMSD was measured for just the DAD backbone during the 100ns simulations. Different colors represent three independent simulations.

Table 2.2: Parameters for bonds to zinc center in active site. Atoms are labeled according to their AMBER atom type (O2: carboxylic acid oxygen on ASP; NB is imidazole nitrogen on HIS).

Bonds	$(k_r, \text{kcal/mol/\AA}^2)$	r_0 (\AA)
O2-Zn	120	1.92
NB-Zn	240	2.00

determine protonation states of residues. His 133 (active site) and 275 were simulated in their positively charged protonated (HIP) states.

2.7 Acknowledgements

We would like to thank Leonardo Boechi for thoughtful discussions and editing of the manuscript. This work was supported by NSF Graduate Research Fellowship, Molecular Biophysics Training Grant GM08326 (M.A.), the National Science Foundation Grant MCB-1020765, NBCR, CTBP, and Howard Hughes Medical Institute (J.A.M.), and National Institutes of Health Grant GM31749 (J.A.M.).

Chapter 2 is a near reprint of the material as it appears in Protein Science 2013, Volume 22, Pages 83-92. The dissertation author was the primary contributor and first author of this work.

Table 2.3: Parameters for angles involving zinc center in active site. Atoms are labeled according to their AMBER atom names.

Angles	(k_{θ} kcal/mol/rad ²)	θ (rad)
O2-Zn-NB	50	105.00
O2-Zn-O2	50	105.00
C-O2-Zn	80	125.00
CR-NB-Zn	80	119.00
CC-NB-Zn	80	131.00

Table 2.4: Charges assigned to atoms near Zn center.

Res Name	Atom	Charge	Res name	Atom	Charge
Zn	Zn	1.24			
ASP	N	-0.5163	HIS	N	-0.4157
	H	0.2936		H	0.2719
	CA	0.0381		CA	-0.0581
	HA	0.0880		HA	0.1360
	CB	-0.4267		CB	-0.4934
	HB2	0.0676		HB2	0.0961
	HB3	0.0676		HB3	0.0961
	CG	0.9542		CG	0.2789
	OD1	-0.7929		ND1	-0.3216
	OD2	-0.7929		CE1	-0.0216
	C	0.5366		HE1	0.2449
	O	-0.5819		NE2	-0.1949
				HE2	0.3406
				CD2	-0.3169
				HD2	0.2226
				C	0.5973
				O	-0.5679

Chapter 3

Inactivating mutation in Histone deacetylase 3 stabilizes its active conformation.

3.1 Introduction

Histone deacetylases (HDACs) were originally discovered and named for their role as keepers of the histone code [71]. Through deacetylation of acetyllysines located on the n-termini, or “tails”, of histone proteins, HDACs repress transcription of certain genes [25]. It is now known, however, that HDACs also modulate the activity of many other non-histone proteins, like p53 and NF κ B [71]. HDAC3 is a deacetylase that has quickly become a popular drug target as it not only promotes cell proliferation and metastasis in cancer cells, but is also implicated in several neurodegenerative diseases as well as in HIV-1 latency [102, 39, 35, 44].

Like nearly all other histone deacetylases, HDAC3 is only active upon recruitment to a multi-protein nuclear receptor co-repressor complex. Specifically, HDAC3 must be minimally bound to the deacetylase-activating domain (DAD) of either the silencing mediator for retinoid and thyroid hormone receptors (SMRT) or the homologous nuclear receptor co-repressor 1 (NCoR) complex[73, 64, 38].

Upon resolving the crystal structure of HDAC3, bound to DAD, Watson et al.

also discovered that a molecule of inositol tetraphosphate (IP4) rested at the protein-protein interface (Fig 3.1)[99]. Although it was established that IP4 is necessary for both formation of the HDAC3:DAD complex and for deacetylase activation, the underlying structural mechanism was not clear. Mutagenesis data has suggested that a delicate interdependence between IP4 and DAD is important in the activation of HDAC3 [20, 99].

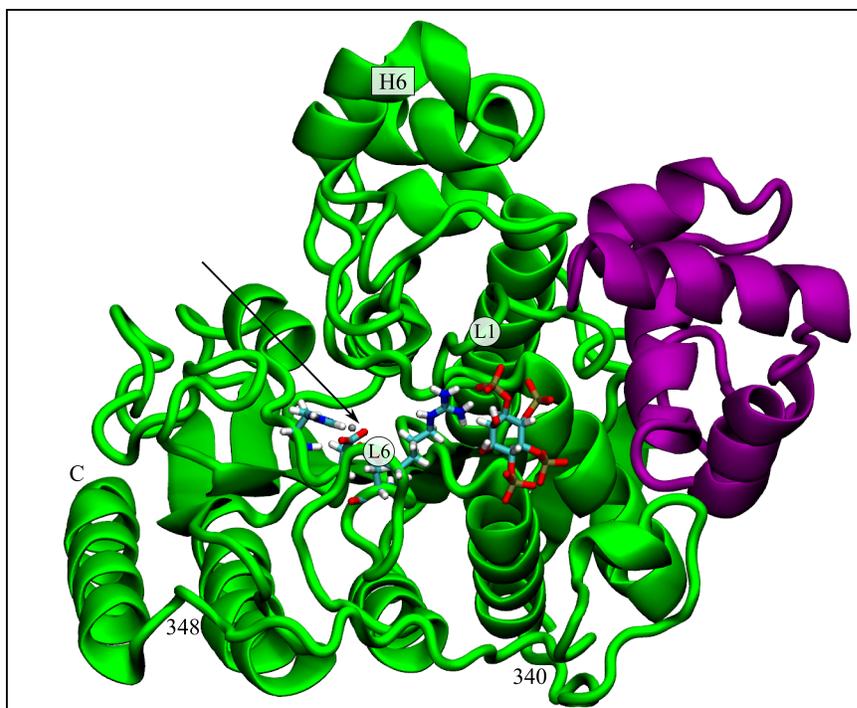


Figure 3.1: Cartoon representation of HDAC3 (green ribbon) bound to DAD (purple ribbon) and IP4 (sticks), from Protein Data Bank (PDB) identification code: 4A69. Arrow indicates the location of the active site. For reference, some loops and helices are identified, as well as the location of the R265 residue.

Recently, using molecular dynamics (MD) simulations, we showed that binding of either DAD or IP4 to HDAC3 results in a restriction of its conformational space, which is further exemplified once both are present in the active ternary (HDAC3:IP4:DAD) complex [5]. We proposed that this rigid conformation is important for the activation of HDAC3, likely by priming the deacetylase for substrate recognition.

Here, we report an HDAC3 mutant that mimics the rigid ternary complex

deacetylase conformation, without the need for DAD or IP4. A single arginine-to-proline substitution at position 265 (R265P) is capable of biasing the HDAC3 conformational ensemble towards its active state, identifying the 265 position as a critical hot spot in the allosteric activation of HDAC3. Intriguingly, despite its structural dynamics that suggest it to be constitutively active, the HDAC3^{R265P} mutant was actually shown to be *inactive* [99]. We discuss the implications of these results on our understanding of the general activation mechanism of HDAC3 by DAD and IP4.

3.2 Methods

System set-up and equilibration. The systems were built using the AMBER ff99SB force field, with the ildn modification for ILE, LEU, ASP, and ASN residues[65]. Parameters for both IP4 and the active site Zn center, as well as protonation states were taken from previous work[5]. Each system was solvated in a 10.0 Å TIP3P octahedron, and potassium ions were added to neutralize the charge in each case: HDAC (9), HDAC:IP4 (17), HDAC:IP4:DAD (13)[51]. Additional potassium and chloride ions were added to simulate physiological salt concentration (0.15 M). The updated Lennard Jones parameters for ions were used[52]. Systems were heated to 300K at constant volume (NVT) with restraints on the protein that were gradually reduced from 200 to 0 kcal/mol/Å² over a period of 150ps. The Langevin thermostat was used with a collision frequency of 1.0 ps⁻¹. The SHAKE algorithm was used to constrain bonds to non-polar hydrogens, and a 1.0 fs time-step was used during dynamics[82]. An 8.0 Å cutoff was used for non-bonded interactions. 100ps equilibration runs were done in the NPT ensemble, using isotropic pressure scaling and a pressure relaxation time of 2.0 ps.

Production MD simulations For each system, three independent simulations were started from different snapshots from the equilibration simulations (taken at arbitrary intervals after density of water box was equilibrated) with randomized velocities. All production runs were done in the same NVT conditions described above, with the exception of a 2.0 fs time-step. All simulations were done using the PMEMD module within the Amber11 simulation package[14]. Production runs were performed on GPUs using Amber11 pmemd.CUDA, each for 100ns.

Trajectory analysis. We used the ptraj module from the AmberTools package to perform RMSD, RMSF, and PC analyses. In all analyses the protein backbone atoms were aligned to the coordinates in the crystal structure (PDB ID 4A69). For the PC analysis, we build the PC1:PC2 space using the first two principal components (over 50% of the total variance) of the three 100ns apo HDAC3^{WT} trajectories, using C_α atoms only. C_α atoms from all other trajectories were first aligned to the crystal structure before projecting into the PC space. For the analysis of the active site channel geometry, we used 100 snapshots taken from equally spaced intervals from the three independent 100ns simulations of each state. We used the Caver algorithm to find tunnels in the protein, using starting points in the active site (residues 132, 255, and 145), a probe radius of 0.9 Å, and a shell radius of 6.0 Å[19]. Clustering of tunnels was done using the average linkage algorithm, with a threshold of 4.

3.3 Results

In our previous work, we showed that the conformational ensemble of apo HDAC3^{WT} is restricted upon binding to DAD and IP4 [5]. Here we use several structural metrics, outlined below, to compare the conformational flexibility of the inactive HDAC3^{R265P} mutant to that of the free and complexed HDAC3^{WT} states, based on triplicate 100-ns molecular dynamics (MD) simulations for each state.

3.3.1 Global stability of apo HDAC3^{R265P} resembles that of activated HDAC3^{WT}

Using principal component (PC) analysis, we characterized how the HDAC3 dynamics change according to its two most dominant low-frequency modes. By projecting the HDAC3:IP4, HDAC3:DAD, and HDAC3:IP4:DAD trajectories onto a PC space built from the apo HDAC3^{WT} simulations, we previously showed a constriction of the deacetylase ensemble upon binding to DAD and IP4 (Fig 3.2, panels A-D)[5]. We note that binding of either DAD or IP4 results in comparable restriction of dynamics along these modes and the deacetylase backbone is further restricted in the active ternary complex.

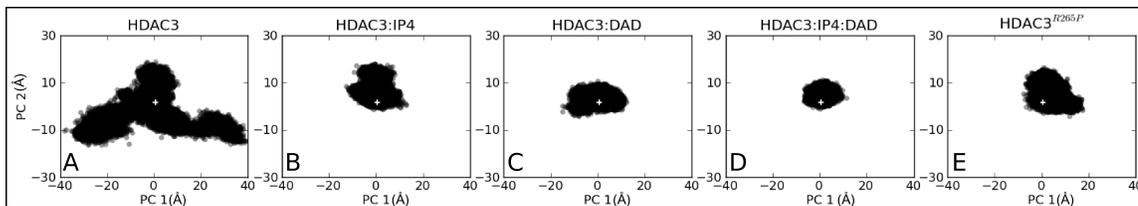


Figure 3.2: Principal component analysis of HDAC3. (A) apo HDAC3, (B) IP4-bound HDAC3, (C) DAD-bound HDAC3, (D) IP4- and DAD-bound HDAC3, and (E) apo HDAC3^{R265P} mutant dynamics are each projected onto PC1:PC2 space of the apo HDAC3^{WT}. A white cross in each panel indicates the projection of the HDAC3:IP4:DAD crystal structure (PDB ID: 4A69).

The projection of the mutant HDAC3^{R265P} trajectories onto the WT apo PC space demonstrate a bias in its conformational ensemble towards the active state (3.2, panel E). We also note that the mutant deacetylase, in its apo form, maintained a surprisingly low RMSD (≈ 1.5 Å) to the crystal structure of the WT deacetylase in the HDAC3:IP4:DAD complex.

PC1 and PC2 (movies available in Ref[5]) together describe a contractile motion around the active site. PC2 is dominated by the dynamics of Loop 1, whereas PC1 primarily results from concerted motions of Helix 6, the C-terminal helix, and Loop 6 (Fig 3.1). The mutated residue, R265, resides on Loop 6, where it interacts with either the solvent or IP4, if present. Nearby K25 on Loop 1 also interacts with IP4, and causes a charge-charge repulsion with R265 in the absence of IP4. The proline at this position could trivially be expected to stabilize Loop 6, both by removing one of the positive charges from this basic region, and by the rigid nature of the proline residue itself. The dramatic constriction of the conformational ensemble observed for apo HDAC3^{R265P}, however, involves other regions quite far from the mutation site, such as the C-terminal helix and Helix 6, indicating several wide-spread allosteric networks in HDAC3 that can be triggered at position 265.

To further dissect the effect of the inactivating R265P mutation on the deacetylase conformational plasticity, we also analyzed the change in the per-residue root mean squared fluctuations (Δ RMSF) for the mutant and the WT complexed states, relative to the apo HDAC3^{WT} fluctuations. In this way, negative Δ RMSF values indicate residues that are stabilized in the mutant or complexed states, relative to HDAC3^{WT} in its apo

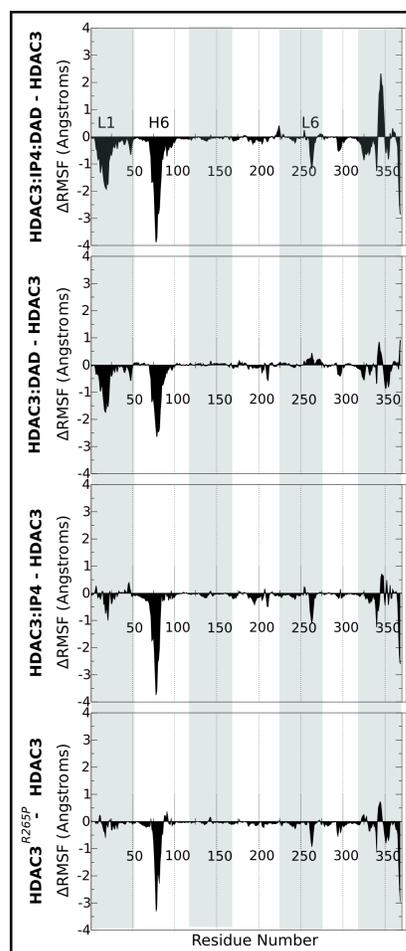


Figure 3.3: $\Delta RMSF$ Analysis of mutant and complexed HDAC3. Differences between per-residue RMSF values (\AA) are shown for (from top to bottom) HDAC3:IP4:DAD, HDAC3:DAD, HDAC3:IP4, and HDAC3^{R265P}, with reference to the apo HDAC3^{WT} RMSF values. Negative $\Delta RMSF$ values correspond to stabilized regions.

form.

First, we note that the R265P mutation results in the stabilization of the same WT deacetylase regions that are stabilized upon complex formation with either DAD, IP4, or both (Fig 3.3). The random coil region at positions 340 to 348 (Fig 3.1) is the only exception to this trend, as it is only destabilized in the ternary complex, but not by DAD, IP4 or the R265P mutation alone. We also identified a striking similarity between the HDAC3^{R265P} and the HDAC3:IP4 $\Delta RMSF$ plots (Fig 3.3). For both the mutant and the IP4-bound, but not DAD-bound deacetylase, we observed negative peaks at the C-terminal helix, which is believed to include the nuclear localization signal (NLS) of

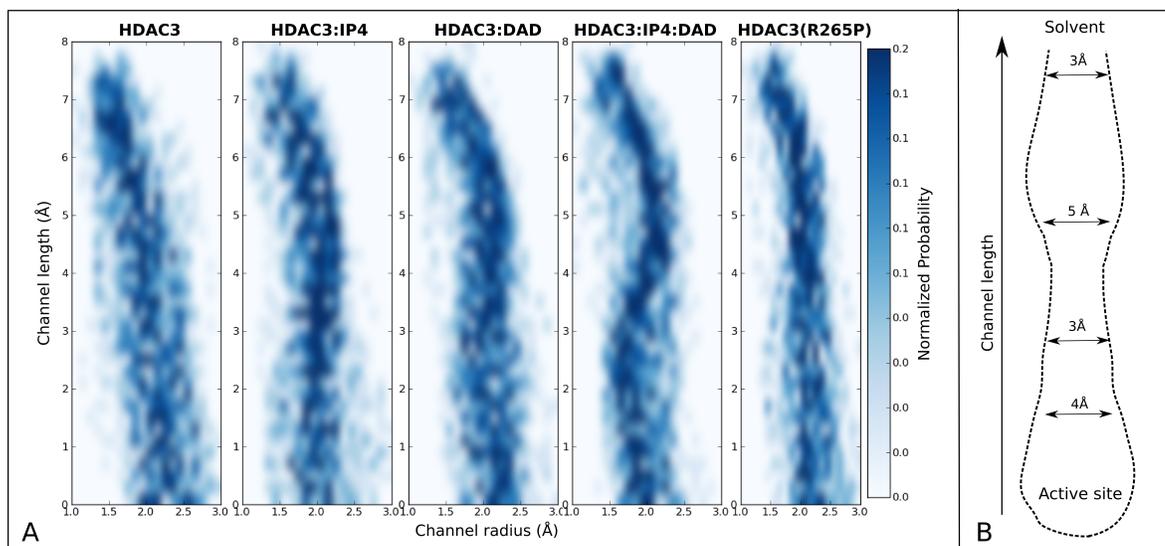


Figure 3.4: Geometry of active site channel. (A) Two-dimensional histograms showing the distribution of radii along the channel length for the apo, IP4-bound, DAD-bound, and ternary WT simulations, as well as for the apo R265P mutant. (B) The stable channel leading to the active site, from simulations of the HDAC3:IP4:DAD complex is shown schematically, with average diameters indicated along the channel length.

HDAC3 [104]. Other motifs that are stabilized to a greater extent by DAD, e.g. Loop 1, were affected to a lesser extent by either the binding of IP4 or by the mutation. The replacement of just one of the three key IP4-interacting residues appears to result in the same curious Δ RMSF signature as seen with IP4 binding.

3.3.2 Geometry of apo HDAC3^{R265P} active site channel resembles that of activated HDAC3^{WT}

Most of the key stabilized regions in HDAC3^{WT} form a channel (≈ 10 Å) to the active site. By analyzing the radii along the length of the channel during the MD simulations, we observed that the channel in the active ternary complex maintains its geometry, whereas the apo HDAC3^{WT} channel has large fluctuations in its geometry (3.4). The apo HDAC3^{R265P} data also indicate a very restrained active site channel geometry, although not identical to that of the ternary complex deacetylase.

The form of this active site channel is likely an important part of the substrate recognition mechanism of HDAC3. An analysis of the key channel-lining residues high-

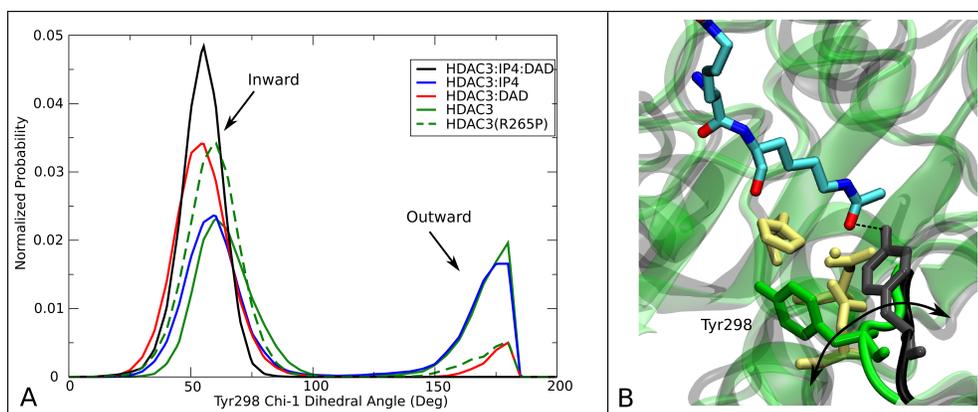


Figure 3.5: Population analysis of Tyr298 χ_1 dihedral angle. (A) Normalized probability of the χ_1 side chain dihedral angle shows two states (inward and outward) sampled in the HDAC3 simulations. (B) Cartoon representation of HDAC3 from the HDAC3:IP4:DAD (black) and apo (green) simulations are superimposed to highlight the two Tyr298 conformations. An arbitrary acetyllysine peptide substrate was docked into the active site to show the functional relevance of the inward Tyr298 conformation (hydrogen atoms are not shown).

lights a cluster of hydrophobic residues at the opening of the channel, followed by more polar residues closer to the active site. This distribution of channel-lining residues is indicative of the current pharmacophore of HDAC inhibitors, as well as of the acetyllysine binding mode [67].

One of the key features of the active site channel is the conserved residue Tyr298, which serves to anchor the acetyllysine substrate (Fig 3.5). From the multiple MD simulations, we found that Tyr298 samples two χ_1 dihedral states, only one of which permits the catalytic interaction with the carbonyl oxygen of the acetyl group (Fig 3.5). For HDAC3^{WT}, the binding of both IP4 and DAD shifts the rotameric population exclusively towards the catalytically relevant conformation, inward, whereas the apo HDAC3^{WT} ensemble reflects nearly equal probability of inward and outward Tyr298 conformations. The R265P mutation, appears to also bias the Tyr298 conformation towards the catalytically relevant state, without the presence of DAD or IP4.

3.4 Discussion

We identified a mutation, R265P, on one of the active site loops of HDAC3, that is capable of reproducing many of the long-range stabilizing effects of IP4 or DAD binding. These results confirm the presence of an allosteric network in HDAC3 that can be triggered at position 265, either by small molecule binding, or as shown here through the R265P replacement. Fluctuations in Loop 6 (Fig 3.1) are directly linked to large fluctuations in Helix 6 and the C-terminal helix, the latter of which contains the NLS of HDAC3 [104]. These fluctuations appear to be transmitted through the protein backbone via the random coil region between residues 340 and 348, as well as the other nearby active site loops, resulting in an overall stabilized backbone upon stabilization of a single amino acid side chain. Even though its conformational ensemble suggests the HDAC3^{R265P} mutant may be constitutively active, however, it was instead shown to result in a loss of all WT deacetylase activity[99].

Activation of HDAC3 may require further stabilization. Although the structural dynamics of the HDAC3^{R265P} mutant are biased towards the deacetylase conformation in the active ternary complex, our results do show that the mutant ensemble more closely resembles either the HDAC3:DAD or HDAC3:IP4 intermediate complexes, neither of which has been shown to be active, to our knowledge. It is possible, then that the deacetylase activity requires additional "fine tuning" of the conformational ensemble. In each of our structural analyses, we observed a synergistic restricting effect on the deacetylase dynamics when both IP4 and DAD were bound, which was not observed in the simulations of the apo HDAC3^{R265P} mutant. The R265P mutation has been shown to preclude complex formation with DAD [99], which may be required to further refine the conformational ensemble towards the active state. We should note, however, that we have also performed simulations of the mutant bound to both IP4 and DAD (not shown here), but did not observe any further stabilization of the deacetylase backbone than was observed in the apo HDAC3^{R265P} case. The HDAC3^{R265P} ensemble, although biased towards the active conformation, may still be lacking necessary, yet subtle, structural rearrangements to be activated.

It is also possible that the physical presence of IP4 and DAD themselves contributes to substrate recognition or stabilization, in addition to their synergistic stabiliz-

ing effects of the deacetylase backbone. In this case, the loss of affinity for DAD and IP4 would be enough to result in a loss of deacetylase activity associated with the R265P mutation. We think this is unlikely, however, because even though DAD is necessary for histone recognition of HDAC3 [105], the HDAC activity assay used to study the mutant utilizes a peptide substrate that can be deacetylated by a number of HDACs, which do not specifically require DAD[99].

R265P mutation may alter substrate specificity It seems, then, that due to rather subtle details in its dynamics, the HDAC3^{R265P} mutant is unable to deacetylate the same substrate that the WT can deacetylate when bound to DAD and IP4. Because of its stability and close resemblance to the WT deacetylase in the ternary complex, however, it is plausible that this mutant may still be active towards some other acetyllsine substrates. In the future, alternative assays with multiple peptide substrates may be used to detect possible changes in deacetylase target specificity as a result of this mutation.

Implications on HDAC3 biology. A shifted substrate specificity for the HDAC3^{R265P} mutant would be of particular biological relevance as this mutation can occur as a result of a known single nucleotide polymorphism (rs467744), not yet associated with any known phenotypes, to our knowledge. Moreover, if this intermediately restrained conformational ensemble of HDAC3 is associated with an altered substrate specificity, this would have profound implications on the role of IP4 in modulating HDAC3^{WT} activity, since the HDAC3:IP4 complex has a comparably restrained conformational ensemble. Because HDAC3 is known to localize to both the nucleus and cytoplasm, and has been shown to modulate cellular processes even when unable to bind to DAD [105], it is possible that IP4 may serve to activate cytoplasmic HDAC3 towards other non-histone targets, and even promote nuclear import of HDAC3 through stabilization of its NLS. The HDAC3^{R265P} mutant, on the other hand, may be unaffected by the inositol signaling pathway and have aberrant sub-cellular localization.

Evolutionary origins of a constitutively active HDAC. The R265 residue is conserved in nearly all of the class I HDACs, suggesting perhaps the same allosteric network is conserved within the catalytic HDAC core [37, 27]. The exception, however, is HDAC8, which happens to also be the only deacetylase in its class that is constitutively active [93]. Intriguingly, in HDAC8, the conserved arginine at position 265 is instead a

proline. Our results on the R265P mutant provide insight into the molecular mechanism through which this constitutive activity may be conferred to HDAC8. Future assays of the HDAC3^{R265P} mutant specificity, among that of other HDAC8-like mutants, would be necessary in order to address this exciting idea.

3.5 Acknowledgements

This work was supported by NSF Graduate Research Fellowship, Molecular Biophysics Training Grant GM08326 (M.A.), the National Science Foundation Grant MCB-1020765, NBCR, CTBP, and Howard Hughes Medical Institute (J.A.M.), and National Institutes of Health Grant GM31749 (J.A.M.).

Chapter 3 is a minimally modified version of a manuscript in press at Protein Science 2013 (DOI: 10.1002/pro.2317). The dissertation author was the primary contributor and first author of this work.

Chapter 4

Structure-based virtual screen for novel inhibitors of histone deacetylase

3.

4.1 Introduction

Histone deacetylases (HDACs) are a family of enzymes that are implicated in a growing number biochemical pathways[25]. Through the deacetylation of acetyllysines on the flexible termini of histone proteins, HDACs tend to repress gene transcription by compacting local chromatin structure[10]. Acetylation, however, is a more general post-translational modification that regulates many other processes involving non-histone proteins as well, for example the assembly and disassembly of tubulin, widening the protein targets of HDACs[34, 47].

HDAC inhibitors are currently used as part of certain cancer treatments like Tcell lymphoma, and are effective treatments for neurodegenerative diseases like Friedrich's Ataxia[35, 102, 71]. Here we use a computational approach to identify novel inhibitors of a class I deacetylase, HDAC3. The family of HDACs has been divided into four classes, based on sequence homology, and inhibition of different HDAC classes can result in a range of physiological effects[25, 39]. The class I deacetylases (HDAC 1,2,3, and 8), for instance, are all involved in cancer cell proliferation and metastasis[102].

Selective inhibition of a deacetylase within a class is ideal in more targeted therapies; inhibition of HDAC3 is sufficient to prompt cells infected with HIV-1 out of latency[44].

In general, the class I HDACs require recruitment to large multi-protein nuclear complexes in order to be activated and deacetylate their substrates[73, 64]. Structural data for these multi-protein active HDAC complexes is sparse, limiting the progress of structure-based drug discovery efforts. Most structure-based HDAC drug discovery efforts rely on the structures of the only constitutively active deacetylase, HDAC8, or of the homologous HDAC-like protein[98]. Because HDAC8 shares the same catalytic core and is at least 30% homologous to the other class I deacetylases, the many crystal structures, both free and inhibitor-bound, of HDAC8 have also been extended to the other class I deacetylases in the literature[96, 92]. Despite their conserved catalytic cores and high homology, however, the dependence of the other deacetylases on co-repressor and accessory proteins suggests distinct activation, and perhaps novel inhibition, mechanisms.

Recently HDAC3 was crystallized in its active form, bound to a portion of the deacetylase activating domain (DAD) of a nuclear receptor co-repressor, and the necessary small molecule inositol tetrakisphosphate (IP4) bound at the protein-protein interface[99]. By deconstructing the active ternary (HDAC3:IP4:DAD) crystal structure and using molecular dynamics (MD) simulations, in previous work we have shown that apo HDAC3 has intrinsically favorable regions that are stabilized in the HDAC3:IP4 and HDAC3:DAD, and ternary complexes[5]. Because many of these stabilized regions also form a channel leading to the catalytic zinc center in the active site (Fig 4.1), we have proposed that the stabilization of this HDAC3 conformation enhances substrate binding, and thus the activity of the deacetylase. Here, we show that the restrained conformational ensembles of DAD-bound or IP4-bound HDAC3 can also be used as predictive receptor models for drug discovery, and we use these receptor models to identify potential HDAC3 inhibitors from an enriched library of drug-like compounds.

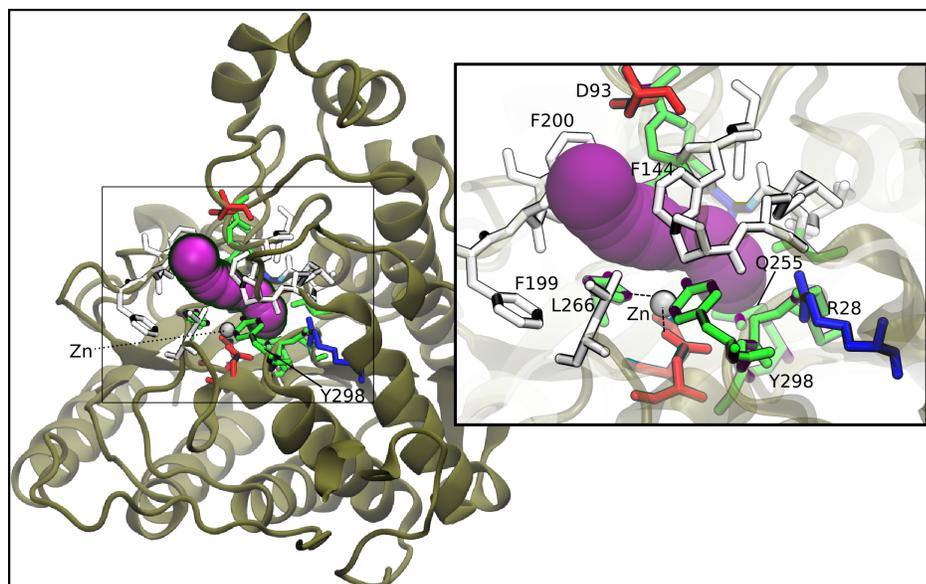


Figure 4.1: Active site channel stabilized by DAD and IP4. One snapshot of HDAC3 is shown (tan ribbons), with the average channel from the HDAC3:IP4:DAD simulations shown in a purple space-filling representation. Residues that line the channel during most of the simulations are shown and colored according to residue type (white: non-polar, green: polar, blue: basic, red: acidic).

4.2 Methods

Compound libraries. Using the ZINC website (www.zinc.docking.org/search/combo), drug-like compounds with at least 70% similarity to the 21 known inhibitors were selected from the ChemBridge and NCI Diversity Set libraries, resulting in a total of 1,384 compounds, downloaded on 2013-04-29[48]. The decoy molecules used were the set of 1000 molecules, with an average molecular weight of 360 g/mol, made available by Schrödinger[32, 40].

Ligand preparation. All compounds were prepared using the LigPrep application in Maestro 9.2 (Schrödinger, LLC), using the OPLS: 2005 force field. Ionization states were generated for a pH of 7 ± 2 , with metal binding states included using Epik. Tautomers were generated for each ligand but chiral centers were not varied.

Receptor preparation. Each receptor model, taken from previous MD simulations[5], was prepared using the Protein Preparation workflow in Maestro 9.2 (Schrödinger, LLC) [83]. Receptor grid was generated using the Glide program, and

the active site was identified as the centroid of residues 134, 254, 297, and 91 to include ligands with length less than or equal to 16 Å. A metal-bond constraint to the zinc center was added as well. Prepped ligands were docked rigidly using Glide single precision, with Epik state penalties for docking to the metal center.

Enrichment analysis. Compounds were ranked according to their Glide scores, and for each ranking the 1% enrichment factor and Robust Initial Enrichment (RIE) metrics were computed using the Enrichment calculator scripts in the Maestro suite. The null distributions for the RIE metric were computed using a Python script that generates 50,000 random rankings of 21 items. RIE values were calculated using Eq. 1, with $\alpha = 20$.

Solvent mapping. The FTMap web-server was used to obtain probe occupancies on the surface of HDAC3[60]. A total of twelve HDAC3 structures were taken from HDAC3 and HDAC3:IP4 MD simulations. Consensus sites were identified using FTProd[95], a plug-in for Visual Molecular Dynamics (VMD) [46].

4.3 Results and Discussion

Library enrichment. To identify potential HDAC3 active site inhibitors, we first constructed an enriched library of compounds. By filtering both the National Cancer Institute (NCI) Diversity set III and the ChemBridge library for drug-like molecules with at least 70% structural identity to at least one of 21 known inhibitors, 3358 compounds were selected for virtual screening.

Receptor refinement. In previous work, we have used molecular dynamics (MD) simulations to study the different conformational ensembles for the apo HDAC3, as well as the IP4- and DAD-bound states. Notably, we had found that the IP4- and DAD-bound HDAC3 ensembles were conformationally restrained in comparison to the apo HDAC3 ensemble[5]. For the purpose of identifying novel inhibitors, we aimed to first distill the MD-generated HDAC3 ensembles into a highly predictive subset of receptor models. Unlike other ensemble-based virtual screens, this approach does not consider the Boltzmann probabilities of the receptor conformations, but still incorporates conformational flexibility of the receptor in the screen. We first generated a naïve

receptor model subset by randomly selecting 30 deacetylase conformations from each of the HDAC3, HDAC3:IP4, HDAC3:DAD, and HDAC3:IP4:DAD ensembles, for a total of 120 naïve receptor models, in addition to the crystal structure itself (PDB id: 4a69). We anticipated that the restrained conformational ensembles sampled in the IP4- and DAD-bound HDAC3 simulations would serve as predictive receptor models, as they retained an active site channel geometry close to that of the crystal structure, seemingly poised for ligand binding (Fig 4.1). The complete virtual screen workflow is described in detail in Methods and summarized in Fig 4.2.

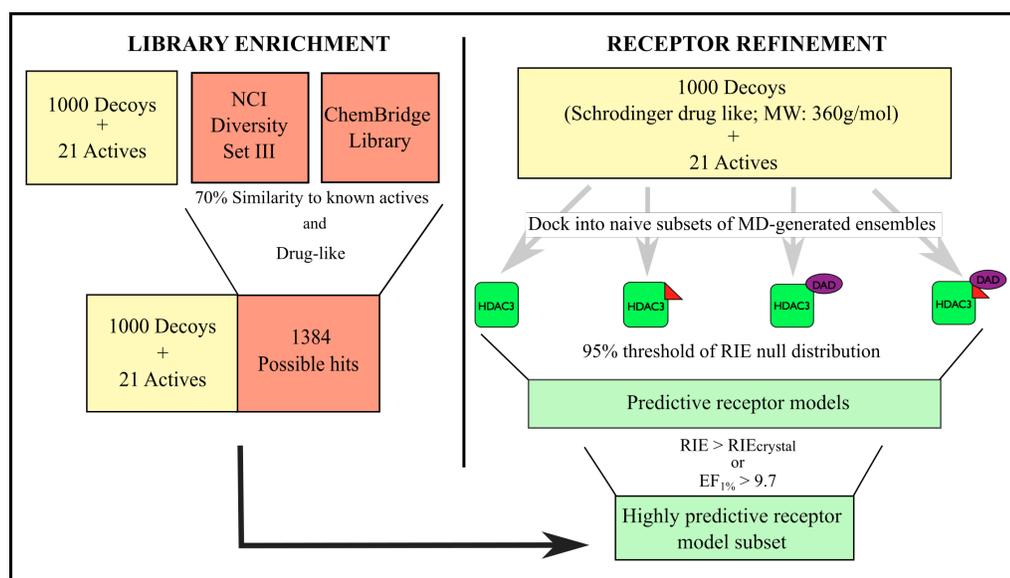


Figure 4.2: Summary of virtual screen workflow.

To evaluate the predictive quality of the naïve receptor model subsets, we seeded the 21 known actives among 1000 drug-like decoy molecules and docked the 1021 molecules into each receptor and computed the Robust Initial Enhancement (RIE) metric for the generated ranking:

$$RIE = \frac{\frac{1}{n} \sum_{i=1}^n \exp(-\alpha \chi_i)}{\frac{1}{N} \frac{1 - \exp(-\alpha)}{\exp(\alpha/N - 1)}} \quad (4.1)$$

In Eq 1, n is the number of actives, out of a total N compounds screened; χ_i is the relative rank of the i^{th} active, and α is a tuning parameter that, at higher values, makes the RIE score a more stringent metric for early enrichment (here $\alpha = 20$, as suggested in

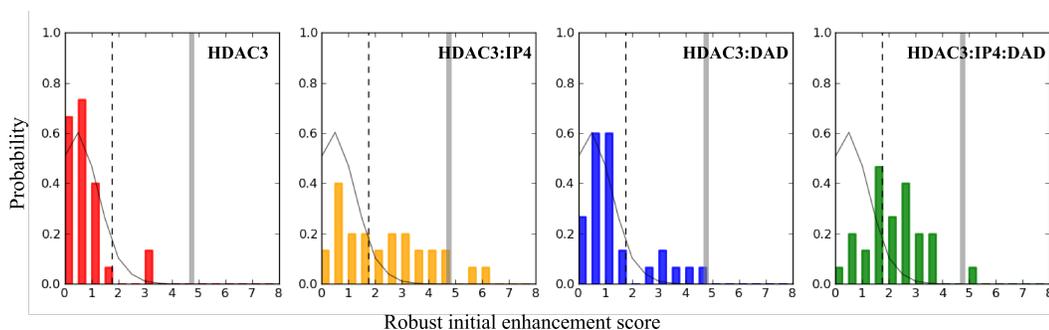


Figure 4.3: Distribution of Robust Initial Enhancement (RIE) metrics for naïve receptor model subsets from the apo HDAC3 (red), HDAC3:IP4 (yellow), HDAC3:DAD (blue), HDAC3:IP4:DAD (green) ensembles. Solid black curve in each panel shows the null distribution of the RIE metric, with the dashed vertical line marking the 95% threshold. Solid grey line indicates the RIE score for the crystal structure (PDB id: 4A69).

literature). We chose to consider the RIE metric here because, unlike the routinely used area under the receptor operating curve (AUC) metric, the RIE is specifically constructed to evaluate early recognition of the active compounds. We empirically generated the null distribution of the RIE metric for the specific case of 1000 decoys and 21 active compounds, and used the 95% threshold of the null distribution as the RIE cutoff to identify statistically predictive receptor models. After evaluating the 30 naïve receptor models from each of the four ensembles as well as the crystal structure conformation, we found that the crystal structure was among the most predictive receptor models. Fig 4.3 shows the distribution of RIE scores obtained using the apo HDAC3, HDAC3:IP4, HDAC3:DAD, and HDAC3:IP4:DAD naïve receptor subsets. In each panel, the crystal structure RIE score is identified as well as the 95% cutoff for the null RIE distribution. More predictive receptor models were indeed selected among the IP4- and DAD-bound HDAC3 trajectories than the conformationally flexible apo HDAC3 ensemble.

Highly predictive subset of receptor models. Because the crystal structure itself achieved a high RIE score, we decided to further refine our receptor models by selecting only receptors that ranked the actives such that their RIE score was greater than or equal to the crystal structure RIE score. In this highly predictive receptor model subset, we also permitted receptors with statistically high RIE scores (but not necessarily higher than that of the crystal structure) that had particularly high 1% enrichment factors (≈ 10), resulting in a subset of 12 receptor models to be used in the virtual screen.

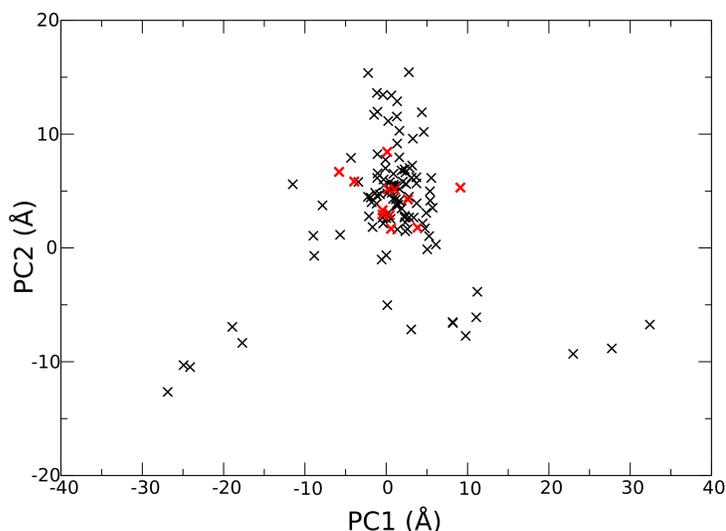


Figure 4.4: Projection of naïve (black) and highly predictive (red) receptor models onto the principal component space of the apo HDAC3 trajectory.

To confirm whether these highly predictive receptor models do indeed correspond to the restrained “active” conformational ensemble observed in the DAD- and IP4-bound HDAC3 simulations, we projected all of the receptor models onto the two-dimensional principal component (PC) space defined by the two dominant low frequency modes of the apo HDAC3 trajectory (Fig 4.4). These two PCs have been described in previous work[5, 4], and comprise a contractile motion about the active site, with the crystal structure and the active conformational ensemble localized near the origin of the plot. We found that the highly predictive receptor models, highlighted in red in Fig 4.4, do cluster near the origin and are representative of the PC subspace identified in the IP4- and DAD-bound ensembles in our previous work. It is reasonable to conclude that the restrained conformational ensemble of HDAC3, observed in the presence of IP4 and DAD, is primed for substrate or inhibitor binding.

Virtual Screen. Having chosen the highly predictive receptor models, we concatenated (i) the enriched library of compounds, (ii) the 21 known actives, and (iii) the decoy set. We then docked the collection of molecules into each of the highly predictive receptors. In this way, we could seed our virtual screen with both positive (known actives) and negative (drug-like decoys) controls. Because the library of compounds was constructed to have at least 70% structural identity with at least one known active,

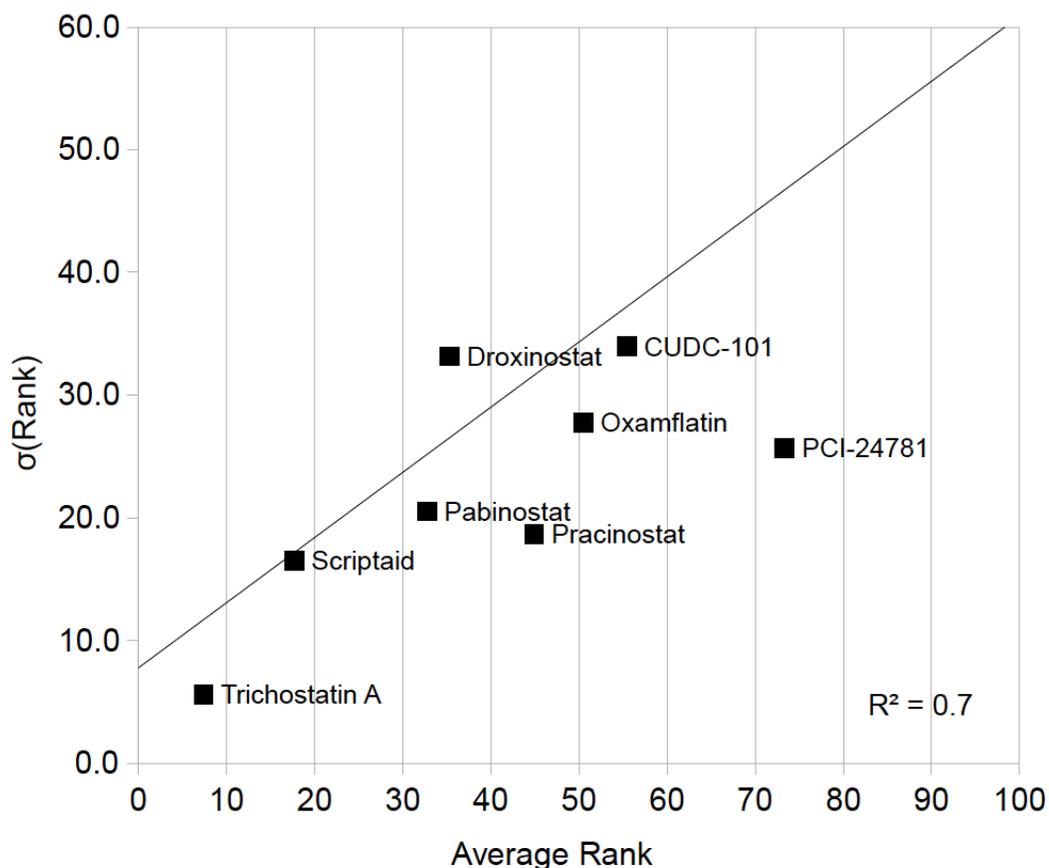


Figure 4.5: Reliability of positive controls. The average rank and standard deviation of rank are shown for the known inhibitors across the twelve highly predictive receptor models. The variance of an inhibitor's rank across receptor model increases linearly with its rank.

the compounds that consistently outrank the known actives can be considered viable candidates as HDAC3 inhibitors.

The top-ranked actives were consistently highly ranked, and the active compounds that were not in the top 100 generally had widely varying ranks among the highly predictive receptor models. Fig 4.5 shows the linear correlation ($R^2 = 0.7$) between average ranks and their variances across receptor models. The top-ranked compound, Trichostatin A, which is not necessarily the most potent inhibitor according to inhibition assays, is in any case the most reliable positive control in our virtual screen. In the second stage of this virtual screen protocol we identified those compounds that consistently out-ranked the most reliable positive control, Trichostatin A. Fig 4.6 shows

the average ranks and standard deviations for the active compounds when seeded with the decoys and after being further combined with the enriched library of compounds. The molecules that were found to outrank Trichostatin A in multiple receptor models are also shown in the right panel of Fig 4.6.

Top-ranking compounds. Although the most common zinc chelating group among the known inhibitors is the hydroxamic acid moiety, only one of the top-ranking compounds (**4**) was indeed a hydroxamic acid. The majority of the top-ranking compounds were chemically similar butyric (**1a-d**) or pentanoic acid (**2a,b** and **3**) derivatives. Notably, one of the top-ranking compounds (**5**) is close analogue to one of the known inhibitors, Scriptaid, with a carboxylic acid replacing the hydroxamic acid moiety. All of the top-ranking compounds bind with zinc chelating groups, but differ in their hydrophobic interactions near the channel entrance.

Alternative druggable sites. We have shown in previous work that the activation of HDAC3, via complex formation with DAD or IP4, involves an impressive restriction of the deacetylase conformational flexibility[5]. In particular, we have further confirmed, through an *in silico* mutational study, that the IP4 binding site is allosterically linked to both the C-terminal helix and Helix 6[4]. The design of allosteric inhibitors of HDAC3 may benefit from disrupting these allosteric pathways and destabilizing the formation of the active ternary complex. To explore this aim, we performed solvent mapping using FTMap, which docks various organic probes along the surface of a protein structure to identify favorable binding sites, and perhaps druggable hot spots. By selecting snapshots from the apo HDAC3 and HDAC3:IP4 conformational ensembles, we identified three consensus druggable sites, present in both ensembles. Because IP4 carries a strong negative charge, any small molecules that would inhibit HDAC3 by competing with IP4 would have to be highly polar, and would not likely serve as suitable pharmacological candidates, so we did not consider the HDAC3:DAD ensemble in the search for druggable sites.

The three consensus sites, shown in Fig 4.7, are (1) near Helix 6, (2) the C-terminal helix, and (3) adjacent to the IP4-binding site, at the DAD-binding interface. Of the three alternative sites identified, Site 1 is most hydrophobic, and farthest from the IP4- and DAD-binding interface. Site 2 has the deepest pocket of the three sites,

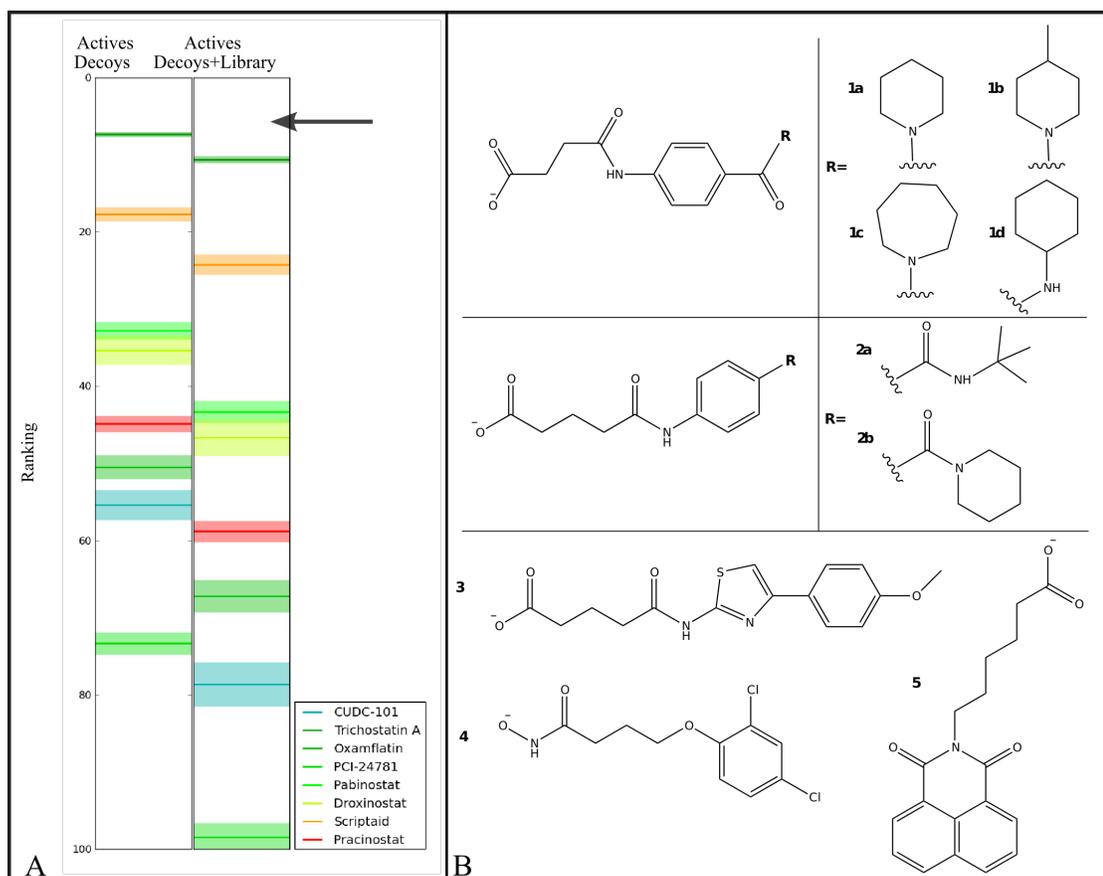


Figure 4.6: Virtual screen results. (A) Average ranks (dark bands, fixed width), along with their standard deviations (light bands, varied widths) are shown for the active compounds that were ranked among the top 100 in virtual screens. The left column shows rankings when actives were mixed with decoy molecules, and the right panel shows the rankings when actives were mixed with both decoys and possible inhibitors. Black arrow indicates region from which potential inhibitors were chosen. (B) Chemical structures of the potential inhibitors that tend to out-rank the highest ranking active compound are shown.

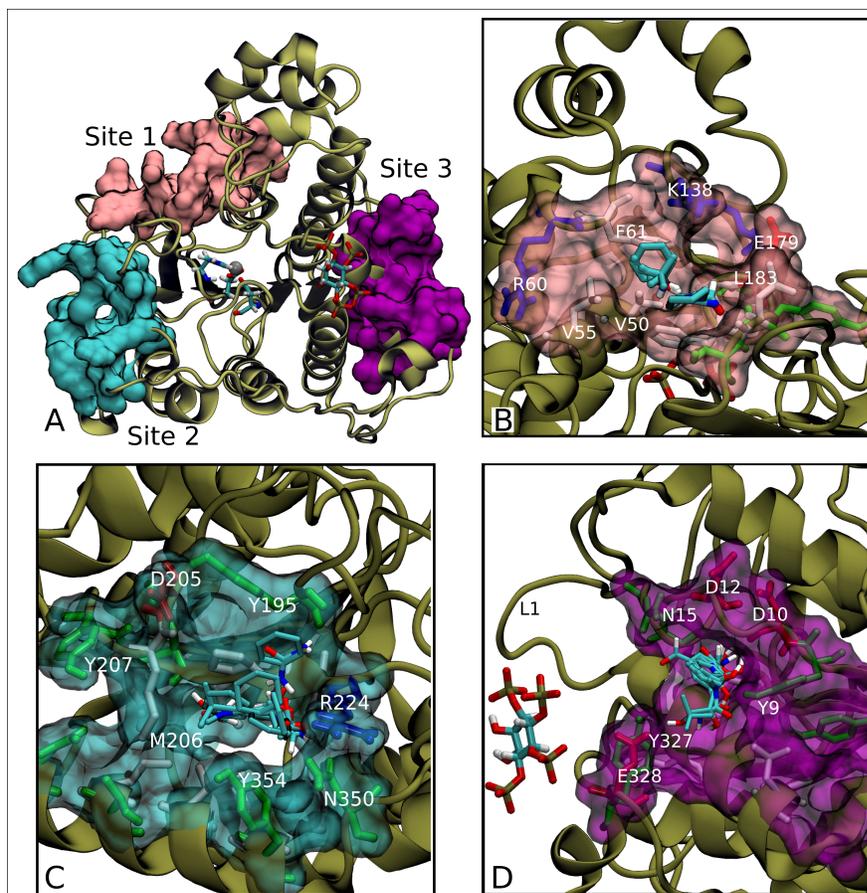


Figure 4.7: Alternative druggable sites identified by solvent mapping. Site 1 is adjacent to Helix 6, Site 2 is located at the C-terminal helix, and Site 3 is adjacent to the IP4 binding site, at the DAD-binding interface. Panels B-D show closer views of each binding site, with surrounding residues shown as sticks, colored by residue type (green: polar; red: acidic; blue: basic; white: hydrophobic).

with hydrophobic residues in the pocket core and polar residues lining the rim of the pocket; many of the probes also formed electrostatic interactions with a basic residue, R224, in Site 2. Our previous work suggests that the HDAC3 backbone dynamics near Sites 1 and 2 are allosterically linked to the IP4 binding site, so small molecule that bind to either site may serve to destabilize the binding of IP4 or disrupt the IP4-induced conformational selection. Site 3 is the most polar of the three sites and small molecules that bind to Site 3 would directly compete with DAD binding.

4.4 Conclusions

In an attempt to identify novel HDAC3 active site inhibitors, we have devised a virtual screen protocol that considers HDAC3 conformations in its apo, and IP4/DAD-complexed states. In the first phase of our virtual screen we assessed the predictive power of the receptor structures and the reliability of ranks of known inhibitors. Specifically, we were interested in receptor models that performed best in early enrichment; that is, the receptor models that result in known actives ranked close to the top of the list. An assessment of early enrichment is of particular importance when experimental inhibition assays can only screen a small number of compounds. In our enrichment analysis the crystal structure conformation itself performed quite well, however other more predictive receptor models identified from simulations of the HDAC3:IP4 and HDAC3:IP4:DAD complexes. In our analysis of the reliability of known active rankings, we found that the top-ranked actives were consistently top-ranked across predictive receptor models, and the poorly ranked actives had a high variability in their rank across receptor models. Overall, we found that the receptor models obtained from simulations of HDAC3 in its apo state were least predictive, supporting our hypothesis that the presence of IP4 and DAD biases HDAC3 towards a conformation more suitable for the binding of its substrate or a competitive inhibitor. In the second phase of our virtual screen, we identified ten potential inhibitors as the compounds that consistently out-ranked the most reliable positive control, Trichostatin A.

As an alternative inhibition mechanism, we have also considered possible allosteric inhibition sites. Using solvent mapping, we have identified three consensus sites, (1) close to the C-terminal helix, (2) Helix 6, and (3) the IP4-binding site, from among the apo HDAC3 and HDAC3:IP4 conformational ensembles that are likely to be promising candidates in future drug discovery efforts. Two of these sites are near regions that were found to be allosterically linked to the IP4-binding site in previous work, and the third site could be leveraged to directly disrupt the binding of DAD. These druggable hot spots may be exploited in future virtual screening efforts.

4.5 Acknowledgements

This work was supported by NSF Graduate Research Fellowship, Molecular Biophysics Training Grant GM08326 (M.A.), the National Science Foundation Grant MCB-1020765, NBCR, CTBP, and Howard Hughes Medical Institute (J.A.M.), and National Institutes of Health Grant GM31749 (J.A.M.).

Chapter 4 is a minimally modified version of a manuscript being prepared for submission. The dissertation author was the primary contributor and first author of this work.

Chapter 5

w-REXAMD: A Hamiltonian replica exchange approach to improve free energy calculations for systems with kinetically-trapped conformations

5.1 Introduction

The change in the Gibbs free energy is the thermodynamic property that drives equilibrium aspects of complex biological processes like the folding of proteins, drug binding, and molecular recognition of proteins in cell signaling pathways.[12, 90, 59] A great deal of progress has been made in the area of free energy calculations in recent years [50, 17, 6, 75, 66].

One of the obstacles in using computer simulation to calculate free energy changes, for example in the solvation of small molecules like ibuprofen, or in comparing protein-drug binding interactions, is the limit of conformational sampling[58, 62]. A transformation of interest, perhaps changing a substituent on a drug-like molecule, or a mutation in a protein active site, may involve a conformation change that is inaccessible on the timescale of the simulation. Without considering the mutation in both possible conformations, the calculated free energy change for the process is

inaccurate. This sampling problem can be particularly difficult to solve if the two (or more) conformations are not known beforehand.

Enhanced sampling techniques are often leveraged to mitigate the sampling problem[45, 63, 70]. Accelerated molecular dynamics (aMD) [41] is a Hamiltonian-modifying technique that has been used extensively in our group to simulate micro- to millisecond timescale conformational changes in biomolecules [79, 24, 13, 42, 68, 36, 101, 80, 23]. The original form of aMD adds a bias to potential energy minima, making barriers easier to cross, without defining a reaction coordinate *a priori*. While this method has been shown to be a useful tool in exploring the conformational space of large proteins, the simulation must be re-weighted to recover the Boltzmann distribution. In this exponential re-weighting the conformations at energy minima make significantly larger contributions than all other conformations to the average value and the effective number of data points is reduced, introducing statistical noise in ensemble averages. By using the original aMD equation in Hamiltonian replica exchange (REXAMD), Fajer et al. showed that the exponential re-weighting problem could be ameliorated by simulating replicas at various states of acceleration, including an unaccelerated "ground" state, and collecting data only from the unaccelerated state.[30].

Recently a new "windowed" aMD (w-aMD) method has been introduced in which the potential energy barriers are scaled down while preserving the original features of the landscape (eqs 1 and 2) [88]. The advantage of this implementation relative to the original aMD implementation is that the structures already at energy minima remain on the unbiased potential energy surface, and energy barriers are scaled without affecting the overall shape of the energy landscape. While this method was shown to be an improvement on the previous barrier-lowering aMD method[22], one limitation of the windowed aMD method is the introduction of additional parameters that determine the degree of acceleration: E_1 and E_2 define the window in which the dynamics are accelerated, while α_1 and α_2 are tuning parameters (see Supplemental Information for details). Here we implement the windowed aMD equation in a Hamiltonian replica exchange framework (w-REXAMD), in which multiple sets of acceleration parameters distinguish the Hamiltonians of exchanging replicas.

$$V^*(r) = \begin{cases} V(r) - \Delta V^c(r) & V(r) > E_1 \\ V(r) & V(r) \leq E_1 \end{cases} \quad (5.1)$$

where

$$\Delta V^c(r) = \frac{(V(r) - E_1)^2}{(\alpha_1 + V(r) - E_1) \left(1 + \exp\left(\frac{V(r) - E_2}{\alpha_2}\right)\right)} \quad (5.2)$$

In our discussion of w-REXAMD simulations, a “state” refers to a particular set of aMD parameters introduced in eq 2. A “replica” refers to an evolving configuration which may exchange its state with that of its neighbor’s, according to the Metropolis criterion. In an analogous temperature replica exchange simulation, the “state” would be defined by the temperature, while a “replica” may visit different temperatures.

We anticipate that the additional conformational sampling from multiple replicas at different acceleration states will enrich the sampling in the unaccelerated state ensemble. As with any replica exchange method though, we must acknowledge the increase in computational cost associated with the multiple replicas. To account for both the advantage in sampling simply from multiple independent simulations and the limitation of increased computational cost, we compare w-REXAMD to conventional MD in the same replica exchange framework, which we refer to as REGREX. In a REGREX simulation, multiple independent conventional MD replicas exchange their configurations at the same frequency at which exchanges are attempted in a w-REXAMD simulation. The states in a REGREX simulation are equivalent to each other, and each state contains equivalent portions of independent MD trajectories. One REGREX state, then, may also benefit from conformational space sampled in multiple replica trajectories, but will still contain the same number of structures as one w-REXAMD state.

We first compare the extent of conformational sampling in w-REXAMD and REGREX simulations of two systems: butane and cyclohexane, both in explicit water solvent. For the case of butane, conventional MD, and thus REGREX simulations, can exhaustively sample the easily characterizable conformational space of the system (Fig 5.1A) and provide a reference for a converged result. In the case of cyclohexane, however, a large (≈ 10 kcal/mol) barrier separates the two stable chair conformations (Fig 5.5A), making the interconversion between the two conformations, so-called "chair-

flipping", a rare event that occurs on the microsecond timescale [15]. Since cyclohexane is a model system in which the timescales are no longer accessible for reference REGREX simulations, we instead use umbrella sampling calculations as an appropriate reference to which we compare ensembles generated in w-REXAMD simulations.

We then present an application of w-REXAMD to free energy calculations using an alchemical identity transformation. As mentioned earlier, multiple kinetically-trapped conformations can be a major obstacle in free energy calculations. Here we discuss a simple example of such asymmetry in the transformation of bromocyclohexane to itself. This identity transformation, in which the bromine is displaced from one carbon to another (shown in Fig 5.3), should result in a free energy change of zero and serves as a force-field independent benchmark. With the bromine substituent, the two chair conformations of the cyclohexane are no longer equivalent. The bromine occupies either an equatorial or axial position, and interconversion of these two conformers occurs on the microsecond timescale. In our alchemical self-transformation, the bromine in equatorial-substituted bromocyclohexane is displaced to the axial position at another carbon. The axial and equatorial conformers exist in equilibrium, so constructing the transformation in this way requires sampling of both conformers according to their thermodynamic probabilities in order to reach the analytical result of zero. On the typical nanosecond timescale of MD simulations, the axial and equatorial configurations are kinetically trapped, and the identity transformation is a good test case for the convergence problem encountered when multiple conformations are separated by high barriers. We show that the enhanced sampling in w-REXAMD simulations is useful in mitigating this convergence issue in the free energy calculation.

5.2 Methods

Hamiltonian Replica Exchange. Each thermodynamic state in the w-REXAMD simulation corresponds to a set of four parameters (E_1 , E_2 , α_1 , and α_2) that determine the degree of acceleration of a replica at that state. Four replicas were used in each simulation, and the least accelerated state was always the unmodified Hamiltonian. Every 300 MD steps (0.3 ps), exchanges between adjacent replicas were

attempted. A description of how parameters were selected can be found in Supplemental Information. Only torsional potential energy terms were accelerated in these test cases. The Amber ff99SB force field[43] was used for all simulations, with the TIP3P water model[91].

Thermodynamic Integration. All atoms being coupled or decoupled according to the parameter λ were treated with soft-core potentials. Nine evenly spaced λ points were chosen between 0 and 1 for each transformation and 1.5ns of REGREX and w-REXAMD was run at each λ point (NVT) after an equilibration period of 150 ps (NPT). $\langle dV/d\lambda \rangle_{\lambda_i}$ from one state (the unaccelerated state in w-REXAMD simulations) was used for the free energy calculations. Free energies were calculated by numerically integrating the $\langle dV/d\lambda \rangle_{\lambda_i}$ values from 0 to 1 using the Trapezoid rule. The uncertainty associated with each of the nine $\langle dV/d\lambda \rangle_{\lambda_i}$ values was calculated as:

$$\delta^2 = \frac{\sigma^2}{N/g} \quad (5.3)$$

where N is the number of $dV/d\lambda$ values used to compute the average and g is the statistical inefficiency, as defined by Chodera et al.[18], which contains the integrated autocorrelation time τ ($g = 1 + 2\tau$). The variance of the free energy from one transformation was estimated by propagating the uncertainty at each of the nine lambda values. Three independent $C_{1,ax}$ -to- $C_{4,eq}$ and $C_{1,eq}$ -to- $C_{4,ax}$ transformations were performed using REGREX, and three $C_{1,eq}$ -to- $C_{4,ax}$ transformations were done using w-REXAMD. The final free energies reported are averages weighted by each transformation's uncertainty:

$$\Delta\bar{G} = \frac{\sum_{j=1}^3 \Delta G_j w_j}{\sum_{j=1}^3 w_j} \quad (5.4)$$

where $w_j = \frac{1}{\delta_j^2}$

Umbrella Sampling. The umbrella sampling simulations were performed in NAMD[78] using the Amber ff99SB[43] forcefield. At each of sixty-three evenly spaced ϵ centers, ranging from $\epsilon = -62$ degrees to $\epsilon = 62$ degrees, a 2ns MD simulation was performed with a harmonic potential centered at the value of ϵ . A force constant of 1.0 kcal/mol/deg² was used for all simulations, and only the last 1.5ns were used

to construct the free energy profile. Initial structures were taken from pre-equilibrated structures from cyclohexane NVT simulations. The free energy profile was constructed using the python implementation of MBAR.[87]

5.3 Results and discussion

5.3.1 Conformational Sampling

We distinguish between state trajectories (corresponding to the ensemble of structures sampled with a certain set of acceleration parameters) and replica trajectories (corresponding to the trajectory of a replica as it visited different states).

We use the dihedral angle ($C_1-C_2-C_3-C_4$) of butane during the simulations to monitor the interconversion between gauche(-/+) and trans conformations of butane (Fig 5.1A). In order to probe convergence in conformational sampling in butane, we clustered multiple independent unaccelerated state trajectories into gauche(+), gauche(-), or trans conformations and plot the cumulative fraction for each conformation during the timescale of the w-REXAMD, REGREX, and w-aMD simulations. In the w-REXAMD simulations, all conformation populations converge to fractional occupancies that lie within the error intervals reached in reference REGREX simulations (Fig 5.1B and C). The two equivalent gauche conformations quickly converge in the w-REXAMD simulations in considerably fewer steps than in the REGREX simulations. For the non-exchanging w-aMD simulations the equivalent gauche conformations are sampled exhaustively (Fig 5.1D) but the less favorable eclipsed conformation is sampled to a greater extent, indicating the effectiveness of the windowed aMD equation in scaling down the barriers, but illustrating the need for Boltzmann re-weighting in order to attain information about the equilibrium ensemble.

Unlike those of butane, the two stable conformations of cyclohexane are separated by a 10 kcal/mol free energy barrier, making interconversion between the two conformations a rare event[15]. Using a conformation coordinate ε (eq 3), which is a function of the six correlated cyclohexane dihedral angles (τ_i), we analyzed the sampled

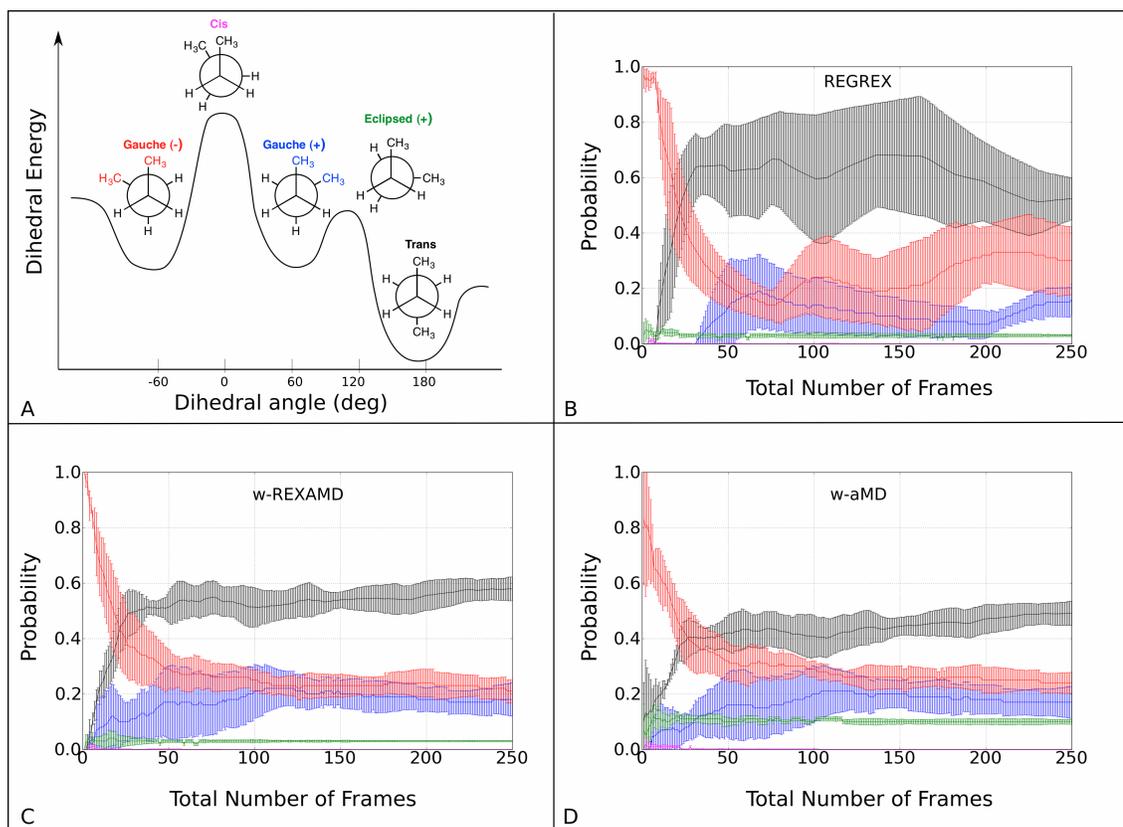


Figure 5.1: Conformational sampling of butane. (A) A schematic of butane conformational landscape according to C₁-C₂-C₃-C₄ dihedral angle (the cis and eclipsed (+) structures are shifted slightly so that the opposed groups are visible). Fractions of state ensembles in the trans (black series), gauche(-) (red series), gauche(+) (blue series), and eclipsed (green series) are shown for (B) REGREX, (C) w-REXAMD, and (D) w-aMD simulations. Total number of frames corresponds to 500ps of simulation. Error bars show standard deviation from three independent simulations.

conformations of cyclohexane (chair₁, boat-like, and chair₂)[13].

$$\varepsilon = (\tau_1 - \tau_2 + \tau_3 - \tau_4 + \tau_5 - \tau_6)/6 \quad (5.5)$$

By clustering the ε coordinate over multiple trajectories, we were able to observe the fractional occupancies of the two chair conformations ($\varepsilon \approx \pm 50$) of cyclohexane. Chair-flipping cannot be observed in reasonable timescales of cMD (0.5 μ s, data not shown), and REGREX simulations similarly only sample the initial chair conformation (Fig 5.5B, magenta series). The probability of each chair conformation for the unaccelerated state from w-REXAMD simulations agrees well with expected probability from reference umbrella calculations (Fig 5.5A and B, black series). The use of the w-aMD equation allows for enhanced sampling of cyclohexane conformations (Fig 5.5C, red series). Without the replica exchange framework, w-aMD simulations must be re-weighted to obtain the same equilibrium populations (Fig 5.5B, red series). Although this re-weighting is quite successful in recovering equilibrium populations for small systems such as these[88], it has been shown to introduce statistical noise in ensemble averages[86]. This issue may be more prominent in more complex systems.

5.3.2 Alchemical Free Energy Calculations

In the previous section we showed that w-REXAMD is a reasonable method to rapidly explore conformational space, particularly in the presence of high barriers separating relevant conformations. We expect, then, that w-REXAMD should also improve the convergence of free energy calculations. We consider the transformation of bromocyclohexane to itself, which has a free energy change of zero. Similar to cyclohexane, bromocyclohexane has a considerable barrier separating its two chair conformations, making sampling of the conformations a microsecond timescale event, presently inaccessible to routine alchemical free energy calculations. The presence of the bromine substituent creates a difference between the two chair conformations; one chair conformation is the axial conformer while the other is the equatorial conformer. The equatorial conformer is thermodynamically favored, as it results in less steric hindrance between the bulky bromine and neighboring hydrogens, but bromocyclohexane exists in equilibrium between both conformers. To highlight the sampling problem, we let the two

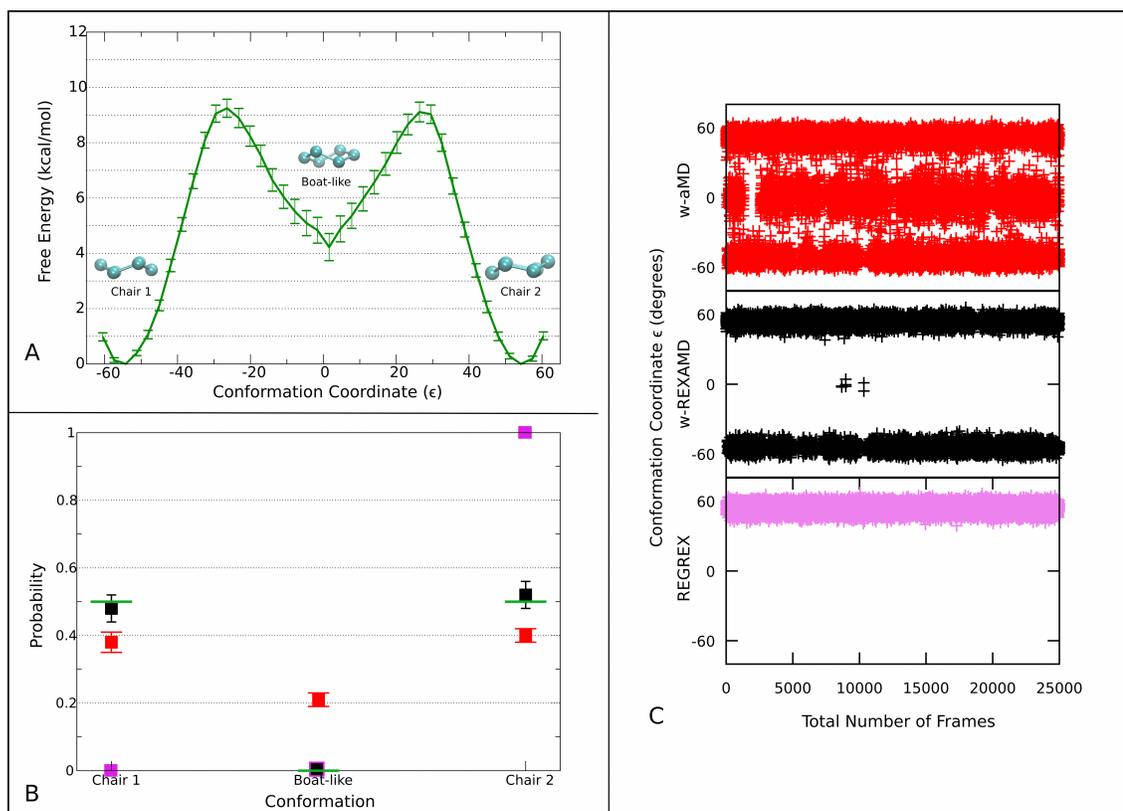


Figure 5.2: Conformational sampling of cyclohexane. (A) Umbrella sampling along ϵ coordinate shows high barrier between chair conformations. (B) Observed probability of chair and boat-like conformations from w-REXAMD (black series), windowed-aMD (red series), and REGREX (magenta series) simulations. Error bars show standard deviation from three independent simulations. Green bars mark expected probability from umbrella sampling calculation in (A) as a reference. (C) ϵ conformation coordinate during windowed-aMD (red series), w-REXAMD (black series), and REGREX (magenta series) shows interconversion of cyclohexane conformations.

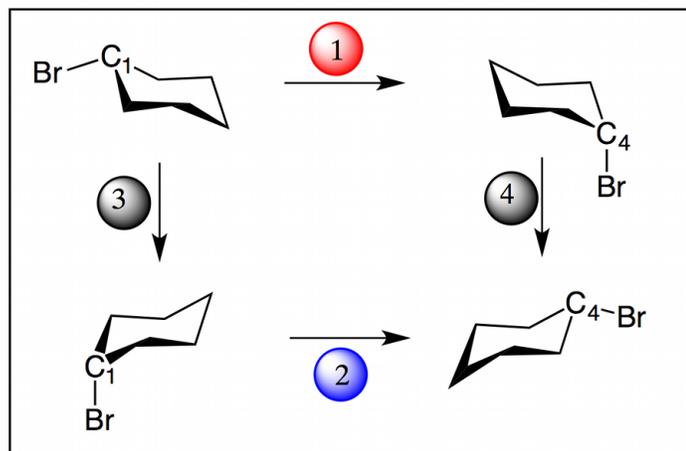


Figure 5.3: Thermodynamic cycle of alchemical identity transformation of bromocyclohexane. In alchemical transformations 1 and 2, the bromine substituent is displaced from C_1 to C_4 , either from an equatorial to an axial position (transformation 1: " $C_{1,eq}$ -to- $C_{4,ax}$ ") or vice versa (transformation 2: " $C_{1,ax}$ -to- $C_{4,eq}$ "). Transformations 3 and 4 represent the slow conformational change isolating the two conformers in conventional MD. $\Delta G_1 + \Delta G_4 = \Delta G_3 + \Delta G_2 = 0$ kcal/mol.

conformers be the initial and final states, and consider the bromocyclohexane identity transformation, where the bromine is displaced from one carbon to another, in both directions: $C_{1,eq}$ -to- $C_{4,ax}$, and $C_{1,ax}$ -to- $C_{4,eq}$. In the $C_{1,eq}$ -to- $C_{4,ax}$ transformation, for example, the equatorial bromine at carbon 1 and an axial hydrogen at carbon 4 are decoupled from the rest of the system, while a bromine atom appears at the axial position on carbon 4, and a hydrogen atom appears at the equatorial position initially occupied by the bromine, at carbon 1 (Fig 5.3).

Using the REGREX simulations at each of nine lambda intermediates along the alchemical pathway, we found that the observed free energy change for the transformation did not result in zero, and that it depended upon the directionality of the transformation (see Fig 5.4A). The REGREX $C_{1,ax}$ -to- $C_{4,eq}$ bromocyclohexane transformations converged to a value of approximately -0.2 kcal/mol (Fig 5.4A, red series), while the $C_{1,eq}$ -to- $C_{4,ax}$ self transformations converged to approximately +0.3 kcal/mol (Fig 5.4A, blue series). Because of the presence of the high free energy barrier separating the two conformers, bromocyclohexane did not flip into its equatorial conformation during the alchemical transformation, and instead remained exclusively in the axial con-

formation (see Fig 5.4B, blue series). In this case, where the relevant conformations are kinetically trapped, we find that the alchemical identity transformation results in a non-zero free energy change. Without any enhanced sampling, the alchemical self-transformations only complete either transformation 1 or 2 of the thermodynamic cycle in Fig 5.3, with $\Delta G_1 = -\Delta G_2$, within statistical uncertainty. In order for the analytical result to be reached with conventional MD, the conformational changes depicted by transformations 3 and 4 of the thermodynamic cycle in Fig 5.3 must also be completed. On the other hand, with the w-REXAMD method the initial and final states are able to sample both axial and equatorial conformations, and as a result the identity transformation quickly converges to the analytical result of 0 kcal/mol (Fig 5.4A, black series). It is important to note that for this small system, re-weighting of the most accelerated w-aMD state trajectory also converges to the analytical result (Fig 5.4A, maroon series).

5.4 Conclusions

We presented an enhanced sampling method, w-REXAMD, which is capable of sampling conformations that would normally interchange on microsecond timescales. We showed that w-REXAMD could be used to improve alchemical free energy calculations, particularly when the conformation of the initial or final state is kinetically trapped. Similar to the use of accelerated Molecular Dynamics, an advantage of the w-REXAMD method is the gain in sampling without having to define any reaction coordinate *a priori*. The replica exchange framework offers the additional advantage of (1) avoiding any re-weighting of data in obtaining equilibrium ensemble averages relevant for free energy calculations, and (2) using multiple sets of acceleration parameters. To optimize the w-REXAMD method, future studies should focus on the distribution of the sets of acceleration parameters in order to both maximize sampling and retain efficient diffusion rates among replicas.

The replica exchange framework does come with the need for additional parallel computing resources, and this need will be amplified as system size increases. In systems with multiple, slow-exchanging conformational states though, a common protocol is to calculate the change in free energy for some transformation, e.g. a mutation in an

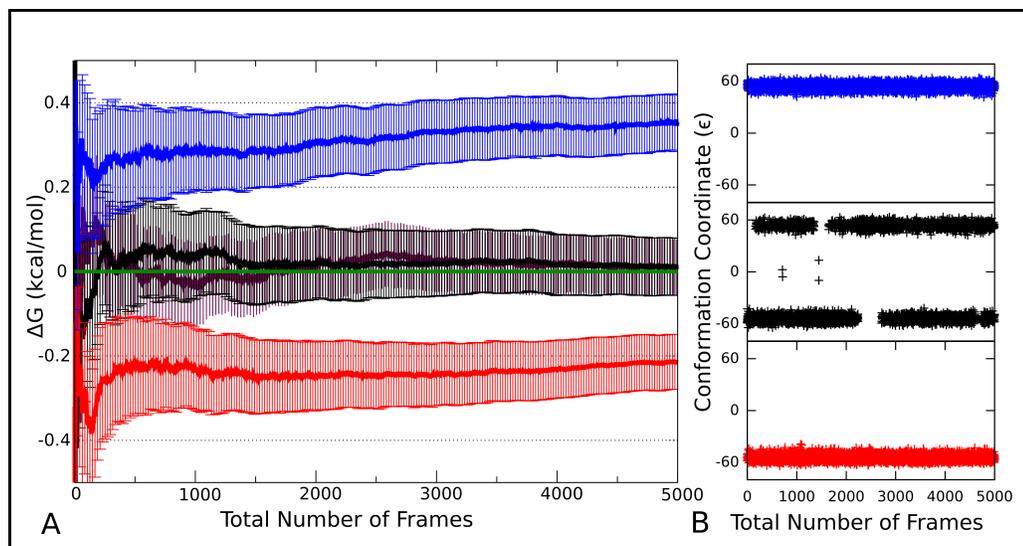


Figure 5.4: Free energy of identity transformation for bromocyclohexane. (A) REGREX transformations were done in both the $C_{1,eq}$ -to- $C_{4,ax}$ (blue series) and $C_{1,ax}$ -to- $C_{4,eq}$ (red series) directions. w-REXAMD transformations were performed for just the $C_{1,eq}$ -to- $C_{4,ax}$ direction. The free energy from the unaccelerated state of the w-REXAMD simulations is shown in the black series, and the re-weighted most accelerated state is shown in maroon. Each series shows the average of three independent transformations, weighted by their statistical uncertainties. (B) Representative values for ϵ at $\lambda = 0.7$ for the transformations show that only one chair conformation is sampled in REGREX $C_{1,eq}$ -to- $C_{4,ax}$ (blue series) and $C_{1,ax}$ -to- $C_{4,eq}$ (red series) transformations, whereas both chair conformations are sampled in the w-REXAMD unaccelerated state of the $C_{1,eq}$ -to- $C_{4,ax}$ transformation (black series).

enzyme active site, for each constrained conformation separately, and then combine the transformations in a thermodynamic cycle to estimate the actual free energy of the mutation. We believe that in such cases, and particularly in cases where the conformational landscapes are not well characterized, w-REXAMD will be a useful enhanced sampling method.

5.5 Supporting Information

5.5.1 Accelerated MD Approach

The windowed aMD equation (Equation 1) was recently introduced [88] and we will summarize it here.

$$V^*(r) = \begin{cases} V(r) - \Delta V^c(r) & V(r) > E_1 \\ V(r) & V(r) \leq E_1 \end{cases} \quad (5.6)$$

where

$$\Delta V^c(r) = \frac{(V(r) - E_1)^2}{(\alpha_1 + V(r) - E_1) \left(1 + \exp\left(\frac{V(r) - E_2}{\alpha_2}\right)\right)} \quad (5.7)$$

When the potential energy $V(r)$ is higher than a fixed threshold E_1 , it is lowered by $\Delta V^c(r)$, allowing the simulation to sample conformations according to the modified Hamiltonian $V^*(r)$. The parameter E_2 creates a window of acceleration, protecting high barriers since $\Delta V^c(r) \rightarrow 0$ when $V(r) \gg E_2$. In this way it is possible to smooth the roughness of the potential energy surface, but avoid increased sampling of high energy conformations. The α parameters tune the transitions between modified and real potential energy surfaces. Although this approach could be used to accelerate according to total potential energy, in this work only the torsional potential energy was modified.

5.5.2 Hamiltonian Replica Exchange

Each thermodynamic state in the w-REXAMD simulation corresponds to a set of four parameters (E_1 , E_2 , α_1 , and α_2) that determine the degree of acceleration of a replica at that state. Four replicas were used in each simulation, and the least accelerated

Table 5.1: Summary table of acceleration parameters

	E_1 (kcal/mol)	E_2 (kcal/mol)	α_1 (kcal/mol)	α_2 (kcal/mol)
Butane	1.0	5	1, 0.5, 0.1	1
Cyclohexane	3.5	15	1, 0.5, 0.1	1
Bromocyclohexane	8.0	50	1, 0.5, 0.1	1

state was always the unmodified Hamiltonian. Neighbor exchanges were attempted every 300 MD steps (0.3 ps).

Selection of Parameters. We chose the parameters for the set of replicas empirically. We used the average dihedral energy from a 1ns conventional MD simulation (cMD) of butane in water as a benchmark for E_1 and tested values for E_2 which were multiples of the maximum dihedral energy in the cMD simulation. We then did a series of short w-aMD simulations with $\alpha_2 = 1$ kcal/mol, which is approximately $0.2 * (E_2 - E_1)$ and varied α_1 from low to high acceleration: 1, 0.5, and 0.1 kcal/mol. The choice of parameters was made by analyzing both the distributions of ΔV^c and the dihedral angle sampling in each simulation. This process was repeated for cyclohexane. For bromocyclohexane the sum of the dihedral energy of the common atoms, and both initial and final states’ soft-core atoms was used to determine the acceleration parameters. 5.1 summarizes the parameters used. The w-REXAMD simulations were performed using a modified version of the AMBER11 sander module[14]. The alchemical transformations were done using a python wrapper and a modified AMBER10 source code.

MD Details. All three molecules were built using the GAFF force field[97], with HF/6-31G* RESP charges obtained using Gaussian 03[33]. Antechamber was used to prepare the molecules and LEaP was used to solvate them[14]. The systems were solvated in 10.0 angstrom TIP3P [91] water boxes (ff99SB force field[43]) which were first heated to 300 K and equilibrated until the the correct density of water at 1 atm was reached. Pressure was kept constant during equilibration using the Berendsen barostat[9]. The simulations were then simulated in the NVT ensemble with temperature regulated using the Anderson thermostat[2]; velocities were randomized every 1000 steps. A 1 fs time-step was used. An 8.0 Angstrom cutoff was used for non-bonded interactions.

The w-REXAMD simulations were compared to identical replica exchange sim-

ulations in which all replicas were run on the unmodified Hamiltonian (simulations are called REGREX). This way any improvement in sampling is separated from the advantage of running multiple trajectories in parallel. In the REGREX simulations, the four CMD trajectories exchanged at every attempt.

5.5.3 Thermodynamic Integration

All atoms being coupled or decoupled according to the value of the parameter λ were treated with soft-core potentials.[106]. Because of the implementation of the soft-core potential in the AMBER sander module, no simulations were done at $\lambda = 0.0$ or $\lambda = 1.0$. Nine evenly spaced lambda points [0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9] were used, and 1.5ns (5000 frames saved) of REGREX and w-REXAMD was run at each lambda point, after an equilibration period of 150 ps.

We used thermodynamic integration to calculate the change in free energy between initial and final states:

$$\Delta G_{0 \rightarrow 1} = \int_0^1 d\lambda \left\langle \frac{\delta V}{\delta \lambda} \right\rangle_{\lambda} \quad (5.8)$$

The integral in (3) was evaluated using the trapezoid rule:

$$\Delta G \approx \left\langle \frac{\delta V}{\delta \lambda} \right\rangle_{\lambda_1} \lambda_1 + \frac{1}{2} \sum_{i=2}^N \left[\left\langle \frac{\delta V}{\delta \lambda} \right\rangle_{\lambda_i} + \left\langle \frac{\delta V}{\delta \lambda} \right\rangle_{\lambda_{i-1}} \right] (\lambda_i - \lambda_{i-1}) + \left\langle \frac{\delta V}{\delta \lambda} \right\rangle_{\lambda_N} (1 - \lambda_N) \quad (5.9)$$

We calculated $\left\langle \frac{\delta V}{\delta \lambda} \right\rangle_{\lambda_i}$ as the average:

$$\left\langle \frac{\delta V}{\delta \lambda} \right\rangle_{\lambda_i} = \frac{1}{N_i} \sum_n^{N_i} \frac{\delta V}{\delta \lambda_{i,n}} \quad (5.10)$$

where N_i is the number of configurations in a state trajectory from a simulation at λ_i . Because the series of $N_i \frac{\delta V}{\delta \lambda_{i,n}}$ values may be correlated, we approximated the statistical uncertainty $\delta^2 \left\langle \frac{\delta V}{\delta \lambda} \right\rangle_{\lambda_i}$ at each λ_i as described by Chodera et al. [18]:

$$\delta_{\lambda_i}^2 = \frac{\sigma^2}{N_i/g} \quad (5.11)$$

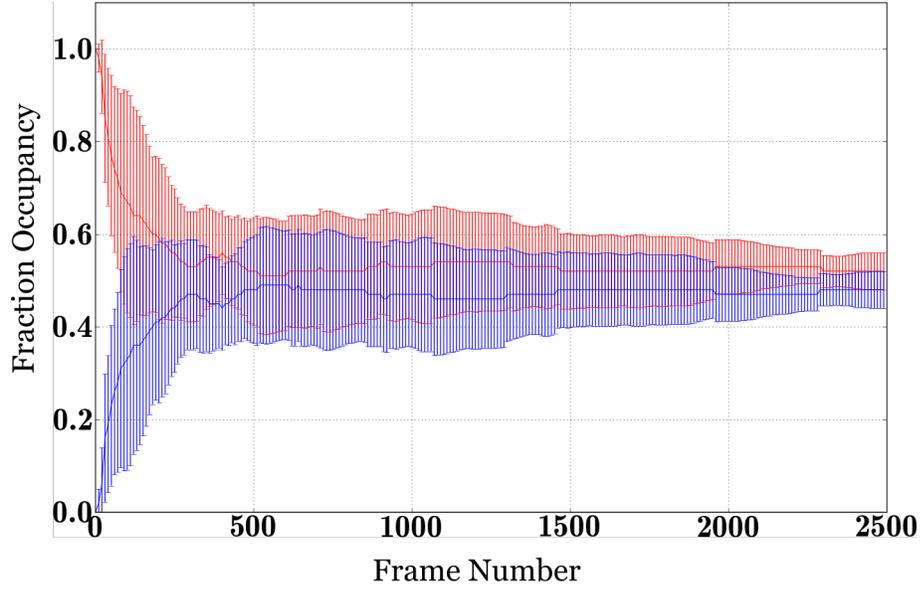


Figure 5.5: Convergence of chair 1 (blue) and chair 2 (red) fraction occupancies in unaccelerated state of w-REXAMD simulations (5ns total). Error bars show standard deviation from three independent simulations.

where σ^2 is the variance of the $\frac{\delta V}{\delta \lambda}$ series, and g is the statistical inefficiency, which incorporates the integrated autocorrelation time τ :

$$g = 1 + 2\tau \quad (5.12)$$

The error propagation formula was used with eq 4 to calculate the variance for the free energy change of the transformation. We repeated the bromocyclohexane self transformation in triplicate and estimate the free energy change $\Delta\bar{G}$ as the average of the three independent ΔG_j values, weighted by their associated statistical uncertainty δ_j^2 .

$$\Delta\bar{G} = \frac{\sum_j \Delta G_j w_j}{\sum w_j} \quad (5.13)$$

where $w_i = \frac{1}{\delta_j^2}$, and the uncertainty of the weighted average $\Delta\bar{G}$ is

$$\sigma = \frac{1}{\sqrt{\sum w_j}} \quad (5.14)$$

Table 5.2: Summary table of mixing among states in replica exchange simulations. Mean-first passage times from each state to either the most accelerated state (τ_{state3}) or down to the unaccelerated state ($\tau_{unaccelerated}$) in w-REXAMD alchemical self-transformation of bromocyclohexane. Error indicates 90% confidence interval.

	τ_{state3} (exchange attempts)	$\tau_{unaccelerated}$ (exchange attempts)
Unaccelerated state	9.610 ± 0.713	-
State 1	3.688 ± 0.063	3.561 ± 0.107
State 2	2.158 ± 0.020	5.159 ± 0.149
State 3	-	6.729 ± 0.193

5.6 Acknowledgements

The authors congratulate Professor Wilfred van Gunsteren for his many contributions on the occasion of his 65th birthday. The authors thank Denis Bucher for helpful discussions. This work was supported by NSF Graduate Research Fellowships (MA and MF), Molecular Biophysics Training Grant GM08326 (WS), the National Science Foundation Grant MCB-1020765, NBCR, CTBP, and Howard Hughes Medical Institute (J.A.M.), and National Institutes of Health Grant GM31749 (J.A.M.).

Chapter 5 is a near reprint of the material as it appears in J. Chem. Theory Comput. 2013, Volume 9, Pages 18-23. The dissertation author was the primary contributor and first author of this work.

Chapter 6

General conclusions

Allostery and entropy in HDAC3 activation. In Chapters 2 and 3 of this dissertation, molecular dynamics simulations were performed to further our understanding of the activation mechanism of a histone deacetylase, HDAC3, which depends upon the enzyme's recruitment to a larger multi-protein complex. We found that upon binding to the deacetylase activating domain (DAD) and necessary small molecule inositol tetraphosphate (IP4), flexible regions of HDAC3 are stabilized in a single conformation. We have also identified a mutation, near the IP4 binding site on HDAC3, that can trigger a very similar long range stabilization of the deacetylase.

There are many examples in biology in which the protein-ligand complex is generally less dynamic than the apo protein itself, thus associating ligand binding with an unfavorable loss of configurational entropy[53, 16, 94, 11]. Allosteric modulators of such proteins, however, can ameliorate this entropic penalty by paying a portion of the cost themselves, provided that the binding of the allosteric modulator also results in some compensatory enthalpic or entropic changes[31, 21, 57]. In the case of HDAC3, it is reasonable to assume that IP4, DAD, and the acetyllysine substrate bind in some sequence and not concurrently. Binding of either DAD or IP4 would already result in a considerable loss of configurational entropy, such that binding of the second modulator would at least be entropically more favorable. The binding of both DAD and IP4 would thus favor the subsequent binding of the acetyllysine substrate, as is also suggested by the observed re-orientation of the catalytic Tyr298 side chain and stabilization of the active site channel.

The finding, in both Chapters 2 and 3, that the IP4 binding site is allosterically linked to the nuclear localization signal of HDAC3 suggests a novel role for the inositol signalling pathway, in addition to specifically activating the HDAC3:DAD complex, the IP4 molecule may indeed serve to localize cytoplasmic HDAC3 to the nucleus, where the DAD and its histone targets are already localized.

Novel inhibitors of HDAC3. In Chapter 4 we utilized the conformational ensembles of the HDAC3, HDAC3:IP4, HDAC3:DAD, and HDAC3:IP4:DAD complexes, generated by molecular dynamics simulations, to develop a new protocol for the identification of HDAC3 inhibitors. Using a library of decoy drug-like molecules and a collection of known HDAC3 inhibitors, we first identified a subset of highly predictive receptor conformations. In this way, we could incorporate receptor flexibility into the docking protocol by including multiple conformations, but still bias our virtual screen to receptor models that perform well in ranking the known active compounds from a large set of decoy compounds. Using this protocol we have identified novel active site inhibitors for future *in vitro* testing.

Development of enhanced sampling methods. In using molecular dynamics simulations to study such biological phenomena, a problem that is often encountered is the non-ergodicity of the MD trajectories, which causes any thermodynamic property, such as a free energy difference, to be inaccurate. In Chapter 5, we address this so-called sampling problem by implementing a new enhanced sampling technique, w-REXAMD, and applying it to free energy calculations for a system with kinetically trapped conformations. We find that this method dramatically enhances the sampling of conformational space for this simple test system and consequently serves to accurately compute the free energy difference for the self-transformation, whereas conventional molecular dynamics simulations are inadequate.

Bibliography

- [1] C. V. P. G. P. J. M. M. A. C. R. D. F. C. S. S. D. M. Alessandro Vannini. Substrate binding to histone deacetylases as shown by the crystal structure of the HDAC8–substrate complex. *EMBO Reports*, 8(9):879, Sept. 2007.
- [2] H. C. Andersen. Molecular dynamics simulations at constant pressure and/or temperature. *J. Chem. Phys.*, 72(4):2384–2393, 1980.
- [3] N. M. Archin, A. L. Liberty, A. D. Kashuba, S. K. Choudhary, J. D. Kuruc, A. M. Crooks, D. C. Parker, E. M. Anderson, M. F. Kearney, M. C. Strain, D. D. Richman, M. G. Hudgens, R. J. Bosch, J. M. Coffin, J. J. Eron, D. J. Hazuda, and D. M. Margolis. Administration of vorinostat disrupts HIV-1 latency in patients on antiretroviral therapy. *Nature*, 487(7408):482–485, July 2012.
- [4] M. Arrar, C. Oliveira, and J. A. McCammon. Inactivating mutation in histone deacetylase 3 stabilizes its active conformation. *Protein Sci.*, 2013.
- [5] M. Arrar, R. Turnham, L. Pierce, C. A. F. de Oliveira, and J. A. McCammon. Structural insight into the separate roles of inositol tetrakisphosphate and deacetylase-activating domain in activation of histone deacetylase 3. *Protein Sci.*, 22(1):83–92, Jan. 2013.
- [6] R. Baron and J. A. McCammon. . *Annu. Rev. Phys. Chem.*, page in press, 2013.
- [7] D. C. Bas, D. M. Rogers, and J. H. Jensen. Very fast prediction and rationalization of pKa values for protein-ligand complexes. *Proteins*, 73(3):765–783, Nov. 2008.
- [8] K. A. Beauchamp, Y.-S. Lin, R. Das, and V. S. Pande. Are Protein Force Fields Getting Better? A Systematic Benchmark on 524 Diverse NMR Measurements. *J Chem Theory Comput*, 8(4):1409–1414, Apr. 2012.
- [9] H. Berendsen and J. Postma. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, 81:3684–3690, 1984.
- [10] S. Bhaskara, S. K. Knutson, G. Jiang, M. B. Chandrasekharan, A. J. Wilson, S. Zheng, A. Yenamandra, K. Locke, J.-L. Yuan, A. R. Bonine-Summers, C. E. Wells, J. F. Kaiser, M. K. Washington, Z. Zhao, F. F. Wagner, Z.-W. Sun, F. Xia,

- E. B. Holson, D. Khabele, and S. W. Hiebert. Hdac3 is essential for the maintenance of chromatin structure and genome stability. *Cancer Cell*, 18(5):436–447, Nov. 2010.
- [11] D. D. Boehr, R. Nussinov, and P. E. Wright. The role of dynamic conformational ensembles in biomolecular recognition. *Nat Chem Biol*, 5(11):789–796, Oct. 2009.
- [12] J. D. Bryngelson, J. N. Onuchic, N. D. Socci, and P. G. Wolynes. Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins*, 21(3):167–195, Mar. 1995.
- [13] D. Bucher, L. C. T. Pierce, J. A. McCammon, and P. R. L. Markwick. On the Use of Accelerated Molecular Dynamics to Enhance Configurational Sampling in Ab Initio Simulations. *J Chem Theory Comput*, 7(4):890–897, Apr. 2011.
- [14] D. A. Case, T. A. Darden, T. E. Cheatham, III, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, R. C. Walker, W. Zhang, K. M. Merz, B. P. Roberts, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossvai, K. F. Wong, F. Paesani, J. Vanicek, J. Liu, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M. J. Hsieh, G. Cui, D. R. Roe, D. H. Mathews, M. G. Seetin, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and P. A. Kollman. Amber 11. *University of California, San Francisco*, 2010.
- [15] D. Chandler. Barrier crossings: classical theory of rare but important events. In B. Berne, G. Ciccotti, and D. Coker, editors, *Computer Simulation of Rare Events and Dynamics of Classical and Quantum Condensed-Phase Systems – Classical and Quantum Dynamics in Condensed Phase Simulations*, pages 3–23. Barrier crossings: classical theory of rare but important events, Singapore, 1998.
- [16] C. e. A. Chang, W. Chen, and M. K. Gilson. Ligand configurational entropy and protein binding. *Proc. Natl. Acad. Sci. U.S.A.*, 104(5):1534–1539, Jan. 2007.
- [17] J. D. Chodera, D. L. Mobley, M. R. Shirts, R. W. Dixon, K. Branson, and V. S. Pande. Alchemical free energy methods for drug discovery: progress and challenges. *Curr. Opin. Struct. Biol.*, 21(2):150–160, Apr. 2011.
- [18] J. D. Chodera, W. C. Swope, J. W. Pitera, C. Seok, and K. A. Dill. Use of the Weighted Histogram Analysis Method for the Analysis of Simulated and Parallel Tempering Simulations. *J Chem Theory Comput*, 3(1):26–41, Jan. 2007.
- [19] E. Chovancova, A. Pavelka, P. Benes, O. Strnad, J. Brezovsky, B. Kozlikova, A. Gora, V. Sustr, M. Klvana, P. Medek, L. Biedermannova, J. Sochor, and J. Damborsky. CAVER 3.0: A Tool for the Analysis of Transport Pathways in Dynamic Protein Structures. *PLoS Comput. Biol.*, 8(10):e1002708, Oct. 2012.

- [20] A. Codina, J. D. Love, Y. Li, M. A. Lazar, D. Neuhaus, and J. W. R. Schwabe. Structural insights into the interaction and activation of histone deacetylase 3 by nuclear receptor corepressors. *Proc. Natl. Acad. Sci. U.S.A.*, 102(17):6009–6014, Apr. 2005.
- [21] A. Cooper and D. T. F. Dryden. Allostery without conformational change. *Eur. Biophys. J.*, 11(2):103–109, Oct. 1984.
- [22] C. de Oliveira and D. Hamelberg. Coupling accelerated molecular dynamics methods with thermodynamic integration simulations. *J Chem Theory Comput*, pages 1516–1525, 2008.
- [23] C. A. F. de Oliveira, B. J. Grant, M. Zhou, and J. A. McCammon. Large-scale conformational changes of Trypanosoma cruzi proline racemase predicted by accelerated molecular dynamics simulation. *PLoS Comput. Biol.*, 7(10):e1002178, Oct. 2011.
- [24] C. A. F. de Oliveira, D. Hamelberg, and J. A. McCammon. Estimating kinetic rates from accelerated molecular dynamics simulations: alanine dipeptide in explicit solvent as a case study. *J. Chem. Phys.*, 127(17):175105, Nov. 2007.
- [25] A. J. M. de Ruijter, A. H. van Gennip, H. N. Caron, S. Kemp, and A. B. P. van Kuilenburg. Histone deacetylases (HDACs): characterization of the classical HDAC family. *Biochem. J.*, 370(Pt 3):737–749, Mar. 2003.
- [26] K. Debacker, A. Frizzell, O. Gleeson, L. Kirkham-McCarthy, T. Mertz, and R. S. Lahue. Histone deacetylase complexes promote trinucleotide repeat expansions. *PLoS Biol.*, 10(2):e1001257, Feb. 2012.
- [27] D. P. Dowling, L. Costanzo, H. A. Gennadios, and D. W. Christianson. Evolution of the arginase fold and functional diversity. *Cell. Mol. Life Sci.*, 65(13):2039–2055, Mar. 2008.
- [28] D. P. Dowling, S. L. Gantt, S. G. Gattis, C. A. Fierke, and D. W. Christianson. Structural studies of human histone deacetylase 8 and its site-specific variants complexed with substrate and inhibitors. *Biochemistry*, 47(51):13554–13563, Dec. 2008.
- [29] S. E Nichols, R. V Swift, and R. E Amaro. Rational Prediction with Molecular Dynamics for Hit Identification. *Curr Top Med Chem*, 12(18):2002–2012, Sept. 2012.
- [30] M. Fajer, D. Hamelberg, and J. A. McCammon. Replica-Exchange Accelerated Molecular Dynamics (REXAMD) Applied to Thermodynamic Integration. *J Chem Theory Comput*, 4(10):1565–1569, Oct. 2008.

- [31] A. T. Fenley, H. S. Muddana, and M. K. Gilson. Entropy-enthalpy transduction caused by conformational shifts can obscure the forces driving protein-ligand binding. *Proceedings of the National Academy of Sciences*, 109(49):20006–20011, Dec. 2012.
- [32] R. A. Friesner, J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, M. Shelley, J. K. Perry, D. E. Shaw, P. Francis, and P. S. Shenkin. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *Journal of Medicinal Chemistry*, 47(7):1739–1749, Mar. 2004.
- [33] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, A. M. A. Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople. Gaussian 03, Revision C.02. 2003.
- [34] M. A. Glozak, N. Sengupta, X. Zhang, and E. Seto. Acetylation and deacetylation of non-histone proteins. *Gene*, 363:15–23, Dec. 2005.
- [35] J. M. Gottesfeld and M. Pandolfo. Development of histone deacetylase inhibitors as therapeutics for neurological disease. *Future neurology*, 2009.
- [36] B. J. Grant, J. A. McCammon, and A. A. Gorfe. Conformational Selection in G-Proteins: Lessons from Ras and Rho. *Biophys. J.*, 99(11):L87–L89, Dec. 2010.
- [37] I. Gregoretto, Y.-M. Lee, and H. V. Goodson. Molecular Evolution of the Histone Deacetylase Family: Functional Implications of Phylogenetic Analysis. *J. Mol. Biol.*, 338(1):17–31, Apr. 2004.
- [38] M. G. Guenther, O. Barak, and M. A. Lazar. The SMRT and N-CoR Corepressors Are Activating Cofactors for Histone Deacetylase 3. *Mol Cell Biol*, pages 6091–6101, 2001.
- [39] A. Hagelkruys, A. Sawicka, M. Rennmayr, and C. Seiser. The Biology of HDAC in Cancer: The Nuclear and Epigenetic Components. pages 13–37. *Histone*

- Deacetylases: the Biology and Clinical Implication, Berlin, Heidelberg, Aug. 2011.
- [40] T. A. Halgren, R. B. Murphy, R. A. Friesner, H. S. Beard, L. L. Frye, W. T. Pollard, and J. L. Banks. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *Journal of Medicinal Chemistry*, 47(7):1750–1759, Mar. 2004.
- [41] D. Hamelberg, J. Mongan, and J. A. McCammon. Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules. *J. Chem. Phys.*, 120(24):11919, 2004.
- [42] D. Hamelberg, T. Shen, and J. A. McCammon. Relating kinetic rates and local energetic roughness by accelerated molecular-dynamics simulations. *J. Chem. Phys.*, 122(24):241103, 2005.
- [43] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins*, 65(3):712–725, Nov. 2006.
- [44] K. Huber, G. Doyon, J. Plaks, E. Fyne, J. W. Mellors, and N. Sluis-Cremer. Inhibitors of histone deacetylases: correlation between isoform specificity and reactivation of HIV type 1 (HIV-1) from latently infected cells. *J. Biol. Chem.*, 286(25):22211–22218, June 2011.
- [45] T. Huber, A. E. Torda, and W. F. van Gunsteren. Local elevation: a method for improving the searching properties of molecular dynamics simulation. *J Comput Aided Mol Des*, 8(6):695–708, Dec. 1994.
- [46] W. Humphrey, A. Dalke, and K. Schulten. VMD: Visual molecular dynamics. *J Mol Graph*, 14(1):33–38, 1996.
- [47] S. Inoue. Inhibition of Histone Deacetylase Class I but not Class II Is Critical for the Sensitization of Leukemic Cells to Tumor Necrosis Factor-Related Apoptosis-Inducing Ligand-Induced Apoptosis. *Cancer Research*, 66(13):6785–6792, July 2006.
- [48] J. J. Irwin, T. Sterling, M. M. Mysinger, E. S. Bolstad, and R. G. Coleman. ZINC: A Free Tool to Discover Chemistry for Biology. *J Chem Inf Model*, 52(7):1757–1768, July 2012.
- [49] H. Jia, J. Pallos, V. Jacques, A. Lau, and B. Tang. Histone deacetylase (HDAC) inhibitors targeting HDAC3 and HDAC1 ameliorate polyglutamine-elicited phenotypes in model systems of Huntington’s disease. . . . *of disease*, 2012.
- [50] W. L. Jorgensen. The Many Roles of Computation in Drug Discovery. *Science*, 303(5665):1813–1818, Mar. 2004.

- [51] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79(2):926–935, 1983.
- [52] I. S. Joung and T. E. Cheatham. Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *J Phys Chem B*, 112(30):9020–9041, July 2008.
- [53] K. Kappel, J. Wereszczynski, R. T. Clubb, and J. A. McCammon. The binding mechanism, multiple binding modes, and allosteric regulation of *Staphylococcus aureus* Sortase A probed by molecular dynamics simulations. *Protein Sci.*, 21(12):1858–1871, Nov. 2012.
- [54] M. Karplus. Molecular dynamics of biological macromolecules: A brief history and perspective. *Biopolymers*, 68(3):350–358, Feb. 2003.
- [55] M. Karplus and J. A. McCammon. Molecular dynamics simulations of biomolecules. *Nat Struct Mol Biol*, 9(9):646–652, Sept. 2002.
- [56] M. Karplus and D. L. Weaver. Protein-folding dynamics. *Nature*, 260(5550):404–406, Apr. 1976.
- [57] D. Kern and E. R. Zuiderweg. The role of dynamics in allosteric regulation. *Curr. Opin. Struct. Biol.*, 13(6):748–757, Dec. 2003.
- [58] P. V. Klimovich and D. L. Mobley. Predicting hydration free energies using all-atom molecular dynamics simulations and multiple starting conformations. *J Comput Aided Mol Des*, 24(4):307–316, Apr. 2010.
- [59] P. Kollman. Free energy calculations: Applications to chemical and biochemical phenomena. *Chem. Rev.*, 93(7):2395–2417, Nov. 1993.
- [60] D. Kozakov, D. R. Hall, G.-Y. Chuang, R. Cencic, R. Brenke, L. E. Grove, D. Beglov, J. Pelletier, A. Whitty, and S. Vajda. Structural conservation of druggable hot spots in protein–protein interfaces. *Proceedings of the ...*, 2011.
- [61] T. J. Lane, D. Shukla, K. A. Beauchamp, and V. S. Pande. To milliseconds and beyond: challenges in the simulation of protein folding. *Curr. Opin. Struct. Biol.*, 23(1):58–65, Feb. 2013.
- [62] M. Lawrenz, J. Wereszczynski, R. Amaro, R. Walker, A. Roitberg, and J. A. McCammon. Impact of calcium on N1 influenza neuraminidase dynamics and binding free energy. *Proteins*, 78(11):2523–2532, Aug. 2010.
- [63] H. Li, M. Fajer, and W. Yang. Simulated scaling method for localized enhanced sampling and simultaneous “alchemical” free energy simulations: A general method for molecular mechanical, quantum mechanical, and quantum mechanical/molecular mechanical simulations. *J. Chem. Phys.*, 126(2):024106, 2007.

- [64] J. Li, J. Wang, J. Wang, Z. Nawaz, J. M. Liu, J. Qin, and J. Wong. Both corepressor proteins SMRT and N-CoR exist in large protein complexes containing HDAC3. *EMBO J.*, 19(16):4342–4350, Aug. 2000.
- [65] 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. *Proteins*, 78(8):1950–1958, June 2010.
- [66] H. Liu, A. Mark, and W. F. van Gunsteren. Estimating the relative free energy of different molecular states with respect to a single reference state. *J. Phys. Chem.*, 100(22):9485–9494, 1996.
- [67] P. Marks. Histone deacetylases. *Curr Opin Chem Biol*, 3(4):344–351, Aug. 2003.
- [68] P. R. L. Markwick, G. Bouvignies, and M. Blackledge. Exploring Multiple Timescale Motions in Protein GB3 Using Accelerated Molecular Dynamics and NMR Spectroscopy. *J. Am. Chem. Soc.*, 129(15):4724–4730, Apr. 2007.
- [69] J. A. McCammon, B. R. Gelin, and M. Karplus. Dynamics of folded proteins. *Nature*, 267(5612):585–590, June 1977.
- [70] D. Min and W. Yang. A divide-and-conquer strategy to improve diffusion sampling in generalized ensemble simulations. *J. Chem. Phys.*, 128(9):094106, 2008.
- [71] S. Minucci and P. G. Pelicci. Histone deacetylase inhibitors and the promise of epigenetic (and more) treatments for cancer. *Nat. Rev. Cancer*, 6(1):38–51, Jan. 2006.
- [72] J. MONOD, J. WYMAN, and J. P. Changeux. ON THE NATURE OF ALLOSTERIC TRANSITIONS: A PLAUSIBLE MODEL. *J. Mol. Biol.*, 12:88–118, May 1965.
- [73] J. Oberoi, L. Fairall, P. J. Watson, J.-C. Yang, Z. Czimmerer, T. Kampmann, B. T. Goult, J. A. Greenwood, J. T. Gooch, B. C. Kallenberger, L. Nagy, D. Neuhaus, and J. W. R. Schwabe. Structural basis for the assembly of the SMRT/NCoR core transcriptional repression machinery. *Nat Struct Mol Biol*, 18(2):177–184, Jan. 2011.
- [74] M. Olsson and C. R. Søndergaard. PROPKA3: consistent treatment of internal and surface residues in empirical p K a predictions. *Journal of Chemical ...*, 2011.
- [75] D. S. Palmer, A. I. Frolov, E. L. Ratkova, and M. V. Fedorov. Toward a universal model to calculate the solvation thermodynamics of druglike molecules: the importance of new experimental databases. *Mol. Pharm.*, 8(4):1423–1429, Aug. 2011.

- [76] M. F. Perutz, M. G. ROSSMANN, A. F. CULLIS, H. MUIRHEAD, G. WILL, and A. C. NORTH. Structure of haemoglobin: a three-dimensional Fourier synthesis at 5.5-Å resolution, obtained by X-ray analysis. *Nature*, 185(4711):416–422, Feb. 1960.
- [77] M. B. Peters, Y. Yang, B. Wang, L. Füsti-Molnár, M. N. Weaver, and K. M. Merz. Structural Survey of Zinc Containing Proteins and the Development of the Zinc AMBER Force Field (ZAFF). *J Chem Theory Comput*, 6(9):2935–2947, Sept. 2010.
- [78] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé, and K. Schulten. Scalable molecular dynamics with NAMD. *J Comput Chem*, 26(16):1781–1802, 2005.
- [79] L. C. T. Pierce, P. R. L. Markwick, J. A. McCammon, and N. L. Doltsinis. Accelerating chemical reactions: exploring reactive free-energy surfaces using accelerated ab initio molecular dynamics. *J. Chem. Phys.*, 134(17):174107, May 2011.
- [80] L. C. T. Pierce, R. Salomon-Ferrer, C. Augusto F de Oliveira, J. A. McCammon, and R. C. Walker. Routine Access to Millisecond Time Scale Events with Accelerated Molecular Dynamics. *J Chem Theory Comput*, 8(9):2997–3002, Sept. 2012.
- [81] B. G. Pogo, V. G. Allfrey, and A. E. Mirsky. RNA synthesis and histone acetylation during the course of gene activation in lymphocytes. . . . *of Sciences of the United States . . .*, 1966.
- [82] J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen. Numerical integration of the Cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *J. Comput. Phys.*, 23:327–341, 1977.
- [83] G. M. Sastry, M. Adzhigirey, T. Day, R. Annabhimoju, and W. Sherman. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J Comput Aided Mol Des*, 27(3):221–234, Mar. 2013.
- [84] J. Shao, S. W. Tanner, and N. Thompson. Clustering molecular dynamics trajectories: 1. Characterizing the performance of different clustering algorithms. *Journal of Chemical . . .*, 2007.
- [85] D. E. Shaw, K. J. Bowers, E. Chow, M. P. Eastwood, D. J. Ierardi, J. L. Klepeis, J. S. Kuskin, R. H. Larson, K. Lindorff-Larsen, P. Maragakis, M. A. Moraes, R. O. Dror, S. Piana, Y. Shan, B. Towles, J. K. Salmon, J. P. Grossman, K. M. Mackenzie, J. A. Bank, C. Young, M. M. Deneroff, and B. Batson. Millisecond-scale molecular dynamics simulations on Anton. In *SC '09*, pages 1–11, New York, New York, USA, 2009. ACM Press.

- [86] T. Shen and D. Hamelberg. A statistical analysis of the precision of reweighting-based simulations. *J. Chem. Phys.*, 129(3):034103, July 2008.
- [87] M. R. Shirts and J. D. Chodera. Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.*, 129(12):124105, Sept. 2008.
- [88] W. Sinko, C. A. F. de Oliveira, L. C. T. Pierce, and J. A. McCammon. Protecting High Energy Barriers: A New Equation to Regulate Boost Energy in Accelerated Molecular Dynamics Simulations. *J Chem Theory Comput*, 8:17–23, Nov. 2011.
- [89] C. C. Spurling, C. A. Godman, E. J. Noonan, T. P. Rasmussen, D. W. Rosenberg, and C. Giardina. HDAC3 overexpression and colon cancer cell proliferation and differentiation. *Mol. Carcinog.*, 47(2):137–147, Feb. 2008.
- [90] T. P. Straatsma and J. A. McCammon. Computational Alchemy. *Annu. Rev. Phys. Chem.*, 43(1):407–435, Oct. 1992.
- [91] C. Tan, Y.-H. Tan, and R. Luo. Implicit Nonpolar Solvent Models. *J Phys Chem B*, 111(42):12263–12274, Oct. 2007.
- [92] S. Thangapandian, S. John, S. Sakkiyah, and K. W. Lee. Ligand and structure based pharmacophore modeling to facilitate novel histone deacetylase 8 inhibitor design. *European Journal of Medicinal Chemistry*, 45(10):4409–4417, Oct. 2010.
- [93] I. Van den Wyngaert, W. de Vries, A. Kremer, J. Neefs, P. Verhasselt, W. H. Luyten, and S. U. Kass. Cloning and characterization of human histone deacetylase 8. *FEBS Lett.*, 478(1-2):77–83, July 2000.
- [94] B. F. Volkman, D. Lipson, D. E. Wemmer, and D. Kern. Two-state allosteric behavior in a single-domain signaling protein. *Science*, 291(5512):2429–2433, Mar. 2001.
- [95] L. Votapka and R. E. Amaro. Multistructural hot spot characterization with FT-Prod. *Bioinformatics*, 29(3):393–394, Feb. 2013.
- [96] D.-F. Wang, P. Helquist, N. L. Wiech, and O. Wiest. Toward Selective Histone Deacetylase Inhibitor Design: Homology Modeling, Docking Studies, and Molecular Dynamics Simulations of Human Class I Histone Deacetylases. *Journal of Medicinal Chemistry*, 48(22):6936–6947, Nov. 2005.
- [97] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case. Development and testing of a general amber force field. *J Comput Chem*, 25(9):1157–1174, July 2004.
- [98] B. Waszkowycz. Towards improving compound selection in structure-based virtual screening. *Drug Discov. Today*, 13(5-6):219–226, Mar. 2008.

- [99] P. J. Watson, L. Fairall, G. M. Santos, and J. W. R. Schwabe. Structure of HDAC3 bound to co-repressor and inositol tetrakisphosphate. *Nature*, 481(7381):335–340, Jan. 2012.
- [100] Y. D. Wen, V. Perissi, L. M. Staszewski, W. M. Yang, A. Krones, C. K. Glass, M. G. Rosenfeld, and E. Seto. The histone deacetylase-3 complex contains nuclear receptor corepressors. *Proc. Natl. Acad. Sci. U.S.A.*, 97(13):7202–7207, June 2000.
- [101] J. Wereszczynski and J. A. McCammon. Statistical mechanics and molecular dynamics in evaluating thermodynamic properties of biomolecular recognition. *Q. Rev. Biophys.*, pages 1–25, Nov. 2011.
- [102] O. Witt, H. E. Deubzer, T. Milde, and I. Oehme. HDAC family: What are the cancer relevant targets? *Cancer Letters*, 277(1):8–21, May 2009.
- [103] R. Wu, S. Wang, N. Zhou, and Z. Cao. A proton-shuttle reaction mechanism for histone deacetylase 8 and the catalytic role of metal ions. *Journal of the American ...*, 2010.
- [104] W.-M. Yang, S.-C. Tsai, Y.-D. Wen, G. Fejer, and E. Seto. Functional domains of histone deacetylase-3. *J. Biol. Chem.*, 277(11):9447–9454, Mar. 2002.
- [105] S.-H. You, H.-W. Lim, Z. Sun, M. Broache, K.-J. Won, and M. A. Lazar. Nuclear receptor co-repressors are required for the histone-deacetylase activity of HDAC3 in vivo. *Nat Struct Mol Biol*, 20(2):182–187, Jan. 2013.
- [106] M. Zacharias, T. P. Straatsma, and J. A. McCammon. Separation-shifted scaling, a new scaling method for Lennard-Jones interactions in thermodynamic integration. *J. Chem. Phys.*, 100(12):9025–9031, 1994.
- [107] J. Zhang, M. Kalkum, B. T. Chait, and R. G. Roeder. The N-CoR-HDAC3 nuclear receptor corepressor complex inhibits the JNK pathway through the integral subunit GPS2. *Molecular Cell*, 9(3):611–623, Mar. 2002.
- [108] Y. Zhang, H. Fang, J. Jiao, and W. Xu. The structure and function of histone deacetylases: the target for anti-cancer therapy. *Curr. Med. Chem.*, 15(27):2840–2849, 2008.