Flexible Receptor Docking Method Development and Molecular Dynamics Studies Towards Targeting Dynamic Protein Surfaces

by

Jessica K. Gagnon

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Chemistry)
The University of Michigan
2015

Doctoral Committee:
Professor Charles L. Brooks III, Chair
Professor Heather A. Carlson
Professor Kevin J. Kubarych
Professor Anna K. Mapp

Dedication

For my loving parents,

Agnes and Jamie Gagnon

Acknowledgements

I would like to begin by thanking my incredible parents, Agnes and Jamie Gagnon, who have always been my biggest fans. I could not have succeeded in my degree ambitions without their love and support. I cannot thank them enough.

I would like to express my deepest gratitude to my PhD advisor, Charlie Brooks III (not 3rd) for being an amazing mentor, and challenging me to become a better scientist. He put up with my growing pains during my graduate school tenure and meticulously developed my skill set, both technical and in general. I would like to thank him for the invaluable training I received in his lab. I cannot say enough good things about him; he is truly my favorite PhD advisor I have ever had.

The Brooks research group came as part of the package of choosing Charlie as my advisor and I could not have asked to work with a better group of people. I would like to thank all of the past and present Brooks group members. I want to take a moment to thank Roger Armen, who first introduced me to my thesis project, and was always incredibly enthusiastic. I would not be at this point in my graduate experience without the mentorship by Dr. Jen Knight and Dr. Sean Law, who made incredible impact in my research and taught me what being a professional scientist means. Dr. Karunesh Arora also made huge contributions to my education, encouraged me to investigate ideas that I found interesting, and I must thank him for all his assistance. A special thanks to Dr.

Garrett Goh, who over my time in the Brooks lab has become my friend in addition to being my colleague, and graduate school would not have been nearly as interesting without him.

I have been very fortunate to have wonderful collaborators both within the Brooks group and outside. I would like to thank all of my collaborators along the way. I must send a special thanks to Prof. Anna Mapp and Prof. Malani Raghavan, for allowing me to work with them on exciting biological questions. Within their research groups I had the pleasure to work with many people, but I would like to specifically thank: Dr. Ningkun Wang, who was a friend as well as a collaborator, Dr. Chinmay Majmudar, Dr. William Pomerantz, Jean Lodge, Dr. James Clayton, and Dr. Sanjeeva Wijeysakere.

I would like to thank my wonderful committee, who had the difficult role of advising me through out graduate school. I appreciate the time they took to give me suggestions and help guide me throughout my tenure at Michigan. Thank you Prof. Anna Mapp, Prof. Heather Carlson and Prof. Kevin Kubarych for all of your assistance.

Finally, I would like to thank my support network outside of research and lab. I have had the benefit of many friends who have supported me along the way. Thank you, my good friends Dr. Lauren Soblosky and Amber Koenig, for commiserating with me regardless of the hour of day. I must thank my undergraduate advisor Prof. Jeff Schwehm, who continued to give me advice long after I graduated from Lakeland College. I want to give a very special thanks to my better half, Dr. Brenden Arruda, who has lent me his love and strength throughout this roller coaster that is graduate school.

Table of Contents

List of Figures

Chapter 5

List of Tables

Table

Chapter 1

Introduction

1.1 Computer aided drug design

Drug-discovery is important but extremely challenging. Between the years 1950-2008 only 1,200 new biologically active small-molecules were approved by the FDA at the incredible cost of $50 billion.[1-3] There has been an immense multidisciplinary effort to improve the attrition rates along the process, where approximately 35% discovery projects succeed in delivering dugs to the clinical development stage where chances are around 10% of reaching market.[4-6] Failures at different phases of the clinical trials are fairly consistent in their reason. For example, the lack of sufficient efficacy leads to about two-thirds of Phase III failures, more than half for Phase II and about sixteen percent of Phase I failures.[7-10] Discovery projects, however, fail for a variety of reasons, such as lack of leads and poor potency, or unexpected animal toxicity.[11] As the understanding of the complexities and intricacies involved in successful drug design increased the value of a multidisciplinary approach became more apparent. Drugs today arise through discovery programs that begin with an identification of a target with extensive biological, biochemical and structural studies to assess the therapeutic value.[2] Computational efforts are a large part of those programs.

In the infancy of computer-aided drug design the expectations for the new-technology exceeded the capabilities at the time. Regardless of challenges, over the past few decades computational chemistry has become entrenched in the drug discovery process. Frequently applied methods include quantative-structure activity relationships (QSAR), ligand or structure-based drug design, and there is great value in the prolific nature of available databases.[12-18] QSAR was one of the earliest contributions to the drug design process, with frontrunners such as Corwin Hansch and Yvonne Martin demonstrating a consistent relationship between a small-molecule's *in vivo* biological activity and the log of the octanol/water partition coefficient (logP).[19,20] Predicting *in vivo* activity is important, as the most potent drug *in vitro* is useless if it is never able to get to its target. Since the original logP relationship was outlined an incredible number of different QSAR equations have been reported in the literature. A wide variety of characteristics have been investigated such as molecular volume, surface area, and polar surface area.[21-23] This method continues to be important in drug discovery process.[2,11]

Predicting drug likeness using QSAR methods are a part of the drug discovery process but there is also a huge amount effort put into structure-based drug design (SBDD). The availability of structural data continues to increase at an incredibly rapid rate, as seen in the astonishing number of structures now available in the Protein Data Bank (PDB). SBDD methodology depends on the availability of high quality structural and biochemical data and efforts to collect such information in one place have been made. A plethora of databases have been procured such as the Ligand-Protein Databank (LPDB), the Protein-Protein Interaction Inhibitor (2P2I) database, and the Mother of All Databases (MOAD).[12,13,16,17] Recently, there was an entire issue of Nucleic Acids

Research that was dedicated to the databases available online as of the beginning 2015.[14] With the advancements in structural information and computer hardware, advancements in computational software have attempted to keep pace. A variety of methodologies have been developed to predict possible high-affinity small-molecules for targets and their orientation within a binding pocket. Methods to enumerate possible modifications to a lead have become quite advanced, with abilities to predict relative affinities to a known compound within a few kcal/mol error or calculate free energies of binding with great accuracy.[24-33]

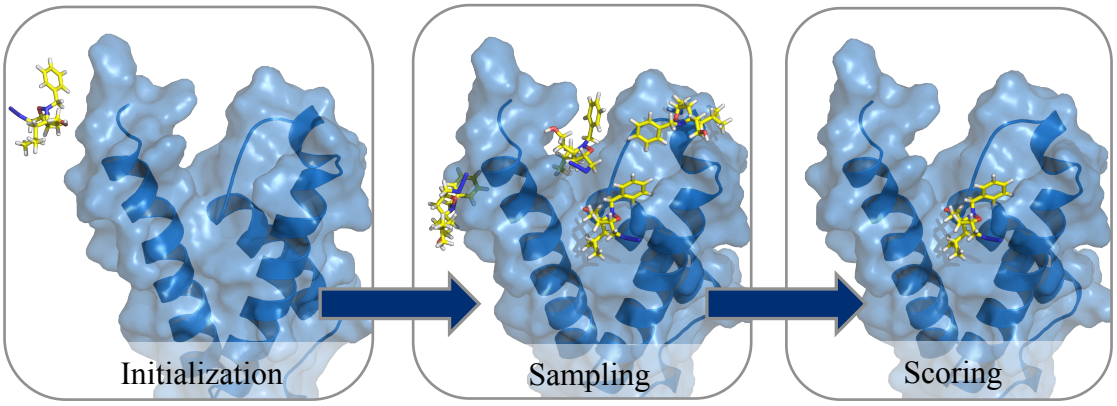


Figure 1.1: Docking methodologies generally can be broken down into these three steps. Initialization describes the steps required to prepare the docking trials such as gather parameters and coordinates for the complex. The sampling step is the approach taken to explore the conformational space, and the scoring step ranks the resulting conformations from the sampling.

There has been much advancement in different areas of SBDD and the area of docking is no exception. Among first docking software to become available was called DOCK from the Kuntz research group in the 1980s, which still is at the forefront of docking software today with a recent release of DOCK6.[34,35] Other software packages include, but are by no means limited to: Autodock, ICM, Glide, ROSETTALIGAND and CDOCKER.[36-43] Structure-based docking uses available structural data for the protein

and attempts to identify compounds have high affinity for a target and, generally, how those small-molecules would orient themselves on the receptor surface. A schematic of the three basic components of docking methodology is illustrated in Figure 1.

A frequently overlooked, but extremely important part of a docking trial is the initialization step. This first step is where the coordinates are prepared for the ligand and receptor, protonation states are declared, and parameters are gathered. The execution of this step can seriously impact the docking accuracy of any docking methodology being implemented where a small error can cause fluctuations of greater than 20% in the docking success rates.[44-46] For assessing docking software limitations in this area and in general there is great need for high quality structural and biochemical data.[47] Challenges arise in generating high quality data sets, due to the need for consistency in the experimental conditions across the set in addition to the need for high resolution structural data. There have been several docking and scoring challenges put out recently, including two by Nicholis and Jain in 2007 and 2011.[48,49] It was a goal of the CSAR benchmarking exercises to provide the docking community with high-quality datasets that were also blind trials for the participants.[50-53]

The Docking and Scoring Challenges and CSAR benchmarking exercises provided grounds to test different docking software sampling, and scoring methodologies.[48-53] The sampling step is a challenging balance of search all the conformational space available to the small molecule while maintaining computational efficiency. There are many different approaches to achieve the efficiency; the primary method is to use some reduction on the degrees of freedom for the receptor. Traditionally, in docking protocols the receptor was maintained as a rigid entity. For example,

CDOCKER uses a series of nonbonded grids to represent the receptor. The grids are pre-calculated for a single receptor prior to the docking trials, and while it is initially a time consuming calculation it only has to be performed once for each receptor. These grids then replace the computationally expensive calculations of van der Waals and electrostatic energies with essentially simple look-up tables. Autodock uses a very similar method where grids of various interaction energies are pre-calculated prior to the docking trials while the program DOCK uses affinity maps as their receptor representation.[34-36,41]

This approximation of the receptor saves much on computational cost for the docking calculation, but there are still an enormous number of degrees of freedom available to a ligand that must in some way be sampled. The major classes of sampling methods fall into different flavors of Monte Carlo (MC), evolutionary or genetic algorithms (EA/GA), and molecular dynamics (MD). MC methods are popular in a wide variety of applications, from politics and economics to a wide array of scientific fields. MC methods have a long history, but much of the modern MC implementations have roots in the MC methods using Markov chains pioneered by the Los Alamos team of Fermi, Von Neumann, Ulam, Metropolis, Teller, and others. The most famous algorithm is that of Metropolis, A.W. Rosenbuth, M. N. Rosenbuth, A.H. Teller and E. Teller in 1953.[54,55] Modern MC methods efficiently generate conformational ensembles, and methods that incorporate a simulated annealing sampling protocol to heat up and cool the system assist in identifying low-energy regions. Autodock adopted a simulated annealing MC method very early on incorporated into docking programs like Autodock and ICM.[36,37,56]

Genetic algorithms are another stochastic approach that was established around the 1970s and has become quite popular in docking methodologies.[57-59] A GA approach for docking can be thought of as evolving a population of ligand conformations towards low energy as defined by a fitness function.[60] Genes encode translational, rotational and internal degrees of freedom for the ligand. Each chromosome then represents a ligand conformation, which in practice is a string of numbers encoding for the degrees of freedom. A new population is generated from the previous generation through genetic operators: mutation, crossover exchanges and migration. A single mutation randomly changes a single gene, or degree of freedom. Crossover exchanges switch one or more genes between chromosomes, conformations, in the same population. Migration takes the least fit conformation from one population and switches it with the most fit individual from another population. These events are used at differing weights to obtain new ligand conformations, and have been implemented in the CHARMM docking methodology as well as Autodock.[41,42]

The CHARMM docking methodology (CDOCKER) is the primary focus of the development portion, Chapter 2, of this dissertation. GA and MC have been implemented in previous versions of CDOCKER, but the primary sampling method currently is MD.[40,42] MD has evolved from the study of simple hard spheres in the 1950s to massive biological systems with numerous advanced capablities.[61,62] The CHARMM simulation package contains a wide variety of capabilities with many solvent approximations, methodology to perform alchemical transformations, and a plethora of advanced sampling techinques.[40,63-70] MD is rooted in the fundamental laws of physics to shed light on structure, function and properties of complex systems. MD simulations take detailed

force field and integrates over Newton's equations of motion to evolve the bio-macromolecule system with time.[70,71] The CDOCKER methodology utilizes simulated annealing MD protocol to sample the ligand on a grid representation of the receptor.[40,42,43] The heating phase of the annealing protocol allows for the ligand to escape a minimum, and the cooling relaxes the system to, ideally, a new minimum. This sampling method and its analog in MC have shown reasonable success in various docking programs. Contained in Chapter 2 is more on CDOCKER docking methods and development.

Once conformations of the small molecule are sampled they must be scored, both for selecting the "native" conformation and to predict the affinity of the molecule in a virtual screening application. Scoring functions for docking methodologies can generally be broken down into three classifications: force field/physics based, empirical, and knowledge based.[72-74] Early on there was no scoring functions specifically developed for the purpose of ranking protein-ligand interactions and combinations of different components of the force field were used. In versions of DOCK and Autodock a hydrogen bonding term was added to the electrostatic interaction.[34-36,41] There was addition of continuum and implicit solvent models for inclusion of solvation terms, where previously potential energies were calculated in gas phase.[64,75-77] Methods were developed using a fit of force field energies with known binding energies of ligands to account for systematic error, like the unfortunately named Linear Interaction Energies (LIE).[78,79]

The empirical scoring functions are frequently difficult to distinguish from a physics based scoring function, as both often decompose the components to protein-ligand binding in some way. Popular empirical methods include PLP, ChemScore, X-

Score, and Glide-Score.[80-84] Most empirical scoring functions rely on a regression analysis to derive the weighted factors for various energy contributions or penalties. The major differences from a physics-based scoring method is the empirical scoring is composed of a functional form built up using different contributions, while physics based methods borrows the complete theoretical framework, including the parameters and energy function.[73] Empirical scoring functions tend to include only common protein-ligand interactions, as less common interactions are ignored because they are not significant in the regression analysis. These types of scoring functions tend to do well when the ligands of interest have frequently observed properties.

The final major category is knowledge-based scoring functions. These scoring functions sum pairwise statistical potentials between protein and ligand and include scoring functions such as DrugScore, IT-Score, and KECSA. The distance dependent pairwise interaction is obtained through comparison to large data sets, if a particular pairwise interaction occurs with greater frequency than random it is scored as favorable; if it occurs less frequently that pair is scored as disfavorable. These scoring functions tend to be computationally inexpensive due to the pairwise distance nature of the calculation.

1.2 Outline of chapters

The CDOCKER methods in the future chapters use a physics-based scoring function, utilizing the CHARMM general force field (CGenFF) with the pre-calculated energies on the grid representation of the receptor.[40,70,71,85-87] Outlined in Chapter 2 are developments to the CDOCKER sampling methodology to improve docking accuracy

and incorporate flexible receptor. The previous implementation of CDOCKER tried to accommodate for some receptor variability by using a soft grid representation, where steric clash penalties a severely reduced.[40] However, with the increase in biophysical and structural data, it has become apparent that this soft grid representation is not sufficient for receptors that do not fit into the lock and key model of binding ligands.[88-93] To accommodate local rearrangements of a receptor surface given various ligands we have implemented explicit side chain representation of select residues in the presence of the grid, and is termed Flexible CDOCKER. This methodology should account for local rearrangements and larger scale motions that may occur upon ligand binding must be addressed in some other manner.[94-96] The Flexible CDOCKER implementation outperforms its rigid counterpart and is competitive with other docking software such as Glide and Autodock.[36,38,41,81]

Contained in Chapters 3 and 4 are application projects of CDOCKER and other CHARMM MD methods in collaboration with experimental groups to investigate protein-ligand interactions. Chapter 3 outlines efforts to identify a putative ATP binding site on calreticulin (CRT), an important chaperone protein involved in MHC Class I assembly. This work was done in collaboration with the Raghavan research group at the University of Michigan. Initial docking trials narrowed down target regions on the surface of CRT and the experimental group took the docking predictions as a guide to perform point mutations. The mutations narrowed down the likely location for ATP binding, and more localized computational docking study followed by MD simulations identified other key residues for both binding and catalysis. These computational

predictions were confirmed by experiment, demonstrating for the first time that CRT both binds and catalyzes ATP.

Collaborative efforts to target the GACKIX domain of CREB binding protein (CBP) with small molecules are described in Chapter 4. For these projects I worked with the Mapp research group at the University of Michigan.[97,98] CBP is a coactivator in transcription regulation, a process integral to many cell processes. To better understand transcription regulation and towards developing biochemical tools we investigated two major different classes of small-molecules to target this domain of CBP. We identified natural product small-molecules that were the first to show the ability to inhibit transcriptional activator domain peptides at two separate binding sites on the GACKIX domain. The docking simulations of this ligand were challenging and MD simulations were undertaken to better understand possible binding interactions with GACKIX. These simulations demonstrated that the small-molecule, natural product sekikiac acid could interact with GACKIX as a helical mimetic. Additionally, through the innovative experimental technique of tethering the Mapp research group was able to crystalize this dynamic protein for the first time. The small molecule when tethered is covalently bonded to the receptor through a disulfide cross-link. Docking was undertaken of another tethered ligand identified as promising candidate followed by MD simulations. The closest to average structure of this ligand in complex with GACKIX showed it interacted with GACKIX in a different orientation than the crystalized ligand, and this could disrupt crystal contacts, suggesting a reason why they were unsuccessful in obtaining quality crystals from that complex.

Chapter 5 continues on the theme of the GACKIX domain of CBP, however instead of investigating ligand-protein interactions, we investigated protein-protein interactions.[99] GACKIX is an interesting domain as it interacts with a large number of different transcriptional activator domains and has positive allostery between its two binding sites. We used Gō-like models tuned to the binary complexes of GACKIX to investigate the origin the allostery in the ternary complex. The results from these simulations showed that the binding of a ligand to one site results in a rigidifying of certain regions on the GACKIX domain. This yields a reduction in the states available to GACKIX causing a reduction in entropy of a binary complex, resulting in longer off rates in the ternary complex and an apparent increase in affinity of the second ligand. This work highlighted the importance of entropy, and a different view from the traditional structural/enthalpy based allostery.

Overall, these studies highlight the importance of receptor flexibility in binding ligands, both small-molecule and bio-macromolecule based. The development of the CDOCKER methodology to incorporate receptor flexibility was essential for targeting receptors such as the GACKIX domain. SBDD is an important part of drug discovery process, and consideration of receptor flexibility has become more computationally affordable with hardware and software development. The drug discovery process also depends on investigations towards understanding fundamental biochemical questions such as the role of ATP in MHC Class I assembly and the immune response or allosteric signaling.

1.3 References

(1) Munos, B. *Nat. Rev. Drug Discovery* **2009**, *8*, 959.
(2) Jorgensen, W. L. *Science* **2004**, *303*, 1813.
(3) Marshall, G. R. In *George, R. And R. Okun* 1987, p 193.
(4) Paul, S. M.; Mytelka, D. S.; Dunwiddie, C. T.; Persinger, C. C.; Munos, B. H.; Lindborg, S. R.; Schacht, A. L. *Nat. Rev. Drug Discovery* **2010**, *9*, 203.
(5) Corey, E. J.; Wipke, W. T. *Science* **1969**, *166*, 178.
(6) Firestone, R. A. *Nature reviews. Drug discovery* **2011**, *10*, 963.
(7) Arrowsmith, J. *Nat. Rev. Drug Discovery* **2011**, *10*, 1.
(8) Arrowsmith, J. *Nat. Rev. Drug Discovery* **2011**, *10*, 1.
(9) Kola, I.; Landis, J. *Nat. Rev. Drug Discovery* **2004**, *3*, 711.
(10) Schaefer, S.; Kolkhof, P. *Drug Discovery Today* **2008**, *13*, 913.
(11) Borhani, D. W.; Shaw, D. E. *J. Comput.-Aided Mol. Des.* **2012**, *26*, 15.
(12) Benson, M. L.; Smith, R. D.; Khazanov, N. A.; Dimcheff, B.; Beaver, J.; Dresslar, P.; Nerothin, J.; Carlson, H. A. *Nucleic Acids Res.* **2008**, *36*, D674.
(13) Bourgeas, R.; Basse, M.-J.; Morelli, X.; Roche, P. *PLoS One* **2010**, *5*.
(14) Galperin, M. Y.; Rigden, D. J.; Fernandez-Suarez, X. M. *Nucleic Acids Res.* **2015**, *43*, D1.
(15) Hendlich, M.; Bergner, A.; Gunther, J.; Klebe, G. *J. Mol. Biol.* **2003**, *326*, 607.
(16) Hu, L. G.; Benson, M. L.; Smith, R. D.; Lerner, M. G.; Carlson, H. A. *Proteins-Structure Function and Bioinformatics* **2005**, *60*, 333.
(17) Roche, O.; Kiyama, R.; Brooks, C. L. *J. Med. Chem.* **2001**, *44*, 3592.
(18) Wang, R. X.; Fang, X. L.; Lu, Y. P.; Wang, S. M. *J. Med. Chem.* **2004**, *47*, 2977.
(19) Van Drie, J. H. *J. Comput.-Aided Mol. Des.* **2007**, *21*, 591.
(20) Martin, Y. C. *J. Med. Chem.* **1981**, *24*, 229.
(21) Brown, R. D.; Martin, Y. C. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 572.
(22) Eriksson, L.; Jaworska, J.; Worth, A. P.; Cronin, M. T. D.; McDowell, R. M.; Gramatica, P. *Environ. Health Perspect.* **2003**, *111*, 1361.
(23) Martin, Y. C. *J. Med. Chem.* **1992**, *35*, 2145.
(24) Bash, P. A.; Singh, U. C.; Langridge, R.; Kollman, P. A. *Science* **1987**, *236*, 564.
(25) Bennett, C. H. *J. Comput. Phys.* **1976**, *22*, 245.
(26) Chodera, J. D.; Mobley, D. L.; Shirts, M. R.; Dixon, R. W.; Branson, K.; Pande, V. S. *Curr. Opin. Struct. Biol.* **2011**, *21*, 150.
(27) Guo, Z.; Brooks, C. L.; Kong, X. *J. Phys. Chem. B* **1998**, *102*, 2032.
(28) Hansen, N.; van Gunsteren, W. F. *J. Chem. Theory Comput.* **2014**, *10*, 2632.
(29) Jorgensen, W. L. *Acc. Chem. Res.* **1989**, *22*, 184.
(30) Knight, J. L.; Brooks, C. L. *J. Comput. Chem.* **2009**, *30*, 1692.
(31) Kollman, P. *Chem. Rev.* **1993**, *93*, 2395.
(32) Shivakumar, D.; Deng, Y. Q.; Roux, B. *J. Chem. Theory Comput.* **2009**, *5*, 919.
(33) Wang, L.; Wu, Y. J.; Deng, Y. Q.; Kim, B.; Pierce, L.; Krilov, G.; Lupyan, D.; Robinson, S.; Dahlgren, M. K.; Greenwood, J.; Romero, D. L.; Masse, C.; Knight, J. L.; Steinbrecher, T.; Beuming, T.; Damm, W.; Harder, E.; Sherman, W.; Brewer, M.; Wester, R.; Murcko, M.; Frye, L.; Farid, R.; Lin, T.; Mobley, D. L.; Jorgensen, W. L.; Berne, B. J.; Friesner, R. A.; Abel, R. *J. Am. Chem. Soc.* **2015**, *137*, 2695.
(34) Allen, W. J.; Balius, T. E.; Mukherjee, S.; Brozell, S. R.; Moustakas, D. T.; Lang, P. T.; Case, D. A.; Kuntz, I. D.; Rizzo, R. C. *J. Comput. Chem.* **2015**, *36*, 1132.

(35) Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. *J. Mol. Biol.* **1982**, *161*, 269.
(36) Trott, O.; Olson, A. J. *J. Comput. Chem.* **2010**, *31*, 455.
(37) Neves, M. A. C.; Totrov, M.; Abagyan, R. *J. Comput.-Aided Mol. Des.* **2012**, *26*, 675.
(38) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. *J. Med. Chem.* **2004**, *47*, 1739.
(39) Meiler, J.; Baker, D. *Proteins-Structure Function and Bioinformatics* **2006**, *65*, 538.
(40) Wu, G. S.; Robertson, D. H.; Brooks, C. L.; Vieth, M. *J. Comput. Chem.* **2003**, *24*, 1549.
(41) Goodsell, D. S.; Morris, G. M.; Olson, A. J. *J. Mol. Recognit.* **1996**, *9*, 1.
(42) Vieth, M.; Hirst, J. D.; Dominy, B. N.; Daigler, H.; Brooks, C. L. *J. Comput. Chem.* **1998**, *19*, 1623.
(43) Vieth, M.; Hirst, J. D.; Kolinski, A.; Brooks, C. L. *J. Comput. Chem.* **1998**, *19*, 1612.
(44) Corbeil, C. R.; Williams, C. I.; Labute, P. *J. Comput.-Aided Mol. Des.* **2012**, *26*, 775.
(45) Corbeil, C. R.; Moitessier, N. *J. Chem. Inf. Model.* **2009**, *49*, 997.
(46) Warren, G. L.; Andrews, C. W.; Capelli, A. M.; Clarke, B.; LaLonde, J.; Lambert, M. H.; Lindvall, M.; Nevins, N.; Semus, S. F.; Senger, S.; Tedesco, G.; Wall, I. D.; Woolven, J. M.; Peishoff, C. E.; Head, M. S. *J. Med. Chem.* **2006**, *49*, 5912.
(47) Carlson, H. A.; Dunbar, J. B. *J. Chem. Inf. Model.* **2011**, *51*, 2025.
(48) Jain, A. N.; Nicholls, A. *J. Comput.-Aided Mol. Des.* **2008**, *22*, 133.
(49) Spitzer, R.; Jain, A. N. *J. Comput.-Aided Mol. Des.* **2012**, *26*, 687.
(50) Dunbar, J. B., Jr.; Smith, R. D.; Damm-Ganamet, K. L.; Ahmed, A.; Esposito, E. X.; Delproposto, J.; Chinnaswamy, K.; Kang, Y.-N.; Kubish, G.; Gestwicki, J. E.; Stuckey, J. A.; Carlson, H. A. *J. Chem. Inf. Model.* **2013**, *53*, 1842.
(51) Damm-Ganamet, K. L.; Smith, R. D.; Dunbar, J. B., Jr.; Stuckey, J. A.; Carlson, H. A. *J. Chem. Inf. Model.* **2013**, *53*, 1853.
(52) Dunbar, J. B., Jr.; Smith, R. D.; Yang, C.-Y.; Ung, P. M.-U.; Lexa, K. W.; Khazanov, N. A.; Stuckey, J. A.; Wang, S.; Carlson, H. A. *J. Chem. Inf. Model.* **2011**, *51*, 2036.
(53) Smith, R. D.; Dunbar, J. B., Jr.; Ung, P. M.-U.; Esposito, E. X.; Yang, C.-Y.; Wang, S.; Carlson, H. A. *J. Chem. Inf. Model.* **2011**, *51*, 2115.
(54) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087.
(55) Metropolis, N.; Ulam, S. *J. Am. Stat. Assoc.* **1949**, *44*, 335.
(56) Goodsell, D. S.; Olson, A. J. *Proteins-Structure Function and Genetics* **1990**, *8*, 195.
(57) Clark, D. E.; Westhead, D. R. *J. Comput.-Aided Mol. Des.* **1996**, *10*, 337.
(58) Holland, J. H. "Complex adaptive systems adaptation in natural and artificial systems an introductory analysis with applications to biology controls and artificial intellegence" 1992.

(59) Judson, R. S.; Jaeger, E. P.; Treasurywala, A. M.; Peterson, M. L. *J. Comput. Chem.* **1993**, *14*, 1407.
(60) Davis, L. "Handbook of genetic algorithms" 1991.
(61) Alder, B. J.; Wainwright, T. E. *J. Chem. Phys.* **1957**, *27*, 1208.
(62) de Groot, B. L.; Grubmuller, H. *Science* **2001**, *294*, 2353.
(63) Dickson, A.; Brooks, C. L., III *J. Phys. Chem. B* **2014**, *118*, 3532.
(64) Dominy, B. N.; Brooks, C. L. *J. Phys. Chem. B* **1999**, *103*, 3765.
(65) Feig, M.; Karanicolas, J.; Brooks, C. L. *J. Mol. Graphics Modell.* **2004**, *22*, 377.
(66) Spassov, V. Z.; Yan, L.; Szalma, S. *J. Phys. Chem. B* **2002**, *106*, 8726.
(67) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141.
(68) Wu, X.; Brooks, B. R. *J. Chem. Phys.* **2011**, *134*.
(69) Wu, X.; Brooks, B. R. *J. Chem. Phys.* **2011**, *135*.
(70) Brooks, B. R.; Brooks, C. L., III; Mackerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. *J. Comput. Chem.* **2009**, *30*, 1545.
(71) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.
(72) Bohm, H. J.; Stahl, M. In *Reviews in Computational Chemistry, Vol 18*; Lipkowitz, K. B., Boyd, D. B., Eds. 2002; Vol. 18, p 41.
(73) Liu, J.; Wang, R. *J. Chem. Inf. Model.* **2015**, *55*, 475.
(74) Muegge, I.; Rarey, M. *Reviews in Computational Chemistry, Vol 17* **2001**, *17*, 1.
(75) Lee, M. S.; Salsbury, F. R.; Brooks, C. L. *J. Chem. Phys.* **2002**, *116*, 10606.
(76) Gilson, M. K.; Davis, M. E.; Luty, B. A.; McCammon, J. A. *J. Phys. Chem.* **1993**, *97*, 3591.
(77) Gilson, M. K.; Given, J. A.; Head, M. S. *Chem. Biol.* **1997**, *4*, 87.
(78) Almlof, M.; Brandsdal, B. O.; Aqvist, J. *J. Comput. Chem.* **2004**, *25*, 1242.
(79) Rahaman, O.; Estrada, T. P.; Doren, D. J.; Taufer, M.; Brooks, C. L.; Armen, R. S. *J. Chem. Inf. Model.* **2011**, *51*, 2047.
(80) Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 425.
(81) Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. *J. Med. Chem.* **2006**, *49*, 6177.
(82) Murray, C. W.; Auton, T. R.; Eldridge, M. D. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 503.
(83) Verkhivker, G.; Appelt, K.; Freer, S. T.; Villafranca, J. E. *Protein Eng.* **1995**, *8*, 677.
(84) Wang, R. X.; Lai, L. H.; Wang, S. M. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 11.
(85) Yesselman, J. D.; Price, D. J.; Knight, J. L.; Brooks, C. L., III *J. Comput. Chem.* **2012**, *33*, 189.
(86) Mackerell, A. D.; Feig, M.; Brooks, C. L. *J. Comput. Chem.* **2004**, *25*, 1400.
(87) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; MacKerell, A. D., Jr. *J. Comput. Chem.* **2010**, *31*, 671.

(88) Carlson, H. A. *Curr. Opin. Chem. Biol.* **2002**, *6*, 447.
(89) Carlson, H. A.; McCammon, J. A. *Mol. Pharmacol.* **2000**, *57*, 213.
(90) Kuhn, L. A. *Strength in flexibility: Modeling side-chain conformational change in docking and screening*, 2008.
(91) Zavodszky, M. I.; Kuhn, L. A. *Protein Sci.* **2005**, *14*, 1104.
(92) Jorgensen, W. L. *Science* **1991**, *254*, 954.
(93) Bowman, G. R.; Geissler, P. L. *J. Phys. Chem. B* **2014**, *118*, 6417.
(94) Lorber, D. M.; Shoichet, B. K. *Protein Sci.* **1998**, *7*, 938.
(95) Bowman, A. L.; Nikolovska-Coleska, Z.; Zhong, H.; Wang, S.; Carlson, H. A. *J. Am. Chem. Soc.* **2007**, *129*, 12809.
(96) Lexa, K. W.; Carlson, H. A. *Q. Rev. Biophys.* **2012**, *45*, 301.
(97) Majmudar, C. Y.; Hojfeldt, J. W.; Arevang, C. J.; Pomerantz, W. C.; Gagnon, J. K.; Schultz, P. J.; Cesa, L. C.; Doss, C. H.; Rowe, S. P.; Vasquez, V.; Tamayo-Castillo, G.; Cierpicki, T.; Brooks, C. L., III; Sherman, D. H.; Mapp, A. K. *Angewandte Chemie-International Edition* **2012**, *51*, 11258.
(98) Wang, N.; Majmudar, C. Y.; Pomerantz, W. C.; Gagnon, J. K.; Sadowsky, J. D.; Meagher, J. L.; Johnson, T. K.; Stuckey, J. A.; Brooks, C. L., III; Wells, J. A.; Mapp, A. K. *J. Am. Chem. Soc.* **2013**, *135*, 3363.
(99) Law, S. M.; Gagnon, J. K.; Mapp, A. K.; Brooks, C. L. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 12067.

Chapter 2

CHARMM Docking Method Development

Text adapted from: J.K. Gagnon, S.M. Law, C.L. Brooks III. Flexible CDOCKER: Development and application of a pseudo-explicit structure-based docking method within CHARMM. *Submitted.*

2.1 Introduction

Structure-based protein-ligand docking is an important methodology in the repertoire of tools for drug discovery and design. This approach is often used to rapidly predict ligand orientation and affinity for virtual screening in lead identification or optimization applications.[1,2] A plethora of docking software has become available over the years, beginning with DOCK and including other programs such as Autodock, Glide and ICM.[3-6] There have also been developments toward improving the ability to predict affinities of possible lead molecules through improved knowledge-based scoring functions such as Drugscore, IT-Score, and KECSA.[7-10]

Traditionally, structure-based docking uses a rigid approximation of the receptor to reduce the number of degrees of freedom needing to be sampled and thus maintain computational efficiency. Speed in these calculations is desirable for lead discovery through virtual screening to achieve feasibility in screening the immense chemical space of small molecules. For example, the ZINC Database contains approximately 34 million purchasable compounds.[11] Additionally, this type of rigid receptor model represents a reasonable approach in the context of the lock and key type understanding of ligand binding as is important in many protein-ligand binding mechanisms.[12] However, with ever increasing structural and biophysical data available, the importance of protein flexibility has become more apparent.[13-17] Proteins exist as an inter-converting ensemble of states, and the highest populated state in the apo protein is often not the dominant state when a ligand is present. This has been captured experimentally, where throughout a series of crystallographic structures of the same protein side chain conformations can vary substantially.[17-20] For example, in alpha-Momorcharin (alpha-MMC) (Figure 1) there are different rotameric states for many of the side chains depending on the ligand in complex. Predicting the ligand conformation consistent with the experimental holo structure would not be successful when docking to the experimental apo structure with a rigid receptor docking method due to steric clashes.

Incorporation of protein flexibility has become more feasible due to advancements in computational resources as well as recent software development towards incorporating flexibility in docking.[1,16,21,22] Existing methods try to balance computational cost and the enormous size of the configurational space that needs to be sampled through a variety of approaches. These approaches include docking to a soft receptor, sampling the ligand conformations relative to an ensemble of protein conformers, or representing the protein as a single averaged grid.[23-25] Others

take advantage of rotamer libraries to optimize the receptor after sampling the ligand on a rigid representation of the receptor, or use a soft representation of the protein, where terms for steric clash penalties are reduced.[6,26-29] Though advances have been made in these areas, incorporation of receptor flexibility into docking remains at the forefront of challenges in the field.[1,23,24]

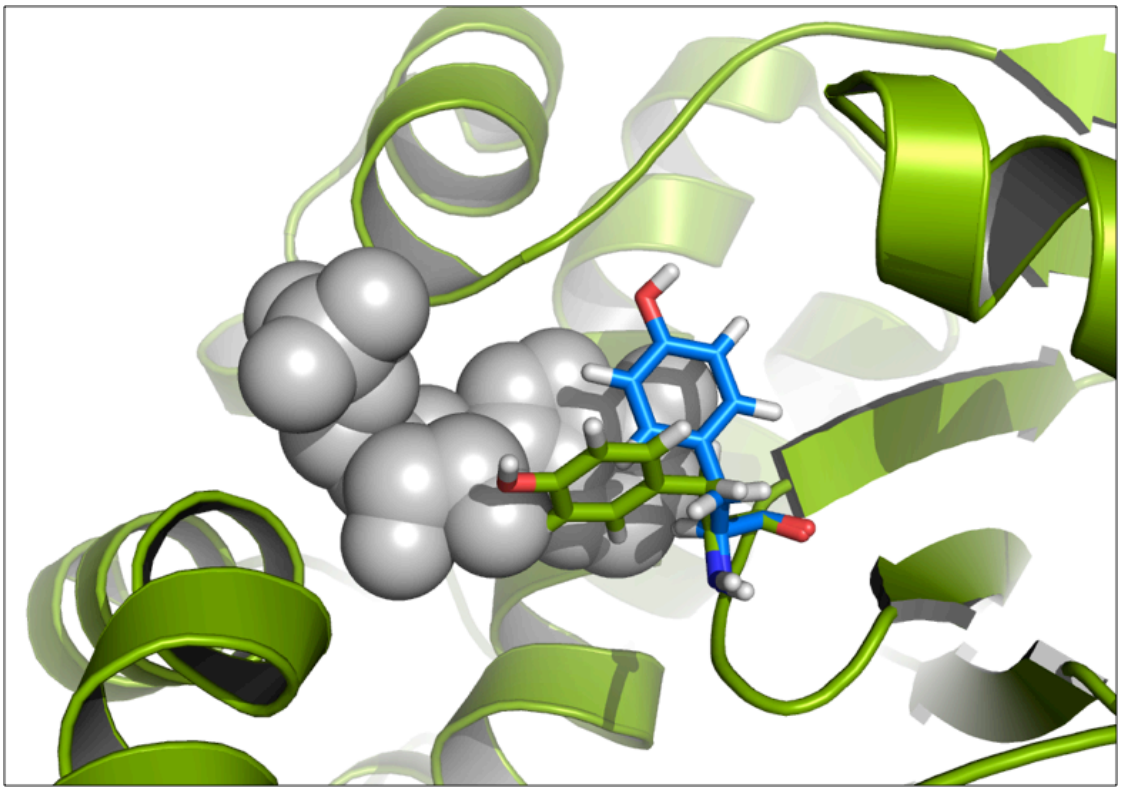


Figure 2.1: The importance of receptor flexibility has become more obvious with the increase of structural data. For instance, pictured here is an overlay of an apo (PDBID: 1AHC in green) and holo (PDBID: 1AHB in blue and grey) structure of alpha-Momorcharin. There is a steric clash between the ligand and the apo structure that would prevent rigid docking methods from predicted the crystallographic conformation.

In this paper we present improvements to the sampling in the CHARMM based docking method to include receptor flexibility. Previously, CHARMM docking, or CDOCKER, achieved ligand sampling via simulated annealing on a rigid representation of the receptor, although genetic algorithm methods have also been explored in this context.[29,30] In this setup, the protein

is represented as a series of non-bonded grids, where the non-bonded interactions are softened to accommodate small differences in protein structure when the ligand it bound to facilitate ligand sampling.[29,31] This representation works well for rigid pockets but may not be sufficient for a case like that portrayed in Figure 1. Nevertheless, the CDOCKER approach has served as a standard in the field for a number of years.[28-30,32]

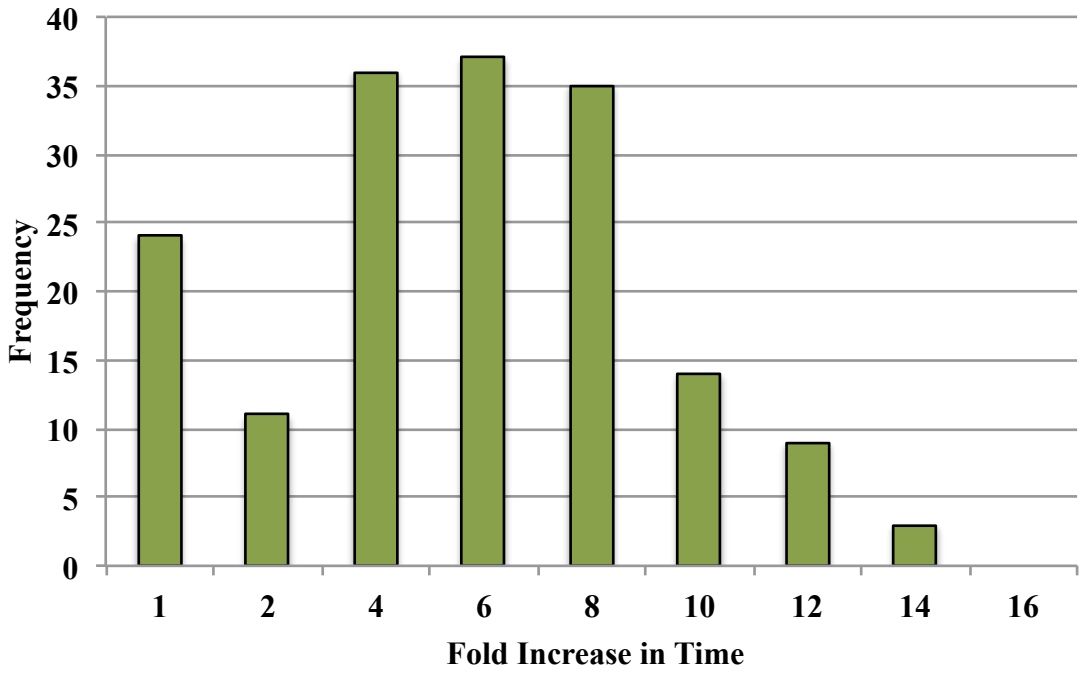


Figure 2.2: The distribution of the fold increase of docking time compared to rigid receptor docking method.

To extend the capabilities and scope of CDOCKER to accommodate a wider variety of receptors, we have incorporated receptor flexibility by including explicit side-chains while maintaining the rest of the receptor interactions using a grid representation. This allows the protein and ligand to sample configurational space simultaneously while having a minimal impact on the computational cost (Figure 2), which scales linearly with the number of explicit receptor atoms included in the sampling (Figure 3). In addition to incorporating receptor

flexibility we have implemented a docking protocol that employs enhanced sampling molecular dynamics followed by minimization (MD+Minimization) utilizing self-guided Langevin dynamics, which further improves CDOCKER docking accuracy.[33,34] These improvements to the CHARMM docking methods lead to docking accuracy that is competitive with other highly successful docking programs, such as Glide and Autodock.[5,6]

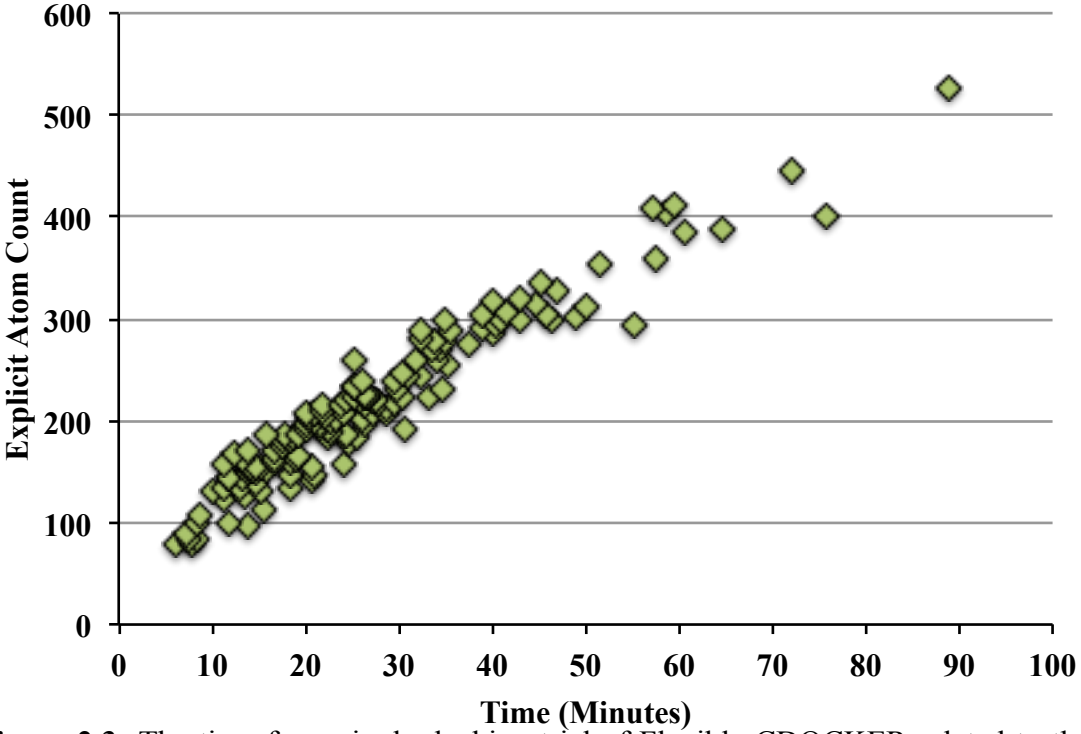


Figure 2.3: The time for a single docking trial of Flexible CDOCKER related to the number of explicit receptor atoms.

2.2 Methods

Setup of protein-ligand test set

For the docking trials a subset of 161 protein-ligand complexes from the CCDC/ASTEX set was used.[35] All structures were obtained directly from the Protein Data Bank (see Appendix A, Table 1 for PDBIDs).[36] The set was selected for direct comparison against results presented recently in publications from the Gohlke research group.[37] From the full set examined by Gohlke and co-workers, the entries in complex with heme or those for which parameters were not available within the CHARMM CGENFF36 parameter library were removed.[38-40] Single complexes were extracted from the downloaded PDBs using the MMTSB Toolset and all crystallographic waters and small molecules not representing the target ligand were removed. Only ions found in the region of the binding site were retained as part of the receptor.[41] The hydrogen atoms for the receptors and ligands were prepared independently of one another. The protonation states of the titratable residues in the protein were predicted using PROPKA version 3.1.[42] The hydrogen atoms were built using Babel version 2.3.2 for both the protein and ligand of each entry.[43] Ligand parameters were obtained using MATCH, which assigns atom types, charges and force field parameters through a chemical pattern-matching engine.[44] Symmetric models of the ligand, to assure uniquely named atoms that are of the same type do not contribute to higher RMSD, were generated using an in-house developed Perl script, mapPDB. This script is now in the MMTSB Toolset for use in root mean square deviation (RMSD) calculations.[41]

Docking protocol: Initialization of ligands and receptor grids

Prior to docking the small molecule into the receptor the system must be initialized by generating parameters, building the grid and generating initial configurations of the ligand. We developed a Perl script as an automated tool to interface user input with CHARMM for docking initialization. During initialization the receptor representation may be set up as either rigid or flexible. The ligand parameters are generated using MATCH at the beginning of docking setup. The grid representation of the protein employs a set of van der Waals and electrostatic grid-based potentials that are used in the sampling and ranking steps of our docking protocol. The van der Waals grids are built from a set of test particles with 20 different van der Waals radii that are centered on the most highly populated radii found in the CHARMM general force-field (CGenFF).[40] This grid is calculated with a mesh spacing of 1Å spanning 30Å centered at the specified center of mass for the target-binding pocket. In these trials the center of mass of the crystallographic conformation of the ligand was used. This grid size is sufficiently large to allow for a ligand to sample different rigid body rotations at the target binding site without a high energetic penalty. The "softness" of the grid can be varied for each non-bonded interaction: van der Waals, electrostatic attraction and repulsion, and follow the general form as given in equation 1.

$$E_{\mathrm{ij}} = E_{\max} - \mathrm{a}\left(\mathrm{r}_{\mathrm{ij}}^{\mathbf{b}}\right) \quad \text{if } \left|E_{\mathrm{ij}}^{*}\right| > \frac{E_{\max}}{2} \tag{1}$$

Where E_{ij}^{*} is the energy of the regular non-bonded potential at a given distance $\mathrm{r_{ij}}$ for either electrostatic or van der Waals.[28] Outlined in Table 1 are three different CDOCKER sampling protocols, each utilizing alternative grid representations with differing $E_{\max}$ values. The a and b coefficients are extracted from equations that express the potential and forces at the switching

distance. Grids were generated with a distance dependent dielectric with the relative dielectric constant, ε, equal to 3 as previously outlined by Vieth et al.[29] The grid soft-core potentials for the protocol with two grids, called here "SA- 2 Grids" for the simulated annealing protocol with two grids, were previously outlined in Wu, et. al.[28] The other implementations of the grid use a single grid with simulated annealing, "SA – 1 Grid", or our newly implemented molecular dynamics (MD) followed by minimization protocol, "MD+Minimization", which is described later in this section.

Protocol	$\boldsymbol{E_{max}}$ $\boldsymbol{(vdW)}$	$\boldsymbol{E_{max}}$ $\boldsymbol{(attractive)}$	$\boldsymbol{E_{max}}$ $\boldsymbol{(repulsive)}$
SA- 1 Grid	15.0	-120.0	-2.0
SA- 2 Grids	0.6	-0.4	8.0
	3.0	-20.0	40.0
MD+Minimization	15.0	-120.0	-2.0

Table 2.1: The parameters that determine the softness of the grid potential, all $E_{\max}$ values are in units of kcal/mol. The methods outlined here use a either a single grid of different sampling methods, simulated annealing (SA- 1 Grid) or molecular dynamics followed by minimization (MD+Minimization), or a two grids with a simulated annealing protocol as outlined by Wu, et. al.[28]

In previous versions of CDOCKER, the grid was generated with the entire protein held fixed, creating a rigid receptor, and will be referred to as rigid-docking (Figure 4A). To incorporate receptor flexibility, side-chains selected around the target binding-site are removed prior to grid generation. These residues are kept as explicit side chains during the sampling step and they are not present in the grid representation of the receptor (Figure 4B). The backbone atoms of these selected explicit residues are held fixed during all sampling. However, the side-chains of the explicit residues are allowed to sample configuration space simultaneously with the

ligand, and both ligand and side chains interact with the grid representation of the receptor. The residues that remain flexible can be user specified or the residues may be selected to be within a specified cut-off distance from the reference small-molecule, 3Å in these docking trials. These residues allow for changes in the surface of the binding site to better accommodate the small molecule. Incorporating explicit side chains allows for receptor flexibility while adding a minimal number of degrees of freedom to the sampling calculations. Additionally, side chain conformational sampling should provide sufficient receptor flexibility to overcome the clashes between the apo receptor structure and the small molecule crystallographic conformation, as illustrated in Figure 1. The targeting of large-scale conformational changes involving the receptor backbone that occur upon ligand binding in some situations would need to be addressed in some other manner, such as ensemble docking, and is beyond the scope of this flexible receptor docking approach.[22,25,45-47]

The initial configurations of the small molecule are generated through rigid body rotations and random rotations around the rotatable bonds. The energy for each configuration is calculated on the grid and all configurations below an energy cutoff of 1,000 kcal/mol are clustered. This removes the extremely high-energy conformations and also yields a structurally diverse set of starting conformations for docking. The clustering tool within the MMTSB toolset was used for the k-means based clustering of the configurations. The cutoff RMSD for the clustering was determined so that at least 50 clusters were generated. The lowest energy conformation from each of the lowest energy clusters was selected for the sampling steps. This protocol provides a diverse set of ligand conformations and also represents an ensemble of relatively low energy starting configurations from which to initialize sampling. Each docking

trial generates 50 ligand conformations starting from 50 unique, low energy conformations of the small molecule.

Docking protocol: Sampling

Flexible CDOCKER was implemented with two different sampling schemes; one uses a simulated annealing protocol while the other uses a series of short molecular dynamics simulations followed by minimization (MD+Minimization). Both implementations also utilize self-guided Langevin dynamics (SGLD) to accelerate the sampling.[33,34] SGLD uses an average of local velocities to calculate a guiding force to accelerate dynamics of lower frequency motions. There are two key parameters that determine how aggressive the acceleration to the system is: the average time and the guiding factor. The larger the average time, 0.2ps in these simulations, the slower the motion that is enhanced. The guiding factor is related to the guiding force and governed by the target self-guiding temperature, 700K for these simulations. The larger the guiding factor the larger the energy barriers that can be overcome.

For the re-docking trials using SA the protocol as previously outlined by Wu, et al was used.[28] For details on the temperatures and lengths of phases for the SA protocols see Appendix A, Table 2. SGLD is implemented for use with this protocol but was not used in the simulated annealing protocol re-docking trials, as it tends to cause the ligand to move out of the pocket because the sampling becomes too aggressive. Each small-molecule starting configuration was subjected to five repetitions of this protocol, where the end result of one simulated annealing step provided the starting configuration for the next. This protocol allows the ligand, and in the case

of flexible receptor docking the side chains, to sample different conformations and rotameric states.

The MD+Miniminization protocol was inspired by the successful ICM implementation of Monte Carlo steps followed by minimization, following the concept to guide the ligand down the energy landscape towards a low energy state corresponding to the crystallographic configuration.[4] Each ligand configuration from the clustering results is subjected to 20 rounds of 500 steps of molecular dynamics at 300K with SGLD, followed by 500 steps of Adopted Basis Newton-Raphson (ABNR) minimization. Before the start of each round an energy comparison between the current and the previous ligand configuration is made, and the lowest energy configuration is selected to continue the sampling, driving the system towards lower energy states. Each MD step is 2fs of enhanced sampling on the grid representation of the receptor. We found that in order to prevent the ligand from being ejected from the binding site fewer than 2,000 MD steps must be used. Additionally, to achieve high acceptance rates between rounds of sampling the number of MD steps fewer than 1,000 steps should be used.

Each docking trial for the MD+Minimization and the simulated annealing protocol was limited to 2.6×10^6 energy evaluations to allow the fairest comparison with the benchmark studies by Gohlke and co-workers.[37] The docking trials were repeated 10 times, with the same input structures for the protein and ligand. However, each trial generated a new set of random orientations and conformations of the small molecule followed by sampling on the grid representation of the receptor.

Docking protocol: Ranking and RMSD calculation

As the focus of this work is to develop a new docking search strategy, the energy evaluation on the grid representation of the receptor is used for sorting the resulting configurations from the sampling step. The RMSD between the docked configurations and their respective crystallographic configurations was calculated with consideration for the symmetry of the small molecule. A configuration of the ligand is considered "docked" when it has an RMSD less than 2Å. This metric, as is the convention, will be used to assess docking accuracy.[48]

Molecular dynamics simulations for investigating side-chain sampling

To investigate the ability of side chains to sample alternative rotameric states we carried out explicit solvent simulations of a few selected receptors as well as performed sampling in the presence of the grid. Simulations of two structures from the MMC alpha protein (PDBIDs: 1AHC and 1AHB) were run for comparison. The explicit solvent MD simulations were carried out using a GPU capable CHARMM version c39a2 with an OpenMM interface using OpenMM version 5.2.[49-51] The simulations are 5ns in length using a 2fs time step and were run at constant pressure and temperature, 300K, with the CHARMM27 force field.[38,39] Langevin dynamics was used to provide the thermal heat bath with a friction coefficient of 10 ps^{-1}. Particle mesh ewald with a non-bonded cutoff of 9Å was used with a FFT grid size of 72Å. SHAKE was applied to the hydrogen atoms and the parameters for the small molecules were obtained with MATCH using CGenFF.[40,44] The simulations run on the grid were also 5ns in duration with flexible residues selected within a 3Å cutoff of the ligand. This is the same as the cutoff as used for the docking trials. The grid parameters used were identical to those described in the

MD+Minimization protocol outlined in Table 1, as this is the "hardest" grid and would be the most constraining. To investigate side-chain sampling, both the explicit solvent and grid-based simulations were run in the absence of the ligand. Analysis of the χ angles was performed using CHARMM.

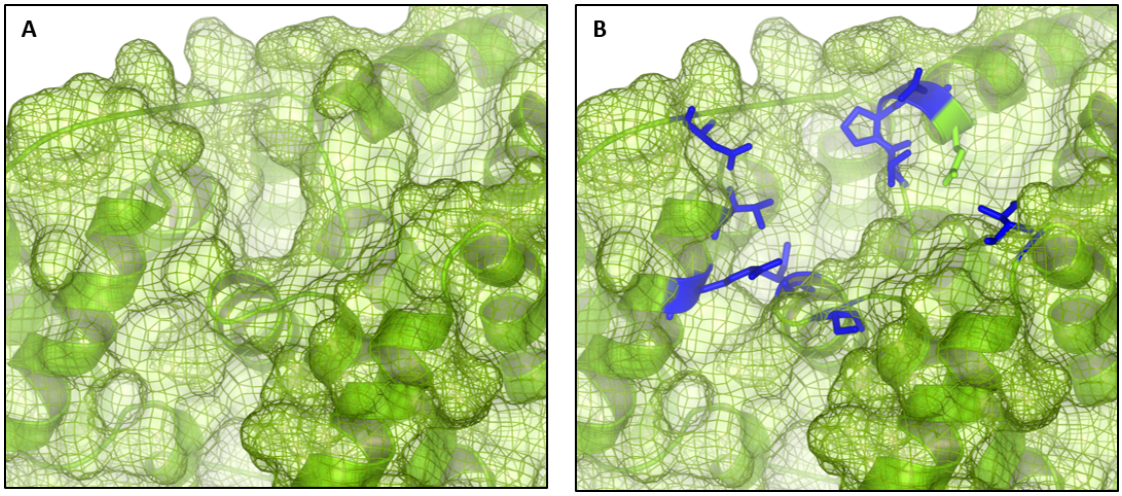


Figure 2.4: Rigid CDOCKER represents the receptor as a series of non-bonded grids (A). To include receptor flexibility, residues around a target binding site are kept as explicit side chains (B) allowing them to sample conformational space with the ligand simultaneously in the presence of the grids.

2.3 Results & discussion

Grid-embedded explicit side chains sample well the side chain rotameric states

To investigate the side chain sampling in the presence of the grid we compared the sampling of the side chains to that of an explicit solvent simulation. It was important to ensure that the side chains were able to sample alternate conformations and not be constrained by the surrounding grid. As has been noted in Figure 1, with the alpha MMC protein, a clash would occur between the tyrosine and the ligand in a cross-docking trial, and it is thus critical to ensure that both side-chain conformers are attainable during the course of a simulation on the grid-embedded side chain representation of the receptor. In Figure 5, we illustrate the $\chi 1$ and $\chi 2$ angles sampled in both explicit solvent and grid based simulations for the essential tyrosine residue noted in Figure 1. From the displayed configurational history maps we see more sampling on the grid than in the all-atom simulation. The triangles represent the angles observed in the starting structures. Indeed, independent of the starting conformation of the side chain, both conformations observed in the apo and holo-crystal structures are sampled. The grid-based simulations are able to sample the various conformations of the protein side chains without the enhanced sampling that is implemented in the docking protocol. As to be expected, even more sampling of alternate side-chain rotameric states occurs with the SGLD enhanced sampling. This sampling is important to consider because most applications of the docking methodology will be in cross-docking experiments one is trying to either improve the potency of a lead compound or identify a new lead compound.

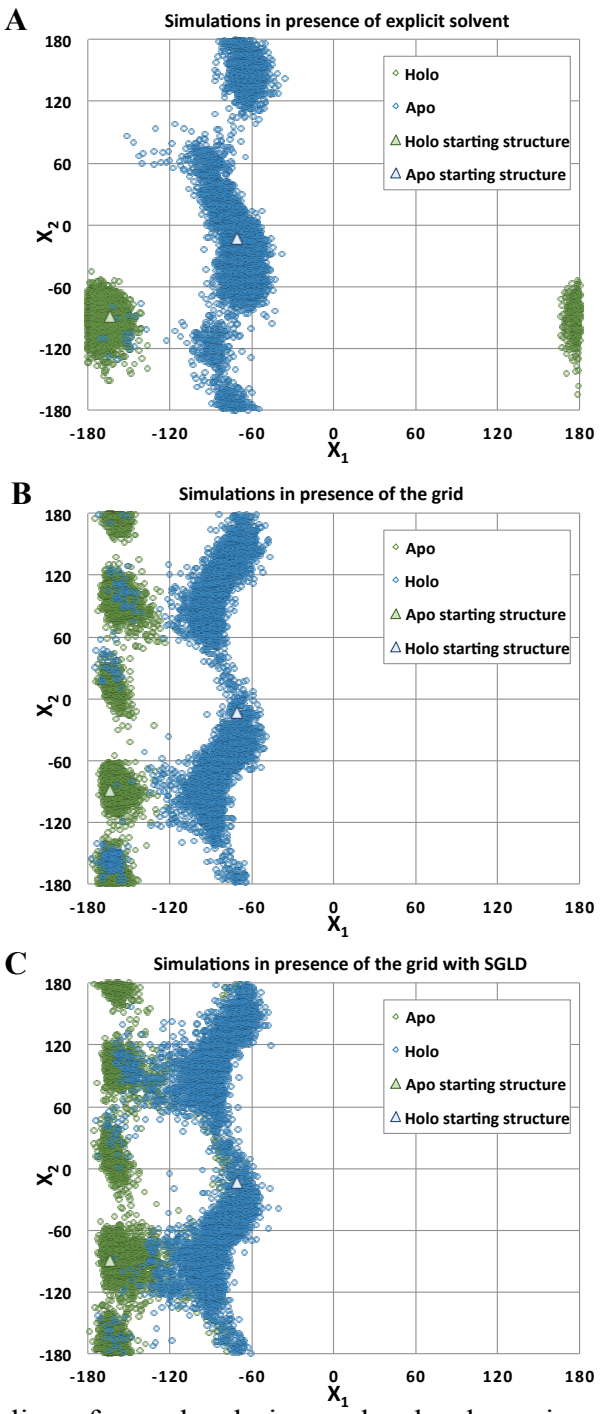


Figure 2.5: Sampling of χ angles during molecular dynamics simulation of a single tyrosine of alpha-Momorcharin protein in the presence of explicit solvent (A) or the grid (B) with enhanced sampling (C). The simulations started from the apo structure (PDBID: 1AHC) begin at the green triangle and sample the rotameric states shown in green circles. Alternatively, the blue triangle was the starting rotameric state for the holo structure (PDBID: 1AHB) and the states sampled are shown in blue circles.

Incorporation of SGLD MD and minimization in sampling protocol improves docking accuracy

CDOCKER originally used a simulated annealing protocol, where the ligand was heated in the presence of the grid, allowed to sample at the higher temperature and then cooled. This would push the ligand out of minima and, ideally, into new ones, occasionally sampling the desired low-energy, docked conformation. While such an approach has worked well in many situations, for a single docking trial, a native-like pose (closer than 2Å) was sampled in less than 60% of the complexes, and this is without scoring the conformations (see Table 2). To improve this step we implemented a MD and minimization approach, where a single starting ligand conformer is subjected to a short round of enhanced sampling SGLD molecular dynamics simulation and is followed up with a round of minimization. Each subsequent round of MD and minimization is started from either the conformation at the beginning or end of the previous round of sampling; whichever conformation is the lowest in energy. This protocol aims to reduce the time spent in high energy, irrelevant states that may be sampled during a simulated annealing procedure. This method improves docking accuracy for a single docking trial by about 16% when flexible side-chain sampling is employed (Table 2).

Table 2.2:

	Rigid CDOCKER percent docked	**Flexible CDOCKER percent docked**
Simulated Annealing	54.50%	56.10%
MD + Minimization	56.10%	72.40%

Docking to flexible receptor greatly improves re-docking accuracy

Re-docking trials are commonly used to assess the quality of a docking method. Re-docking takes a ligand and docks it into the same receptor structure with which it was originally crystallized. Presented here are re-docking trials of a subset of the CCDC/Astex set to compare various implementations of sampling and receptor flexibility within our novel extension of the CHARMM docking methodology.

The docking accuracy results for the re-docking trials are presented in Figure 6, and include a percentage found and percentage scored. When a conformation that is "docked", less than 2Å RMSD from the crystallographic conformation in the heavy atoms of the ligand, exists in the docking output it is considered "found". If a docked conformation is chosen to be the lowest energy conformation it is considered "scored", where the energy is that of the grid interacting with the ligand and, in the case of flexible docking, explicit receptor side-chains. Results are presented for 10 docking trials. To be included in the "found" percentage, at least one structure for a given complex in at least a single trial must be identified as docked. However, to be "scored" it must be the lowest predicted energy of all of the trials. For the rigid receptor docking, finding at least one docked conformation across all of the sampling methods occurred with high frequency, all greater than 81%. The docked conformation is selected out of the results at approximately half the frequency that the docked conformation is found. The MD and minimization (MD+Minimization) protocol improves docking accuracy in both finding and identifying a docked conformation with 41% scored with simulated annealing methods and 53.4% when MD+Minimization is used. For the single grid, the incorporation of flexibility in both simulated annealing and the MD and minimization protocols improves the scoring step by 8.7 and 9.3 percentage points respectively. This is likely due to some reorganization of the

protein side-chains to better accommodate the ligand and improve the overall energy. The two grid simulated annealing (SA – 2 Grids) protocol suffers greatly from the incorporation of the flexible side chains with a loss in the docked conformations found of 43.5 percentage points, which is also reflected in the scoring of 19.9%. Both of the grids in this SA – 2 Grids method are much softer than those used in the other protocols. While this allows for sampling of side chain and ligand conformations without large energetic penalties, it also causes loss of sensitivity in flexible receptor docking. It should be noted that this reduction of docking accuracy with the inclusion of flexible side chains is not unlikely for a re-docking trial, where additional receptor degrees of freedom added allow for sampling away from the crystallographic conformation predefined in the rigid docking case. However, the loss of docking accuracy seen in the SA – 2 Grids case was not observed in the MD+Minimization or the SA – 1 Grid protocol, which have "harder" grids and larger penalties for high energy conformations. In general, the addition of the flexible receptor further improves the percent "scored" and does not diminish the percent "found" for the single grid sampling protocols. The MD+Minimization protocol improves re-docking accuracy over the simulated annealing protocols, an improvement of 11.5 and 12.4 percentage points for the rigid and flexible docking respectively.

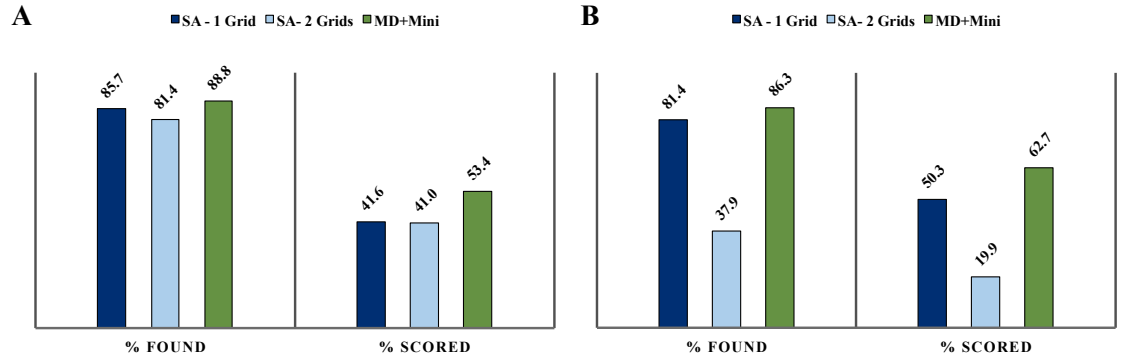


Figure 2.6: Re-docking trial results by metrics of finding at least one conformation in the ten trials that is less than 2Å RMSD from crystallographic pose (found) and the lowest energy conformation is docked (scored). Rigid CDOCKER method results are shown in panel A and flexible CDOCKER results are in panel B.

Flexible CDOCKER performs comparably with current docking software

Numerous docking methodologies have been developed in recent years and many useful comparisons of the re-docking ability of these software packages have been published.[37,52-59] Recently, the Gohlke research group published a comparative study of current docking programs for re-docking of ligand-receptor pairs in the CCDC/Astex dataset.[35] Taking the results reported by Krueger, et al. for the subset of complexes from the CCDC/Astex set reported above, comparisons between the popular docking methods and CDOCKER can be made (Figure 7).[37] In addition to reporting the Autodock results, these researchers reported Autodock results that were rescored using a knowledge based scoring function, DrugScore.[37] Comparing the best performing CDOCKER, the flexible receptor with the MD+minimization sampling, and the other docking methods that were reported, CDOCKER is very competitive with the other methods. Rigid CDOCKER with the MD+Minimization protocol is competitive, with 53.4% of the complexes scored, which is on par with Autodock 3 and Autodock 4 rescored with DrugScore, which show 52.4%, docked. Only Autodock Vina and Flexible CDOCKER outperform Rigid CDOCKER with this subset of CCDC/Astex set. Flexible receptor CDOCKER outperforms the

other docking programs, selecting the docked conformation for 62.7% of the complexes, while the next most successful docking software reported, Autodock Vina, has a success rate of 57.7%.

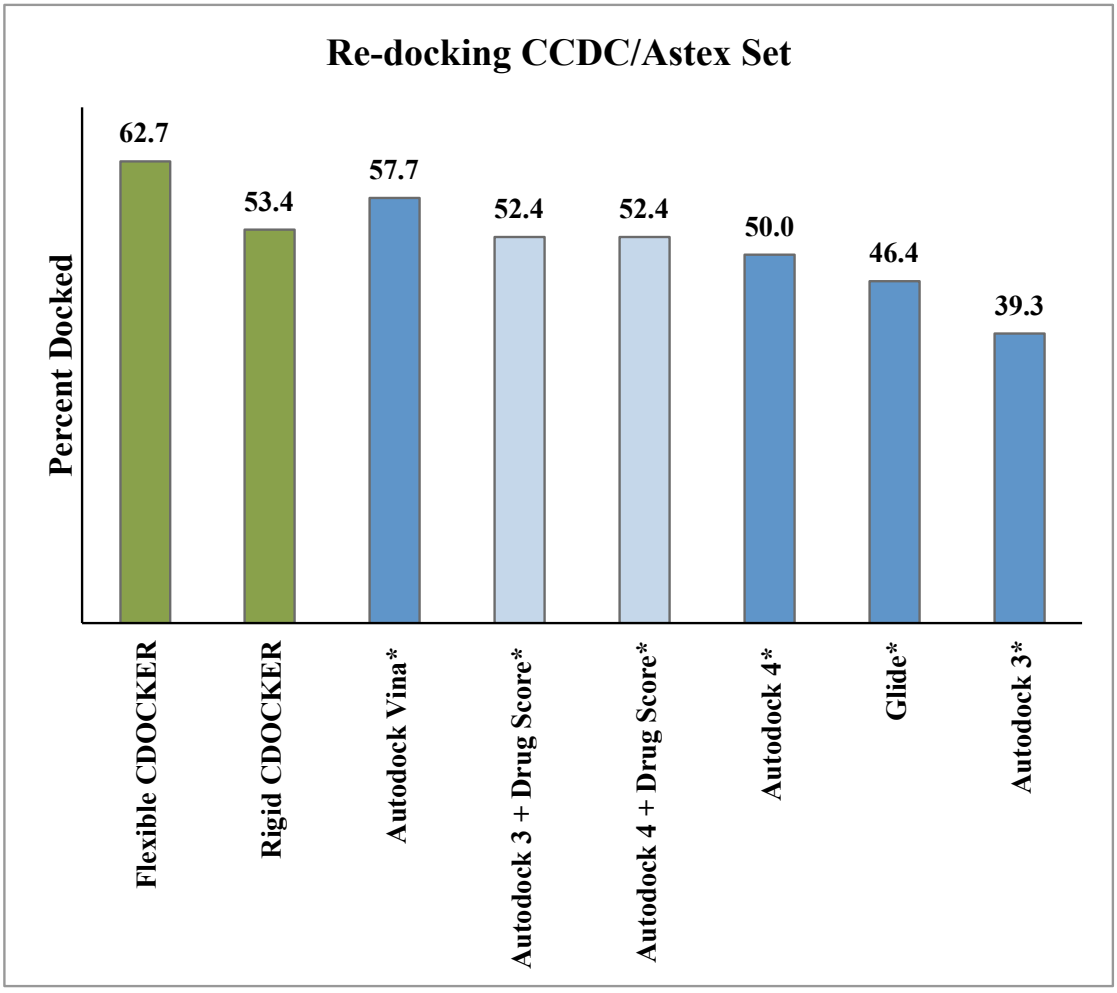


Figure 2.7:
Docking accuracy comparison of widely used docking programs and Flexible CDOCKER and CDOCKER for a subset of the CCDC/Astex. Docked is defined as the lowest predicted energy conformation is within 2Å RMSD of the crystallographic conformation. Results for the non-CDOCKER based methods are taken from the published results of Krueger et al, and are marked with an asterix.[37]

Flexible CDOCKER improves cross-docking accuracy: a case study

The primary motivation for incorporating receptor flexibility into docking was to target a receptor structure that requires side chain conformational flexibility to accommodate its ligand (Figure 1). As a case study, we performed re-docking and cross-docking experiments using the structures from Figure 1, utilizing the MD+Minimization implementation of CDOCKER. For proper comparison identical starting conformations of the small molecule were used for the sampling step of the docking trial with either rigid or flexible receptor representations and for both re-docking and cross-docking trials. Using the same small molecule conformations allows for investigation of the benefits of the sampling without concern for the variability in initial conditions. The re-docking trial docked the small-molecule to the receptor conformation from which it was derived (PDBID: 1AHB). The cross-docking trial docked the small-molecule to the experimentally determined apo receptor conformation (PDBID: 1AHC).

The results for these trials are presented in Table 3, and show the percentage of the resulting conformations, from the ten docking trials, that have a RMSD of less than or equal to 2Å, as well as the RMSD of the lowest predicted free energy conformation. The percentages docked shown in Table 3 are different from the percentages shown in Figure 7, which shows the percent of complexes that the lowest predicted energy conformation is docked. Consistent with the results observed for the CCDC/Astex set, rigid and flexible CDOCKER sample docked conformations with similar frequency, 10.6% and 10.2% of the resulting conformations (500 total) docked for rigid and flexible CDOCKER respectively for the re-docking trial. Flexible CDOCKER is superior in identifying the docked conformation; selecting a conformation with a 1.69Å RMSD while the rigid version's low energy conformation had an RMSD of 6.99Å. As expected, there is a drop in the frequency of sampling native-like poses for the cross-docking

trial. We observed docked conformations with a frequency of 5.6% for rigid CDOCKER and 7.2% for flexible CDOCKER of 500 total conformations. Interestingly, in the cross-docking trial rigid CDOCKER was able to sample a docked conformation of the small-molecule 5.6% of the time. This suggests the softness of the grid mimics some aspects of receptor flexibility and enables the sampling of docked conformations. However, for scoring flexible CDOCKER was superior, identifying a docked conformation of 1.42Å RMSD compared to 5.0Å RMSD for the rigid implementation. Flexible CDOCKER, in this test case, has no loss in sampling for the re-docking trial despite an increase in the number of degrees of freedom. It also shows improved sampling compared to the rigid implementation for the cross-docking trial. Flexible CDOCKER is superior to the rigid implementation in ranking docked conformations, leading to improved docking accuracy.

Table 2.3:

	Docking Method	**Percent of all conformations docked**	**Low Energy RMSD(Å)**
Re-docking	Rigid CDOCKER	10.6	6.99
	Flexible CDOCKER	10.2	1.69
Cross-docking	Rigid CDOCKER	5.6	5.00
	Flexible CDOCKER	7.2	1.42

Flexible CDOCKER cross-docking HIV Reverse Transcriptase

Flexible CDOCKER is capable sampling and scoring a docked conformation in a cross-docking test case of a single receptor. Cross-docking trials are more challenging than re-docking trials, as the receptor is not in the optimal configuration to interact with the small molecule. Results from the CSAR Benchmark Exercise 2011-2012, hosted by the Carlson research group, demonstrated a wide range of docking accuracy for different targets.[53] The overall docking accuracy for the different receptors found that 33% of the top scoring results were docked,

however some complexes were as low as 5.4% scored.[53] As a challenge for the newly implemented CDOCKER methodology we performed a cross-docking trial of 131 unique complexes of HIV reverse transcriptase (Appendix A, Table 3). Consistent with previous results, the flexible and rigid receptor implementations had similar ability to sample docked conformations with 47.3% and 48.1% of the complexes docked respectively. Again, Flexible CDOCKER outperformed the rigid implementation in its ability to score the docked conformation as lowest in energy, with 19.1% docked compared to 10.7% for rigid. These docking percentages are low compared to redocking trials, which is to be expected, as the receptor structure is not in an optimal conformation for ligand binding. Histograms of the low energy conformations demonstrate the improvement of docking success with Flexible CDOCKER over rigid CDOCKER, as illustrated in Figure 8 left-side y-axis. It is especially apparent that the inclusion of flexible side-chains improves the sampling of docked conformations when considering the total structures docked, which is plotted in dashed lines along the right-side y-axis of Figure 8. For Flexible CDOCKER, 26.2% of all the conformations sampled were docked while the rigid implementation only achieved 14.2% docked.

While adding the receptor flexibility improves the docking accuracy, it is still fairly low. Recent results published for this crossdocking set using DOCK 6 show 66.7% of the 61 overlapping complexes were scored (Appendix A, Table 4).[6] The initial placement of the small molecule for CDOCKER is somewhat naïve, generating many random ligand conformations targeting a specific binding site, neglecting much information concerning the receptor surface, and perhaps improvements to the initial placement would lead to increased docking accuracy. DOCK 6 for example, uses pre-computed site points for excluded volume and chemical matching for initial ligand placement. There was an overlap of 61 complexes, while the

remaining 70 complexes were considered non-viable pairings by Rizzo et al. A non-viable pairing was defined as a ligand minimized in a receptor structure that had an RMSD of greater than 2Å.[6] For the complexes not attempted by Rizzo et al, Flexible CDOCKER had about twice the docking accuracy as the rigid implementation with 14.3% and 7.1%, respectively, scored and 32.9% of the complexes were found by both methods (Appendix A, Table 4). CDOCKER was able to find docked conformations of these more challenging complexes. However, many of the docked conformations that are being found by the docking protocol are not being scored as low energy conformations. This suggests a need for improved methods of ranking ligand conformations.

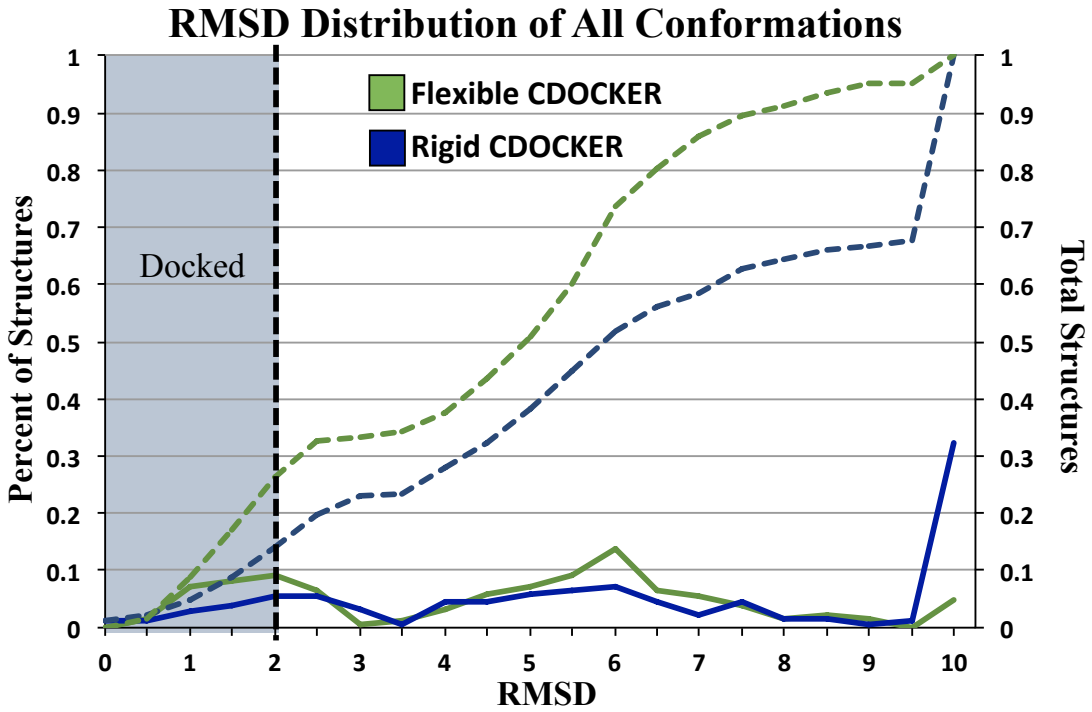


Figure 2.8: Histogram of RMSD (solid lines) of all low energy conformations from the HIV reverse transcriptase for both rigid (blue) and flexible (green) CDOCKER. The cumulative sums (dashed lines) of the number of conformations for rigid (blue) and flexible (green) CDOCKER for a given RMSD show flexible side chains inclusion leads to a greater number of docked conformations sampled.

2.4 Conclusions

Incorporating receptor flexibility into docking methods is an essential step in more accurately modeling protein-ligand interactions, however, maintaining efficiency requires this to be done in a limited manner. Flexible CDOCKER represents the majority of the receptor as a rigid entity, representing side chains around a target binding pocket explicitly to minimize computational costs (Figure 1). When integrated with a grid representation of the remaining receptor that is softened with respect to its interactions one finds that the sampling of flexible side chains is maintained similarly to what is observed with conventional solvated molecular dynamics. Compared to the previously reported rigid receptor CDOCKER method, the inclusion of flexibility with a new MD+Minimization sampling protocol leads to improvements in docking accuracy, particularly in the scoring step of the docking process.[28] This approach to incorporating receptor flexibility does not lead to loss of docking accuracy in re-docking trials over the rigid implementation, as may occur when including additional degrees of freedom. The computational cost of includiing the receptor flexibility increases linearly with the number of atoms (Appendix A, Figure 2). In a test case of cross-docking, the flexible receptor version of CDOCKER was superior to the rigid implementation in both sampling and scoring of docked conformations of the receptor's paired small-molecule. Finally, comparing our new docking protocols to results reported by the Gohlke research group shows that Flexible CDOCKER outperforms other widely used docking methodologies in re-docking trials of the CCDC/Astex clean set.[37]

2.5 Development of automated covalent ligand docking method

Covalent, irreversable drugs have become increasingly popular category. In fact a surprising number of approved drugs can covalently inhibit enzymes.[60-64] Asprin, developed and manufactured in the 1800s by Bayer acts by a covalent and irreversable inhibition. Though the mode of action was not identified until the 1970s.[65] Antibiotics also act as covalent inhibitiors, such as penicillins and cephalasporins.[66] There are at least 39 approved drugs that are covalent inhibitors, many targeting GPCRs.[62,67] Three of the top ten best selling drugs in the United States in 2009 were covalent inhibitors to their respective targets: clopidogrel, lansoprazole and esomeprazole.[67] There are two major classes of colvalent ligands, those that have prexisting reactive electrophilic functionality (penicilin) and compounds that undergo a transformation in viio creating reactive metabolites (acetaminophen).[68] The second class has great safety concerns due to possibility of highly reactive intermediates that may react non-specifically. For this reason, a drug discovery project does not generally set out to develop a covalent drug.[67,69] Though the theraputic value of these compounds is still great, much study of the possible off target interactions must be undertaken.

In addition to being used as drugs, covalent ligands have become a powerful biochemical and biophysical tool. Development by the Wells research group of tethering methodology for screening a library of small-molecule fragments for a targeting a specific binding pocket with a cysteine mutation has yeilded impressive results.[70,71] This tethering technique is a rapid and reliable screening method of small-molecule fragments to be used for rational design benifiting from the use of a covalent disulfide bond. A non-covalent fragment screen often suffers from non-specific binding of the small molecule fragments at many sites on the receptor, while the tethering techinque assures localization to the target binding site.[70-72] This method was employed

by the Mapp research group to target KIX domain of CBP, a transcriptional coactivator domain.[72,73] The resulting small molecule hits stabilized the KIX domain sufficently to obtain the first ever crystal structure of this receptor (see Chapter 4).[72,73] The tethering method has been used to target other protein-protein interfaces including the identification of an IL-2 inhibitor.[74] Furthermore, tethering was used as a basis for rational design by the Mann research group, who developed a non-covalent inhibitor based on the fragment hits in the tethering screen.[75]

Targeting challegning surfaces with covalent ligands is becoming increasingly popular, with a recent review demonstrating an increase in the percentage of colvalent inhibitors compared to non-covalent inhibitors. However, functionality to investigate covalent ligands with docking methodology remains fairly limited. CovalentDock is available as an interface with AutoDock creating molecular geometry based restraints to keep the ligand in the desired region on the receptor.[76] Also available is the CovDock suite within Schrodinger, which also mimics covalent docking through restraining the ligand to the target vicinity on the receptor.[77] We developed a docking methodology that creates a physcial bond between the ligand and receptor in an automated fashion, named CHARMM Docker with crosslink, or CDOCKER with XLink.

2.6 XLink: A tethering docking method

The tethering methodology that is currently implemented within CDOCKER is based off of the disulfide linkers utilized in the fragment screening method developed by the Wells research group. However, CHARMM is a very versitile software package, and tethers other than disulfide bonds could be implemented.[49,50] The XLink functionality is structured as a MMTSB tool, as is CDOCKER, using a set of Perl scripts to interface user input with CHARMM.[41] These scripts generate ligand parameters in an automated fashion using MATCH and the CHARMM CGENFF36 parameter library.[38-40,44] Additonally, there is functionality to generate a cystiene point mutation on a receptor, allowing for screening of various target regions on the receptor. The framework of the XLink functionality is incorporated into CDOCKER, so that the sampling methodologies already implemented within CDOCKER including receptor flexibility are available with XLink functionality.

A cysteine on the receptor and a small molecule structure are required for the crosslink parameters to be generated (Figure 9). XLink uses MATCH to generate the parameters of just the small molecule and of the small molecule bound to the cystiene.[44] The ligand-cystiene tethered structure is generated in the absence of the rest of the receptor and as such the orientation has no relation to, for example, the lowest energy conformation in the prescence of the receptor. MATCH utilizes a chemical pattern-matching engine to assign atom types, charges and force field parameters through comparision to a list of chemical fragments, and as such the initial orientation need not be optimized in the context of the rest of the receptor. XLink takes the differences between the ligand and ligand-cystiene MATCH generated parameters to create a CHARMM readable patch.[49,50] This patch accounts for the changes in partial charges of the ligand and cystiene when they are tethered, as well as adding angles, dihidrals etc.

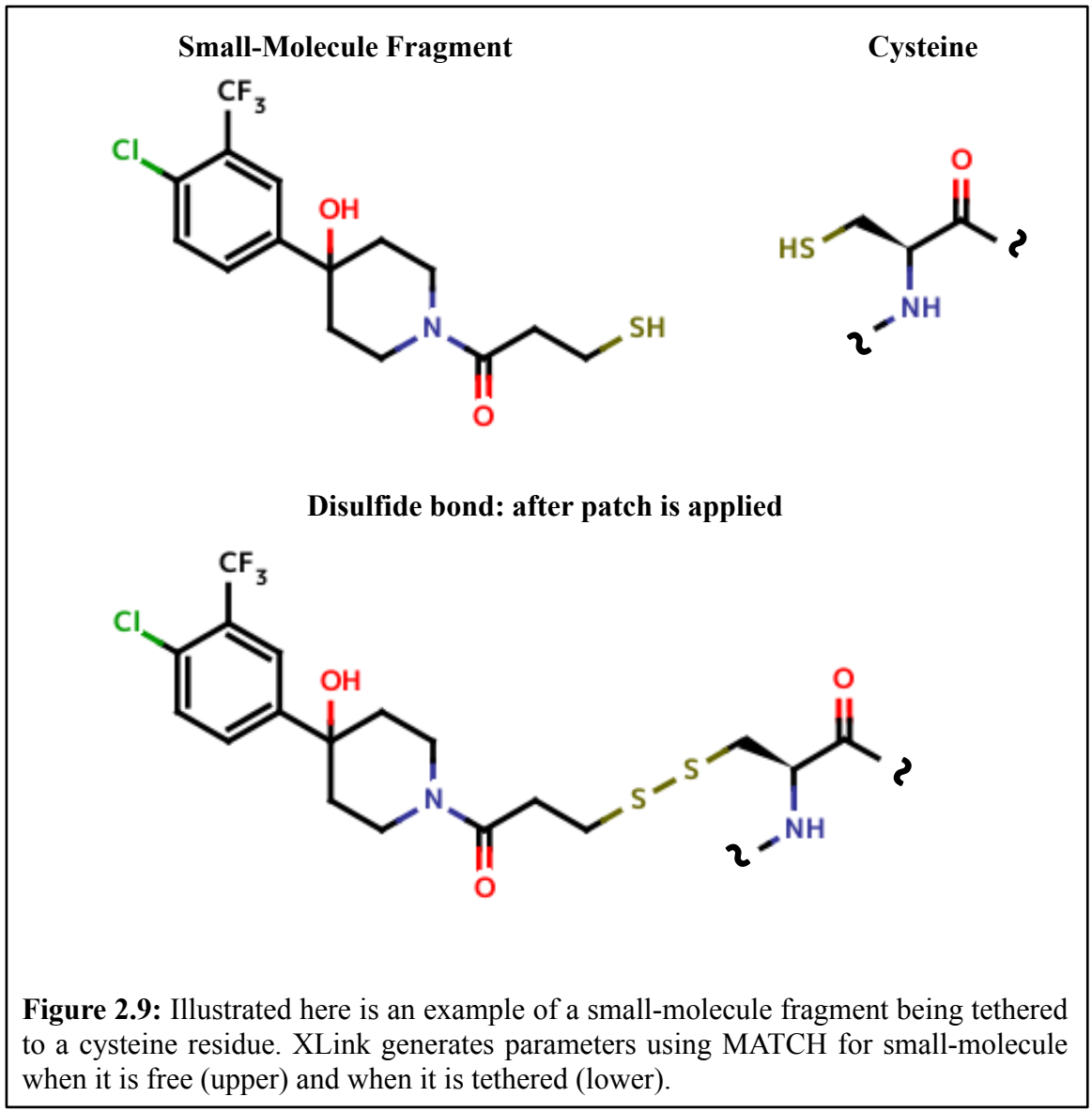


Figure 2.9: Illustrated here is an example of a small-molecule fragment being tethered to a cysteine residue. XLink generates parameters using MATCH for small-molecule when it is free (upper) and when it is tethered (lower).

In addition to having an extra initialization step, XLink also must sample the ligand conformational space differently than non-covalent CDOCKER. Instead of performing rigid body rotations of the ligand about a center of mass, we rotate the ligand around the disulfide bond. Like non-covalent CDOCKER, we also randomly rotate the rotatable bonds contained within the ligand. The resulting low energy conformations are clustered and the low energy member of the low energy clusters is subjected to simulated annealling sampling as outlined

above. Additionally, the flexible receptor and MD+Minimization sampling can be employed with XLink in a fairly straightforward fashion.

This CDOCKER with XLink was utilized to investigate tethered fragments binding to KIX domain of CBP (See Chapter 4.3). Flexible CDOCKER with XLink was able to predict a small-molecule conformation that is similar to the crystallographic conformation in a cross-docking trial. Additionally, the resulting CHARMM parameters can be used in straigthforward MD simulations for further investigation of the dynamics of the receptor-ligand complex. This implementation of tethering in docking is, to my knowledge, the first that generates a physical bond as opposed to using constraints to keep the ligand within a certain region of the receptor.

2.7 References

(1) Borhani, D. W.; Shaw, D. E. *Journal of Computer-Aided Molecular Design* **2012**, *26*, 15.
(2) Jorgensen, W. L. *Science* **2004**, *303*, 1813.
(3) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. *Journal of Medicinal Chemistry* **2004**, *47*, 1739.
(4) Neves, M. A. C.; Totrov, M.; Abagyan, R. *Journal of Computer-Aided Molecular Design* **2012**, *26*, 675.
(5) Trott, O.; Olson, A. J. *Journal of Computational Chemistry* **2010**, *31*, 455.
(6) Allen, W. J.; Balius, T. E.; Mukherjee, S.; Brozell, S. R.; Moustakas, D. T.; Lang, P. T.; Case, D. A.; Kuntz, I. D.; Rizzo, R. C. *Journal of Computational Chemistry* **2015**, *36*, 1132.
(7) Huang, S.-Y.; Zou, X. *J. Comput. Chem.* **2006**, *27*, 1866.
(8) Liu, J.; Wang, R. *Journal of Chemical Information and Modeling* **2015**, *55*, 475.
(9) Velec, H. F. G.; Gohlke, H.; Klebe, G. *Journal of Medicinal Chemistry* **2005**, *48*, 6296.
(10) Zheng, Z.; Merz, K. M., Jr. *Journal of Chemical Information and Modeling* **2013**, *53*, 1073.
(11) Irwin, J. J.; Sterling, T.; Mysinger, M. M.; Bolstad, E. S.; Coleman, R. G. *Journal of Chemical Information and Modeling* **2012**, *52*, 1757.
(12) Jorgensen, W. L. *Science* **1991**, *254*, 954.
(13) Alvarez-Garcia, D.; Barril, X. *Journal of Chemical Theory and Computation* **2014**, *10*, 2608.
(14) Baron, R.; McCammon, J. A. *Annual Review of Physical Chemistry, Vol 64* **2013**, *64*, 151.
(15) Carlson, H. A. *Current Opinion in Chemical Biology* **2002**, *6*, 447.
(16) Cozzini, P.; Kellogg, G. E.; Spyrakis, F.; Abraham, D. J.; Costantino, G.; Emerson, A.; Fanelli, F.; Gohlke, H.; Kuhn, L. A.; Morris, G. M.; Orozco, M.; Pertinhez, T. A.; Rizzi, M.; Sotriffer, C. A. *Journal of Medicinal Chemistry* **2008**, *51*, 6237.
(17) Zavodszky, M. I.; Kuhn, L. A. *Protein Science* **2005**, *14*, 1104.
(18) Bowman, G. R.; Geissler, P. L. *Journal of Physical Chemistry B* **2014**, *118*, 6417.
(19) Teague, S. J. *Nature Reviews Drug Discovery* **2003**, *2*, 527.
(20) Kuhn, L. A. *Strength in flexibility: Modeling side-chain conformational change in docking and screening*, 2008.
(21) Durrant, J. D.; McCammon, J. A. *Current Opinion in Pharmacology* **2010**, *10*, 770.
(22) Lexa, K. W.; Carlson, H. A. *Quarterly Reviews of Biophysics* **2012**, *45*, 301.
(23) Carlson, H. A.; McCammon, J. A. *Molecular Pharmacology* **2000**, *57*, 213.
(24) Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. *Proteins-Structure Function and Bioinformatics* **2006**, *65*, 15.
(25) Miranker, A.; Karplus, M. *Proteins-Structure Function and Genetics* **1991**, *11*, 29.
(26) Armen, R. S.; Chen, J.; Brooks, C. L., III *Journal of Chemical Theory and Computation* **2009**, *5*, 2909.
(27) Meiler, J.; Baker, D. *Proteins-Structure Function and Bioinformatics* **2006**, *65*, 538.
(28) Wu, G. S.; Robertson, D. H.; Brooks, C. L., III; Vieth, M. *Journal of Computational Chemistry* **2003**, *24*, 1549.
(29) Vieth, M.; Hirst, J. D.; Dominy, B. N.; Daigler, H.; Brooks, C. L., III *Journal of Computational Chemistry* **1998**, *19*, 1623.

(30) Vieth, M.; Hirst, J. D.; Kolinski, A.; Brooks, C. L., III *Journal of Computational Chemistry* **1998**, *19*, 1612.
(31) Jiang, F.; Kim, S. H. *Journal of Molecular Biology* **1991**, *219*, 79.
(32) Roche, O.; Kiyama, R.; Brooks, C. L., III *Journal of Medicinal Chemistry* **2001**, *44*, 3592.
(33) Wu, X.; Brooks, B. R. *Journal of Chemical Physics* **2011**, *134*.
(34) Wu, X. W.; Brooks, B. R. *Chemical Physics Letters* **2003**, *381*, 512.
(35) Nissink, J. W. M.; Murray, C.; Hartshorn, M.; Verdonk, M. L.; Cole, J. C.; Taylor, R. *Proteins-Structure Function and Genetics* **2002**, *49*, 457.
(36) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Research* **2000**, *28*, 235.
(37) Krueger, D. M.; Jessen, G.; Gohlke, H. *Journal of Chemical Information and Modeling* **2012**, *52*, 2807.
(38) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *Journal of Physical Chemistry B* **1998**, *102*, 3586.
(39) Mackerell, A. D.; Feig, M.; Brooks, C. L., III *Journal of Computational Chemistry* **2004**, *25*, 1400.
(40) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; MacKerell, A. D., Jr. *Journal of Computational Chemistry* **2010**, *31*, 671.
(41) Feig, M.; Karanicolas, J.; Brooks, C. L., III *Journal of Molecular Graphics & Modelling* **2004**, *22*, 377.
(42) Olsson, M. H. M.; Sondergaard, C. R.; Rostkowski, M.; Jensen, J. H. *Journal of Chemical Theory and Computation* **2011**, *7*, 525.
(43) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. *Journal of Cheminformatics* **2011**, *3*.
(44) Yesselman, J. D.; Price, D. J.; Knight, J. L.; Brooks, C. L., III *Journal of Computational Chemistry* **2012**, *33*, 189.
(45) Bowman, A. L.; Nikolovska-Coleska, Z.; Zhong, H.; Wang, S.; Carlson, H. A. *Journal of the American Chemical Society* **2007**, *129*, 12809.
(46) Lorber, D. M.; Shoichet, B. K. *Protein Science* **1998**, *7*, 938.
(47) Zavodszky, M. I.; Ming, L.; Thorpe, M. F.; Day, A. R.; Kuhn, L. A. *Proteins-Structure Function and Bioinformatics* **2004**, *57*, 243.
(48) Fleischman, S. H.; Brooks, C. L., III *Proteins-Structure Function and Genetics* **1990**, *7*, 52.
(49) Brooks, B. R.; Brooks, C. L., III; Mackerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. *Journal of Computational Chemistry* **2009**, *30*, 1545.
(50) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *Journal of Computational Chemistry* **1983**, *4*, 187.

(51) Eastman, P.; Friedrichs, M. S.; Chodera, J. D.; Radmer, R. J.; Bruns, C. M.; Ku, J. P.; Beauchamp, K. A.; Lane, T. J.; Wang, L.-P.; Shukla, D.; Tye, T.; Houston, M.; Stich, T.; Klein, C.; Shirts, M. R.; Pande, V. S. *Journal of Chemical Theory and Computation* **2013**, *9*, 461.
(52) Bursulaya, B. D.; Totrov, M.; Abagyan, R.; Brooks, C. L., III *Journal of Computer-Aided Molecular Design* **2003**, *17*, 755.
(53) Damm-Ganamet, K. L.; Smith, R. D.; Dunbar, J. B., Jr.; Stuckey, J. A.; Carlson, H. A. *Journal of Chemical Information and Modeling* **2013**, *53*, 1853.
(54) Dunbar, J. B., Jr.; Smith, R. D.; Yang, C.-Y.; Ung, P. M.-U.; Lexa, K. W.; Khazanov, N. A.; Stuckey, J. A.; Wang, S.; Carlson, H. A. *Journal of Chemical Information and Modeling* **2011**, *51*, 2036.
(55) Guthrie, J. P. *Journal of Physical Chemistry B* **2009**, *113*, 4501.
(56) Kim, R.; Skolnick, J. *Journal of Computational Chemistry* **2008**, *29*, 1316.
(57) Skillman, A. G.; Geballe, M. T.; Nicholls, A. *Journal of Computer-Aided Molecular Design* **2010**, *24*, 257.
(58) Smith, R. D.; Dunbar, J. B., Jr.; Ung, P. M.-U.; Esposito, E. X.; Yang, C.-Y.; Wang, S.; Carlson, H. A. *Journal of Chemical Information and Modeling* **2011**, *51*, 2115.
(59) Taufer, M.; Crowley, M.; Price, D. J.; Chien, A. A.; Brooks, C. L., III *Concurrency and Computation-Practice & Experience* **2005**, *17*, 1627.
(60) Robertson, J. G. *Biochemistry* **2005**, *44*, 5561.
(61) Johnson, D. S.; Weerapana, E.; Cravatt, B. F. *Future Medicinal Chemistry* **2010**, *2*, 949.
(62) Kalgutkar, A. S.; Dalvie, D. K. *Expert Opinion on Drug Discovery* **2012**, *7*, 561.
(63) Potashman, M. H.; Duggan, M. E. *Journal of Medicinal Chemistry* **2009**, *52*, 1231.
(64) Wilson, A. J.; Kerns, J. K.; Callahan, J. F.; Moody, C. J. *Journal of Medicinal Chemistry* **2013**, *56*, 7463.
(65) Roth, G. J.; Stanford, N.; Majerus, P. W. *Proceedings of the National Academy of Sciences of the United States of America* **1975**, *72*, 3073.
(66) Wright, P. M.; Seiple, I. B.; Myers, A. G. *Angewandte Chemie-International Edition* **2014**, *53*, 8840.
(67) Singh, J.; Petter, R. C.; Baillie, T. A.; Whitty, A. *Nature Reviews Drug Discovery* **2011**, *10*, 307.
(68) Fry, D. W.; Bridges, A. J.; Denny, W. A.; Doherty, A.; Greis, K. D.; Hicks, J. L.; Hook, K. E.; Keller, P. R.; Leopold, W. R.; Loo, J. A.; McNamara, D. J.; Nelson, J. M.; Sherwood, V.; Smaill, J. B.; Trumpp-Kallmeyer, S.; Dobrusin, E. M. *Proceedings of the National Academy of Sciences of the United States of America* **1998**, *95*, 12022.
(69) Guterman, L. *Chemical & Engineering News* **2011**, *89*, 2.
(70) Erlanson, D. A.; Wells, J. A.; Braisted, A. C. *Annual Review of Biophysics and Biomolecular Structure* **2004**, *33*, 199.
(71) Erlanson, D. A.; Braisted, A. C.; Raphael, D. R.; Randal, M.; Stroud, R. M.; Gordon, E. M.; Wells, J. A. *Proceedings of the National Academy of Sciences of the United States of America* **2000**, *97*, 9367.
(72) Lodge, J. M.; Rettenmaier, T. J.; Wells, J. A.; Pomerantz, W. C.; Mapp, A. K. *Medchemcomm* **2014**, *5*, 370.
(73) Wang, N.; Majmudar, C. Y.; Pomerantz, W. C.; Gagnon, J. K.; Sadowsky, J. D.; Meagher, J. L.; Johnson, T. K.; Stuckey, J. A.; Brooks, C. L., III; Wells, J. A.; Mapp, A. K. *Journal of the American Chemical Society* **2013**, *135*, 3363.

(74) Braisted, A. C.; Oslob, J. D.; Delano, W. L.; Hyde, J.; McDowell, R. S.; Waal, N.; Yu, C.; Arkin, M. R.; Raimundo, B. C. *Journal of the American Chemical Society* **2003**, *125*, 3714.
(75) Nonoo, R. H.; Armstrong, A.; Mann, D. J. *Chemmedchem* **2012**, *7*, 2082.
(76) Ouyang, X.; Zhou, S.; Su, C. T. T.; Ge, Z.; Li, R.; Kwoh, C. K. *Journal of Computational Chemistry* **2013**, *34*, 326.
(77) Warshaviak, D. T.; Golan, G.; Borrelli, K. W.; Zhu, K.; Kalid, O. *Journal of Chemical Information and Modeling* **2014**, *54*, 1941.

Chapter 3

Application of CDOCKER: Identification of a putative ATP binding site on calreticulin

Text adapted from: S.J. Wijeyesakere, J.K. Gagnon, K.A. Arora, C.L. Brooks III, M. Raghavan. Regulation of calreticulin-major histocompatibility complex (MHC) class I interactions by ATP. *Submitted.*

3.1 Introduction

In the protein-dense environment of the ER, macromolecular chaperones help maintain protein homeostasis.[1,2] These specialized proteins support the folding of nascent or mis-folded proteins by interacting with exposed hydrophobic regions on these proteins, thereby preventing their aggregation and stabilizing folding intermediates.[3] Protein chaperone cycles are often regulated by conformational changes that occur upon nucleotide binding and hydrolysis.[2,4-8] Most well characterized chaperones such as GroEL, HSP70, and HSP90, employ nucleotide-associated conformational changes to

regulate and guide their interactions with client proteins.[3,6,8-10] Calreticulin (CRT) is a multi-domain lectin-binding chaperone that is believed to play an important role in the assembly of the peptide loading complex and proper folding of proteins in endoplasmic reticulum (ER).[7,11-13]

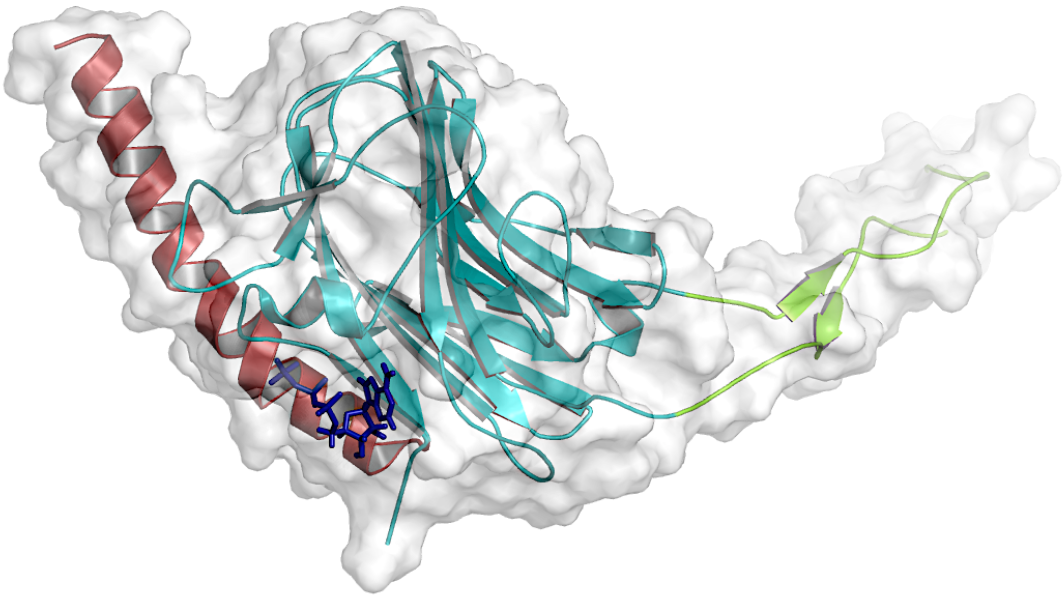


Figure 3.1: CRT globular domain, containing the glycan binding site, is shown in teal. The acidic domain has low affinity calcium binding sites (red), and ERp57 binds the with p-loop (green). (PDBID:3o0v with p-loop modeled in)

CRT is a soluble dual function ER protein. It behaves as a lectin, binding monoglucosylated glycans present on nascent glycopeptides and as a chaperone, preventing aggregation of proteins through glycan-independent interactions.[14-16] CRT is recruited into the peptide-loading complex within the ER, where, with other proteins, MHC Class-I molecules that mediate immune surveillance by CD8+T-cells and natural killer (NK) cells are folded.[13,17] Structurally, CRT consists of three domains (Figure 1). One domain is a lectin fold globular domain containing interaction sites for glycans and peptides.[15,18,19] Another consists of a proline-rich domain termed the P-domain where

interactions with co-chaperones like ERp57 and cyclophillin-B occur.[12,20,21] The final domain is a C-terminal acidic region that acts as a calcium buffer due to numerous low affinity calcium binding-sites.[7,16]

While it has been shown *in vitro* that ATP can induce conformational changes in CRT and enhance its chaperone activity towards both glycosylated and non-glycosylated substrates the nature and location of the nucleotide-binding site of CRT remains unknown.[15,19] Contained in this chapter is a collaborative effort with the Raghavan research group at the University of Michigan Medical School towards identifying a putative binding site for ATP on CRT and the role of ATP on the CRT chaperone cycle. All wet-lab experimental work reported in this chapter was carried out by the Raghavan group. I would also like to acknowledge Karunesh Arora, from the Brooks research group, for his assistance and guidance in this project.

3.2 Results and Discussion

Towards characterizing the nucleotide dependence of CRT we employed fluorescence quenching to determine the steady-state affinity of both ATP and ADP with murine calreticulin (mCRT(WT)). The affinity for binding ATP to mCRT(WT) is 480 ± 69 μM and 1.26 ± 0.28 μM for ADP. These affinities are similar to those reported for other nucleotide binding chaperones such as HSP90.[6,8] Using differential scanning calorimetry (DSC) we find that there is significant decrease in thermostability of mCRT(WT) in the presence of 4mM ATP (ΔT_{trans} = 2.5 ± 0.2 °C; $p > 0.001$) which is consistent with previous findings from the Raghavan research group.[11] Furthermore, removal of the P-domain (mCRT(ΔP)) or C-terminal region (mCRT(ΔC)) leads to similar loss of thermostability (ΔT_{trans} = 5.9 ± 0.7 °C; $p > 0.001$ and ΔT_{trans} = 4.6 ± 0.5 °C; $p > 0.001$ respectively) indicating the ATP binding site is located on the globular domain (Figure 1, teal). Using this information from the DSC experiments we employed a blind, global docking trial of ATP and the crystal structure of CRT globular domain (PDBID: 3o0v) to predict the location of the nucleotide binding site. This initial trial was undertaken using a consensus approach, utilizing both Autodock Vina and CDOCKER in CHARMM simulation packages.[22-25] The version of CDOCKER applied to this trial did not include the flexible receptor options, and both the Autodock Vina and CDOCKER trials were rigid receptor structure-based docking trials. The initial docked ligand poses obtained from the Autodock suite were rescored using an all-atom protein and ligand representation with Generalized Born Molecular Volume (GBMV) implicit solvent in CHARMM and compared to the docked poses identified by CDOCKER.[22,23,26-30] These

results yielded 4 distinct clusters of ATP binding site locations with similar predicted energies ranging from 1.8 to -1.1 kcal/mol (Figure 2).

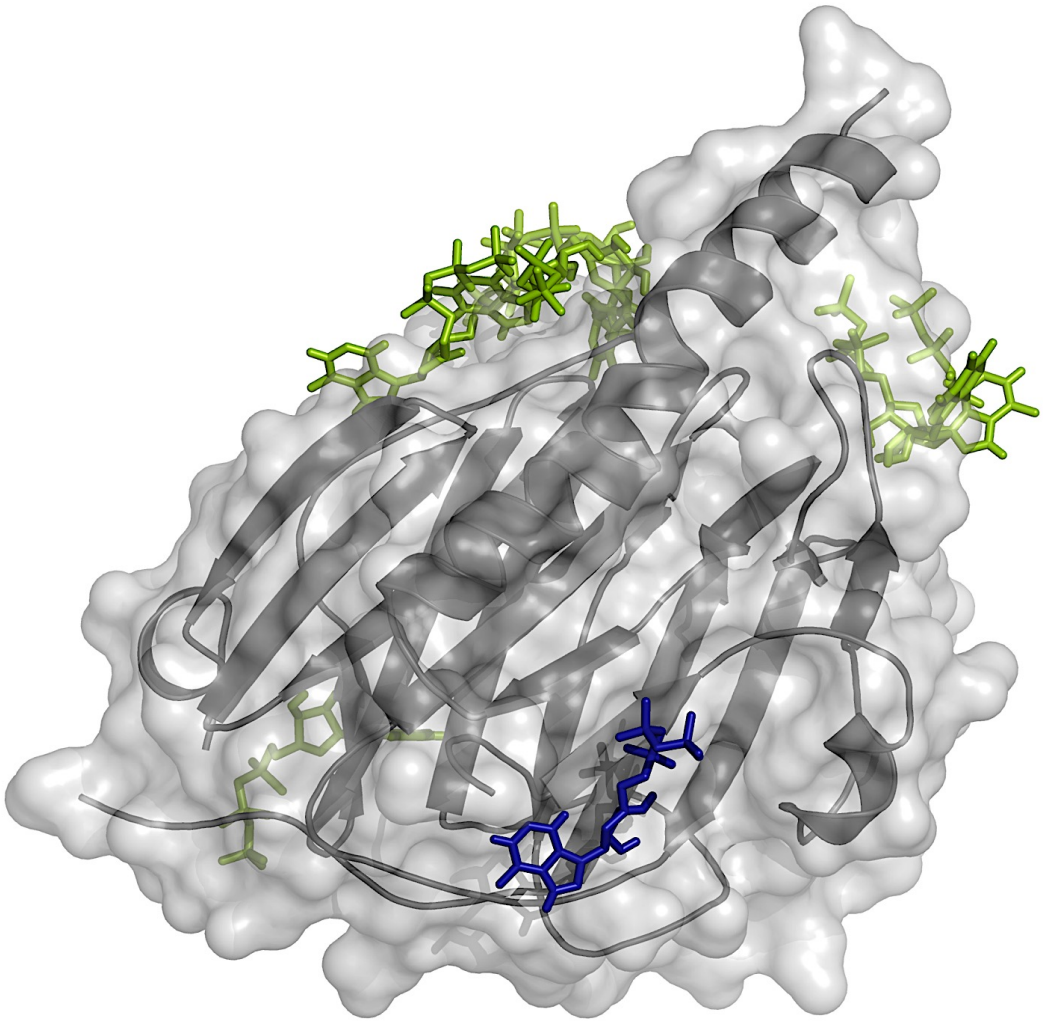


Figure 3.2: Structure-based docking results showed 4 major clusters, shown in sticks. The predicted lowest energy conformation is shown in blue, which corresponds to a binding site near Lys7.

Guided by these docking results we undertook site-directed mutagenesis, mutating residues that were in contact with the predicted ATP pose. Using DSC we identified residues inhibiting ATP binding based on the lack of significant loss of thermostability of mCRT(WT) in the presence of ATP as seen in the initial DSC scans (Figure 3). All initial

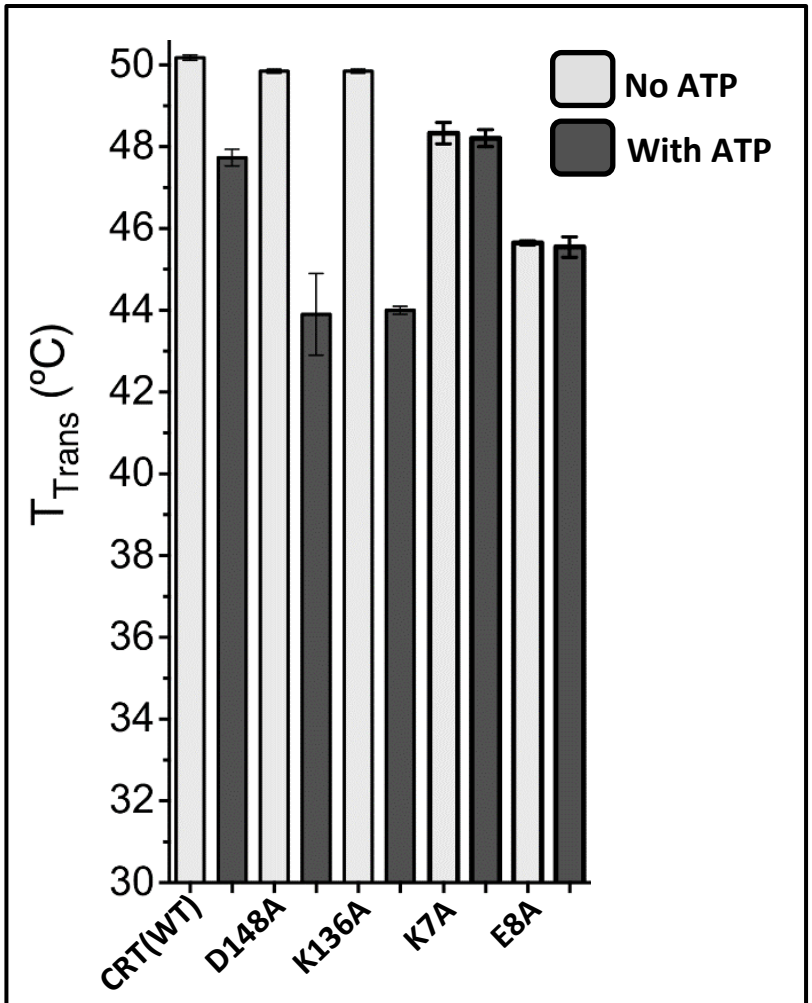


Figure 3.3: Wild-type CRT experiences a thermostability loss upon binding ATP as measured with DSC. Residues that were near predicted binding sites by docking that disrupted this thermostability loss, K7A and E8A, are likely residues important for ATP binding.

mutants tested displayed the loss in thermostability as shown with mCRT(WT) upon interacting with ATP with the exception of the K7A and E8A mutants, indicating that this residue is key in ATP binding (Figure 3). These data informed a more focused docking trial centered on Lys7. The resulting predicted low energy ATP conformation from this docking trial was subjected to a 20 ns all-atom explicit solvent molecular dynamic (MD) simulation. The structure-based docking performed held the protein fixed while ligand conformations were sampled, this allows for rapid and computationally inexpensive sampling. The advantage of the MD simulations is that the receptor conformation is allowed to relax around the ligand and adapt side-chain orientations to optimally interact with the

nucleotide. The resulting average structure from the MD simulations allowed us to identify additional residues important for ligand binding and catalysis (Figure 4).

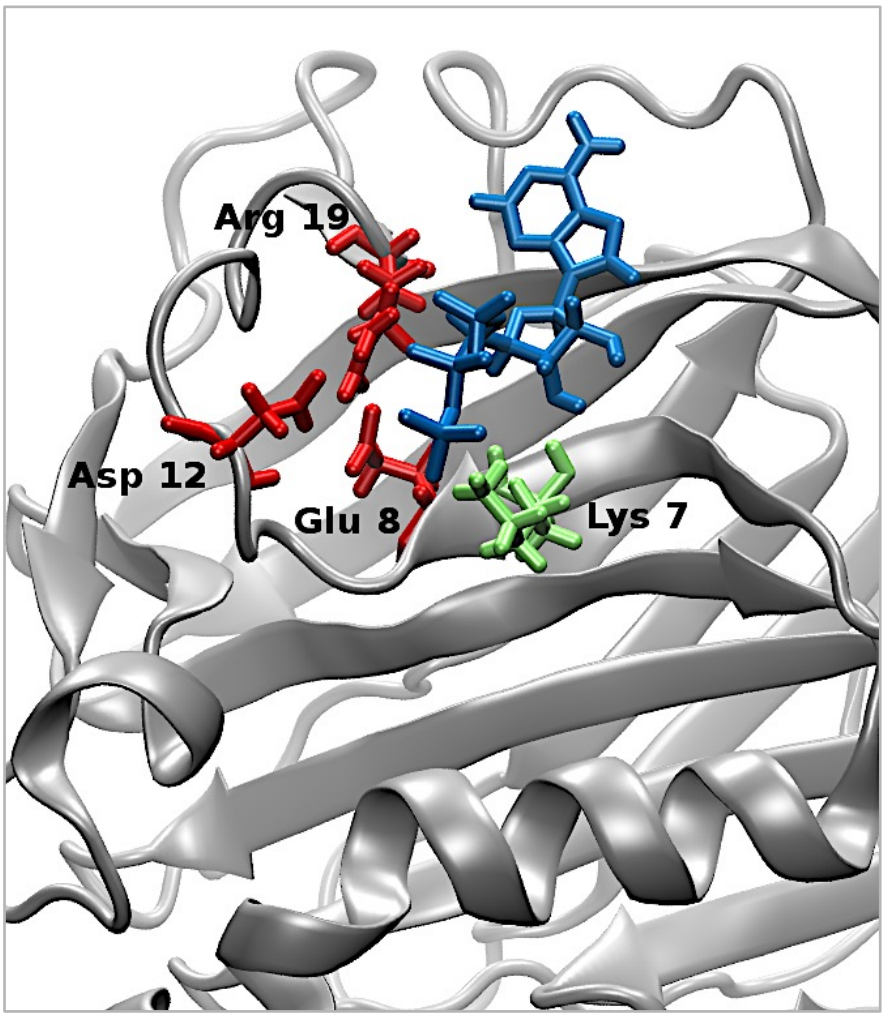


Figure 3.4: Closest to average structure from explicit solvent simulation started from predicted lowest energy conformation of ATP (blue) bound to CRT at the Lys7 (green) binding surface.

The structure shows Lys7 interacting with the alpha-phosphate of ATP and another lysine, Lys63, interacts with the base of ATP. Additional residues interacted with the γ-phosphate moiety of ATP, Asp12 and Arg19, suggesting these to be putative catalytic residues that hydrolyze ATP, forming ADP and Pi. Interestingly during the

course of the MD simulations, sodium ions from the bulk solvent come in and interact with the γ-phosphate of ATP and orient this moiety with the Asp12 and Arg19 residues. It should be noted that while CRT lacks the consensus sequences for common ATP-binding and hydrolysis motifs, such as the Walker A and Walker B motifs, but Asp12 is contained in a DxD motif that has been suggested as an alternative to the Walker B motif.[31] Additionally, we mutated the residue 7 from a lysine to alanine *in silico* and simulated the mutant (mCRT(K7A)) with ATP started in its predicted conformation from the structure-based docking. While the simulation of mCRT(WT) remained stable throughout the simulation the ATP dissociated from mCRT(K7A) within 10ns of the simulation further supporting Lys7 is essential for ATP binding.

Lead by the computational predictions we used fluorescence quenching to assess the ATP binding ability of CRT containing single point alanine mutations at Lys7, Glu8, Asp12, or Arg19 positions. mCRT(K7A) and mCRT(E8A) showed a decrease in their steady state affinities for ATP (K_D = 8.1 ± 0.2 mM and K_D = 5.1 ± 1.8 mM respectively, Figure 4). Conversely, the steady state affinities for mCRT(D12A) and mCRT(R19A) towards ATP are similar to that observed for the wild-type protein (K_D = 1.3 ± 0.3 mM and K_D = 1.0 ± 0.3 mM respectively, Figure 4). The loss of affinity for ATP with mutations at the Lys7 or Glu8 positions was encouraging. The small change in affinity with alanine mutations at Asp12 and Arg19 suggest these residues do not play a large role in ATP binding to CRT, however as shown in Figure 3, the simulation results predict these residues interact with ATP.

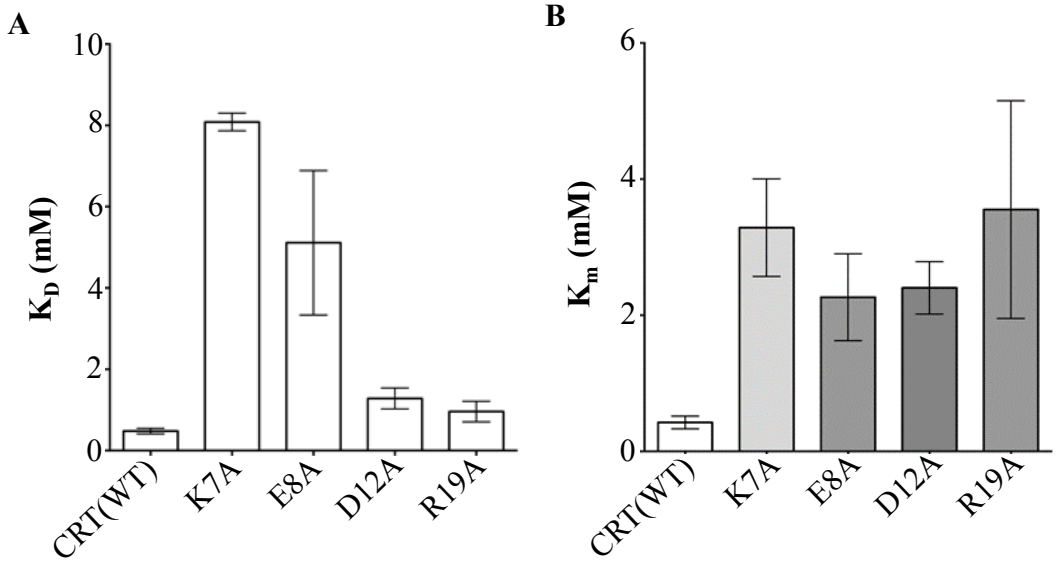


Figure 3.5: Dissociation (A) and Michaelis–Menten (B) constants for ATP binding to wild-type CRT and mutants. Residues Lys7 and Glu8 are important for ATP binding while Asp12 and Arg19 are key for catalysis.

In well-defined chaperones such as HSP90, ATP hydrolysis is a key factor in driving the chaperone cycle. Previous studies have shown recombinant CRT to have a weak ATPase activity but the lack of a CRT mutant deficient in ATP binding has made it difficult to rule out contamination with endogenous ATPases. To this end, we measured the ATPase activity of mCRT(WT) and the identified ATP-binding site mutants: mCRT(K7A), mCRT(E8A), mCRT(D12A), mCRT(R19A). We find the Michaelis–Menten constant (K_m) of ATP hydrolysis for mCRT(WT) was 424.5 ± 9.3 μM (Figure 5), a value similar to the calculated steady-state affinity, K_D, of CRT for ATP (480 ± 69 μM). Furthermore, significantly higher K_m values are observed for mCRT(K7A) and mCRT(E8A) relative to mCRT(WT) (Figure 5), findings consistent with the reduced ability of these mutants to interact with nucleotides. Similarly, significantly higher K_m values is observed for mCRT(D12A) and mCRT(R19A) relative to mCRT(WT), suggesting that these constructs are deficient in their ability to hydrolyze ATP. These data

with data presented in Figure 4, showing no loss of affinity upon mutating residues D12A or R19A, support that the residues Asp12 and Arg19 are key to the hydrolysis of ATP.

In chaperone systems such as HSP70, both nucleotide binding and hydrolysis play an allosteric role in the chaperone's interactions with various substrates.[5,12,32] It is known that ATP increases the chaperone activity of CRT towards glycosylated and non-glycosylated substrates as measured via its ability to prevent the thermal aggregation of proteins containing and lacking monoglucosylated glycans.[15,19] However, the effect of nucleotide binding on the ability of CRT to interact with the glycan moiety of a glycoprotein is currently unknown. To investigate these interactions we chose a synthetic monoglucosylated glycan Glcα1–3Manα1–2Manα1–2Man (G1M3) that has been used previously as a model substrate to investigate CRT+glycan interactions.[19,33-35] We performed MD simulations of CRT in complex with G1M3, as well as a ternary complex of CRT with G1M3 and either ATP or ADP. From these simulations we calculated a free energy of binding using MM/GBMV method with quasiharmonics to predict entropy.[22,27] The calculated $\Delta G_{binding}$ for G1M3 in the absence of nucleotides is -55.12 kcal/mol. Interestingly, there is a 2.8 fold reduction in affinity predicted when ATP is bound ($\Delta G_{binding}$ = -19.4 kcal/mol) while the reduction in affinity is only 1.8 fold in the presence of ADP ($\Delta G_{binding}$ = -35.4 kcal/mol).

To better understand the basis for ATP-mediated reduction in the glycan-binding affinity of calreticulin, we performed covariance analyses on MD trajectories of mCRT-G1M3 complexes in the presence or absence of nucleotides (ATP or ADP). Covariance analysis reports on the coupled motions between distant regions of proteins.[36] Results of the covariance analyses show that the distant residues located at the ATP and glycan

binding sites are coupled, exhibiting correlated motions (Figure 6). Relative to the ATP-bound state, the correlated motions are reduced in the ADP bound state and in the absence of nucleotide (Figure 6A and 6C vs. 6B). These findings suggest that dynamic coupling between the remote glycan and nucleotide-binding sites of calreticulin following ATP binding could cause a reduction in glycan-binding affinity.

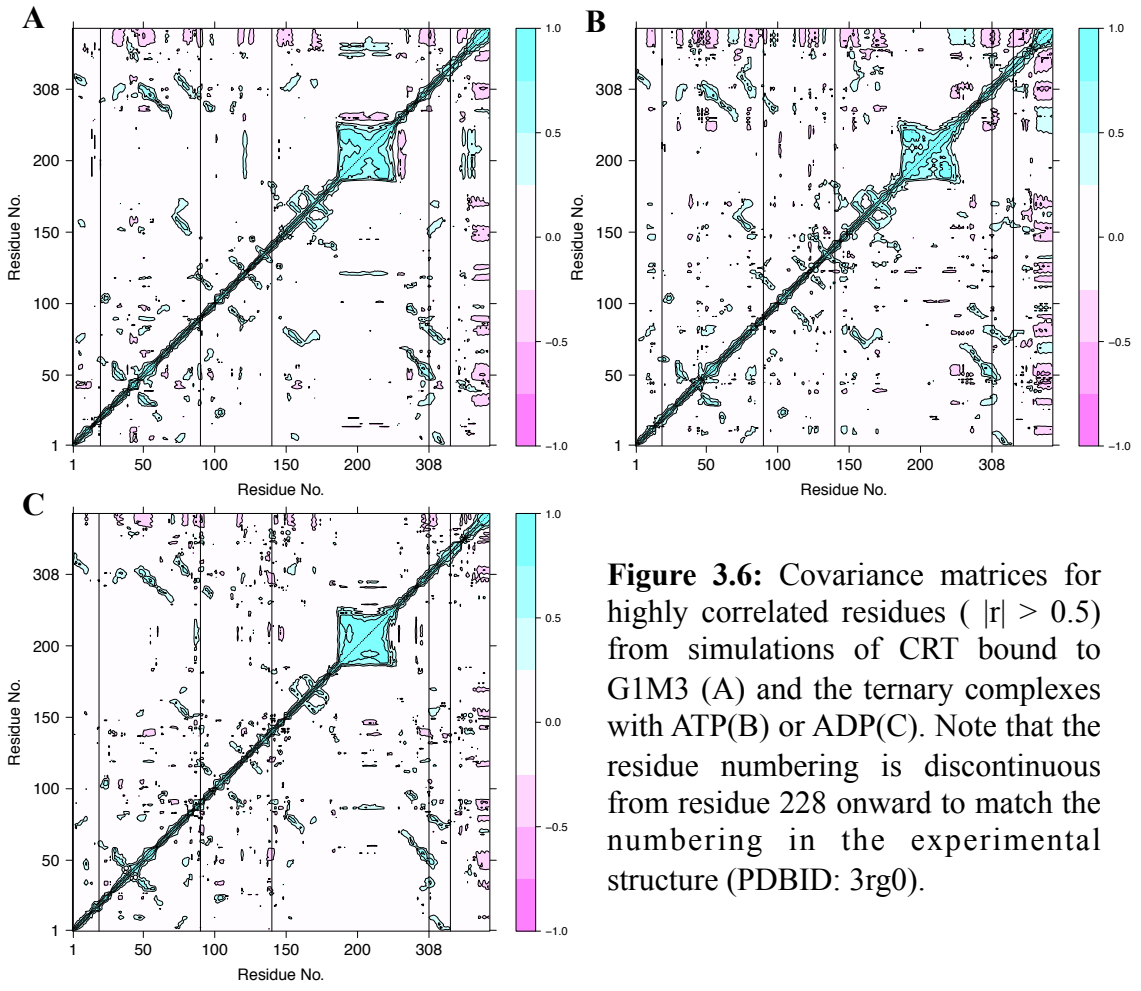


Figure 3.6: Covariance matrices for highly correlated residues (|r| > 0.5) from simulations of CRT bound to G1M3 (A) and the ternary complexes with ATP(B) or ADP(C). Note that the residue numbering is discontinuous from residue 228 onward to match the numbering in the experimental structure (PDBID: 3rg0).

In conclusion, these biochemical experiments guided with molecular modeling identify a putative ATP binding site on the globular domain of CRT defined by the residues Lys7, Glu8, Asp12, and Arg19. The residues Lys7 and Glu8 are essential for ATP binding (Figure 5A) while Asp12 and Arg19 (Figure 5B) are shown to be key for ATP hydrolysis. The MD simulations demonstrate that there is allosteric communication between the identified ATP binding site and the glycan binding site, suggesting ATP may trigger release of glycosylated ligands. The computationally identified residues lead to the ATP hydrolysis deficient mutants, allowing experimental conformation that CRT has ATPase activity, ruling out endogenous ATPase contamination.

3.3 Methods

Initial docking runs were performed in AutoDock Vina 1.1.1 using the crystal structure of the globular domain of murine calreticulin (PDB ID 3o0v) as the receptor.[24,25] The AutoDock results were scored in GBMV implicit solvent using CHARMM and compared to results from the CHARMM docking protocol, CDOCKER.[22,23,26,27] Subsequent docking trials were constrained within a simulation cell (a = b = c = 15 Å) centered on the C-atom of Lys7. The lowest scoring configuration of ATP was subjected to molecular dynamics (MD) simulations using the CHARMM force field within NAMD.[37-41] Six different conditions were simulated: calreticulin bound to ATP, ATP + G1M3, ADP, ADP + G1M3, G1M3, and no ligand. G1M3 was docked to the glycan-binding site based on the crystal structure of the calreticulin-G1M3 complex (PDB ID 3o0w). Each simulation lasted 20ns and was run in triplicate with a Ca^{2+} ion in the high affinity calcium binding site of calreticulin. The simulations were run in triplicate using constant volume Langevin dynamics with a 2fs time-step at 310K and a total of 50ns of production simulation was analyzed. Systems were solvated in an octahedral water box with 0.15M NaCl to neutralize charges.

Additional simulations were run using Mg^{2+} as a counter ion for ATP-bound calreticulin. MM/GBMV calculations were performed in CHARMM to estimate free energies of binding for G1M3 in the presence and absence of ATP or ADP. MM/GBMV free energy calculations take the energies from the molecular dynamics simulation, an estimate of solvation energies in implicit solvent (GBMV), and a quasi-harmonic entropy calculation to estimate free energies of binding. Ion lifetimes were also

calculated using CHARMM.

For covariance analysis of MD trajectories, the normalized covariance (C_{ij}) between two atoms i and j of a protein is defined as:

$$C_{ij} = \frac{\langle (\vec{r_i} - \langle \vec{r_i} \rangle) \cdot (\vec{r_j} - \langle \vec{r_j} \rangle) \rangle}{\langle \sqrt{(\vec{r_i} - \langle \vec{r_i} \rangle)^2} \cdot \sqrt{(\vec{r_j} - \langle \vec{r_j} \rangle)^2} \rangle}$$

r_i and r_j represents the instantaneous fluctuations of the position of the atoms i and j with respect to their mean positions. When two atoms i and j move concertedly in the same direction, C_{ij}=1, and, when the atoms move in opposite directions, C_{ij}=-1.

In this study, covariance analyses of Cα atoms were performed on 60 ns MD trajectories obtained by combining three individual 20ns MD trajectories for each system (calreticulin + G1M3, calreticulin + G1M3 + ATP and calreticulin + G1M3 + ADP) using the crystal structure of murine calreticulin (PDB ID 3RG0) docked to G1M3 and nucleotides. For the covariance analyses, each individual structure in the MD trajectory was superimposed onto the globular domain of calreticulin in the starting structure. This procedure removes rotational and translational motions of the proteins that occur over the course of the MD trajectories.

3.4 References

(1) Braakman, I.; Bulleid, N. J. *Annu Rev Biochem* **2011**, *80*, 71.
(2) Peters, L. R.; Raghavan, M. *J Immunol* **2011**, *187*, 919.
(3) Jackson, S. E. *Top Curr Chem* **2012**.
(4) Clare, D. K.; Vasishtan, D.; Stagg, S.; Quispe, J.; Farr, G. W.; Topf, M.; Horwich, A. L.; Saibil, H. R. *Cell* **2012**, *149*, 113.
(5) Kityk, R.; Kopp, J.; Sinning, I.; Mayer, M. P. *Mol Cell* **2012**.
(6) Nadeau, K.; Das, A.; Walsh, C. T. *J Biol Chem* **1993**, *268*, 1479.
(7) Nakamura, K.; Zuppini, A.; Arnaudeau, S.; Lynch, J.; Ahsan, I.; Krause, R.; Papp, S.; De Smedt, H.; Parys, J. B.; Muller-Esterl, W.; Lew, D. P.; Krause, K. H.; Demaurex, N.; Opas, M.; Michalak, M. *J Cell Biol* **2001**, *154*, 961.
(8) Prodromou, C.; Roe, S. M.; O'Brien, R.; Ladbury, J. E.; Piper, P. W.; Pearl, L. H. *Cell* **1997**, *90*, 65.
(9) Barril, X.; Brough, P.; Drysdale, M.; Hubbard, R. E.; Massey, A.; Surgenor, A.; Wright, L. *Bioorg Med Chem Lett* **2005**, *15*, 5187.
(10) Didenko, T.; Duarte, A. M.; Karagoz, G. E.; Rudiger, S. G. *Biochim Biophys Acta* **2012**, *1823*, 636.
(11) Del Cid, N.; Jeffery, E.; Rizvi, S. M.; Stamper, E.; Peters, L. R.; Brown, W. C.; Provoda, C.; Raghavan, M. *J Biol Chem* **2010**, *285*, 4520.
(12) Ellgaard, L.; Frickel, E. M. *Cell Biochem Biophys* **2003**, *39*, 223.
(13) Raghavan, M.; Wijeyesakere, S. J.; Peters, L. R.; Del Cid, N. *Trends Immunol* **2012**.
(14) Pocanschi, C. L.; Kozlov, G.; Brockmeier, U.; Brockmeier, A.; Williams, D. B.; Gehring, K. *J Biol Chem* **2011**, *286*, 27266.
(15) Saito, Y.; Ihara, Y.; Leach, M. R.; Cohen-Doyle, M. F.; Williams, D. B. *EMBO J* **1999**, *18*, 6718.
(16) Wijeyesakere, S. J.; Gafni, A. A.; Raghavan, M. *J Biol Chem* **2011**, *286*, 8771.
(17) Gao, B.; Adhikari, R.; Howarth, M.; Nakamura, K.; Gold, M. C.; Hill, A. B.; Knee, R.; Michalak, M.; Elliott, T. *Immunity* **2002**, *16*, 99.
(18) Chouquet, A.; Paidassi, H.; Ling, W. L.; Frachet, P.; Houen, G.; Arlaud, G. J.; Gaboriaud, C. *PLoS One* **2011**, *6*, e17886.
(19) Kozlov, G.; Pocanschi, C. L.; Rosenauer, A.; Bastos-Aristizabal, S.; Gorelik, A.; Williams, D. B.; Gehring, K. *J Biol Chem* **2010**, *285*, 38612.
(20) Frickel, E. M.; Riek, R.; Jelesarov, I.; Helenius, A.; Wuthrich, K.; Ellgaard, L. *Proc Natl Acad Sci U S A* **2002**, *99*, 1954.
(21) Kozlov, G.; Bastos-Aristizabal, S.; Maattanen, P.; Rosenauer, A.; Zheng, F.; Killikelly, A.; Trempe, J. F.; Thomas, D. Y.; Gehring, K. *J Biol Chem* **2010**, *285*, 35551.
(22) Brooks, B. R.; Brooks, C. L., III; Mackerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. *J Comput Chem* **2009**, *30*, 1545.
(23) Wu, G.; Robertson, D. H.; Brooks, C. L., 3rd; Vieth, M. *J Comput Chem* **2003**, *24*, 1549.

(24) Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. *J Comput Chem* **2009**, *30*, 2785.
(25) Trott, O.; Olson, A. J. *J Comput Chem* **2010**, *31*, 455.
(26) Lee, M. S.; Feig, M.; Salsbury, F. R., Jr.; Brooks, C. L., 3rd *J Comput Chem* **2003**, *24*, 1348.
(27) Lee, M. S.; Salsbury, F. R.; Brooks, C. L. *J Chem Phys* **2002**, *116*, 10606.
(28) Vieth, M.; Hirst, J. D.; Dominy, B. N.; Daigler, H.; Brooks, C. L. *J Comput Chem* **1998**, *19*, 1623.
(29) Vieth, M.; Hirst, J. D.; Kolinski, A.; Brooks, C. L. *J. Comput. Chem.* **1998**, *19*, 1612.
(30) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.
(31) Aravind, L.; Koonin, E. V. *Nucleic Acids Res* **1999**, *27*, 1609.
(32) Mayer, M. P. *Molecular Cell* **2010**, *39*, 321.
(33) Jeffery, E.; Peters, L. R.; Raghavan, M. *J Biol Chem* **2011**, *286*, 2402.
(34) Kapoor, M.; Ellgaard, L.; Gopalakrishnapai, J.; Schirra, C.; Gemma, E.; Oscarson, S.; Helenius, A.; Surolia, A. *Biochemistry* **2004**, *43*, 97.
(35) Kapoor, M.; Srinivas, H.; Kandiah, E.; Gemma, E.; Ellgaard, L.; Oscarson, S.; Helenius, A.; Surolia, A. *Journal of Biological Chemistry* **2003**, *278*, 6194.
(36) Radkiewicz, J. L.; Brooks, C. L. *Journal of the American Chemical Society* **2000**, *122*, 225.
(37) Kumar, S.; Huang, C.; Zheng, G.; Bohm, E.; Bhatele, A.; Phillips, J. C.; Yu, H.; Kale, L. V. *Ibm Journal of Research and Development* **2008**, *52*, 177.
(38) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *Journal of Physical Chemistry B* **1998**, *102*, 3586.
(39) Mackerell, A. D.; Feig, M.; Brooks, C. L. *J Comput Chem* **2004**, *25*, 1400.
(40) Vanommeslaeghe, K.; MacKerell, A. D., Jr. *Journal of Chemical Information and Modeling* **2012**, *52*, 3144.
(41) Vanommeslaeghe, K.; Raman, E. P.; MacKerell, A. D., Jr. *Journal of Chemical Information and Modeling* **2012**, *52*, 3155.

Chapter 4

Targeting KIX domain of CBP with small molecules

4.1: Introduction

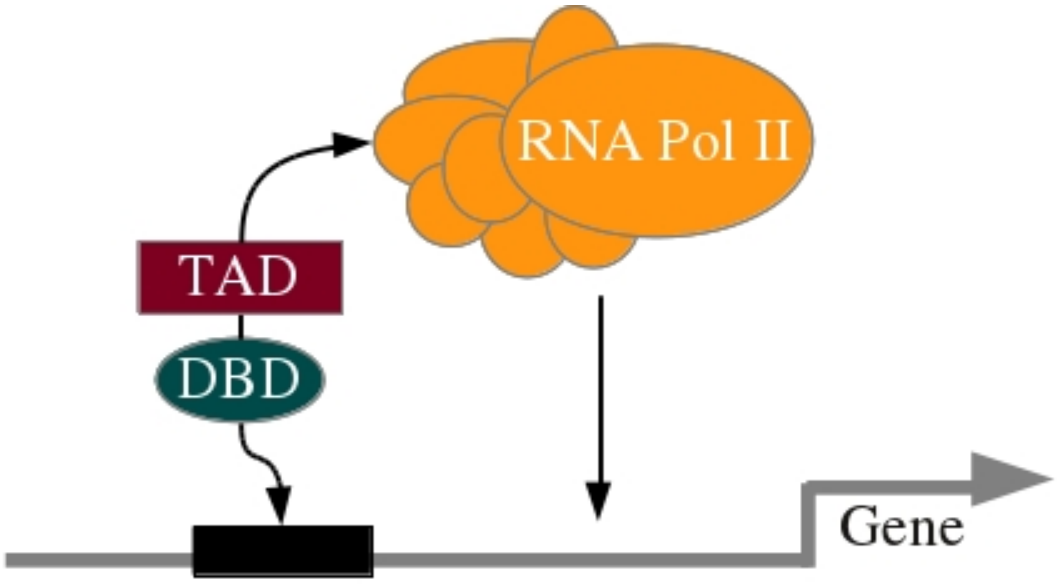


Figure 4.1: Schematic of transcription activation, where a DNA binding domain (DBD) and a transcriptional activation domain (TAD) recruit the transcription machinery to up-regulate a gene. (Adapted from Lee LW, et. al., 2010)

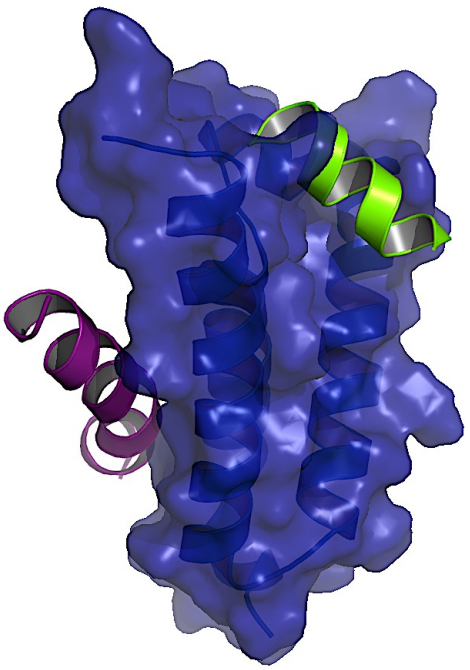


Figure 4.2: GACKIX domain of CREB binding protein (CBP in blue with activation domains from the c-Myb (in purple) and mixed lineage leukemia (MLL in green) transcription factors bound to the two distinct binding sites of this CBP. (PDBID: 2agh)

Transcription is an intricate and highly regulated process that is essential for a myriad of cellular processes including cell proliferation and apoptosis.[1-4] A vast assortment of protein-protein interactions compose the complex machinery that oversees the regulation of transcription.[5] The schematic in Figure 1 is a brief overview of some of the components required for transcription to be up regulated. A DNA binding domain (DBD) interacts with transcriptional activator domains (TAD).[5,6] The TADs recruit the transcription machinery, and bring it into proximity to the DNA upstream of the gene because of the DBD interaction with the DNA, up-regulating transcription of that gene.

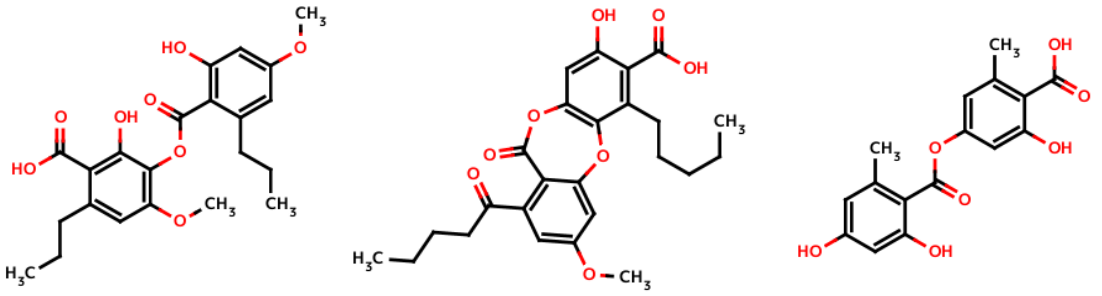
One protein involved in transcription regulation is the co-activator CREB binding protein (CBP) and its paralog p300. The GACKIX domain of CBP interacts with greater than ten distinct sequences at two unique binding sites (Figure 2).[7-10] The ability of this domain to interact with such a variety of peptide sequences makes it difficult to target with small molecules due to the necessary plasticity of these molecules to recognize so many unique binding partners.[11,12] Additionally, specificity is can be encouraged through allosteric regulation, where the binding of one peptide will effect the affinity of the

second, adding further complexity to inhibiting these interactions.[13-16] While there have been some here have been successes in targeting protein-protein interactions with small molecules targeting interactions that demonstrate such high promiscuity continue to be fairly illusive.[17-21]

This challenging system was an interesting application for collaborative efforts that I undertook with the Mapp research group at University of Michigan. Contained in this chapter are efforts towards identifying possible small molecule inhibitors for the GACKIX domain of CBP.

4.2 Natural products inhibitors as possible helical mimetic

The GACKIX domain of CBP is a plastic surface, and has the ability to interact with a variety of different peptide sequences. This domain has further complexity with the ability have an allosteric response between its two binding pockets. To overcome this challenging system my experimental collaborators developed a high throughput screen to capture possible inhibitors of this domain. The high throughput screen utilized a high-throughput fluorescence polarization assay with a fluorescein-labeled version of the MLL transcriptional activation domain. Two different compound sets were screened experimentally. One was a drug-like set of 50,000 commercially available compounds, which yielded no hits. A set of approximately 15,000 natural product extracts was also screened. This set consisted of extracts isolated from cyanobacteria, sponges and sediments, and yielded 64 hits that inhibited the MLL interaction with the GACKIX domain. The active ingredients were identified using NMR and mass spectrometry and included sekikaic acid (Figure 3). Sekikaic acid is a depside which was first identified by Emil Fischer as polypeptide-like small molecules consisting of a series of phenol carboxylic acids units.[22] Depsides have been reported to have antibiotic, anti-oxidant, and anti-HIV activity and have recently been shown to affect protein-protein interactions. [23-26]



Sekikaic acid **Lobaric acid** **Lecanoric acid**

Figure 4.3: Sekikaic acid is a small molecule natural product and was identified by high-throughput screening as an inhibitor for peptide interactions with GACKIX domain of CBP. The success of sekikaic acid lead to investigated the structurally relevant small molecules lecanoric and lobaric acids.

Sekikaic acid inhibits the MLL-GACKIX complex in a dose dependent fashion with an IC_{50} of 34 mM based on fluoresce polarization based inhibition assays. As shown in Figure 2 in section 4.1, GACKIX has two unique binding sites, and to investigate the second binding site a FP-based inhibition experiment was performed with KID. Sekikaic acid also inhibited the KID-GACKIX interaction with an IC_{50} of 64 μM. These experiments demonstrate that the small-molecule sekikaic acid is capable of blocking activator peptides at both binding sites on the GACKIX domain, and is the first reported small molecule that can effectively perform this function.

Due to this unique property of sekikaic acid it was important to investigate the structural details of the small-molecule's interaction with the GACKIX domain. Circular dichroism and 1D- ^{1}H-NMR experiments demonstrated that sekikaic acid binding is reversible and not inducing protein misfolding or aggregation. Additionally, ^{1}H,^{15}N-HSQC NMR experiments with ^{15}N-labeled GACKIX in the presence and absence of sekikaic acid revealed significant chemical shift perturbations in several regions of the

receptor (Figure 4). The most significant shifts were around the flexible loop region

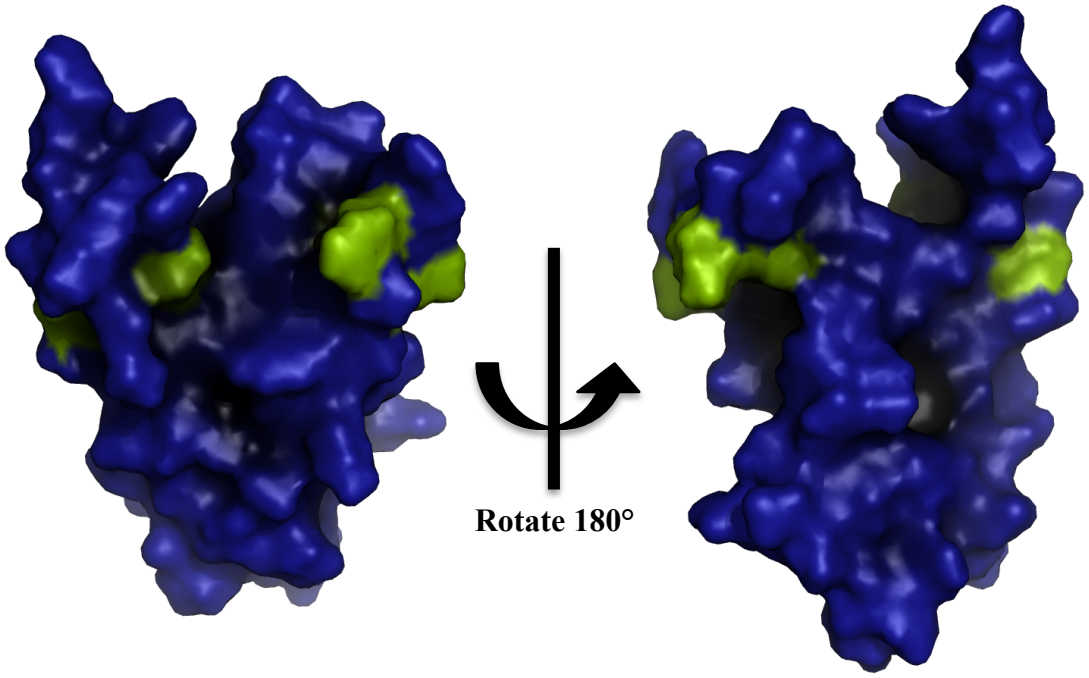


Figure 4.4: Residues (green) measured to be highly shifted upon binding of sekikaic acid via HSQC-NMR. (PDBID:2agh) Regions containing these residues were targeted as possible binding surfaces during docking trials.

where MLL interacts with GACKIX and defined by residues: Val 608, Leu 620, Lys 621, Arg 624, Met 625, and Glu 626.[27] Residues near the binding region for CREB also experienced significant shifts upon sekikaic acid binding: Leu 664 and Lys 662, which is a key residue in binding CREB to GACKIX.[28] These shifts suggest a model where sekikaic acid binds to the MLL binding site and disrupts the interaction of the allosterically connected KID-binding site. This experimental data was used to inform the computational protein-ligand docking experiments with aims to obtain an atomistic-structural model of the sekikaic-GACKIX interaction. The highly shifted residues, shown in green in Figure 4, were used as centers for the grid generation of the different docking trials. Sekikaic acid conformations were randomly generated and placed at centers of mass selected near each target binding site. However, the generated ligand

conformations had near identical energies, making it difficult to distinguish small-molecule binding modes from misdocked conformations.

While unable to identify in atomistic detail the binding orientation of sekikaic acid to the GACKIX domain other structural questions could be addressed. TAD-binding motifs like those on the GACKIX domain tend to be hydrophobic and conformationally dynamic so they can interact with many distinct amphipathic TADs.[27,29-32] Sekikaic acid has unique ability to interact with two unique sites that bind these helical, amphipathic TADs and it was hypothesized that this small-molecule could interact with GACKIX in a helical mimetic manner. To investigate this we performed molecular dynamics (MD) simulations of sekikaic acid and two structurally similar compounds lecanoric and lobaric acids (Figure 3).

The MD simulations were run in implicit solvent and begun from geometry-optimized coordinates of the small molecules. All three small molecules have a convenient property of containing two substituted phenyl rings with an ester linker. This allows for an equivalent metric to inspect the conformational rigidity across these small molecules (Figure 5). The trajectories of the small-molecules were aligned to one phenyl ring and the angle between the two planes defined by the phenyl rings was followed throughout the trajectory (Left panel of Figure 5) and the probability density of the angles sampled was calculated (Right panel Figure 5). Lecanoric acid has the largest variation in angles sampled, as seen by its broad probability distribution, which spans approximately 160 degrees. Sekikaic acid has a more narrow distribution, peaking at about 95 degrees, indicating it is not as free to rotate around its ester bond, and is likely hindered by its

larger functional groups. Lobaric acid is limited to sampling angles distributed about 50 degrees as it is restrained by the bonds between the phenyl rings.

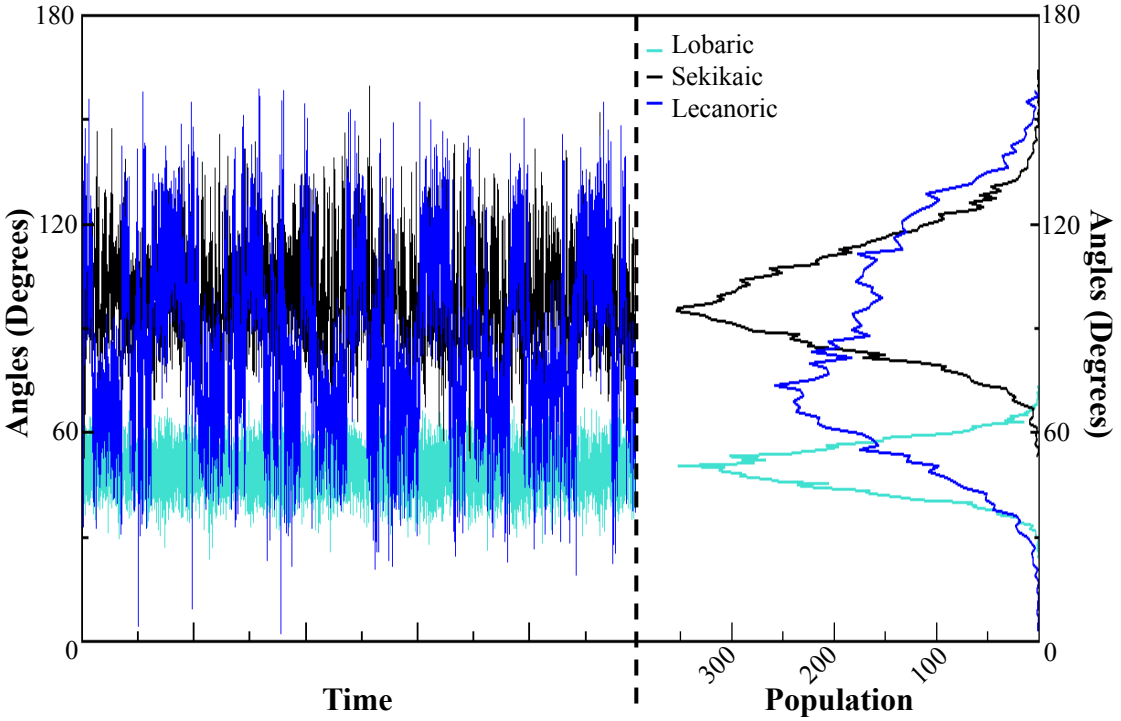


Figure 4.5: Angles between the phenyl rings during MD simulation in implicit solvent for three natural products that are structurally similar. Lobaric acid (cyan) rings are constrained by bonds between them and has very narrow angles that may be sampled. Sekikiac acid (black) has larger functional groups that limit the angles sampled unlike the unhindered lecanoric acid (blue).

While sekikaic and lobaric acids are not as free to rotate about the ester bond as lecanoric acid, it does not show the positioning of the functional groups off the phenyl rings. For a helical mimetic, looking down the z-axis of the small molecule there should be functional groups occupying space in regular intervals. In Figure 6 (A-C), the small-molecule phenyl rings were aligned to the first frame and the position of the last heavy atom for a given functional group was plotted. Interestingly, sekikaic acid has functional groups that stay in a fairly constrained area, in a circular pattern around the z-axis when

compared to the other two small-molecules. Lecanoric acid, consistent with the relative angle analysis in Figure 5, had the largest spread of positions of its functional groups. Lobaric acid was similar to sekikaic acid. However, due to the functional groups containing longer chains the spread of possible positions is greater than that of sekikaic acid. Sekikaic acid demonstrates a relative structural rigidity to the other two small molecules, and has functional group able to occupy space as would side chains of peptide helix.

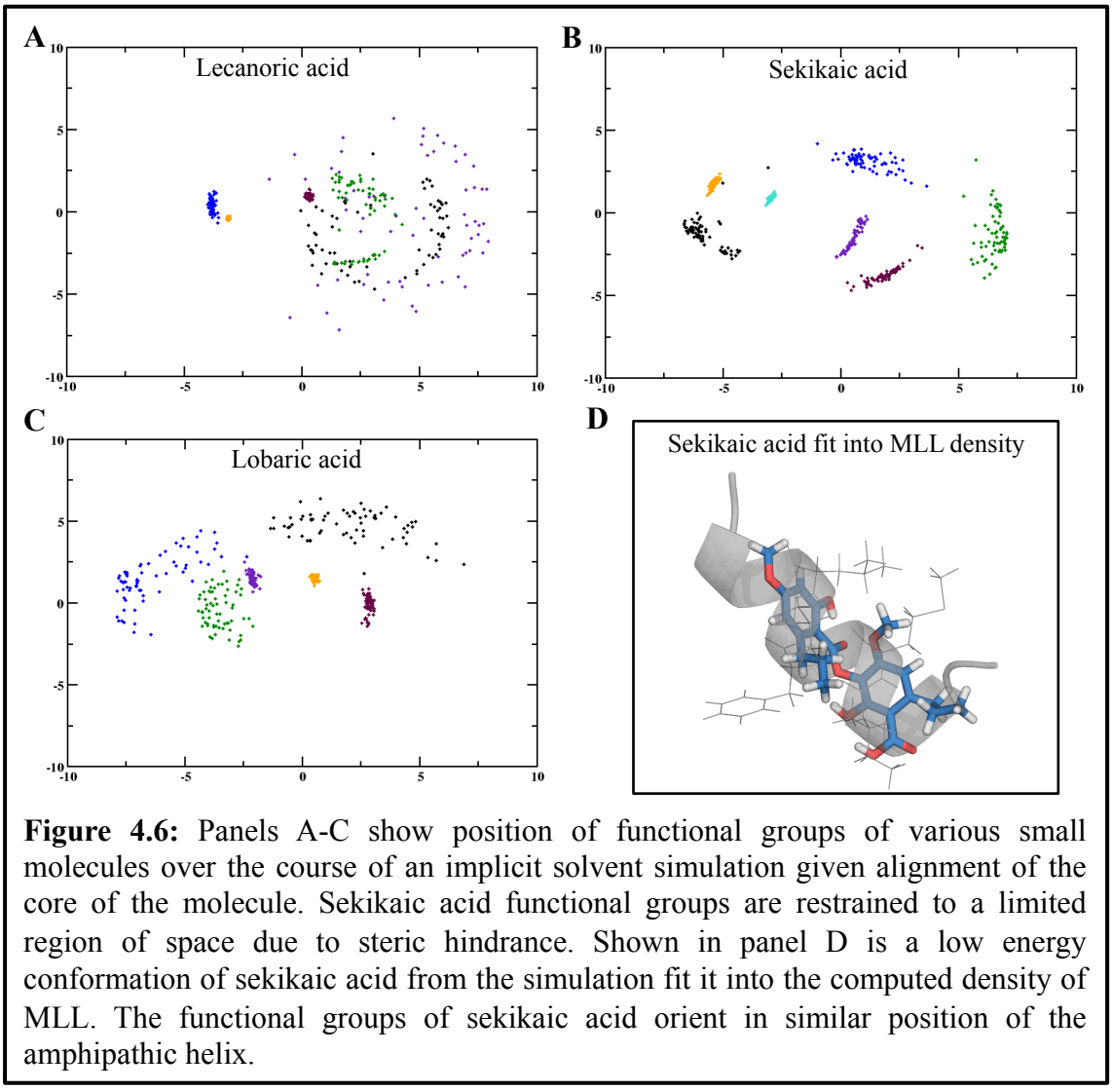


Figure 4.6: Panels A-C show position of functional groups of various small molecules over the course of an implicit solvent simulation given alignment of the core of the molecule. Sekikaic acid functional groups are restrained to a limited region of space due to steric hindrance. Shown in panel D is a low energy conformation of sekikaic acid from the simulation fit it into the computed density of MLL. The functional groups of sekikaic acid orient in similar position of the amphipathic helix.

To visualize sekikaic acid as a possible helical mimetic, low energy members of highly populated clusters of conformations from the implicit solvent trajectories were fit into a computed pseudo-density map of MLL (Figure 6D). The density map was generated with a 5Å resolution using the Situs tool and the clustered conformations were rigidly fit using Chimera.[33,34] The low energy, highly sampled conformations of sekikaic acid fits well into the pseudo-density map, with functional groups oriented in similar positions as the side-chains of MLL supporting that this small molecule would interact with GACKIX as helical mimetic.

To investigate the acids that are structurally similar to sekikaic acid were subjected to competitive binding assays. Interestingly, the highly flexible lecanoric acid was unable to inhibit binding of both peptides, MLL and KID, which bind to separate binding sites on GACKIX. However, the more constrained lobaric acid was able to inhibit both activators, as did sekikaic acid, with IC_{50}s of 17 μM for MLL and 25 μM for KID. These data suggest that identification of possible helical mimetic could be used as a tool towards targeting challenging, dynamic surfaces like those on GACKIX.

These natural product GACKIX inhibitors were the first shown to modulate two distinct binding sites of such a dynamic receptor surface. Sekikaic acid shows conformational rigid and functional group orientations in solvent to suggest a helical mimetic. These properties can be used as a basis for the development of future small-molecule inhibitors to target dynamic surfaces. Targeting transcription co-activator domains with small-molecule transcriptional modulators could be an invaluable biochemical probe and, potentially, have powerful therapeutic implications.

4.3: Methods

Molecular dynamics simulations of the natural products in implicit solvent were performed using CHARMM.[35,36] Simulations were run at 300K for 1 nanosecond after an equilibration of 500 picoseconds using the Leapfrog Verlet integrator and a 1 femtosecond timestep.[37] The CHARMM27 all-atom force field was used while the parameters for the small molecules were produced using the atom typing toolset MATCH and the CHARMM Generalized Force Field (CGenFF).[38-40] The implicit solvent model used was GBMV, with a constant dielectric and a 16 angstrom nonbonded cutoff.[41,42] The simulations were started from geometry optimized ligand coordinates using Hartree-Fock with the 6-31G* basis set in Guassian03.[43-45] Clustering was performed using the MMTSB Tool Set with a 2 angstrom heavy atom RMSD radius.[46]

The simulations revealed that these small molecules prefer to adopt a helical-like conformation in solution. Specifically, it was observed that the two phenyl rings of the small-molecules had a tendency to orient themselves out of register with respect to the other. To quantify this observation we measured the angle between the two planes that pass through the carbons on each of the phenyl rings. A time series of these angles is plotted in the right panel of Figure 4, while the probability density is plotted on the left panel of the same figure. The simulations were then aligned to the core of the small molecule such that the core was along the z-axis. For Figure 5(A-C), the x and y coordinates of the last heavy atom of a given function group (see Figure 1) was plotted at each simulation snapshot.

To visually demonstrate that sekikaic acid takes on a helical mimetic-like

conformation in solution, the small molecule conformations generated from the simulations were clustered and a low energy member from each of the five largest clusters was rigidly fit using Chimera into a pseudo-density map of the MLL peptide generated using the Situs tool with an approximate resolution of 5 angstroms. As shown in Figure 5D, a preferred conformation of sekikaic acid fits well into the pseudo-density map, and supports the idea that it behaves as a helical mimic.

4.4 Tethered small-molecule fragments and further docking methodology development

Text adapted from: N. Wang, C.Y. Majmudar, W.C. Pomerantz, JK Gagnon, JD Sadowsky, JL Meagher, TK Johnson, JA Stuckey, CL Brooks III, JA Wells and AK Mapp. Ordering a dynamic protein via a small-molecule stabilizer. *J Am Chem Soc*, 2013, **135**:3363-6.

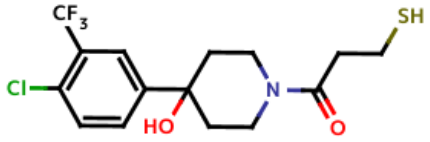
The GACKIX domain of co-activator CBP/p300 recognizes a diverse set of transcriptional activators at two separate binding sites. This ability to reorganize to accommodate so many different binding partners makes it especially challenging for structural characterization. Here we use a covalently tethered small-molecule to capture an angstrom-resolution snapshot of the conformationally dynamic GACKIX domain. This is the first time a crystal structure of this coactivator was obtained, suggesting that the ligand discovery strategy of tethering to stabilize a dynamic protein could be a general approach for structural characterization of these challenging targets.[47-50]

We screened small molecule fragments using the tethering approach to identify small-molecule fragments that interact with the GACKIX domain.[50] GACKIX has two unique binding sites, and the site that binds the Mixed Lineage Leukemia (MLL) activator consists of a deeper more distinct binding pocket than the other site that binds KID activator.[51,52] A residue was selected near the edge of the MLL binding surface, Leu 664, and was mutated to a cysteine. Small molecule fragments were screened in the presence of a competitor, β-mercaptoethanol, where molecules with some affinity for the receptor would remain through the experiment to be characterized. Two fragments

(fragment 1-10 and 2-64) were identified in this screen to have high tethering efficiency to GACKIX L664C by DR_{50} (dose response) values of 2-8μM (Figure 7).

Given the promising tethering efficiency we investigated the fragments' ability to inhibit binding of activators at the two binding surfaces of GACKIX. Employing fluorescent polarization binding assays the affect the fragments had on GACKIX L664C binding MLL or pKID. Fragment 1-10 and 2-64 both decreased the binding of MLL by approximately 7-16 fold. Fragment 2-64 has no affect on the ability of GACKIX L664C to bind pKID, which is the distal site from the MLL site targeted by the initial high throughput screen.[53] However, pKID affinity is reduced by approximately 2-fold when GACKIX is bound to the fragment 1-10. This suggests that 1-10 is disrupting the allosteric communication between the two binding surfaces on GACKIX.[9,54,55]

Frag 1-10



Frag 2-64

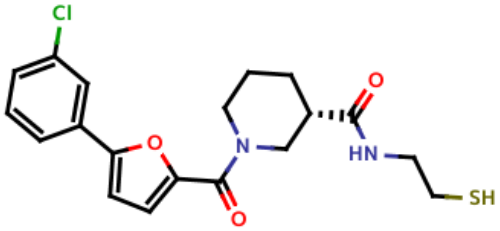


Figure 4.7: Small-molecule fragments identified with a high throughput screen that utilizes a disulfide tethering technique.

The tethered fragments significantly altered the stability of the GACKIX domain. This was assessed for each of the fragment-protein pairs by measuring changes in CD-monitored thermal melting temperature, the extent of amide hydrogen-deuterium (H-D) exchange and the extent of proteolysis. The protein-small molecule complexes demonstrate a 15-18°C (≥20%) increase in melting temperature. Improved protection in

hydrogen-deuterium exchange showing that 40-55% of the amides were protected compared to the apo protein.[47] The proteolytic stability (half life) of the tethered complex increased 5-37 fold compared to a T½ range of 0.7-2.1 min of the untethered protein.[15,56] These data encouraged the pursuit of crystallization of fragment—GACKIX L664C complexes.

High quality crystals were obtained of the fragment 1-10/GACKIX L664C complex when dissolved in 1.8 M ammonium sulfate and 0.1 M Tris, pH 7.0 at 25° C. However, only microcrystals of 2-64 tethered to GACKIX L664C were obtained and were of too poor quality to solve. Initial attempts at molecular replacement strategies were undertaken using the NMR structures of GACKIX bound to native transcriptional activation domains but were unsuccessful.[53,57] Therefore, a selenomethionine mutant of GACKIX tethered to 1-10 was prepared and the X-ray structure solved. Using these data, the structure of 1-10—GACKIX L664C was determined to 2.0 Å resolution (Figure 8). The tethered fragment sits in the MLL binding surface of GACKIX and is oriented towards the core of the

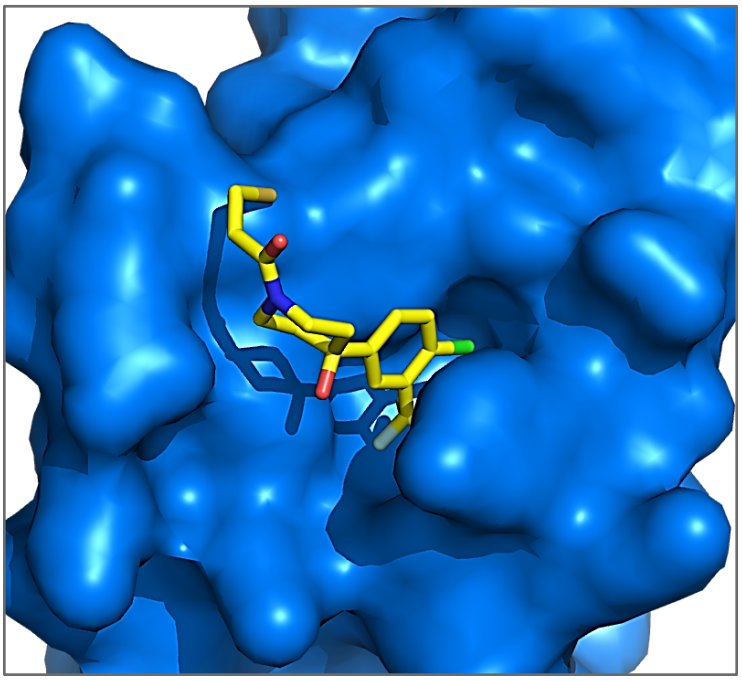


Figure 4.8: The first crystal structure of the GACKIX domain of CBP was obtained with the stability gained from binding tethered small-molecule fragment 1-10. (PDBID:4i9o)

protein. Interestingly, the aromatic ring of 1-10 is positioned relatively deep in a hydrophobic pocket that is created by a rotameric shift of Tyr631 which has previously been shown to be involved in GACKIX interacting with MLL.[55,57]

The prevailing structural model of amphipathic activator/coactivator complexes is that the activator forms an amphipathic helix upon binding to the surface of the coactivator.[58-60] While only a few different surfaces have been characterized to date, the available structures suggest that the binding surfaces are often shallow and broad making them particularly challenging to target with small molecules that have far less volume and surface area than the typical helix of a transcriptional activator.[61-63] Considering an overlay of the 1-10/GACKIX L664C structure with the averages of the previously reported NMR structures of GACKIX demonstrates a high similarity of backbone structure with a RMSD of the backbone ranging between 1.07-1.81Å.[57,64,65] These structures vary in the loop region between helices α1 and α2 that has a backbone RMSD range of 2.73-3.11Å. These variations in the loop region are consistent with literature suggesting the conformational loop being

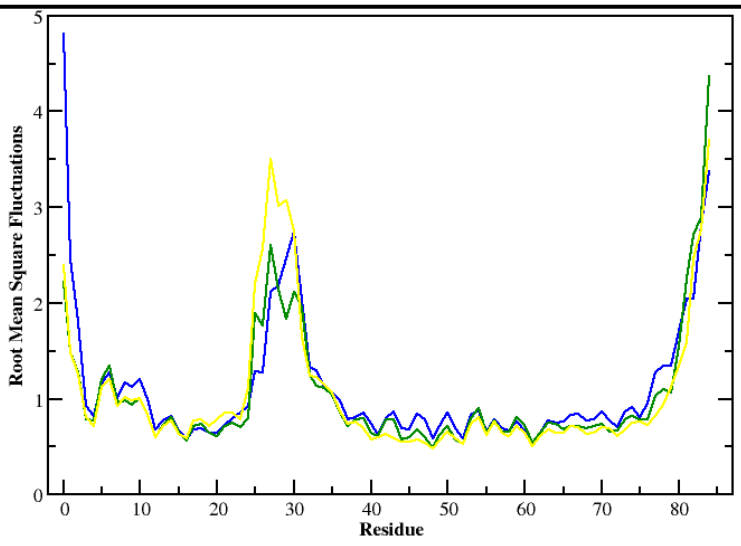


Figure 4.9: Cα RMSF for untethered GACKIX L664C (blue), fragment 1-10 tethered (yellow), and fragment 2-64 tethered (green).

integral in the ability of GACKIX to accommodate such a variety of binding partners.[53,64,66,67]

To dissect in more detail how the GACKIX surface remodels itself to recognize fragment 1-10 we carried out 40 ns MD simulations of the GACKIX solid-state structure in the presence and absence of ligand. A gross comparison of the amide backbone reveals that a change in the loop confirmation is the most significant, as shown in the root mean square fluctuations (Figure 9) and in the average structure overlay (Figure 10). These changes are often difficult to visualize by solution methods because the loop region contains several proline residues, but mutagenesis and NMR methods have suggested that conformational plasticity in this region underlies the ability of GACKIX to recognize diverse amphipathic sequences.[54,57,67] It is this movement of the loop and a rotation of helix α1 that enable the formation of a narrower binding surface to

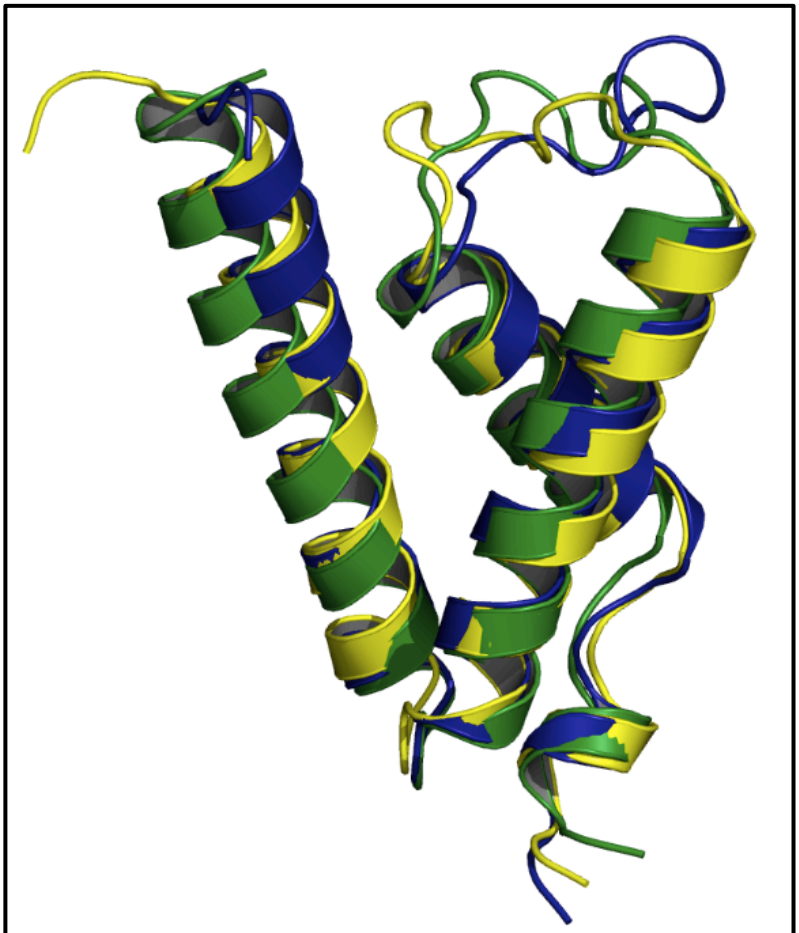


Figure 4.10: Overlay of the structures closest to the calculated average from the simulations with untethered GACKIX L664C shown in blue, fragment 1-10 tethered in yellow, and fragment 2-64 tethered in green.

accommodate a molecule that is considerably smaller than a peptidic helix (~77% smaller volume). The binding surface that is targeted by 1-10 is also significantly different, both as a result of loop conformational changes and because of side chain motions as demonstrated by the change in solvent accessible surface area of the residues when the fragment is tethered (Figure 11). For example, the liganded GACKIX shows a population shift in the Tyr 631 side chain chi angles relative to the untethered protein, leading to a deeper hydrophobic binding surface for the small molecule.

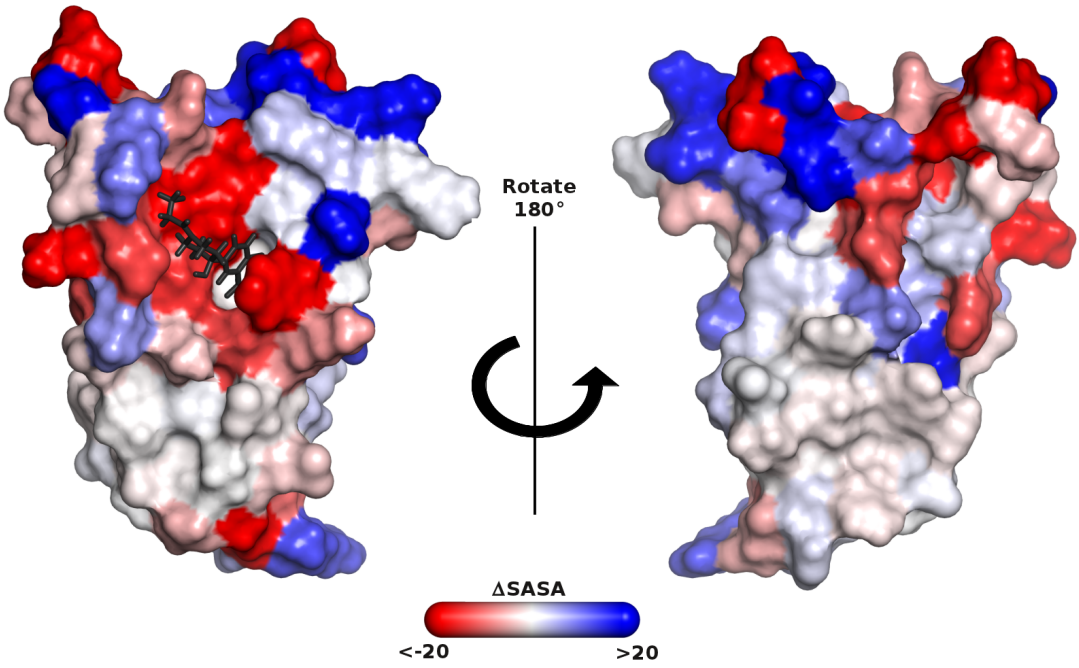


Figure 4.11: Difference in the average solvent-accessible surface area (SASA, in Å^2) calculated by residue between simulations of untethered and 1-10-tethered GACKIX L664C. A residue colored red is less solvent-exposed in the 1-10-tethered structure, with the color intensity indicating the extent of the change; blue residues are more solvent-exposed in the 1-10-tethered structure.

While the larger fragment 2-64 demonstrated greater stabilization to the GACKIX domain in thermal melt, and greater protection in the H/D exchange the crystals produced for the 2-64 tethered to GACKIX complex were of poor quality. To identify a conformation of fragment 2-64 when tethered to GACKIX we developed methodology within the CHARMM docking package, CDOCKER.[68-70] The methodology is in the style of MMTSB tool and consists of a set of Perl scripts to interface user input with CHARMM.[35,36,46] The scripts generate a disulfide bond between the fragment and protein receptor allowing for sampling of physical dihedral angle around that bond. Docking trials of fragment 2-64 onto GACKIX L664C were performed in triplicate, and the resulting conformations clustered. The lowest energy member of the lowest energy cluster was simulated using a GPU capable version of CHARMM. The resulting simulations of 2-64 tethered to GACKIX L664C predict the binding mode of this ligand to be similar to that of 1-10 and further demonstrates the ability of this protein to adapt to different binding partners (Figure 12). The α3 and α2 helices must open to accommodate this larger ligand and correspond to

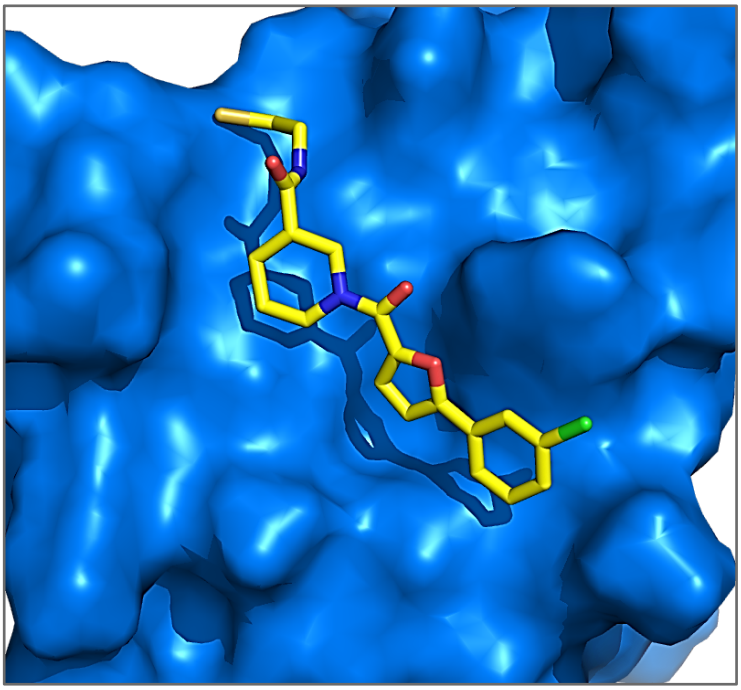


Figure 4.12: Predicted low energy docking result of Fragment 2-64 into L664C KIX domain of CBP suggests that binding of this fragment disrupts the crystal packing through rearrangement of key residues.

changes in the NMR chemical shifts of residues involved in this opening. Additionally, many side chains shift to accommodate the larger fragment 2-64 including Tyr 631 which makes important crystal contacts (Figure 12). This ligand binding orientation offers a hypothesis to the lack of high-quality crystals for the fragment 2-64/GACKIX complex.

Having a crystal structure of such a dynamic receptor as GACKIX domain of CBP was a very interesting trial for Flexible CDOCKER. As previously mentioned we developed the docking methodology to have flexible side-chains to sample with the ligand conformations during docking simultaneously. In addition to the flexible receptor docking, we developed a docking methodology that includes physical disulfide bond between the small molecule fragment and GACKIX. A cross-docking trial of docking fragment 1-10 into a structure of GACKIX solved in the presence of another ligand

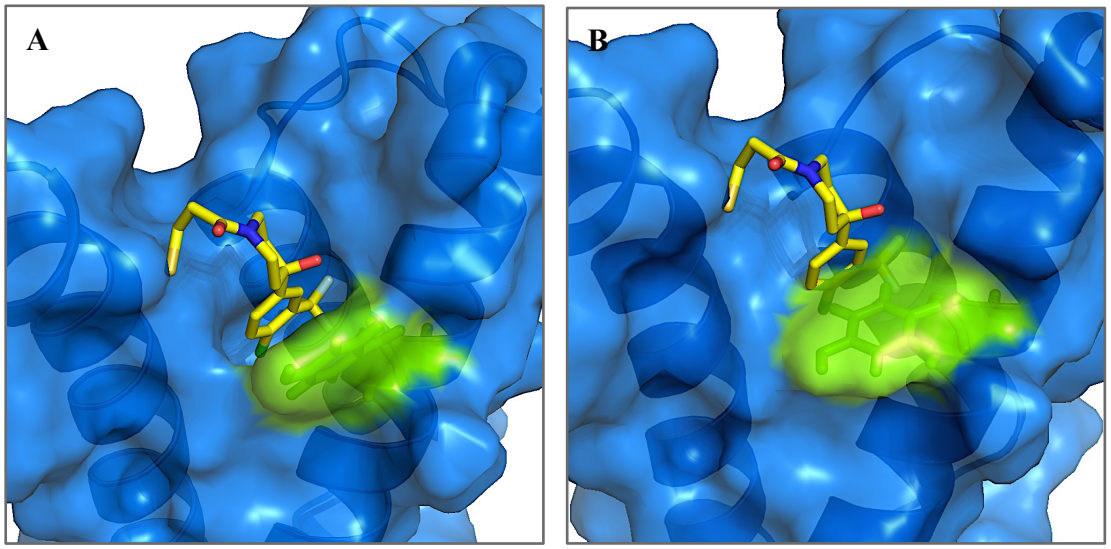


Figure 4.13: Cross-docking of fragment 1-10 to the NMR structure of KIX domain of CBP. (PDBID:1kdx) During docking the tyrosine is able to move out of the way to accomidate the small molecule (A). For comparison the same ligand conformation overlay with the starting NMR structure (B).

would be a challenging test for this new docking methodology. Towards this end, we

docked to the first set of coordinates from the NMR structure of GACKIX (PDBID: 1kdx), which was experimentally determined in the presence of phosphorylated kinase-inducible domain (pKID). The resulting low energy conformation is shown in Figure 13 (left panel), which has a RMSD from the crystallographic ligand pose of about 3Å. While the RMSD is greater than that of the docking convention of a RMSD of < 2Å to be considered docked it should be noted that the Tyr631 (shown in green) shifts rotameric orientation to reveal the deeper hydrophobic pocket (Figure 13 right panel). Also notable are the energies for the lower RMSD to the crystallographic orientation of fragment 1-10 tend to be lower than those conformations far from the native state (Figure 14). The results from this test are encouraging test and suggest Flexible CDOCKER with tethered ligands can be a powerful tool for investigating fragments binding at alternate locations or possibly other receptors.

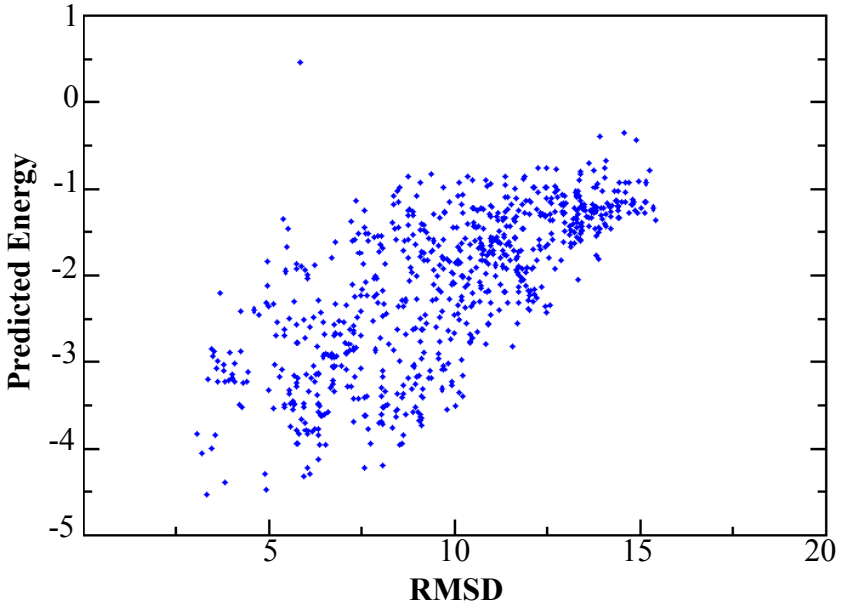


Figure 4.14: The predicted energies for fragment 1-10 of various docking poses and the RMSD from the crystallographic conformation. The lower predicted energy conformations trend towards lower RMSD.

In conclusion, we have obtained an angstrom-resolution snapshot of the conformationally dynamic co-activator GACKIX domain complexed with a small molecule. Development in the CDOCKER docking protocol to include a physical disulfide bond and a flexible receptor were used to investigate the conformation of the larger fragment 2-64 and a cross-docking trial of fragment 1-10. In the cross-docking trial, the receptor conformation shifted in a way that allowed the fragment to sit in a deeper hydrophobic pocket seen in the crystal structure (Figure 13). Additionally, the docking results followed by simulation suggest that the fragment 2-64/GACKIX complex may not have crystalized due to the binding mode of fragment 2-64 disrupting key crystal contacts formed in the fragment 1-10/GACKIX complex.

From a broader perspective, these results suggest that tethering may be an exceptionally enabling approach to obtain long-sought solid state data of conformationally dynamic proteins. This includes transcriptional coactivators such as CBP/p300 targeted here, but also members of other cellular machineries that rely upon conformationally dynamic interfaces to recognize binding partners.[71-72,73]

4.4 Tethering methods

The MD simulations were carried out using a GPU capable CHARMM version c37b1 with an OpenMM interface and run on two nVidia Tesla C2050 GPUs using OpenMM version 4.0.[35,36,38,74] The simulations are 40 ns in length, and were run at constant pressure and temperature (298K) with the CHARMM27 forcefield, using Langevin dynamics to provide the thermal heat bath with a friction coefficient of 10 ps-1.[38,75] A variable time step was implemented, where the time step was minimally bound to 2 fs, but could increase to a time step that satisfied the error tolerance of 0.002. Particle mesh ewald and a non-bonded cutoff of 9Å were used with an FFT grid size of 72. SHAKE was applied to the hydrogen atoms and the parameters for the small molecule fragments were obtained using MATCH.[39,40] For the fragment 1-10 simulation, the starting structure was the crystal structure reported in this manuscript. Similarly, the untethered GACKIX L664C simulation began from the crystal structure. However, the small molecule fragment was deleted, leaving only GACKIX L664C.

CDOCKER was used to determine the pose for fragment 2-64 and the predicted lowest energy pose was simulated for a total of 80 ns under the same conditions as above to confirm the ligand was stable in that configuration. The fragment was docked into the crystal structure, as the fragment was a similar small molecule as fragment 1-10. Starting configurations for fragment 2-64 were generated using random rotations around torsions that were then subjected to 100 rounds of simulated annealing, where the ligand and select residues were flexible while the rest of the system was represented as a grid. Each

docking trial generated a total of 2,500 structures and three docking trials were run for this study, yielding a total of 7,500 ligand poses. These poses were clustered using K-means clustering within the MMTSB toolset with a heavy atom cutoff of 3Å.[46] The energy was calculated for each cluster, using the member closest to the cluster centroid. The resulting lowest energy pose was used in the all-atom simulation.

The solvent accessible surface area (SASA), root mean square fluctuations (RMSF), and average structures were calculated using CHARMM.[35,36] The SASA was calculated per residue and averaged over each frame of the trajectory. The average structures were calculated for the heavy atoms, also using the entire trajectory. The calculated average structure was then compared to each simulation frame and the frame with the smallest root mean square deviation (RMSD) from the calculated average was chosen for figures 9 & 10. Images were generated using PyMol and VMD.[76,77]

4.6 References

(1) Darnell, J. E. *Nature Reviews Cancer* **2002**, *2*, 740.
(2) Ptashne, M.; Gann, A. *Nature* **1997**, *386*, 569.
(3) Kasper, L. H.; Boussouar, F.; Ney, P. A.; Jackson, C. W.; Rehg, J.; van Deursen, J. M.; Brindle, P. K. *Nature* **2002**, *419*, 738.
(4) Wood, M. A.; Attner, M. A.; Oliveira, A. M.; Brindle, P. K.; Abel, T. *Learn Mem* **2006**, *13*, 609.
(5) Lemon, B.; Tjian, R. *Genes & Development* **2000**, *14*, 2551.
(6) Lee, L. W.; Mapp, A. K. *Journal of Biological Chemistry* **2010**, *285*, 11033.
(7) Cook, P. R.; Polakowski, N.; Lemasson, I. *Journal of Molecular Biology* **2011**, *409*, 384.
(8) Ernst, P.; Wang, J.; Huang, M.; Goodman, R. H.; Korsmeyer, S. J. *Molecular and Cellular Biology* **2001**, *21*, 2249.
(9) Goto, N. K.; Zor, T.; Martinez-Yamout, M.; Dyson, H. J.; Wright, P. E. *Journal of Biological Chemistry* **2002**, *277*, 43168.
(10) Vendel, A. C.; McBryant, S. J.; Lumb, K. J. *Biochemistry* **2003**, *42*, 12481.
(11) Boehr, D. D.; Nussinov, R.; Wright, P. E. *Nature Chemical Biology* **2009**, *5*, 789.
(12) Singh, G. P.; Ganapathi, M.; Dash, D. *Proteins-Structure Function and Bioinformatics* **2007**, *66*, 761.
(13) Swapna, L. S.; Mahajan, S.; de Brevern, A.; Srinivasan, N. *BMC Struct Biol* **2012**, *12*, 6.
(14) Bruschweiler, S.; Schanda, P.; Kloiber, K.; Brutscher, B.; Kontaxis, G.; Konrat, R.; Tollinger, M. *J Am Chem Soc* **2009**, *131*, 3063.
(15) Korkmaz, E. N.; Nussinov, R.; Haliloglu, T. *PLoS Comput Biol* **2012**, *8*, e1002420.
(16) Thompson, A. D.; Dugan, A.; Gestwicki, J. E.; Mapp, A. K. *ACS Chem Biol* **2012**.
(17) Arkin, M. R.; Wells, J. A. *Nature Reviews Drug Discovery* **2004**, *3*, 301.
(18) Berg, T. *Curr Opin Chem Biol* **2008**, *12*, 464.
(19) Majmudar, C. Y.; Mapp, A. K. *Curr Opin Chem Biol* **2005**, *9*, 467.
(20) Lee, L. W.; Mapp, A. K. *J Biol Chem* **2010**, *285*, 11033.
(21) Cummings, C. G.; Hamilton, A. D. *Curr Opin Chem Biol* **2010**, *14*, 341.
(22) Fischer, E. *J Am Chem Soc* **1914**, *36*, 1170.
(23) Thadhani, V. M.; Choudhary, M. I.; Ali, S.; Omar, I.; Siddique, H.; Karunaratne, V. *Nat Prod Res* **2011**, *25*, 1827.
(24) Fujimoto, A.; Shingai, Y.; Nakamura, M.; Maekawa, T.; Sone, Y.; Masuda, T. *Bioorg Med Chem Lett* **2010**, *20*, 7393.
(25) Sperl, B.; Seifert, M. H.; Berg, T. *Bioorg Med Chem Lett* **2009**, *19*, 3305.
(26) Muller, K. *Appl Microbiol Biotechnol* **2001**, *56*, 9.
(27) Arai, M.; Dyson, H. J.; Wright, P. E. *FEBS Lett* **2010**, *584*, 4500.
(28) Radhakrishnan, I.; Pérez-Alvarado, G. C.; Parker, D.; Dyson, H. J.; Montminy, M. R.; Wright, P. E. *J Mol Biol* **1999**, *287*, 859.

(29) De Guzman, R. N.; Goto, N. K.; Dyson, H. J.; Wright, P. E. *J Mol Biol* **2006**, *355*, 1005.
(30) Dyson, H. J.; Wright, P. E. *Nat Rev Mol Cell Biol* **2005**, *6*, 197.
(31) Ferreira, M. E.; Hermann, S.; Prochasson, P.; Workman, J. L.; Berndt, K. D.; Wright, A. P. *J Biol Chem* **2005**, *280*, 21779.
(32) Hermann, S.; Berndt, K. D.; Wright, A. P. *J Biol Chem* **2001**, *276*, 40127.
(33) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. *Journal of Computational Chemistry* **2004**, *25*, 1605.
(34) Birmanns, S.; Rusu, M.; Wriggers, W. *Journal of Structural Biology* **2011**, *173*, 428.
(35) Brooks, B. R.; Brooks, C. L., III; Mackerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. *Journal of Computational Chemistry* **2009**, *30*, 1545.
(36) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *Journal of Computational Chemistry* **1983**, *4*, 187.
(37) Crippen, G. M.; Havel, T. F. *Journal of Chemical Information and Computer Sciences* **1990**, *30*, 222.
(38) Mackerell, A. D.; Feig, M.; Brooks, C. L. *Journal of Computational Chemistry* **2004**, *25*, 1400.
(39) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; MacKerell, A. D., Jr. *Journal of Computational Chemistry* **2010**, *31*, 671.
(40) Yesselman, J. D.; Price, D. J.; Knight, J. L.; Brooks, C. L., III *Journal of Computational Chemistry* **2012**, *33*, 189.
(41) Lee, M. S.; Feig, M.; Salsbury, F. R.; Brooks, C. L. *Journal of Computational Chemistry* **2003**, *24*, 1348.
(42) Lee, M. S.; Salsbury, F. R.; Brooks, C. L. *Journal of Chemical Physics* **2002**, *116*, 10606.
(43) Rassolov, V. A.; Ratner, M. A.; Pople, J. A.; Redfern, P. C.; Curtiss, L. A. *Journal of Computational Chemistry* **2001**, *22*, 976.
(44) Frisch, M.; Trucks, G. W.; Schlegel, H.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C. **2008**.
(45) Ditchfie.R; Hehre, W. J.; Pople, J. A. *Journal of Chemical Physics* **1971**, *54*, 724.
(46) Feig, M.; Karanicolas, J.; Brooks, C. L. *Journal of Molecular Graphics & Modelling* **2004**, *22*, 377.
(47) Liu, G.-H.; Qu, J.; Shen, X. *Biochimica Et Biophysica Acta-Molecular Cell Research* **2008**, *1783*, 713.
(48) Sadowsky, J. D.; Burlingame, M. A.; Wolan, D. W.; McClendon, C. L.; Jacobson, M. P.; Wells, J. A. *Proceedings of the National Academy of Sciences of the United States of America* **2011**, *108*, 6056.
(49) Hardy, J. A.; Lam, J.; Nguyen, J. T.; O'Brien, T.; Wells, J. A. *Proceedings of the National Academy of Sciences of the United States of America* **2004**, *101*, 12461.

(50) Scheer, J. M.; Romanowski, M. J.; Wells, J. A. *Proceedings of the National Academy of Sciences of the United States of America* **2006**, *103*, 7595.
(51) Erlanson, D. A.; Braisted, A. C.; Raphael, D. R.; Randal, M.; Stroud, R. M.; Gordon, E. M.; Wells, J. A. *Proceedings of the National Academy of Sciences of the United States of America* **2000**, *97*, 9367.
(52) Majmudar, C. Y.; Hojfeldt, J. W.; Arevang, C. J.; Pomerantz, W. C.; Gagnon, J. K.; Schultz, P. J.; Cesa, L. C.; Doss, C. H.; Rowe, S. P.; Vasquez, V.; Tamayo-Castillo, G.; Cierpicki, T.; Brooks, C. L., III; Sherman, D. H.; Mapp, A. K. *Angewandte Chemie-International Edition* **2012**, *51*, 11258.
(53) Pomerantz, W. C.; Wang, N.; Lipinski, A. K.; Wang, R.; Cierpicki, T.; Mapp, A. K. *Acs Chemical Biology* **2012**, *7*, 1345.
(54) Radhakrishnan, I.; PerezAlvarado, G. C.; Parker, D.; Dyson, H. J.; Montminy, M. R.; Wright, P. E. *Cell* **1997**, *91*, 741.
(55) Arai, M.; Dyson, H. J.; Wright, P. E. *Febs Letters* **2010**, *584*, 4500.
(56) Park, C. W.; Marqusee, S. *Nature Methods* **2005**, *2*, 207.
(57) Hoofnagle, A. N.; Resing, K. A.; Ahn, N. G. *Annual Review of Biophysics and Biomolecular Structure* **2003**, *32*, 1.
(58) Uesugi, M.; Nyanguile, O.; Lu, H.; Levine, A. J.; Verdine, G. L. *Science* **1997**, *277*, 1310.
(59) Fuller, J. C.; Burgoyne, N. J.; Jackson, R. M. *Drug Discovery Today* **2009**, *14*, 155.
(60) Thoden, J. B.; Ryan, L. A.; Reece, R. J.; Holden, H. M. *Journal of Biological Chemistry* **2008**, *283*, 30266.
(61) Thompson, A. D.; Dugan, A.; Gestwicki, J. E.; Mapp, A. K. *Acs Chemical Biology* **2012**, *7*, 1311.
(62) Fry, D. C. *Biopolymers* **2006**, *84*, 535.
(63) Wang, F.; Marshall, C. B.; Yamamoto, K.; Li, G.-Y.; Gasmi-Seabrook, G. M. C.; Okada, H.; Mak, T. W.; Ikura, M. *Proceedings of the National Academy of Sciences of the United States of America* **2012**, *109*, 6078.
(64) Zor, T.; De Guzman, R. N.; Dyson, H. J.; Wright, P. E. *Journal of Molecular Biology* **2004**, *337*, 521.
(65) Saalau-Bethell, S. M.; Woodhead, A. J.; Chessari, G.; Carr, M. G.; Coyle, J.; Graham, B.; Hiscock, S. D.; Murray, C. W.; Pathuri, P.; Rich, S. J.; Richardson, C. J.; Williams, P. A.; Jhoti, H. *Nature Chemical Biology* **2012**, *8*, 920.
(66) De Guzman, R. N.; Goto, N. K.; Dyson, H. J.; Wright, P. E. *Journal of Molecular Biology* **2006**, *355*, 1005.
(67) Brueschweiler, S.; Schanda, P.; Kloiber, K.; Brutscher, B.; Kontaxis, G.; Konrat, R.; Tollinger, M. *Journal of the American Chemical Society* **2009**, *131*, 3063.
(68) Vieth, M.; Hirst, J. D.; Dominy, B. N.; Daigler, H.; Brooks, C. L. *Journal of Computational Chemistry* **1998**, *19*, 1623.
(69) Vieth, M.; Hirst, J. D.; Kolinski, A.; Brooks, C. L. *Journal of Computational Chemistry* **1998**, *19*, 1612.
(70) Wu, G. S.; Robertson, D. H.; Brooks, C. L.; Vieth, M. *Journal of Computational Chemistry* **2003**, *24*, 1549.
(71) Basse, N.; Kaar, J. L.; Settanni, G.; Joerger, A. C.; Rutherford, T. J.; Fersht, A. R. *Chemistry & Biology* **2010**, *17*, 46.

(72) Dunker, A. K.; Silman, I.; Uversky, V. N.; Sussman, J. L. *Current Opinion in Structural Biology* **2008**, *18*, 756.
(73) Dyson, H. J.; Wright, P. E. *Nature Reviews Molecular Cell Biology* **2005**, *6*, 197.
(74) Eastman, P.; Friedrichs, M. S.; Chodera, J. D.; Radmer, R. J.; Bruns, C. M.; Ku, J. P.; Beauchamp, K. A.; Lane, T. J.; Wang, L.-P.; Shukla, D.; Tye, T.; Houston, M.; Stich, T.; Klein, C.; Shirts, M. R.; Pande, V. S. *Journal of Chemical Theory and Computation* **2013**, *9*, 461.
(75) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *Journal of Physical Chemistry B* **1998**, *102*, 3586.
(76) The PyMOL Molecular Graphics System, Version 1.7.4 Schrödinger, LLC.
(77) Humphrey, W.; Dalke, A.; Schulten, K. *Journal of Molecular Graphics & Modelling* **1996**, *14*, 33.

Chapter 5

Applying Go-Like models to investigate allosteric signalling in KIX

Text adapted from: S.M. Law, J.K. Gagnon, A.K. Mapp, C.L. Brooks III. Pre-Paying the Entropic Cost for Allosteric Regulation in KIX. *Proc. Natl. Acad. Sci. U.S.A.* 2014.

5.1 Introduction

The term "allosteric" describes the effects of one ligand (or effector) on the catalysis or binding of another ligand at a non-overlapping site.[1,2] This classical view can be characterized by three main attributes: (i) only two states exist (e.g. *active* or *inactive*); (ii) the allosteric signal is relayed through a single well-defined pathway; and (iii) a conformational change occurs in the non-overlapping site. A modern interpretation treats the native state as a conformational ensemble (rather than only two conformational states) whose pre-existing population of conformations can be redistributed as a result of some allosteric perturbation that can be transmitted through multiple signaling pathways.[3-6] This new "unified view" defines allostery in purely structural and thermodynamic terms (i.e. allostery can be controlled by entropy and/or enthalpy) and provides a quantitative framework for classifying allosteric mechanisms.

Allostery is a central pillar in the regulation of gene transcription. The transcriptional coactivator CREB binding protein (CBP), which serves as a keystone in the assembly of the transcriptional machinery, is responsible for regulating transcriptional processes relating to hematopoiesis [7,8], cell differentiation [9,10], etc. Contained within CBP is an 87 residue-long domain called KIX that is critical for regulating transcriptional activity and can be cooperatively targeted by a diverse set of transcription factors [11-16] and small molecules.[17-22,23] The structure of KIX (Figure 1) can be summarized as a three-helix bundle (α_1, residues 597-611; α_2, residues 623-640; α_3, residues 646-669) containing a mobile L_{12}-G_2 loop (residues 614-621) and an N-terminal 3_{10} helix (residues 591-594). More importantly, KIX consists of at least two binding sites that are located on opposite surfaces of the protein, which can simultaneously bind, for example, the activation domain of the mixed lineage leukemia (MLL) protein and the activation domain of the proto-oncogene protein c-Myb (Figure 1), both of which are expressed concurrently during hematopoietic precursor cell development and therefore play an important role in blood cell proliferation.[24-27]

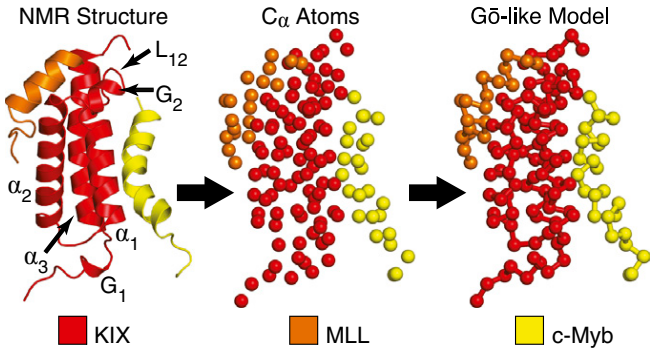


Figure 5.1: Different representations of the MLL-KIX-c-Myb ternary complex (PDB ID code 2AGH).

The binding of either MLL or c-Myb to one of these promiscuous sites can stabilize the C-terminus of the α_3 helix of KIX and/or decrease the mobility of the L_{12}-G_2 loop but the functional role of these changes remains to be delineated.[4,19,28-32]

Furthermore, *in vitro*, KIX alone demonstrates a ~2 fold lower affinity for c-Myb and ~1.6 fold lower affinity for MLL when compared to KIX bound to MLL or c-Myb, respectively.[12] Although numerous efforts have been put forth to clarify the molecular mechanism for cooperative binding in KIX, the current allosteric picture detailing the thermodynamic and structural effects of MLL and/or c-Myb binding remains incomplete. While all-atom MD simulations complement experiment and have provided valuable insight into the allosteric process, they are computationally too expensive to explore timescales where multiple binding events or protein folding/unfolding are necessary to understand the underlying mechanism of action.[19,30,33] To address these limitations, we utilize a topology-based Gō-like model [34,35] that has shown success in examining coupled protein folding and binding [36-44] as well as modeling ligand-induced structural transitions, and allosteric communication.[45-47] We extend its application to the quantitative study of allostery and cooperative binding for the first time here in the context of the MLL-KIX-c-Myb ternary system. To develop a quantitative picture that connects the microscopic changes in structural dynamics to the macroscopic shift in binding affinities and kinetics, we performed multiple simulations for each of KIX-free, MLL-KIX, KIX-c-Myb, and MLL-KIX-c-Myb and carried out a comparative analysis of the potential allosteric communication pathways. Through this approach, we detected similarities in protein folding/unfolding, observed multiple c-Myb/MLL binding events, and estimated thermodynamic and kinetic parameters for these processes. Our results show that the binding of either peptide stabilizes the L_{12}-G_2 loop as well as increases the helicity of the α_3 helix, both of which play a role in controlling the size of the hydrophobic core. Most notably, we find that this reduction in structural dynamics results in an increase in the

population of KIX structures that resemble KIX in the ternary complex and the entropic cost for binding a peptide at the allosteric site is lowered when an effector is already bound to KIX. Collectively, our data demonstrate that these observations are sufficient to explain the increased binding affinity for c-Myb and MLL when the other complementary peptide is bound to KIX and can direct the design of small molecules to target important regions of conformationally dynamic proteins.

Table 5.1:

Ligand	Binding to	Expt. K_d^*, μM	Simulated $K_d^\dagger$, μM	Simulated $k_{off}^\dagger$, s^{-1}	Simulated $k_{on}^\dagger$, $M^{-1}{\cdot}s^{-1}$	$-T\Delta S^\ddagger$, kcal/mol
c-Myb	KIX	10 ± 2	30 ± 8	5.92E+06 ± 0.94E+06	0.20E+12 ± 0.02E+12	0.8
c-Myb	KIX-MLL	4 ± 1	3 ± 1	1.36E+06 ± 0.24E+06	0.48E+12 ± 0.17E+12	1.4
MLL	KIX	2.8 ± 0.4	0.8 ± 0.3	2.46E+06 ± 0.54E+06	3.19E+12 ± 0.60E+12	1.2
MLL	KIX-c-Myb	1.7 ± 0.1	0.1 ± 0.1	0.75E+06 ± 0.10E+06	6.02E+12 ± 0.57E+12	1.4

*Experimental K_d published in ref. 13.
†Averaged over three groups of 20 simulations.
‡Calculated at 300 K relative to KIX-free.

5.2 Results

Binding affinities and configurational entropies

Using a renormalized Gō-like model that was adjusted to capture experimentally reported helical content and binding affinities, extensive MD simulations of the free KIX, KIX-c-Myb, MLL-KIX, and MLL-KIX-c-Myb systems were performed. It is important to emphasize that no additional calibration of the intermolecular interactions was used to simulate the MLL-KIX-c-Myb ternary complex. That is, aside from the native contacts, the interaction energies were not explicitly biased towards ternary complex formation (see Supporting Information). The resulting thermodynamic parameters and kinetic rate constants for each simulation system are presented in Table 1 and show that MLL and c-Myb bind cooperatively to KIX rather than competing for KIX binding. The simulations revealed that c-Myb and MLL bind to KIX with a K_d = 30 ± 8 μM and a K_d = 0.8 ± 0.3 μM, respectively. In the presence of MLL, the affinity of KIX for c-Myb is enhanced by ~10 fold and is accompanied by a ~4.4 fold decrease in k_{off} and ~2.4 fold increase in k_{on}. Similarly, the affinity of KIX-c-Myb for MLL is enhanced by ~8 fold and is accompanied by a ~3.3 fold decrease in k_{off} and ~1.9 fold increase in k_{on}. Relative to free KIX, the difference in configurational entropy ($-T\Delta S$) at 300 K for forming KIX-c-Myb, MLL-KIX, and MLL-KIX-c-Myb was ~0.8 kcal/mol, ~1.2 kcal/mol, and ~1.4 kcal/mol, respectively (see Table 1, Figure 4 and Figure 8).

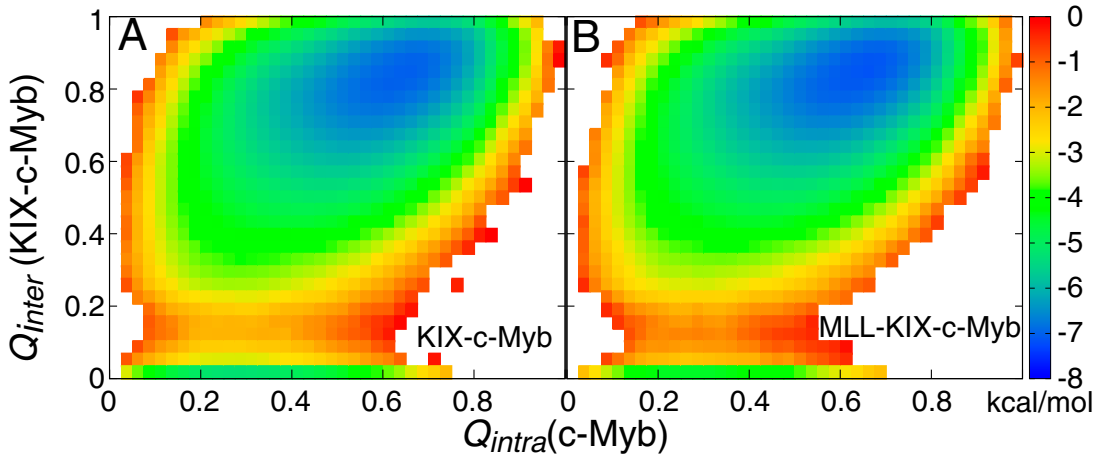


Figure 5.2: Two-dimensional free-energy surfaces as a function of c-Myb folding and binding in (A) the absence of MLL and (B) the presence of MLL. Qintra is the fraction of native intramolecular contacts in c-Myb, whereas Qinter is the fraction of native intermolecular contacts formed between KIX-c-Myb.

IDP binding mechanism

We examined the relationship between c-Myb binding and c-Myb folding by constructing free energy profiles along the fraction of intramolecular (Q_{intra}) and intermolecular (Q_{inter}) native contacts (Figure 2). In both the presence and absence of MLL, the free energy minimum shows c-Myb as being ~40-80% folded and bound to KIX with ~70-95% of its intermolecular native contacts intact. However, when c-Myb loses all of its intermolecular native contacts it is ~10-50% helical. Free c-Myb was not found to exist in a preformed helical state and, instead, followed a folding-and-binding mechanism that appears to be independent of MLL. Similar results that support a folding-and-binding mechanism were also observed for MLL binding to free KIX (Figure 3*).

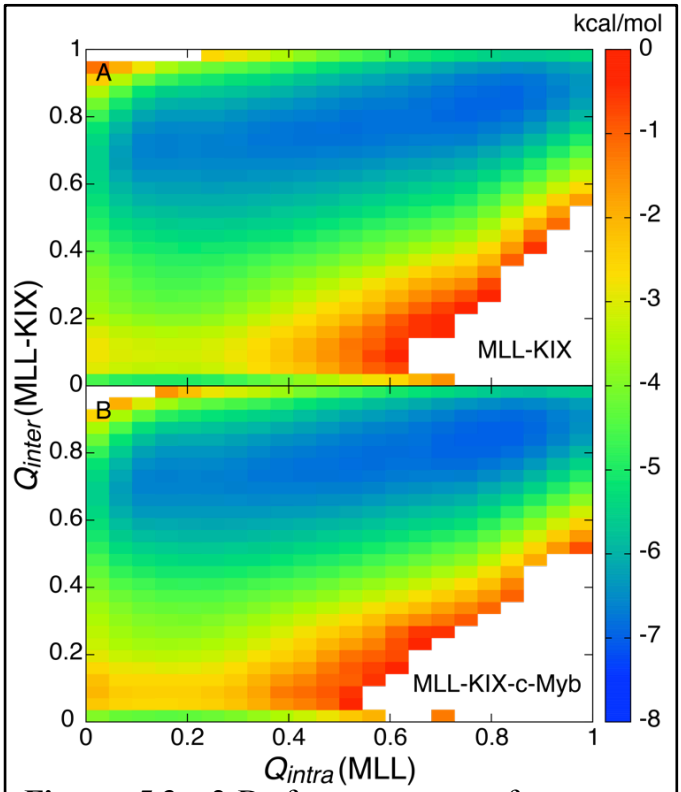


Figure 5.3: 2-D free energy surfaces as a function of MLL folding and binding in (A) the absence of c-Myb and (B) the presence of c-Myb. Q_{intra} is the fraction of native intramolecular contacts in MLL while Q_{inter} is the fraction of native intermolecular contacts formed between MLL-KIX.

To further probe the effects of MLL on c-Myb binding, we compared the order of contact formation between KIX and c-Myb by monitoring the contact appearance order. Leu302 of c-Myb is predominantly first to form its intermolecular native contacts (Figure 4*). Then, other intermolecular native contacts are formed sequentially outward towards both termini of c-Myb. The order in which the intermolecular native contacts were formed between KIX-c-Myb was also found to be independent of MLL binding. Similarly, the order of contact formation between MLL and KIX was also found to be independent of c-Myb binding (Figure 5).

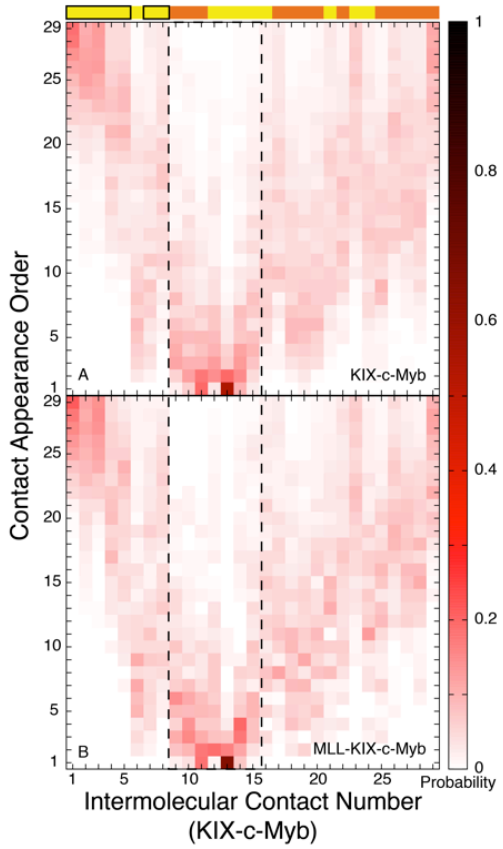
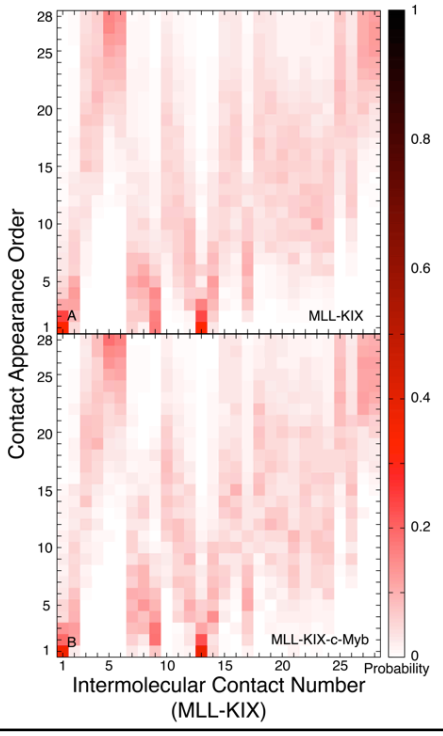


Figure 5.4: Contact appearance order (CAO) for c-Myb binding in (A) the absence of MLL and (B) the presence of MLL. The horizontal axis corresponds to native intermolecular contacts made between KIX-c-Myb and low, medium, and high contact numbers correspond to contacts made by the N-terminus, middle, and C-terminus of c-Myb, respectively. The order in which contacts are formed between KIX-c-Myb is shown along the vertical axis and can be thought of as a time axis. White squares have a zero probability and dark red squares have the highest probability of being the n^{th} contact formed. The colored bar above the figure shows where c-Myb contacts KIX with orange and yellow squares corresponding to helices α_1 and α_3, respectively. The C-terminus of α_3 is denoted by solid black boxes. Intermolecular native contacts made by Leu302 are enclosed by a dashed box (contact numbers 9-15).

Figure 5.5: Contact appearance order (CAO) for MLL binding in (A) the absence of c-Myb and (B) the presence of c-Myb. The horizontal axis corresponds to native intermolecular contacts made between MLL-KIX and low, medium, and high contact numbers correspond to contacts made by the N-terminus, middle, and C-terminus of MLL, respectively. The order in which contacts are formed between MLL-KIX is shown along the vertical axis and can be thought of as a time axis. White squares have a zero probability and dark red squares have the highest probability of being the n^{th} contact formed.



KIX dynamics

Figure 6 shows the root-mean-square fluctuation of the highest ranked mode computed from principal component analysis (PCA) and projected onto the average simulation structure for each system. In free KIX, the C-terminus of the α_3 helix and the L_{12}-G_2 loop were the most mobile. In fact, the average structure shows a completely unfolded α_3 C-terminus beginning at around residue 657. Upon binding c-Myb, the mobility of the α_3 helix is slightly reduced while the L_{12}-G_2 loop remains as flexible as in free KIX. Aside from the C-terminus of the α_3 helix, the average structures of KIX-free and KIX-c-Myb are very similar. In contrast, MLL binding stabilizes the C-terminal residues of α_3 and regains the secondary structure of this helix. Also, the L_{12}-G_2 loop appears to be less mobile than in free KIX or KIX-c-Myb. Finally, the ternary complex shows a rigid α_3 helix and only some flexibility in the L_{12}-G_2 loop, similar to the MLL-KIX binary system. Consistent with these findings, Figure 9 shows that the binding of MLL and/or c-Myb results in a significant increase in α_3 helicity.

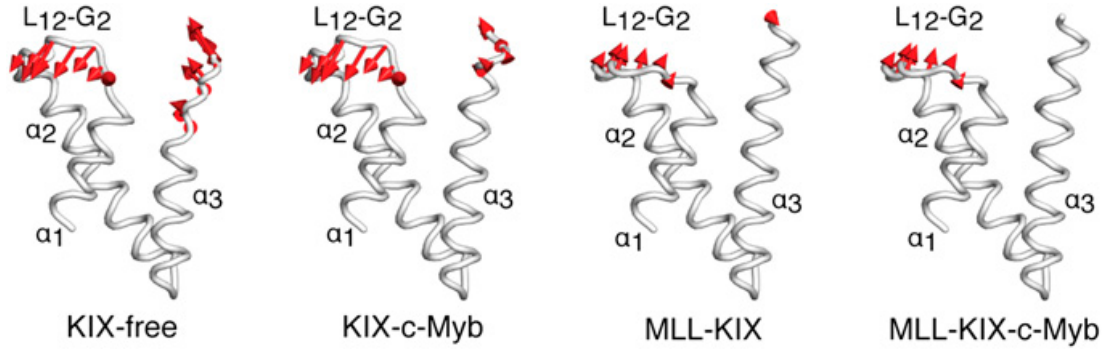


Figure 5.6: Root-mean-square fluctuations from PCA. The average KIX structure from each simulation is shown in white, and the ligands were omitted for clarity. The red arrows point along the direction of the lowest frequency mode (highest ranked eigenvector), and the amplitude is directly pro- portional to the length of the arrow. Only fluctuations greater than 1.0 Å are displayed.

Representative free energy surfaces along the L_{12}-G_2 loop RMSD (with respect to the NMR structure) and the hydrophobic core size, which effectively describes the distance between helices α_1 and α_3, were computed and are displayed in Figure 7.

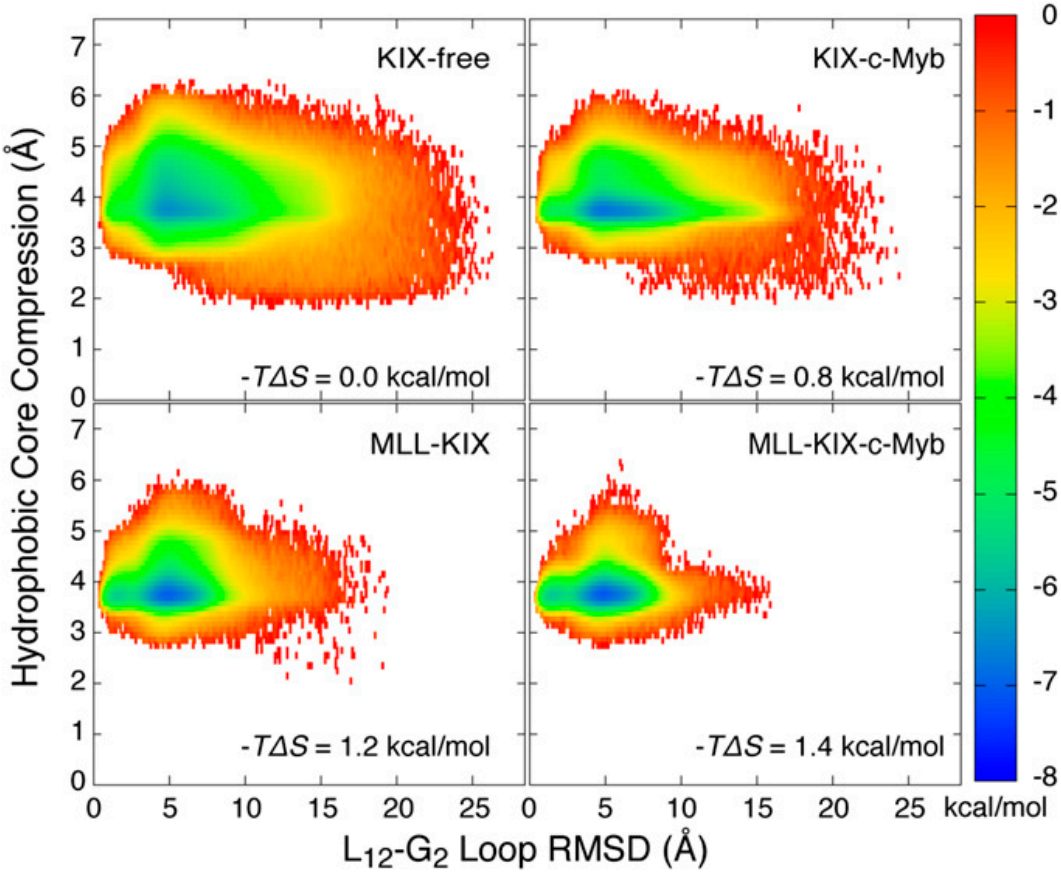


Figure 5.7: Representative 2D free-energy surfaces as a function of the hydro- phobic core size and the L12-G2 loop RMSD. All RMSD values were calculated relative to the NMR structure, which had a hydrophobic core size of ~3.6 Å. All ΔS were calculated with respect to KIX-free at 300 K.

Overall, the conformational sampling of KIX along the two collective variables is the most restricted when both peptides are bound. Relative to KIX-free, the binding of MLL and/or c-Myb decreases the extent of sampling of the loop with the ternary complex being affected the most. The maximum opening of the hydrophobic core is also reduced from ~5.5 Å to ~4 Å when the loop deviates significantly from the reference NMR structure. Again, this trend is most pronounced in the ternary complex. In all cases, the

lowest free energy minimum has a hydrophobic core that is open by ~3.5-4 Å but the loop RMSD within each minimum is restricted to ~4-7 Å when MLL is present, and expands to a wider range (~4-10 Å) when MLL is absent. Additional free energy profiles calculated with respect to the α_3 helix RMSD, which is believed to contribute to the allosteric binding of c-Myb, are reported in Figure S6 and show similar results to Figure 4. Overall, all of the free energy surfaces show a reduction in the conformational space being sampled when one or both peptides are bound to KIX (i.e. the colored area became smaller upon binding MLL and/or c-Myb and narrowed towards a ternary-like KIX).

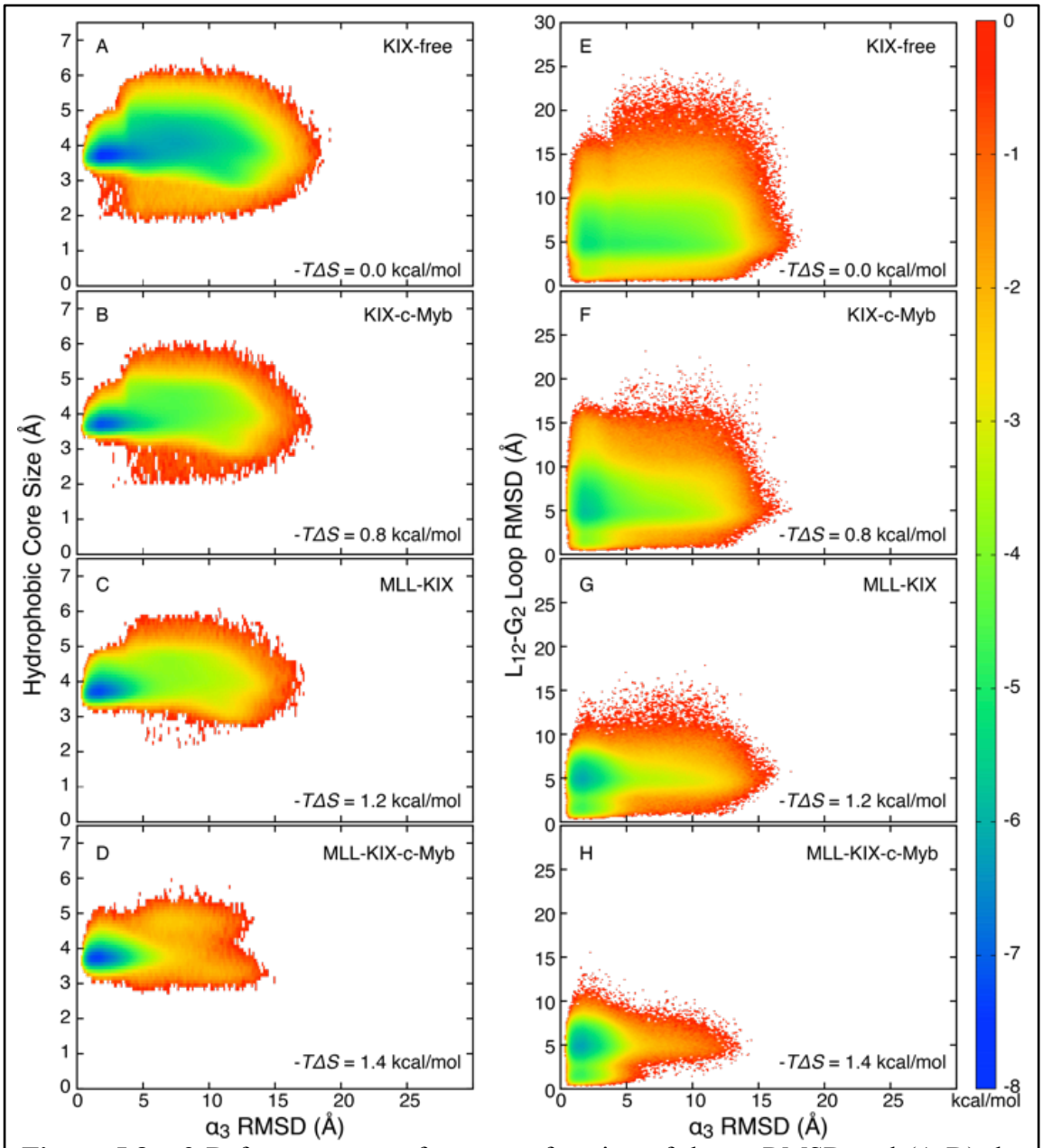


Figure 5.8: 2-D free energy surfaces as a function of the α_3 RMSD and (A-D) the hydrophobic core size or (E-H) the L_{12}-G_2 loop RMSD. All RMSDs were calculated relative to the NMR structure, which had a hydrophobic core size of ~3.6 Å. All ΔS were calculated with respect to KIX-free at 300 K.

5.3 Discussion

It has been hypothesized that protein intrinsic disorder offers numerous benefits such as fast protein turnover [48], high specificity for a diverse set of targets [48-52], and high specificity with low affinity binding [49,50]. In addition, it has been predicted that allosteric coupling between independent sites is maximized in the presence of intrinsic disorder [53]. In IDP systems, like those in the present study, there is an intricate relationship between binding and folding.[36-39,54-60] We observed that both c-Myb and MLL first bind to KIX before forming a folded peptide and the free energy surfaces between binary and ternary KIX complexes were virtually indistinguishable for both MLL and c-Myb (Figure 2 and Figure 3). Additionally, MLL and c-Myb, which were tuned to have a peak helicity of ~5% and ~25%, respectively, to match experiment (Figure 9), achieved over 80% helicity upon binding to KIX. This binding-induced-folding mechanism is consistent with kinetic experiments of KIX-c-Myb (i.e. rapid mixing of KIX with c-Myb) [55,61,62] and our results suggest that effector binding allosterically regulates peptide binding at the second site without altering the binding mechanism.

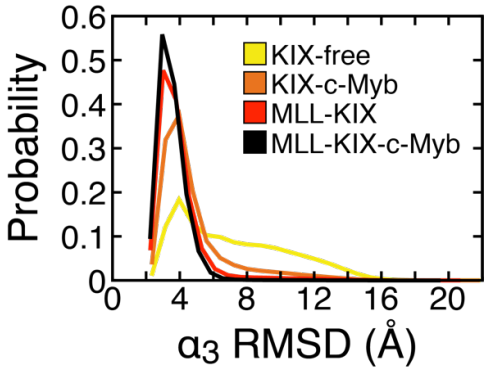


Figure 5.9: Helicity of the α_3 helix. The RMSD is calculated with respect to the fully helical α_3 in the NMR structure and RMSDs > ~5 Å correspond to a significant loss of secondary structure.

We explored this further by examining the intermolecular contact appearance order for KIX and c-Myb to see whether MLL binding alters the sequence in which contacts between c-Myb and KIX are formed with a low probability of breaking. Figure

4*A* shows that Leu302, which is positioned in the middle of the helix and inserts its side chain deep into the hydrophobic core in the native NMR structure [28], is first to form most of its interactions with the α_1 and α_3 helices of KIX. According to Figure 2, c-Myb is mostly unstructured at this stage, with only ~20-40% helicity. Then, binding proceeds outward towards both termini of c-Myb and is concomitant with c-Myb folding (Figure 4A). The N-terminus of c-Myb, which interacts exclusively with the C-terminus of the α_3 helix, has a slightly higher probability than the C-terminus for forming last, which may imply a potential role for α_3 in regulating c-Myb binding. These results were similar for c-Myb binding to the binary complex (Figure 4B) and demonstrate that MLL binding does not have a significant impact on the order in which c-Myb residues contact KIX.

Previous studies have observed changes in the KIX structure resulting from transcription factor binding and both MLL and c-Myb have been shown to allosterically affect the affinity of KIX for the complementary peptide.[4,19,28-32] However, a connection between the changes in structural dynamics in KIX and how this directly leads to shifts in transcription factor binding affinities remain tenuous. To bridge this gap, we employed a coarse-grained simulation model that is capable of capturing both the microscopic structural changes and macroscopic ligand binding events. Consistent with cooperative binding, we found that the binding affinity for c-Myb is increased by an order of magnitude in the presence of MLL while the affinity for MLL is also increased by a factor of 8 when c-Myb is present (Table 1). These observations are in agreement with experiment, albeit with a ~5-fold difference in both cases [12], but given the simplicity of our simulation model, it is remarkable that we are able to capture the same allosteric trends as seen in experiment. We also observed relative decreases in k_{off} (~4.4 fold and

~3.3 fold for c-Myb and MLL, respectively) and small increases in k_{on} (~2.4 fold and ~1.9 fold for c-Myb and MLL, respectively) when an effector was bound first (Table 1). These changes in kinetics combined with our observations of an effector-independent binding-induced-folding mechanism suggest that pre-binding of the first peptide likely stabilizes the flexible/unfolded parts of the KIX structure so that the second peptide can (i) form a native intermolecular contact more quickly (i.e. small increase in k_{on}) and (ii) stay bound for much longer (i.e. decrease k_{off}).

To further understand the allosteric mechanism, we investigated how MLL and/or c-Myb binding can affect the dynamics of KIX (Figure 3, Figure 4, and Figure 8). Overall, we found that binding of MLL and/or c-Myb resulted in an increase in stability (or a reduction in dynamics) of the α_3 C-terminus and the L_{12}-G_2 loop. This is consistent with observations made in previous NMR and computational studies, which have led to the suggestion that the increased structural stability of KIX may play a role in allosteric signaling.[4,19,28-32] For example, NMR backbone order parameter measurements of KIX have demonstrated that the L_{12}-G_2 loop and the C-terminal residues of the α_3 helix are more rigid when MLL is bound to KIX.[29] Thus, our findings are not only supported by experiment but they also provide a direct causative link between changes in the local structural dynamics and its effects on the global binding affinities. From the observations made in Figure 4 and Figure 8, we calculated the configurational entropy for each system (Table 1 and Eq. **Error! Reference source not found.**) by examining the distribution of configurations along all three collective variables (*i.e.* L_{12}-G_2 loop RMSD, α_3 RMSD, and hydrophobic core size). The energetic cost for binding either MLL or c-Myb to KIX-free was $-T\Delta\Delta S$ = 0.6 kcal/mol higher than when binding to its complementary binary

complex (Table 1, Figure 4, and Figure 8). Thus, our analyses show that effector-bound KIX lowers the entropic cost for binding the complementary peptide at the second site and this is achieved by stabilizing the KIX structure. The observed reduction in entropy due to effector binding explains the decrease in k_{off} and the small increase in k_{on} and this notion of "prepaying the entropic cost" in allostery has been found to play a role in the tryptophan RNA binding attenuation protein (TRAP) [63,64] and the catabolite activator protein (CAP).[5,65,66] Describing allostery within a thermodynamic framework [3-6] allows us to organize the information in a quantitative manner and it enables comparisons between entropically-driven and/or enthalpically-driven allosteric mechanisms.

Previous NMR studies have shown that a minor excited state of KIX, which is believed to be a higher affinity conformer for c-Myb and pKID, becomes 7% populated when bound by MLL and has been characterized by a rearrangement of the KIX hydrophobic core.[29,32] In a recent metadynamics-based MD simulation study [31] of the MLL-KIX binary structure [29], it was demonstrated that when the L_{12}-G_2 loop was in the "up" state then this resulted in opening of the KIX hydrophobic core which was said to allow KIX to spontaneously interconvert between a major lower energy state and a minor higher energy state. Essentially, the authors proffered that KIX mediates cooperativity between transcription factors through conformational selection of a higher affinity conformer. This means that if conformational selection for a 7% binding-competent state were at play we should expect to see an observable population increase of a minor state in our free energy surfaces. However, we find no evidence to support this. Instead, upon binding an effector, we observe a narrowing of the pre-existing KIX ensemble towards the highly populated minimum, which corresponds to a ternary-like KIX structure (Figure

4 and Figure 8). We also point out that while the enhanced sampling MD simulations of the MLL-KIX complex were informative, simulations of the binary complex alone are insufficient to show, unequivocally, that the observed changes in structural dynamics due to the presence of an effector would have any direct effect on binding affinities of a second transactivation domain. On the other hand, our simulations of the free KIX, MLL-KIX, KIX-c-Myb, and MLL-KIX-c-Myb systems provide a clear and systematic way to directly relate the microscopic allosteric effects and the macroscopic changes in the binding affinity that result from MLL and/or c-Myb binding to KIX.

In the accompanying paper, Shammas and coworkers have performed fluorescence stopped-flow measurements to investigate the binding kinetics of peptides from five different transcription factors that can all bind to KIX (c-Myb and pKID bind at the c-Myb site while MLL, HBZ, and E2A bind at the MLL site). Consistent with a coupled folding-and-binding mechanism, they found that the helicity for c-Myb (as well as all of the other peptides) increased upon mixing with KIX, which is consistent with our results. They also observed a positive allosteric effect and identified relative decreases in k_{off} of 3.9 ± 0.1 and 4.1 ± 0.1 for c-Myb and MLL, respectively, that are similar to those reported above. They also demonstrated that the reduction in the dissociation rate constant was generally true for four different pairs of KIX binding partners and, based upon their data, they proposed a common allosteric mechanism where the entropic cost of binding the second ligand is paid on binding the first ligand rather than involving the formation of a binding-competent low-populated conformer of KIX.[29,31,32] Our simulation results show good agreement with their kinetics experiments and confirm the hypothesis that MLL binding to KIX lowers the entropic cost for c-Myb binding, and vice versa, by

stabilizing the L_{12}-G_2 loop and α_3 helix. However, they found a relative decrease, rather than a small relative increase, in k_{on} of 2.9 ± 0.1 and 2.8 ± 0.1 for c-Myb and MLL, respectively, and, when the favorable long-range electrostatics was screened with salt, only negligible differences in k_{on} were observed. This can be attributed to differences in temperature or in how a bound state is detected in experiment versus computation. Overall, both bodies of work are in good complementary agreement and support an allosteric mechanism whereby the binding of an effector pre-pays the entropic cost for binding a complementary peptide in the secondary site.

If a change in the structural stability (or dynamics) is what drives the allosteric mechanism, one could explore alternative ways to modulate the stability of KIX in order to affect binding affinity. It was previously proposed that stabilizing the C-terminus of α_3 through MLL binding would result in favorable side chain electrostatic interactions between residues Glu665 and Glu666 of α_3 and Arg294 and Lys291 of c-Myb.[28] Not surprisingly, these electrostatically complementary residues are essentially conserved in all known c-Myb and CBP/p300 sequences.[28] Table 2 reports the top six point mutant candidates that were predicted from AGADIR [67-71] to increase the helical propensity of the α_3 helix and therefore can potentially affect cooperative binding. Out of all the C-terminal residues located in the α_3 helix which make no intermolecular native contacts with either c-Myb or MLL, only six point mutations at either Lys667 or Ser670 caused a significant increase in the predicted helicity of α_3 (ca. 3.5-12.7%) relative to the wild type sequence. Thus, we predict that, even in the absence of MLL, one or more of these strategically selected point mutations could mimic the effects of MLL binding to KIX and result in an increased c-Myb binding affinity. These predictions can be verified

experimentally and could inform the design of small molecules that could stabilize KIX and allow for further structural characterization. For instance, Mapp and coworkers recently solved the first crystal structure of KIX complexed to a small molecule referred to as fragment 1-10.[19] It was found that the L_{12}-G_2 loop had large temperature factors (or B factors), the C-terminal residues of the helix α_3 were unresolved, and the size of the hydrophobic core was reduced in comparison to the NMR ternary structure. This suggests that the entropy of KIX would not be reduced upon binding fragment 1-10 and, in light of the observations made in the current study, we would predict that this would not lead to an increase in affinity of KIX for c-Myb. This strategy of complementing experiment with computation can be applied generally to other conformationally dynamic proteins and potentially lead to a faster discovery of effective small molecule modulators.

In conclusion, we have carried out a set of simulations of KIX-free, MLL-KIX, KIX-c-Myb, and MLL-KIX-c-Myb in order to better understand how cooperative binding is regulated in this particular transcriptional activation system. Remarkably, our simplified model was able to capture relevant changes in thermodynamics, binding kinetics, and structural dynamics, all of which were in good agreement with experiment. Specifically, we discovered that the intermolecular native contact formation between c-Myb and KIX is unaffected by MLL binding, and vice versa, and that the folding-and-binding mechanism is well preserved. Furthermore, our results confirm that KIX occupation by a single transactivation domain was found to stabilize the L_{12}-G_2 loop and increase the helicity of the α_3 helix. This lowered the entropic cost for binding a complementary peptide at the allosteric site and was accompanied a narrowing of the pre-existing KIX ensemble towards a ternary-like KIX conformation. We also proposed that

a well-chosen mutation(s) or small molecule(s) targeted towards stabilizing the α_3 helix and/or the L_{12}-G_2 loop would be effective in controlling cooperative binding. As a whole, our study directly links the effects of the microscopic changes in structural dynamics to the macroscopic binding affinities and provides a thermodynamic description of the allosteric mechanism involved in KIX cooperative binding.

5.4 Methods

Simulation model and set up

An initial Cα-based Gō-like model constructed from the NMR coordinates of the MLL-KIX-c-Myb ternary complex (PDBID: 2AGH) [28] was generated using the Multiscale Modeling Tools for Structural Biology (MMTSB) Gō-Model Builder (http://www.mmtsb.org) [72]. MLL (19 residues ranging from 839-857), KIX (87 residues ranging 586-672), and c-Myb (25 residues ranging from 291-315) contained 22, 198, and 40 intramolecular native contacts, respectively, while the total number of intermolecular native contacts between MLL-KIX and KIX-c-Myb were 28 and 29, respectively. Structural models of the individual monomers and binary complexes were derived directly from the ternary complex model. The coarse-grained Gō-like model was then calibrated to match the helical content and binary dissociation constants reported in previous experiments (Figure 10). Each of the four simulation models (*i.e.* free KIX, MLL-KIX, KIX-c-Myb, and MLL-KIX-c-Myb) was simulated for 15 μs using CHARMM [73,74] and repeated 60 times in order to sample multiple c-Myb/MLL binding events in the presence and absence of MLL.

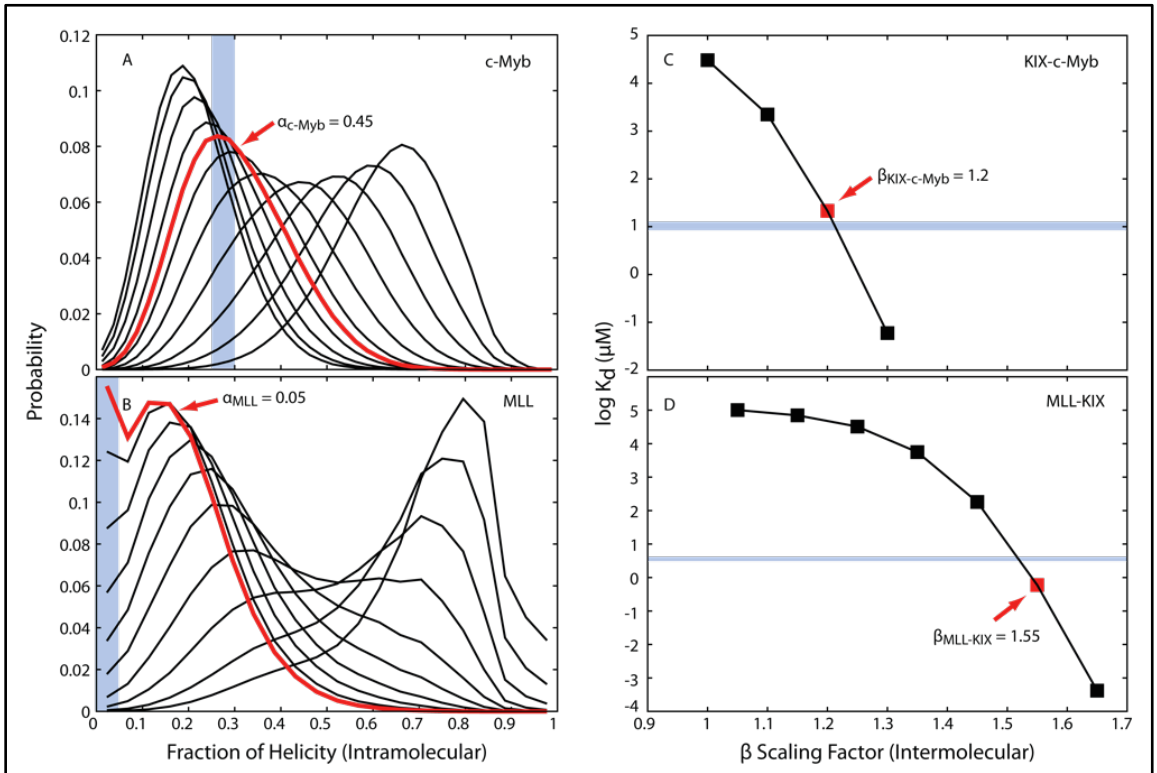


Figure 5.10: Calibration of (A-B) the helical content (α increases from left-to-right)and (C-D) the binding affinities (K_d). The final simulated scaling factors are indicated by red arrows and the target ranges are shown as blue bars (c-Myb: ~25-30% helicity; MLL: ~1-5% helicity; KIX-c-Myb: K_d = 10 ± 2 μM; MLL-KIX: K_d = 2.8 ± 0.4 μM).

Analysis

2-D free energy surfaces were computed at 300 K by constructing 2-D histograms along two reaction coordinates of interest. Q_{intra} and Q_{inter} corresponding to the fraction of native intra- and intermolecular contacts, respectively. The root-mean-square deviation (RMSD) of the L_{12}-G_2 loop (residues 614-621) was calculated after fitting parts of helices α_1, α_2, and α_3 (residues 600-613, 622-662) of KIX from each simulation snapshot to the native NMR structure [31]. Similarly, the RMSD of helix α_3 (residues 646-669) was calculated by pre-fitting the rigid N-terminal portion of helix α_3 (residues 646-657) to the native NMR structure. The size of the hydrophobic core (residues 607-612, 650-661) [29],

whose shape can be approximated by an ellipsoid [31], was calculated from the square root of the second eigenvalue of the gyration tensor [75].

We estimated the configurational entropy, S, using:

$$S = -k_B \sum_{L,A,H} p(L,A,H) \ln p(L,A,H) \quad (1)$$

where k_B is the Boltzmann constant and $p(L,A,H)$ is the probability of occupying a particular state along the three collective variables L, A, and H (corresponding to the L_{12}-G_2 loop RMSD, helix α_3 RMSD, and KIX hydrophobic core size, respectively) in a given simulation. All ΔS were calculated with respect to free KIX at 300 K.

Principal component analysis (PCA) was carried out to study the dynamics of KIX for each system. Each simulation snapshot was superimposed onto the three helices of KIX (residues 600-613, 622-662) in the native NMR structure and the eigenvalues and eigenvectors were obtained by diagonalizing the covariance matrix constructed from the fluctuations of residues 600-669.

Simulation model and details

To study the allosteric binding involved in the MLL-KIX-c-Myb ternary complex, we used the sequence-flavored Gō-like model developed by Karanicolas and Brooks.[34,35] In this model, each protein residue is represented as a single bead with a mass equal to its corresponding amino acid, centered at its Cα atom, and connected to neighboring residues along the protein backbone via virtual bonds. The potential energy function, V, consists of both bonded and non-bonded terms:

$$V = V_{bond} + V_{angle} + V_{dihedral} + V_{non-bonded}^{non-native} + V_{non-bonded}^{native} \quad (2)$$

where V_{bond} and V_{angle} are harmonic potentials with equilibrium values set to those found in the NMR structure.[28] $V_{dihedral}$ is a statistical potential based on probability distributions obtained from Ramachandran plots for each of the 20 x 20 amino acid pairs and, thus, provides additional sequence-specific information while remaining independent of the protein topology and avoiding locally driven folding that results from directly incorporating native dihedral angles into $V_{dihedral}$. Native non-bonded terms mediating the interaction between residue pairs, *i* and *j*, are modeled using a 12-10-6 Lennard-Jones-type potential whose interaction strength is proportional to the statistical potential reported for the residue types of *i* and *j* by Miyazawa and Jernigan.[76] All non-native interactions, defined by residue pairs with all side chain heavy-atoms separated by more than 4.5 Å and that do not form backbone hydrogen bonds, were subject to repulsive interactions. We can also separate $V_{non-bonded}^{native}$ into its intra- and intermolecular components:

$$V_{non-bonded}^{native} = \alpha V_{intramolecuar}^{native} + \beta V_{intermolecular}^{native} \quad (3)$$

where α and β are scaling factors used for renormalizing the intra- and intermolecular interaction energies, respectively (see below).

The cumulative production simulation time for each simulation model is 900 μs (*i.e.*, each set consists of 60 independent x 15 μs-long simulations) and with at least 70 c-Myb binding events and 150 MLL binding events being observed in total. With the exception of KIX-free, which was simulated in a 60 Å per edge cubic volume, each system was simulated with a 120 Å per edge cubic volume using a periodic volume to account for the finite concentration of species in the simulation, e.g., MLL + KIX, etc.

All simulations were performed at 300 K using Langevin dynamics with a 0.1 ps^{-1} friction coefficient along with a 15 fs simulation time step. The cutoff for non-bonded interactions was set to 25 Å and virtual bonds were constrained using SHAKE.[77]

Calibrating the Gō-like model

In accordance with the calibration protocol described in ref. [37], the intramolecular interaction energies of c-Myb and MLL were rescaled by systematically altering α in Eq. (3) above so as to match the experimentally reported helical content. The intermolecular interaction energies between KIX-c-Myb and MLL-KIX were also recalibrated by tuning β so as to match the experimentally reported binary dissociation constants. The intramolecular interaction energies for KIX were unaltered (*i.e.* α_{KIX} = 1.0). All simulations of the MLL-KIX-c-Myb ternary complex use only the tuned α and β parameters from the monomer and binary systems, respectively. No further adjustments are made. 1.5 μs-long simulations were initially used to calibrate α. The percent helicity, measured as the fraction of the native intramolecular contacts formed, was found to be identical even with ten times more sampling. Simulations for calibrating β (eventually used in the production simulations) were 15 μs in length and repeated 30 times in order to provide an accurate estimate of the experimentally reported dissociation constants.

When the native Gō-like model was used (α = 1.0), the fraction of helical content was ~70% and ~80% for c-Myb and MLL, respectively.[78-80] Therefore, the intramolecular interaction energies were reduced by rescaling α in order to match the helical content reported in the literature (~25-30% for c-Myb) [62,78,80] and as predicted empirically by AGADIR (~1-5% for MLL).[67-70] In order to achieve this, the final

intramolecular interaction energies were scaled down to $\alpha_{c\text{-}Myb} = 0.45$ and $\alpha_{MLL} = 0.05$ for c-Myb and MLL, respectively. In contrast, the intermolecular interaction energies for the KIX-c-Myb and MLL-KIX binary complexes were increased to $\beta_{KIX\text{-}c\text{-}Myb} = 1.2$ and $\beta_{MLL\text{-}KIX} = 1.55$, respectively, to match the reported binary dissociation constants.[12]

While all of the intramolecular and intermolecular native contacts were derived from the NMR ternary complex, it is important to reiterate that our recalibration process only targets experimental observables from the monomeric and binary systems. In other words, no additional tuning was made to influence cooperative binding ($\alpha_{c\text{-}Myb} = 0.45$ and $\alpha_{MLL} = 0.05$ were used when scaling β but none of the scaling parameters were tuned to match the ternary dissociation constants). Thus, one would not expect changes in the binding affinities for the ternary complex unless allostery was at play and could be adequately captured by our coarse-grained model. Additionally, unlike traditional all-atom simulations where only a single binding/unbinding event is often observed, we recorded multiple binding/unbinding events from long, continuous trajectories and so any biases caused by the starting structure are minimized as a result of increased statistics.

Fraction of bound and unbound states

If the Cα-Cα distance between any given native interaction (or, equivalently, native contact) was within 1 Å of its native contact distance (found in the NMR structure), then the native contact was considered formed. We defined a bound state as having at least one native intermolecular contact being formed or a configuration that had last visited a bound state (without first visiting an unbound state) and with at least one native intermolecular contact within 25 Å (corresponding to the cutoff distance for non-bonded

interactions). This helps to remove any bias in rebinding coming from long-range interactions and ensures that binding events are uncorrelated. Conversely, an unbound state was defined as having all native intermolecular contact distances greater than 25 Å or a configuration that had last visited an unbound state and had no native intermolecular contacts. Together, these definitions provide a clear delineation between the bound and unbound states.

The dissociation constants, K_d, were calculated from:

$$K_d = \frac{1660}{V_0} \cdot \frac{P_{unbound}^2}{1 - P_{unbound}} \tag{4}$$

where V_0 is the box volume in units of Å^3, 1660 converts the concentration from units of molecules/Å^3 to units of mol/L, and $P_{unbound}$ is the fraction of unbound states.

Contact appearance order

To assess whether or not the binding of one peptide changes the order in which the native intermolecular contacts are formed between a second peptide and KIX, we adopted the contact appearance order metric originally developed to study protein folding [81]. Here, we record the order in which each native intermolecular contact is formed during N independent c-Myb or MLL binding events. Then, we calculate the probability of each native intermolecular contact being formed at a given order in time.

Kinetic rates

By treating the ligand binding process as a simple two-state process with either bound or unbound states, we estimated the kinetic rate constants, k_{off} and k_{on}, by calculating the mean first passage time (MFPT) for the off and on reactions:

$$k_{off} = \frac{1}{MFPT_{off}}; \qquad k_{on} = \frac{1}{MFPT_{on} \cdot \left(\frac{1660}{V_0} \cdot P_{unbound} \right)} \tag{5}$$

where k_{off} has units of s^{-1} and k_{on} is dependent on the concentration (or fraction of unbound states) and so has units of $M^{-1}s^{-1}$.

Principal component analysis

The overall translational and rotational motion was removed by fitting each simulation snapshot to the average KIX structure. Then, a symmetric covariance matrix, *C*, of the positional deviations was constructed:

$$C = \left\langle \left(r(t) - \langle r \rangle \right) \left(r(t) - \langle r \rangle \right)^T \right\rangle \tag{6}$$

where < > denotes an ensemble average and *r(t)* is a 3*N*-dimensional vector of *x*, *y*, and *z* coordinates for all *N* atoms at some simulation time, t . *C* can then be diagonalized by an orthogonal coordinate transformation to obtain the mean square fluctuations (eigenvalues) along each principal component/mode (eigenvectors) of the system. When the eigenvectors are sorted by their corresponding eigenvalues in decreasing order, the total fluctuations of the system can often be described by the first few lowest frequency modes.

In silico mutations

Six C-terminal residues (Glu663, Glu666, Lys667, Ser670, Arg671, and Leu672) of the α_3 helix, which make no intermolecular native contacts with either c-Myb or MLL, were chosen for mutational studies to increase the helicity in this region. The KIX sequence was mutated (to either Ala, Leu, Arg, Met, Lys, Asn, Glu, Ile, Trp, or Ser) at each of the six target positions and the increase in the percent helicity was predicted using AGADIR [67-70] (see Table 2).

Table 5.2:

Point mutation	Percent helicity increase*
K667L	4.4 (1.5)
K667R	9.2 (3.1)
S670L	3.5 (1.2)
S670R	9.5 (3.2)
S670K	7.5 (2.5)
S670N	12.7 (4.2)

*Calculated from α_3 (residues 646–672) and, in parentheses, from the full length of KIX (residues 586–672).

5.5 References

(1) Monod, J.; Jacob, F. In *Cold Spring Harbor Symp. Quant. Biol.*; Cold Spring Harbor Laboratory Press: 1961; Vol. 26, p 389.
(2) Monod, J.; Wyman, J.; Changeux, J.-P. *J. Mol. Biol.* **1965**, *12*, 88.
(3) Tsai, C.-J.; Del Sol, A.; Nussinov, R. *Molecular Biosystems* **2009**, *5*, 207.
(4) del Sol, A.; Tsai, C. J.; Ma, B. Y.; Nussinov, R. *Structure* **2009**, *17*, 1042.
(5) Tsai, C.-J.; del Sol, A.; Nussinov, R. *Journal of Molecular Biology* **2008**, *378*, 1.
(6) Tsai, C.-J.; Nussinov, R. *PLoS Comput. Biol.* **2014**, *10*, e1003394.
(7) Iyer, N. G.; Özdag, H.; Caldas, C. *Oncogene* **2004**, *23*, 4225.
(8) Kimbrel, E. A.; Lemieux, M. E.; Xia, X.; Davis, T. N.; Rebel, V. I.; Kung, A. L. *Blood* **2009**, *114*, 4804.
(9) Kasper, L. H.; Fukuyama, T.; Biesen, M. A.; Boussouar, F.; Tong, C.; De Pauw, A.; Murray, P. J.; Van Deursen, J. M.; Brindle, P. K. *Mol. Cell. Biol.* **2006**, *26*, 789.
(10) Iyer, N.; Xian, J.; Chin, S.; Bannister, A.; Daigo, Y.; Aparicio, S.; Kouzarides, T.; Caldas, C. *Oncogene* **2006**, *26*, 21.
(11) Radhakrishnan, I.; Pérez-Alvarado, G. C.; Parker, D.; Dyson, H. J.; Montminy, M. R.; Wright, P. E. *J. Mol. Biol.* **1999**, *287*, 859.
(12) Goto, N. K.; Zor, T.; Martinez-Yamout, M.; Dyson, H. J.; Wright, P. E. *Journal of Biological Chemistry* **2002**, *277*, 43168.
(13) Campbell, K. M.; Lumb, K. J. *Biochemistry* **2002**, *41*, 13956.
(14) Vendel, A. C.; McBryant, S. J.; Lumb, K. J. *Biochemistry* **2003**, *42*, 12481.
(15) Zor, T.; De Guzman, R. N.; Dyson, H. J.; Wright, P. E. *Journal of Molecular Biology* **2004**, *337*, 521.
(16) Vendel, A. C.; Lumb, K. J. *Biochemistry* **2004**, *43*, 904.
(17) Lee, L. W.; Mapp, A. K. *J. Biol. Chem.* **2010**, *285*, 11033.
(18) Majmudar, C. Y.; Højfeldt, J. W.; Arevang, C. J.; Pomerantz, W. C.; Gagnon, J. K.; Schultz, P. J.; Cesa, L. C.; Doss, C. H.; Rowe, S. P.; Vásquez, V.; Tamayo-Castillo, G.; Cierpicki, T.; Brooks, C. L. I.; Sherman, D. A.; Mapp, A. K. *Angew. Chem., Int. Ed.* **2012**, *51*, 11258.
(19) Wang, N. K.; Majmudar, C. Y.; Pomerantz, W. C.; Gagnon, J. K.; Sadowsky, J. D.; Meagher, J. L.; Johnson, T. K.; Stuckey, J. A.; Brooks, C. L.; Wells, J. A.; Mapp, A. K. *J. Am. Chem. Soc.* **2013**, *135*, 3363.
(20) Bates, C. A.; Pomerantz, W. C.; Mapp, A. K. *Biopolymers* **2011**, *95*, 17.
(21) Pomerantz, W. C.; Wang, N. K.; Lipinski, A. K.; Wang, R. R.; Cierpicki, T.; Mapp, A. K. *ACS Chem. Biol.* **2012**, *7*, 1345.
(22) Lodge, J. M.; Rettenmaier, T. J.; Wells, J. A.; Pomerantz, W. C.; Mapp, A. K. *MedChemComm* **2014**, *10.1039/c3md00356f*.
(23) Thakur, J. K.; Yadav, A.; Yadav, G. *Nucleic Acids Research* **2014**, *42*, 2112.
(24) Jin, S.; Zhao, H.; Yi, Y.; Nakata, Y.; Kalota, A.; Gewirtz, A. M. *The Journal of clinical investigation* **2010**, *120*, 593.
(25) Hess, J. L.; Bittner, C. B.; Zeisig, D. T.; Bach, C.; Fuchs, U.; Borkhardt, A.; Frampton, J.; Slany, R. K. *Blood* **2006**, *108*, 297.
(26) Graf, T. *Curr. Opin. Genet. Dev.* **1992**, *2*, 249.
(27) Kasper, L. H.; Fukuyama, T.; Lerach, S.; Chang, Y. C.; Xu, W.; Wu, S.; Boyd, K. L.; Brindle, P. K. *PLoS One* **2013**, *8*.

(28) De Guzman, R. N.; Goto, N. K.; Dyson, H. J.; Wright, P. E. *Journal of Molecular Biology* **2006**, *355*, 1005.
(29) Brüschweiler, S.; Konrat, R.; Tollinger, M. *ACS Chem. Biol.* **2013**, *8*, 1600.
(30) Korkmaz, E. N.; Nussinov, R.; Haliloglu, Y. *PLoS Comput. Biol.* **2012**, *8*.
(31) Palazzesi, F.; Barducci, A.; Tollinger, M.; Parrinello, M. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 14237.
(32) Brüschweiler, S.; Schanda, P.; Kloiber, K.; Brutscher, B.; Kontaxis, G.; Konrat, R.; Tollinger, M. *J. Am. Chem. Soc.* **2009**, *131*, 3063.
(33) Umezawa, K.; Ikebe, J.; Takano, M.; Nakamura, H.; Higo, J. *Biomolecules* **2012**, *2*, 104.
(34) Karanicolas, J.; Brooks, C. L., III *J. Mol. Biol.* **2003**, *334*, 309.
(35) Karanicolas, J.; Brooks, C. L., III *Protein Sci.* **2002**, *11*, 2351.
(36) Turjanski, A. G.; Best, R.; Hummer, G.; Gutkind, J. S. *FASEB J.* **2007**, *21*, A270.
(37) Ganguly, D.; Chen, J. H. *Proteins: Struct., Funct., Bioinf.* **2011**, *79*, 1251.
(38) De Sancho, D.; Best, R. B. *Mol. BioSyst.* **2012**, *8*, 256.
(39) Ahlstrom, L. S.; Dickson, A.; Brooks, C. L., III *J. Phys. Chem. B* **2013**, *10.1021/jp403264s*.
(40) Clementi, C.; Jennings, P. A.; Onuchic, J. N. *Proceedings of the National Academy of Sciences of the United States of America* **2000**, *97*, 5871.
(41) Levy, Y.; Cho, S. S.; Onuchic, J. N.; Wolynes, P. G. *Journal of Molecular Biology* **2005**, *346*, 1121.
(42) Levy, Y.; Onuchic, J. N. *Accounts of Chemical Research* **2006**, *39*, 135.
(43) Levy, Y.; Wolynes, P. G.; Onuchic, J. N. *Proceedings of the National Academy of Sciences of the United States of America* **2004**, *101*, 511.
(44) Shea, J. E.; Onuchic, J. N.; Brooks, C. L. *Proceedings of the National Academy of Sciences of the United States of America* **1999**, *96*, 12512.
(45) Jana, B.; Morcos, F.; Onuchic, J. N. *Phys. Chem. Chem. Phys.* **2014**, *16*, 6496.
(46) Whitford, P. C.; Gosavi, S.; Onuchic, J. N. *J. Biol. Chem.* **2008**, *283*, 2042.
(47) Hyeon, C.; Jennings, P. A.; Adams, J. A.; Onuchic, J. N. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 3023.
(48) Wright, P. E.; Dyson, H. J. *Journal of Molecular Biology* **1999**, *293*, 321.
(49) Uversky, V. N.; Oldfield, C. J.; Dunker, A. K. *Journal of Molecular Recognition* **2005**, *18*, 343.
(50) Liu, J. G.; Perumal, N. B.; Oldfield, C. J.; Su, E. W.; Uversky, V. N.; Dunker, A. K. *Biochemistry* **2006**, *45*, 6873.
(51) Uversky, V. N. *Protein Science* **2002**, *11*, 739.
(52) Dunker, A. K.; Lawson, J. D.; Brown, C. J.; Williams, R. M.; Romero, P.; Oh, J. S.; Oldfield, C. J.; Campen, A. M.; Ratliff, C. R.; Hipps, K. W.; Ausio, J.; Nissen, M. S.; Reeves, R.; Kang, C. H.; Kissinger, C. R.; Bailey, R. W.; Griswold, M. D.; Chiu, M.; Garner, E. C.; Obradovic, Z. *Journal of Molecular Graphics & Modelling* **2001**, *19*, 26.
(53) Hilser, V. J.; Thompson, E. B. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 8311.
(54) Sugase, K.; Dyson, H. J.; Wright, P. E. *Nature* **2007**, *447*, 1021.
(55) Gianni, S.; Morrone, A.; Giri, R.; Brunori, M. *Biochem. Biophys. Res. Commun.* **2012**, *428*, 205.
(56) Ganguly, D.; Chen, J. H. *Biophys. J.* **2010**, *98*, 257A.
(57) Dyson, H. J.; Wright, P. E. *Curr. Opin. Struct. Biol.* **2002**, *12*, 54.

(58) Wright, P. E.; Dyson, H. J. *Curr. Opin. Struct. Biol.* **2009**, *19*, 31.
(59) Ganguly, D.; Zhang, W.; Chen, J. *PLoS Comput. Biol.* **2013**, *9*, e1003363.
(60) Rogers, J. M.; Steward, A.; Clarke, J. *J. Am. Chem. Soc.* **2013**, *135*, 1415.
(61) Giri, R.; Morrone, A.; Toto, A.; Brunori, M.; Gianni, S. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 14942.
(62) Shammas, S. L.; Travis, A. J.; Clarke, J. *J. Phys. Chem. B* **2013**, *10.1021/jp404267e*.
(63) McElroy, C.; Manfredo, A.; Wendt, A.; Gollnick, P.; Foster, M. *Journal of Molecular Biology* **2002**, *323*, 463.
(64) Heddle, J. G.; Okajima, T.; Scott, D. J.; Akashi, S.; Park, S. Y.; Tame, J. R. H. *J. Mol. Biol.* **2007**, *371*, 154.
(65) Boyer, J. A.; Lee, A. L. *Biochemistry* **2008**, *47*, 4876.
(66) Popovych, N.; Sun, S. J.; Ebright, R. H.; Kalodimos, C. G. *Nat. Struct. Mol. Biol.* **2006**, *13*, 831.
(67) Muñoz, V.; Serrano, L. *J. Mol. Biol.* **1995**, *245*, 275.
(68) Muñoz, V.; Serrano, L. *Nat. Struct. Mol. Biol.* **1994**, *1*, 399.
(69) Muñoz, V.; Serrano, L. *Journal of Molecular Biology* **1995**, *245*, 297.
(70) Muñoz, V.; Serrano, L. *Biopolymers* **1997**, *41*, 495.
(71) Lacroix, E.; Viguera, A. R.; Serrano, L. *J. Mol. Biol.* **1998**, *284*, 173.
(72) Feig, M.; Karanicolas, J.; Brooks, C. L., III *J. Mol. Graphics Modell.* **2004**, *22*, 377.
(73) Brooks, B. R.; Brooks, C. L., III; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. *J. Comput. Chem.* **2009**, *30*, 1545.
(74) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *Journal of Computational Chemistry* **1983**, *4*, 187.
(75) Vymetal, J.; Vondrasek, J. *J. Phys. Chem. A* **2011**, *115*, 11455.
(76) Miyazawa, S.; Jernigan, R. L. *J. Mol. Biol.* **1996**, *256*, 623.
(77) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327.
(78) Parker, D.; Rivera, M.; Zor, T.; Henrion-Caude, A.; Radhakrishnan, I.; Kumar, A.; Shapiro, L. H.; Wright, P. E.; Montminy, M.; Brindle, P. K. *Mol. Cell. Biol.* **1999**, *19*, 5601.
(79) Shammas, S. L.; Travis, A. J.; Clarke, J. *Journal of Physical Chemistry B* **2013**, *117*, 13346.
(80) Zor, T.; Mayr, B. M.; Dyson, H. J.; Montminy, M. R.; Wright, P. E. *Journal of Biological Chemistry* **2002**, *277*, 42241.
(81) Gin, B. C.; Garrahan, J. P.; Geissler, P. L. *J. Mol. Biol.* **2009**, *392*, 1303.

Chapter 6

Conclusions and future outlook

6.1 Summary of chapters

Previously, structure-based docking approximated the receptor as a rigid entity.[1,2] This is a reasonable approximation for receptors that follow a lock and key mechanism of binding their ligands. However, as more biophysical and structural data has become available the importance of flexibility has become extremely apparent.[2-5] This importance of flexibility was really highlighted in Chapter 4 and 5, when considering receptors that are highly dynamic and promiscuous like the GACKIX domain of CBP.[6-8]

Targeting protein-protein interactions with small molecules is important from both a biochemical and drug design point of view. Being able to selectively inhibit a protein-protein interaction would be invaluable for biochemical assays towards understanding of intricate cellular regulatory processes, such as transcription regulation as described in Chapters 4 and 5. Also, from a drug discovery standpoint, having more tools is important for targeting these challenging receptors involved in protein-protein interactions, as they are integral in so many cellular processes.

Towards providing such tools we developed a flexible receptor docking methodology within the CHARMM simulation package. The docking tool was

implemented as an MMTSB style script to interface user input with CHARMM as outlined in Chapter 2.[9,10] This method uses the grid based docking methodology as previously outlined by Wu et al, and achieved receptor flexibility by incorporating explicit side-chains from the receptor.[11-14] While the backbone atoms are held fixed the rest of the residue samples conformational space simultaneously as the small molecule ligand. Addionally, we implemented a new sampling method within CDOCKER that includes advanced sampling techinque of self-guided Langevin dyanmics (SGLD).[15,16] This should account for local rearrangement of side-chain rotamers as frequently seen across receptor structures when bound to different ligands.[17-20] However, this method of incorporating receptor flexibility cannot account for large-scale structural changes that may occur upon ligand binding, and these would have to address in some other manner.[3,21,22]

The new flexible receptor CDOCKER outperforms its rigid counterpart in both redocking and cross docking trials. This was exciting, as by adding receptor degrees of freedom could have reduced the docking accuracy in redocking trials, since the receptor is in the experimental conformation that it would be bound to the ligand. Furthermore, Flexible CDOCKER is competitive with other commonly used docking methodology, Autodock and Glide, in redocking trials.[23,24]

As an application for the CDOCKER methodology I worked on several collaboration projects. One project I worked on with the Raghavan research group at the University of Michigan where we identified a putative ATP binding site on calreticulin (CRT). CRT is a chaperone protein that is involved in MHC Class I assembly, which is responsible for displaying peptide fragments on the cell surface as part of the immune

response.[25-29] We used a consensus docking approach to predict a binding site for ATP on the surface of CRT (Chapter 3).[30,31] These predictions were confirmed by experiment and showed for the first time that the chaperone CRT both binds and catalyzes ATP.

As outlined in Chapter 4, we further developed Flexible CDOCKER to include a tethering technique that is used in experiment to screen a large library of small-molecule fragments.[8,32,33] We modified Flexible CDOCKER to include a physical disulfide bond in an automated fashion between the small-molecule fragments and cysteine on a receptor. Using this Flexible CDOCKER methodology we were able to identify a docked conformation of a small molecule fragment (Frag 1-10) that was very similar to the crystallographic conformer in a cross docking trial that included a large rotameric shift of a tyrosine residue. Encouraged by these results we docked a larger fragment (Frag 2-64) that was identified by experiment but were unable to achieve crystals of high enough quality to use for structure determination. In the docking trials, Frag 2-64 proved to be too large to interact with GACKIX in the same manner as Frag 1-10 had in the experimental structure. Frag 2-64 was predicted to interact in a way that could disrupt key crystal-contacts, suggesting a hypothesis for the lack of high quality crystals of this complex.

In Chapter 5, I moved away from docking, but not from protein-protein interaction systems.[6] Using Gō-like models, we were able to capture an allosteric signal of a ternary complex using simulations parameterized using binary systems.[34,35] We were able to recapitulate both K_D and kinetic rate constants for two peptides, MLL and c-Myb, binding to GACKIX. We were able to show the binding and folding mechanisms of the two peptides that were expected based on previous experimental results, and these

mechanisms were unchanged with the allosteric signal.[36] Most interestingly, we showed that the allostery arises from a slowing of the k_{off}, which was consistent with work done by the Clarke research group published in parallel with our own.[37] This slowing was due to a shrinking of the conformational ensemble after the first peptide binds. This suggests an allosteric signal that is being driven by entropy, the first peptide essentially paying the entropic cost for the second.

6.2 Future outlook

This series of successful studies are encouraging front runners for future studies, which I believe have excellent prospects of being a multi-scale approach. For example, Frag 1-10 binding to GACKIX shows interesting behavior that varies depending where on the protein it is tethered. This small-molecule inhibits a peptide binding at the pKID site when it is tethered to residue 664 (as seen in the crystallographic structure 4i9o.)[8] However, when Frag 1-10 is tethered to residue 627 there is a positive allosteric signal, leading to enhanced affinity for the peptide interacting at the pKID site.[38] To investigate the origin of this allostery a multi-scale computational approach could be undertaken.

Firstly, docking studies would have to be undertaken, as a structure for Frag 1-10 bound to residue 627 has not been solved experimentally. The resulting docking trials and the Frag 1-10//GACKIX (L664C) experimental structure could be subjected to some sort of advanced sampling technique. One such approach could be WExplore, which has been developed by the Brooks Research group.[39,40] This approach allows for rapid sampling of conformational space through tecillating the space as new conformations are sampled, pushing the trajectory away from the starting structure.

This sampling would yield an ensemble of conformations that can give a variety of needed information for subsequent steps. This sampling could confirm that the ligand conformation identified by the docking trials is indeed the native-like low energy conformation. Additionally, these MD studies would identify key residues that interact with the small-molecules and their preferred distances to the small molecule. These residues and distances can be used to develop a Gō-like model potential for the Frag 1-10/GACKIX complex. Simulations similar to those outlined in Chapter 5 could be

performed to investigate the origin of the allosteric signal. Does it stabilize the flexible loop when it binds at residue 627, as the MLL peptide did in the previous studies? Perhaps the advanced sampling will show a structural, or enthalpically driven origin for an allosteric signal. These interesting investigations would lay the groundwork for other multi-scale approaches for similar purposes and improve our understanding of different types of allostery.

With the improvements of computer hardware and computational software, it is an exciting time to be a computational chemist. The ability to work hand-in-hand with experimental groups is rewarding and mutually beneficial. For example, in Chapter 3, the computational predictions were confirmed by experiment, and the resulting experiments fed back into the computational model. Together we were able to make progress in understanding the ATP dependence of CRT, a chaperone involved in the immune response. The advancements in the CHARMM docking methodology is a necessary step towards properly modeling receptors that do not follow the lock and key mechanism of binding, like GACKIX domain of CBP in Chapter 4 and 5. This dynamic receptor is just one example in the mountains of biophysical and structural data that demonstrate just how fluid bio-macromolecules can be. I am positive that the field will continue to develop so that it may continue to make important contributions, both towards understanding fundamental biological mechanisms and in structure-based drug design.

6.3 References

(1) Jorgensen, W. L. *Science* **1991**, *254*, 954.
(2) Carlson, H. A. *Current Opinion in Chemical Biology* **2002**, *6*, 447.
(3) Fischer, M.; Coleman, R. G.; Fraser, J. S.; Shoichet, B. K. *Nature Chemistry* **2014**, *6*, 575.
(4) Lexa, K. W.; Carlson, H. A. *Quarterly Reviews of Biophysics* **2012**, *45*, 301.
(5) Teague, S. J. *Nature Reviews Drug Discovery* **2003**, *2*, 527.
(6) Law, S. M.; Gagnon, J. K.; Mapp, A. K.; Brooks, C. L., III *Proceedings of the National Academy of Sciences of the United States of America* **2014**, *111*, 12067.
(7) Majmudar, C. Y.; Hojfeldt, J. W.; Arevang, C. J.; Pomerantz, W. C.; Gagnon, J. K.; Schultz, P. J.; Cesa, L. C.; Doss, C. H.; Rowe, S. P.; Vasquez, V.; Tamayo-Castillo, G.; Cierpicki, T.; Brooks, C. L., III; Sherman, D. H.; Mapp, A. K. *Angewandte Chemie-International Edition* **2012**, *51*, 11258.
(8) Wang, N.; Majmudar, C. Y.; Pomerantz, W. C.; Gagnon, J. K.; Sadowsky, J. D.; Meagher, J. L.; Johnson, T. K.; Stuckey, J. A.; Brooks, C. L., III; Wells, J. A.; Mapp, A. K. *Journal of the American Chemical Society* **2013**, *135*, 3363.
(9) Brooks, B. R.; Brooks, C. L., III; Mackerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. *J. Comput. Chem.* **2009**, *30*, 1545.
(10) Feig, M.; Karanicolas, J.; Brooks, C. L. *J. Mol. Graphics Modell.* **2004**, *22*, 377.
(11) Armen, R. S.; Chen, J.; Brooks, C. L., III *J. Chem. Theory Comput.* **2009**, *5*, 2909.
(12) Vieth, M.; Hirst, J. D.; Dominy, B. N.; Daigler, H.; Brooks, C. L. *J. Comput. Chem.* **1998**, *19*, 1623.
(13) Vieth, M.; Hirst, J. D.; Kolinski, A.; Brooks, C. L. *J. Comput. Chem.* **1998**, *19*, 1612.
(14) Wu, G. S.; Robertson, D. H.; Brooks, C. L.; Vieth, M. *J. Comput. Chem.* **2003**, *24*, 1549.
(15) Wu, X.; Brooks, B. R. *J. Chem. Phys.* **2011**, *134*.
(16) Wu, X. W.; Brooks, B. R. *Chem. Phys. Lett.* **2003**, *381*, 512.
(17) Cozzini, P.; Kellogg, G. E.; Spyrakis, F.; Abraham, D. J.; Costantino, G.; Emerson, A.; Fanelli, F.; Gohlke, H.; Kuhn, L. A.; Morris, G. M.; Orozco, M.; Pertinhez, T. A.; Rizzi, M.; Sotriffer, C. A. *J. Med. Chem.* **2008**, *51*, 6237.
(18) Kuhn, L. A. *Strength in flexibility: Modeling side-chain conformational change in docking and screening*, 2008.
(19) Zavodszky, M. I.; Kuhn, L. A. *Protein Sci.* **2005**, *14*, 1104.
(20) Zavodszky, M. I.; Ming, L.; Thorpe, M. F.; Day, A. R.; Kuhn, L. A. *Proteins-Structure Function and Bioinformatics* **2004**, *57*, 243.
(21) Bowman, A. L.; Nikolovska-Coleska, Z.; Zhong, H.; Wang, S.; Carlson, H. A. *J. Am. Chem. Soc.* **2007**, *129*, 12809.
(22) Lorber, D. M.; Shoichet, B. K. *Protein Sci.* **1998**, *7*, 938.
(23) Trott, O.; Olson, A. J. *J. Comput. Chem.* **2010**, *31*, 455.

(24) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. *J. Med. Chem.* **2004**, *47*, 1739.
(25) Del Cid, N.; Jeffery, E.; Rizvi, S. M.; Stamper, E.; Peters, L. R.; Brown, W. C.; Provoda, C.; Raghavan, M. *J Biol Chem* **2010**, *285*, 4520.
(26) Ellgaard, L.; Frickel, E. M. *Cell Biochem Biophys* **2003**, *39*, 223.
(27) Gao, B.; Adhikari, R.; Howarth, M.; Nakamura, K.; Gold, M. C.; Hill, A. B.; Knee, R.; Michalak, M.; Elliott, T. *Immunity* **2002**, *16*, 99.
(28) Jeffery, E.; Peters, L. R.; Raghavan, M. *J Biol Chem* **2011**, *286*, 2402.
(29) Raghavan, M.; Wijeyesakere, S. J.; Peters, L. R.; Del Cid, N. *Trends Immunol* **2012**.
(30) Trott, O.; Olson, A. J. *J Comput Chem* **2010**, *31*, 455.
(31) Wu, G.; Robertson, D. H.; Brooks, C. L., 3rd; Vieth, M. *J Comput Chem* **2003**, *24*, 1549.
(32) Erlanson, D. A.; Braisted, A. C.; Raphael, D. R.; Randal, M.; Stroud, R. M.; Gordon, E. M.; Wells, J. A. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 9367.
(33) Hardy, J. A.; Lam, J.; Nguyen, J. T.; O'Brien, T.; Wells, J. A. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 12461.
(34) Karanicolas, J.; Brooks, C. L. *Journal of Molecular Biology* **2003**, *334*, 309.
(35) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; MacKerell, A. D., Jr. *J. Comput. Chem.* **2010**, *31*, 671.
(36) Dyson, H. J.; Wright, P. E. *Current Opinion in Structural Biology* **2002**, *12*, 54.
(37) Shammas, S. L.; Travis, A. J.; Clarke, J. *Proceedings of the National Academy of Sciences of the United States of America* **2014**, *111*, 12055.
(38) Wang, N.; Lodge, J. M.; Fierke, C. A.; Mapp, A. K. *Proceedings of the National Academy of Sciences of the United States of America* **2014**, *111*, 12061.
(39) Dickson, A.; Brooks, C. L., III *Journal of Physical Chemistry B* **2014**, *118*, 3532.
(40) Dickson, A.; Mustoe, A. M.; Salmon, L.; Brooks, C. L., III *Nucleic Acids Research* **2014**, *42*, 12126.

Appendix A

The datasets that were the test sets for Flexible CDOCKER in Chapter 2. This includes, PDBIDs, lowest found RMSDs and the RMSD of the lowest scored ligand conformation.

Table A.1:

The PDBIDs and CDOCKER results for CCDC/Astex subset. Included are lowest RMSD conformation found and the RMSD for the lowest energy conformation for the two simulated annealing implementations and the MD+Minimization protocols.

Rigid CDOCKER results

	SA- 1 Grid		*SA- 2 Grids*		*MD+Minimization*	
PDBID	Lowest RMSD	Lowest Energy	Lowest RMSD	Lowest Energy	Lowest RMSD	Lowest Energy
1a28	0.2	0.3	0.4	0.4	0.2	0.2
1a6w	1.8	2.2	1.5	2.9	1.8	2.1
1abe	1.2	2.9	1.7	2.6	1.1	2.9
1abf	0.9	3.2	2.2	3.4	1.2	3.2
1acj	0.6	0.7	0.3	0.6	0.4	0.5
1acl	2.3	4.2	2.3	5.5	2.5	2.7
1acm	1.1	2.0	1.5	2.3	0.9	1.2
1aco	0.9	3.7	1.7	2.9	0.8	1.4
1aec	1.4	2.5	1.4	1.4	1.6	1.6
1ai5	0.6	1.1	1.0	1.3	0.6	0.6
1aoe	1.9	2.2	1.8	1.9	1.8	2.0
1aqw	0.8	2.5	1.6	2.7	0.7	1.3
1ase	1.6	3.1	1.7	2.9	1.1	3.1
1azm	1.6	5.8	1.7	2.0	1.6	2.0
1b59	0.4	5.3	0.5	6.9	0.5	5.6
1b9v	1.8	2.1	1.9	2.5	1.7	2.0
1baf	1.0	6.1	1.9	4.4	1.1	2.4
1bbp	1.1	8.3	1.2	2.3	1.1	2.7
1bgo	1.8	1.8	1.6	2.1	2.0	2.1
1blh	1.8	1.8	1.4	2.5	1.7	1.8
1bmq	1.6	2.3	1.6	2.0	1.9	1.9
1byg	0.6	2.6	0.5	1.6	0.5	1.7
1c12	1.6	2.3	1.8	2.5	1.7	2.3

1c1e	0.6	4.4	0.7	0.8	0.5	0.6
1c5c	1.7	2.3	1.4	3.4	1.5	2.3
1c5x	1.2	1.3	0.5	0.9	1.2	1.3
1c83	0.6	0.6	1.0	1.0	0.6	0.9
1cbs	0.9	1.1	0.8	1.1	1.0	1.1
1cbx	0.7	1.1	0.9	1.3	0.6	1.1
1cdg	1.8	4.0	1.8	2.9	2.0	2.9
1ckp	0.6	2.9	0.6	2.9	0.6	2.9
1cle	2.1	3.1	1.4	1.9	2.3	2.3
1com	1.7	4.3	0.6	4.3	1.0	1.7
1coy	0.8	0.9	0.7	1.2	0.8	0.9
1cqp	1.1	1.1	1.0	2.0	1.1	1.2
1cvu	1.1	1.7	1.2	2.3	1.1	2.3
1cx2	1.5	1.8	1.5	1.8	1.5	1.7
1d0l	1.2	1.4	1.2	1.8	1.2	3.0
1d3h	1.5	7.3	0.9	1.1	1.4	7.3
1d4p	1.1	1.6	1.3	2.2	1.1	8.5
1dbb	1.8	5.3	1.4	5.2	1.3	4.6
1dbj	0.7	0.8	0.6	0.7	0.7	0.8
1dd7	3.0	3.9	3.1	6.0	1.1	4.6
1dg5	1.5	5.8	1.5	6.0	1.5	2.5
1did	1.2	5.5	1.9	2.6	1.2	5.4
1dmp	3.4	3.4	3.5	4.9	3.5	3.9
1dog	0.4	4.1	1.0	3.1	0.3	3.7
1dr1	0.5	2.5	1.0	2.4	0.6	2.4
1dwb	2.0	2.2	0.9	2.1	0.6	2.2
1dwc	0.9	1.3	2.6	5.6	1.0	1.0
1dwd	2.3	6.1	2.5	5.7	1.2	1.3
1dy9	1.4	1.4	1.8	3.1	1.3	1.4
1eap	2.2	4.3	2.3	3.6	2.3	3.8
1ebg	0.6	3.0	0.9	2.6	0.6	3.0
1ei1	0.6	0.6	1.9	1.9	0.6	0.6
1ejn	1.8	5.5	1.7	4.1	1.6	5.7
1eoc	1.5	1.7	1.3	2.5	1.5	1.5
1epb	1.1	2.8	1.1	1.8	1.1	1.2
1etr	0.9	0.9	0.9	0.9	1.0	1.0
1ets	3.5	4.5	3.2	4.2	3.6	4.6
1f0r	2.0	4.1	1.8	3.8	2.5	3.9
1f0s	1.4	3.9	1.8	4.2	2.4	3.9
1fen	0.6	0.8	0.6	0.6	0.5	0.6
1fgi	0.6	5.7	0.5	0.6	0.7	1.1
1fkg	1.4	2.3	1.0	1.8	1.0	2.0
1fl3	0.6	1.2	0.7	0.7	0.7	0.8

1flr	2.3	4.5	2.3	4.5	2.3	4.5
1frp	1.0	1.0	1.2	2.3	0.7	1.0
1glp	0.9	0.9	1.0	3.0	1.6	1.8
1glq	3.1	4.8	2.7	5.4	3.0	4.5
1hak	1.9	4.0	1.5	7.8	1.8	4.3
1hdc	0.8	1.2	0.8	1.2	0.8	0.9
1hfc	2.3	3.8	2.2	3.5	1.9	3.4
1hiv	1.2	1.7	1.8	1.9	1.9	3.1
1hpv	1.0	1.2	1.1	1.4	1.0	1.1
1hri	0.9	2.0	1.0	1.2	0.9	1.1
1htf	1.8	1.9	1.5	7.4	1.7	3.5
1hyt	2.4	3.2	2.4	2.9	2.5	2.8
1ibg	0.7	0.8	0.6	1.5	0.7	0.7
1imb	2.7	6.7	3.2	6.6	2.6	7.1
1ivb	2.0	3.5	2.4	3.3	2.0	4.5
1kel	1.1	1.4	1.4	1.8	1.2	1.4
1lah	0.5	1.1	1.1	2.7	0.5	1.1
1lcp	0.6	1.0	0.9	1.0	0.5	0.8
1lic	1.3	3.9	1.4	2.2	1.5	2.7
1lpm	0.9	4.8	0.5	0.5	0.6	0.6
1lst	1.1	1.2	1.1	2.2	1.0	2.0
1mcq	0.7	1.1	0.7	1.9	0.7	1.1
1mdr	1.3	2.5	1.5	2.6	1.3	2.5
1mld	1.3	1.5	1.6	2.6	1.4	1.5
1mrg	1.2	4.8	1.0	3.4	1.2	4.8
1mrk	1.4	1.7	1.4	1.5	1.4	1.5
1mts	0.9	2.1	1.1	2.1	1.1	2.0
1mup	1.2	1.4	1.1	4.3	1.1	1.3
1nco	1.1	3.5	1.1	2.6	1.4	3.5
1ngp	1.8	4.1	0.9	4.4	1.2	4.0
1okl	1.7	2.4	1.0	3.0	1.3	2.0
1pdz	1.7	2.2	1.7	4.4	1.4	2.2
1phd	1.8	3.0	2.1	7.0	1.6	3.6
1phg	2.1	4.9	1.3	6.4	1.9	4.9
1poc	1.0	1.0	1.5	1.5	1.0	1.0
1ppc	1.6	3.1	2.3	7.5	1.2	1.2
1pph	4.9	6.2	4.4	6.1	4.9	5.6
1ptv	4.3	7.3	3.9	7.2	4.3	8.6
1qcf	1.3	1.4	0.8	0.8	1.3	1.4
1qpe	1.7	2.2	0.7	0.7	1.4	1.5
1qpq	0.7	3.0	1.1	2.4	0.7	3.0
1rnt	0.6	0.6	0.8	1.1	0.5	0.5
1rob	0.9	1.1	0.7	3.3	0.9	1.1

1rt2	0.6	0.6	0.6	0.7	0.5	0.6
1slt	2.2	3.0	2.1	2.8	1.4	4.8
1snc	1.2	5.3	2.6	5.3	0.7	1.4
1srj	0.9	3.6	0.9	1.2	0.9	2.8
1tdb	1.6	1.8	1.4	3.5	1.4	2.1
1tmn	1.5	1.8	2.3	2.8	1.7	2.4
1tng	0.4	0.4	4.4	4.4	0.3	0.4
1tnh	1.4	1.5	1.4	1.5	1.4	1.6
1tni	1.3	3.0	1.5	2.1	1.6	1.6
1tpp	5.7	7.1	4.8	6.5	5.7	5.9
1tyl	0.8	0.8	0.6	2.1	0.9	0.9
1ukz	0.6	6.3	0.8	4.5	0.6	0.9
1ulb	0.9	0.9	0.6	3.5	0.7	0.9
1uvs	1.8	2.2	1.7	3.9	1.5	1.7
1uvt	3.2	3.8	3.3	3.4	3.1	3.9
1vgc	1.3	2.0	1.3	5.3	1.3	5.3
1xid	1.3	2.1	1.2	2.6	1.2	1.8
1xie	1.6	1.9	3.9	6.5	1.5	1.9
1ydr	0.9	4.0	0.8	3.5	0.9	0.9
1yee	0.9	1.0	1.1	3.2	0.9	1.1
25c8	1.8	2.6	1.6	1.9	1.7	2.0
2aad	0.8	0.9	0.5	1.0	0.8	0.8
2ada	0.8	1.1	0.7	0.8	0.7	1.0
2cht	1.1	3.9	1.3	4.1	1.0	3.4
2cmd	1.1	1.5	1.6	3.4	1.2	1.8
2cpp	0.3	2.7	1.5	2.8	0.3	2.6
2ctc	2.1	2.5	2.1	2.7	1.1	1.6
2fox	1.4	1.4	1.1	1.3	1.2	1.6
2gbp	0.3	0.7	0.7	3.5	0.3	0.7
2h4n	2.0	5.6	1.7	5.8	1.4	5.6
2ifb	1.0	1.3	0.9	1.6	1.0	1.5
2mcp	1.0	3.0	4.5	4.6	0.9	1.0
2pk4	1.2	2.3	1.0	2.8	1.1	2.2
2r07	0.7	0.7	0.7	0.7	0.7	0.7
2yhx	0.8	2.8	1.6	5.4	1.0	1.4
2ypi	1.1	1.5	1.1	1.3	0.7	1.4
3ert	0.6	1.1	1.1	1.5	0.7	1.1
3gpb	1.3	5.5	0.8	2.3	1.2	5.5
3hvt	0.9	0.9	0.8	0.8	0.9	0.9
4aah	0.7	1.0	1.2	5.9	0.6	5.8
4cox	0.6	0.7	0.4	0.7	0.6	0.7
4cts	1.8	2.1	1.7	2.1	1.3	2.2
4dfr	2.9	4.7	1.7	5.6	2.3	3.8

4est	2.3	2.9	2.2	3.8	2.2	3.4
4fbp	0.6	5.7	1.7	2.2	0.4	1.7
4lbd	1.0	1.0	1.0	1.2	1.0	1.0
4phv	3.2	4.8	2.9	5.4	3.2	5.0
5abp	0.8	0.8	0.8	1.4	0.8	0.8
5cpp	2.0	3.0	1.9	2.2	2.0	3.0
6rnt	1.8	2.4	3.0	4.0	1.3	4.1
6rsa	0.8	1.1	0.9	2.1	0.8	0.8
7tim	0.9	1.6	1.3	1.8	0.8	1.6
Averages	1.4	2.7	1.5	2.9	1.3	2.4

Flexible CDOCKER results

	SA- 1 Grid		*SA- 2 Grids*		*MD+Minimization*	
PDBID	Lowest RMSD	Lowest Energy	Lowest RMSD	Lowest Energy	Lowest RMSD	Lowest Energy
1a28	0.2	0.2	0.2	0.3	0.2	0.3
1a6w	1.8	3.0	1.8	2.8	1.7	2.2
1abe	1.1	2.9	2.3	2.9	1.3	2.8
1abf	0.8	3.0	3.3	10.6	1.1	3.0
1acj	0.4	0.7	0.6	1.2	0.3	0.4
1acl	2.1	4.3	2.1	4.2	2.1	3.3
1acm	1.4	2.1	3.3	10.0	0.4	2.4
1aco	0.4	2.9	6.2	7.2	0.6	1.1
1aec	1.7	3.2	4.5	6.9	1.5	1.7
1ai5	0.5	1.0	0.6	8.0	0.6	0.9
1aoe	0.9	1.7	2.6	8.0	1.7	1.8
1aqw	1.1	2.5	1.3	6.5	0.7	2.0
1ase	2.4	4.9	3.1	4.7	1.4	2.7
1azm	1.6	2.2	2.5	4.4	1.5	2.0
1b59	0.4	2.7	2.9	4.0	0.4	2.2
1b9v	1.8	1.9	2.6	8.5	1.6	1.9
1baf	2.3	2.5	3.9	7.3	3.3	3.9
1bbp	1.3	8.1	2.2	3.9	1.4	1.8
1bgo	1.9	2.1	2.4	3.8	1.5	2.2
1blh	1.9	1.9	3.8	4.8	1.8	2.0
1bmq	2.0	2.3	5.0	5.0	1.6	1.9
1byg	0.4	1.8	0.4	0.9	0.3	0.3
1c12	1.5	2.3	1.3	2.8	1.3	2.3
1c1e	0.3	0.3	0.4	0.4	0.3	0.4
1c5c	1.6	4.7	2.6	5.2	1.4	3.9
1c5x	0.9	0.9	0.9	0.9	0.9	0.9
1c83	0.6	1.0	5.5	11.1	0.8	1.0
1cbs	0.8	1.1	1.0	3.5	1.1	1.4
1cbx	0.6	1.1	3.7	5.6	0.7	1.3
1cdg	2.0	2.0	2.0	2.5	2.0	2.6
1ckp	0.9	1.2	3.2	5.2	0.5	1.0
1cle	3.6	5.1	4.2	4.7	3.0	3.1
1com	1.6	3.9	4.8	8.3	0.9	1.9
1coy	0.7	0.8	0.5	0.9	0.6	0.8
1cqp	0.7	1.0	1.1	1.6	1.0	1.5
1cvu	0.9	1.1	1.2	1.7	1.3	1.4
1cx2	1.5	1.7	1.6	1.7	1.5	1.6
1d0l	1.3	1.7	3.3	5.4	1.1	1.5

1d3h	1.5	1.9	1.3	7.6	1.5	1.9
1d4p	1.0	1.3	1.9	7.9	1.0	1.4
1dbb	0.7	4.4	3.0	5.0	0.8	3.5
1dbj	0.4	0.8	0.9	1.3	0.4	0.5
1dd7	2.3	5.1	3.8	6.3	1.8	4.0
1dg5	1.4	2.9	2.5	4.6	0.7	0.9
1did	1.0	5.7	4.9	7.7	1.0	5.7
1dmp	3.5	5.0	3.4	4.5	3.5	4.2
1dog	1.6	2.2	1.9	8.4	0.2	0.3
1dr1	0.4	2.4	3.6	3.6	0.3	2.3
1dwb	0.5	0.5	0.4	2.0	0.4	0.4
1dwc	0.9	0.9	3.3	5.1	0.9	0.9
1dwd	2.2	3.7	3.6	7.5	1.3	1.9
1dy9	1.6	1.6	3.5	5.6	1.4	1.5
1eap	2.0	4.1	2.5	3.7	2.3	4.1
1ebg	0.5	2.0	8.7	9.9	0.4	1.6
1ei1	0.4	0.5	8.2	9.7	0.5	0.5
1ejn	1.5	7.0	1.2	4.7	1.7	3.9
1eoc	2.0	4.5	5.4	7.2	1.8	2.4
1epb	1.0	1.3	0.9	1.2	1.0	1.2
1etr	0.7	0.7	6.1	6.8	0.8	0.9
1ets	3.7	3.7	3.7	4.1	4.0	4.2
1f0r	2.2	4.6	2.2	4.0	2.4	3.3
1f0s	2.5	4.9	1.6	3.4	1.9	5.0
1fen	0.6	0.7	0.5	0.6	0.6	0.7
1fgi	0.4	0.6	2.9	4.7	0.5	0.5
1fkg	1.5	1.6	1.7	2.4	1.3	1.3
1fl3	0.5	1.0	0.8	1.4	0.8	1.2
1flr	2.2	4.5	2.8	4.9	2.2	4.5
1frp	0.4	1.1	5.8	9.8	0.4	0.5
1glp	1.0	5.0	5.3	5.8	1.1	1.4
1glq	2.5	6.7	5.1	6.8	2.0	2.1
1hak	2.2	3.8	2.3	6.2	1.9	3.7
1hdc	0.9	1.2	0.9	1.3	0.8	1.0
1hfc	2.6	3.4	2.4	4.7	2.2	2.4
1hiv	2.2	3.5	1.3	1.3	2.5	2.5
1hpv	1.0	1.4	1.1	1.2	1.3	1.3
1hri	0.9	2.5	0.8	3.8	0.7	0.9
1htf	1.9	3.9	1.8	3.0	1.5	2.2
1hyt	2.5	2.8	2.4	2.9	2.4	3.2
1ibg	0.8	1.8	1.8	2.2	0.7	0.7
1imb	2.8	6.5	5.4	11.8	2.6	4.2
1ivb	1.7	3.8	2.4	4.3	1.8	1.8

1kel	1.0	1.3	1.8	1.8	1.4	1.4
1lah	0.4	1.1	1.5	2.5	0.5	1.1
1lcp	0.8	0.9	1.2	8.2	0.6	0.9
1lic	1.7	2.7	2.7	3.9	1.2	2.6
1lpm	1.9	2.4	1.6	4.2	1.2	2.3
1lst	1.1	1.4	0.9	1.8	1.0	1.6
1mcq	0.7	1.7	0.7	1.2	0.7	2.5
1mdr	1.1	3.0	2.4	4.1	1.2	2.0
1mld	1.5	1.7	3.6	4.1	1.4	1.8
1mrg	1.2	2.5	5.3	10.5	1.1	1.3
1mrk	0.8	1.1	5.7	10.5	0.8	1.2
1mts	0.9	1.3	1.4	2.2	1.1	1.2
1mup	0.9	1.4	1.2	1.5	0.9	1.2
1nco	1.3	2.2	3.6	3.8	0.9	2.1
1ngp	1.1	3.6	2.9	5.0	1.0	3.7
1okl	0.5	2.2	3.4	5.5	1.5	2.2
1pdz	1.4	3.7	2.7	3.2	1.4	2.0
1phd	1.6	4.2	1.5	6.7	1.5	2.6
1phg	2.5	4.2	2.8	6.4	2.4	2.7
1poc	0.8	1.0	1.3	2.1	0.9	0.9
1ppc	2.0	3.3	3.5	7.2	1.2	2.0
1pph	4.9	5.6	5.2	6.1	4.8	5.7
1ptv	4.4	8.0	4.1	9.8	4.1	8.5
1qcf	0.8	1.3	3.8	9.5	1.3	1.3
1qpe	0.7	1.2	1.7	1.7	1.1	1.3
1qpq	1.2	1.5	2.6	12.1	0.8	1.2
1rnt	0.8	1.3	3.9	7.0	0.8	0.8
1rob	0.9	0.9	4.6	8.8	0.7	1.1
1rt2	0.6	0.7	0.5	0.6	0.5	1.1
1slt	1.3	4.7	2.1	6.3	1.0	1.8
1snc	0.9	1.9	4.7	5.8	0.8	4.6
1srj	0.8	1.9	3.4	3.4	0.7	1.0
1tdb	1.6	3.4	3.8	8.7	1.1	1.6
1tmn	1.8	2.1	2.4	2.7	1.4	3.0
1tng	0.1	0.1	3.5	3.7	0.1	0.2
1tnh	1.4	1.8	1.3	1.6	1.4	1.7
1tni	1.4	2.8	1.4	2.0	1.6	2.6
1tpp	5.6	6.0	6.3	11.2	5.5	5.8
1tyl	0.9	1.5	1.5	3.0	0.7	0.8
1ukz	0.5	2.4	5.1	7.7	0.5	0.7
1ulb	0.5	1.3	2.7	7.8	0.4	1.3
1uvs	1.5	1.9	3.8	6.0	1.7	2.4
1uvt	2.9	3.2	3.1	3.5	3.5	4.7

1vgc	1.1	2.0	1.1	1.5	1.1	1.5
1xid	1.4	2.3	1.3	1.6	1.3	1.4
1xie	1.9	4.3	8.2	12.2	1.1	2.2
1ydr	0.5	0.6	1.0	1.6	0.5	0.6
1yee	1.2	1.2	2.3	3.7	1.0	1.2
25c8	1.7	2.1	2.4	4.3	1.5	1.7
2aad	0.6	0.8	3.8	6.9	0.6	0.8
2ada	0.4	0.5	2.9	11.1	0.5	0.5
2cht	1.4	3.1	2.4	8.7	1.1	1.6
2cmd	1.3	1.8	2.8	4.1	1.3	2.1
2cpp	0.5	2.1	1.9	5.5	0.2	0.3
2ctc	2.3	2.6	2.2	2.7	2.3	2.8
2fox	1.1	1.5	6.1	7.9	0.6	1.5
2gbp	0.5	0.7	2.1	14.6	0.3	0.4
2h4n	1.6	1.7	1.8	3.9	1.6	1.8
2ifb	1.0	1.7	1.8	2.6	1.0	1.8
2mcp	0.9	1.6	4.2	6.5	0.8	1.2
2pk4	1.1	2.1	1.2	2.3	1.3	2.5
2r07	0.6	0.9	0.6	1.3	0.6	0.8
2yhx	0.8	2.9	1.8	6.2	0.7	0.8
2ypi	1.0	1.5	3.0	10.3	0.9	1.5
3ert	1.0	1.2	0.9	1.2	0.9	1.4
3gpb	1.1	2.1	3.0	13.0	1.1	2.0
3hvt	0.5	0.7	0.4	0.8	0.7	0.9
4aah	0.5	0.8	4.0	5.6	0.5	3.8
4cox	0.4	0.6	0.7	0.9	0.5	0.5
4cts	2.1	3.7	3.9	8.2	1.5	1.6
4dfr	2.9	8.4	3.5	5.3	2.4	3.5
4est	2.1	2.9	2.3	2.9	2.3	3.5
4fbp	0.3	0.5	5.7	10.8	0.3	0.4
4lbd	1.0	1.0	1.0	1.1	1.0	1.0
4phv	3.7	5.2	3.4	4.9	3.5	4.2
5abp	0.5	0.5	2.1	11.5	0.4	0.5
5cpp	1.9	1.9	1.9	5.4	1.9	1.9
6rnt	1.5	2.1	6.7	10.8	1.0	1.1
6rsa	0.4	0.8	2.6	10.1	0.4	0.9
7tim	1.1	1.7	3.6	4.0	0.8	1.7
Averages	1.3	2.4	2.8	5.2	1.3	1.9

Table A.2:

Results for HIV Reverse Transcriptase cross-docking trial, totaling 173 unique protein-ligand complexes. The MD+Minimization protocol was the only one used in this trial as it was the most successful docking protocol in the redocking trials. Reported are the lowest RMSD structure sampled and the lowest energy RMSD conformation.

		Rigid CDOCKER		*Flexible CDOCKER*	
Receptor PDBID	**Ligand PDBID**	**Lowest RMSD**	**Lowest Energy**	**Lowest RMSD**	**Lowest Energy**
1c1b	1c1c	0.9	6.0	1.0	1.6
1c1b	1ep4	1.7	7.2	1.5	5.5
1c1b	1rth	0.9	6.2	0.9	6.2
1c1b	1vru	1.1	2.2	0.9	5.6
1c1b	2be2	1.3	1.8	1.4	1.5
1c1b	2rf2	2.6	3.4	2.6	2.9
1c1b	2rki	1.9	2.2	1.9	5.7
1c1b	3dlg	1.2	1.3	2.4	8.9
1c1b	3dol	2.5	10.8	3.8	10.6
1c1b	3dya	2.9	3.0	2.9	3.7
1c1c	1c1b	0.8	2.0	0.8	2.1
1c1c	1ep4	2.0	7.0	1.7	6.1
1c1c	1rt2	0.7	0.7	0.8	1.6
1c1c	1rth	1.2	14.6	1.1	6.3
1c1c	1vru	1.5	5.2	1.5	5.5
1c1c	2be2	1.0	2.2	1.0	1.3
1c1c	2rf2	2.2	6.3	2.1	6.1
1c1c	2rki	1.8	2.5	1.7	7.8
1ep4	1c1b	0.9	1.3	1.4	2.8
1ep4	1rt2	1.3	2.0	1.4	5.1
1ep4	1rth	1.2	12.4	1.7	4.6
1ep4	1vru	1.6	5.1	2.1	4.9
1ep4	2be2	1.5	1.6	1.3	2.0
1ep4	2rf2	3.2	5.8	3.3	5.8
1ep4	2rki	2.8	5.0	3.0	5.7
1ikw	1c1b	4.1	5.7	4.3	7.2
1ikw	1c1c	4.2	5.9	4.7	7.0
1ikw	1ep4	3.7	6.6	3.9	6.5
1ikw	1vru	4.9	5.5	4.8	6.5
1ikw	2be2	7.1	10.4	6.7	10.8
1ikw	2rf2	4.4	7.3	4.4	6.7
1ikw	2rki	5.2	8.8	5.2	8.8
1ikw	3dol	7.9	10.0	7.9	10.6

1rt2	1ep4	2.2	13.6	2.1	2.8
1rt2	1rth	1.5	14.9	4.2	14.7
1rt2	1vru	1.8	14.9	1.6	5.1
1rt2	2be2	1.4	13.1	1.3	2.4
1rt2	2rf2	2.1	15.4	2.1	2.2
1rt2	2rki	1.7	13.2	1.8	2.8
1rt2	3dlg	1.0	12.8	0.8	1.6
1rth	1c1b	0.9	4.1	0.9	2.1
1rth	1c1c	1.3	5.4	1.2	1.7
1rth	1rt2	2.4	7.8	2.8	7.3
1rth	1vru	1.3	4.6	0.8	5.3
1rth	2be2	3.1	6.6	2.6	6.3
1rth	2rki	2.8	7.7	2.8	7.9
1s1t	1c1b	0.8	1.2	1.1	1.5
1s1t	1c1c	1.0	6.2	1.1	5.5
1s1t	1rt2	2.6	7.8	2.5	7.0
1s1t	1rth	1.0	13.8	1.1	6.1
1s1t	1vru	1.2	2.3	0.9	5.2
1s1t	2be2	2.3	6.3	2.2	6.8
1s1t	2rf2	1.7	1.9	1.4	4.7
1s1t	2rki	1.7	7.6	2.4	7.4
1s1t	3dlg	3.8	11.4	4.4	8.5
1vrt	1c1b	1.0	5.1	1.0	5.8
1vrt	1c1c	1.1	3.2	1.1	2.9
1vrt	1rt2	2.7	7.3	3.0	7.3
1vrt	1rth	0.7	14.5	0.6	6.1
1vrt	2be2	2.6	6.5	2.4	6.5
1vrt	2rf2	1.2	4.5	1.5	4.5
1vrt	2rki	2.5	8.2	2.8	6.2
1vru	1c1b	1.0	5.0	1.0	1.9
1vru	1c1c	1.2	3.1	1.2	5.0
1vru	1rt2	3.3	7.9	2.9	7.1
1vru	1rth	0.8	13.0	0.7	1.3
1vru	2be2	3.2	6.3	2.9	6.5
1vru	2rf2	1.5	4.5	1.4	5.8
2be2	1c1b	1.3	1.4	1.2	5.0
2be2	1c1c	1.2	1.2	1.2	1.4
2be2	1rt2	1.2	13.1	1.1	1.6
2be2	1rth	1.9	13.4	2.3	5.9
2be2	2rf2	2.2	5.8	2.3	6.8
2hnd	1c1b	1.3	5.7	1.3	1.3
2hnd	1c1c	1.3	5.5	1.1	5.5
2hnd	1rth	1.3	13.9	0.8	1.7

2hnd	2rf2	1.0	4.4	1.0	4.5
2hny	1c1b	1.3	5.8	1.2	5.5
2hny	1c1c	1.2	3.3	1.2	3.1
2hny	1rth	1.2	4.4	1.0	1.3
2hny	2rf2	0.9	4.3	0.9	4.5
2hny	3dlg	4.6	9.5	4.8	7.8
2hny	3dol	6.9	11.4	6.5	11.2
2ops	1c1c	1.3	5.4	1.1	1.6
2ops	1rt2	2.2	6.2	2.7	6.3
2ops	1rth	1.1	14.3	0.7	6.0
2ops	2be2	2.7	6.5	3.1	6.8
2ops	2rf2	1.6	4.6	1.5	4.5
2ops	3dlg	3.9	8.5	4.1	9.4
2rf2	1c1b	1.5	13.9	1.3	5.4
2rf2	1c1c	1.1	5.5	1.3	5.5
2rf2	1ep4	3.4	9.3	3.2	7.3
2rf2	1rt2	2.8	7.6	3.0	7.9
2rf2	1rth	1.4	16.0	1.6	6.0
2rf2	1vru	1.8	1.8	1.6	5.0
2rf2	2be2	3.8	13.7	3.5	7.2
2rf2	2rki	2.7	14.9	2.6	4.2
2rf2	3dlg	4.2	13.1	4.3	10.2
2rf2	3dol	6.9	13.7	6.2	12.3
2rf2	3dya	3.6	14.2	3.6	7.1
2rki	1c1b	1.6	12.2	1.6	2.5
2rki	1c1c	1.6	5.7	1.3	5.5
2rki	1ep4	2.4	10.9	2.3	7.5
2rki	1rt2	1.5	13.3	1.2	1.6
2rki	1rth	2.3	13.7	2.6	6.0
2rki	1vru	2.1	6.6	3.1	4.3
2rki	2be2	1.7	12.2	2.0	2.1
2rki	2rf2	1.5	12.8	1.2	4.7
2rki	3dlg	1.5	4.8	1.5	1.5
2rki	3dol	2.5	10.8	2.3	2.3
2rki	3dya	1.6	11.6	1.5	8.7
2zd1	1c1b	1.5	2.7	1.5	1.6
2zd1	1c1c	1.8	6.7	1.9	6.6
2zd1	1ep4	3.3	6.8	3.3	6.4
2zd1	1rt2	3.0	7.7	3.0	7.9
2zd1	1rth	1.1	5.4	0.9	6.2
2zd1	1vru	1.6	2.3	1.8	2.4
2zd1	2be2	3.1	7.6	2.4	6.0
2zd1	2rf2	2.5	5.0	2.4	5.1

2zd1	2rki	3.0	8.2	3.0	8.1
2zd1	3dlg	4.6	9.8	4.3	9.2
2zd1	3dol	6.5	11.7	6.6	11.4
2zd1	3dya	3.4	6.7	3.4	6.5
3bgr	1c1b	1.4	3.1	1.4	6.1
3bgr	1c1c	1.9	7.0	1.9	6.7
3bgr	1ep4	2.8	8.7	2.8	6.2
3bgr	1rt2	2.5	9.0	3.0	7.8
3bgr	1rth	1.1	6.1	1.0	1.5
3bgr	1vru	1.5	3.3	1.7	4.0
3bgr	2be2	2.2	5.6	2.0	5.6
3bgr	2rf2	3.0	5.2	2.9	5.1
3bgr	2rki	2.8	10.3	3.1	8.2
3bgr	3dol	6.5	13.0	6.2	9.4
3bgr	3dya	2.3	6.4	2.9	6.4
3dle	1c1b	1.2	14.3	1.3	6.0
3dle	1c1c	1.5	6.0	1.3	4.8
3dle	1ep4	2.3	13.2	2.2	2.7
3dle	1rt2	0.7	12.4	0.5	0.8
3dle	1rth	3.7	14.6	4.8	5.0
3dle	1vru	3.2	5.3	3.8	5.2
3dle	2be2	1.4	1.7	1.3	2.5
3dle	2rf2	2.3	5.0	2.7	7.4
3dle	2rki	2.4	15.4	2.0	2.8
3dle	3dol	1.1	4.4	1.7	2.2
3dle	3dya	2.5	11.9	2.4	2.5
3dlg	1c1b	1.3	13.9	1.7	5.2
3dlg	1c1c	1.2	14.6	1.6	6.3
3dlg	1ep4	2.2	13.7	2.1	5.6
3dlg	1rt2	1.0	13.4	1.0	2.1
3dlg	1rth	1.9	12.4	6.0	12.5
3dlg	1vru	2.7	14.8	3.5	4.5
3dlg	2be2	1.2	1.5	1.0	1.5
3dlg	2rf2	2.3	5.0	3.7	5.6
3dlg	2rki	2.2	11.9	2.2	6.8
3dlg	3dol	1.4	2.4	1.7	2.5
3dlg	3dya	2.5	3.6	2.3	2.5
3dol	1c1b	4.7	13.5	2.6	2.6
3dol	1c1c	4.4	5.5	4.8	5.7
3dol	1ep4	3.0	12.8	4.0	8.4
3dol	1rt2	1.9	11.8	2.2	2.6
3dol	1vru	5.4	15.8	4.8	6.8
3dol	2be2	2.4	2.4	2.3	2.9

3dol	2rki	3.4	13.3	3.2	4.3
3dol	3dlg	1.6	2.7	1.4	1.4
3dol	3dya	3.2	4.1	1.5	1.5
3dya	1c1b	2.3	2.5	2.2	2.2
3dya	1c1c	2.1	6.6	1.9	6.5
3dya	1ep4	2.1	3.0	2.0	2.8
3dya	1rt2	2.3	2.9	2.1	2.4
3dya	1rth	3.0	13.6	3.2	6.7
3dya	1vru	3.3	6.0	3.9	5.8
3dya	2be2	1.3	2.3	1.4	2.3
3dya	2rf2	1.9	2.7	1.9	6.4
3dya	2rki	2.2	2.7	2.0	2.3
Averages		2.3	7.7	2.3	5.2