
Theses and Dissertations

Spring 2016

Structural and functional consequences of single mutations at the high affinity binding site of cyanovirin-N

Zhen Li
University of Iowa

Copyright 2016 Zhen Li

This dissertation is available at Iowa Research Online: <http://ir.uiowa.edu/etd/3133>

Recommended Citation

Li, Zhen. "Structural and functional consequences of single mutations at the high affinity binding site of cyanovirin-N." PhD (Doctor of Philosophy) thesis, University of Iowa, 2016.
<http://ir.uiowa.edu/etd/3133>.

Follow this and additional works at: <http://ir.uiowa.edu/etd>

 Part of the [Chemistry Commons](#)

STRUCTURAL AND FUNCTIONAL CONSEQUENCES OF SINGLE
MUTATIONS AT THE HIGH AFFINITY BINDING SITE OF CYANOVIRIN-N

by

Zhen Li

A thesis submitted in partial fulfillment of the
requirements for the Doctor of Philosophy
degree in Chemistry
in the Graduate College of
The University of Iowa

May 2016

Thesis Supervisor: Professor Claudio J. Margulis

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

PH.D. THESIS

This is to certify that the Ph.D. thesis of

Zhen Li

has been approved by the Examining Committee for the
thesis requirement for the Doctor of Philosophy degree in
Chemistry at the May 2016 graduation.

Thesis Committee: _____
Claudio J. Margulis, Thesis Supervisor

Adrian H. Elcock

Alexei V. Tivanski

Amanda J. Haes

Renee S. Cole

ACKNOWLEDGEMENTS

First of all, I would like to thank my adviser Professor Claudio J. Margulis. He has fully supported me on academics and at the same time he is also a life mentor. He encourages and supports me to explore the world. He teaches me how to ask valuable questions and how to solve problems efficiently. The conversations with him are inspiring and I really appreciate it.

In addition, I would like to thank the Department of Chemistry at the University of Iowa. There are many professors, such as Professor Haes, Professor Shaw, and Professor Cole, that I would like to thank for their precious advice and the help they have given. I cherish the memories of outreach activities that I have participated in with them. Moreover, I would like to especially thank Professor Haes for her advice during hard times. I would also like to thank the professors in my committee including Professor Haes, Professor Cole, Professor Tivanski, and Professor Elcock. In addition, I would like to thank all the professors and staff in the department.

I would also like to thank my friends and my parents. I am really glad that I have them to complain at hard times and to share my happiness at good times. Thanks for being around me rain or shine. The volunteer activities and other events we have done together really brought me a lot of pleasure.

I would also like to thank the lab members and countless people who have taught me a lot. I am lucky to have many supportive lab members such as Dr. Ramadugu, Dr. Hettige, Dr. Araque, and Dr. Hu. Their expertise enriched my

knowledge enormously. I would like to thank our collaborators, Professor Ghirlanda, Professor Ozkan, and Dr. Bolia, from Arizona State University as well. It was really nice to work with them.

Finally, I would also like to thank the people around in Iowa City for their friendliness and kindness. Iowa is really a blessed state and I did enjoy my time at Iowa.

ABSTRACT

This thesis focuses mainly on the consequences that single mutations have on structural, functional and energetic aspects of the protein cyanovirin-N. In order to estimate the free energy of single mutations, we have applied thermodynamics integration and Bennett acceptance ratio techniques. Replica exchange molecular dynamics has been applied to accelerate simulations for complicated scenarios. Our studies suggest that certain single mutations may be promising to improve binding affinity to $\text{Man}\alpha 1 \rightarrow 2\text{Man}\alpha$ but we also learned that the simplistic view of a strong hydrogen bond correlating to a high binding affinity may not always be correct. Finally, we explored in detail the widely used mutation P51G for its impact on protein rigidity at the very important hinge region as well as for its possible effect on glycan binding.

TABLE OF CONTENTS

LIST OF TABLES	vii
LIST OF FIGURES	viii
CHAPTER	
1 INTRODUCTION	1
2 ROLE PLAYED BY GLU41 OF P51G-M4-CVN	9
2.1 Introduction	9
2.2 Methods	13
2.2.1 System Preparation	13
2.2.2 Free Energy Simulations	15
2.3 Results and Discussion	18
2.4 Conclusions	32
3 COMPUTATIONALLY DESIGNED MUTATIONS: AN ATTEMPT AT IMPROVING THE BINDING AFFINITY OF CVN TO DIMAN- NOSE	34
3.1 Introduction	34
3.2 Methods	36
3.2.1 System Preparation and Free Energy Calculation	36
3.2.2 Replica Exchange Molecular Dynamics	39
3.3 Results	40
3.3.1 Thr57Ser	40
3.3.2 Ser52Thr	48
3.3.3 Glu41Gln	55
3.3.4 Val43Ala	57
3.3.5 Glu56Asp	60
3.4 Conclusion	65
4 OPEN AND CLOSED BINDING POCKET IN DOMAIN B AFFECT- ING THE BINDING ABILITY OF CVN TO DIMANNOSE	67
4.1 Introduction	67
4.2 Methods	71
4.3 Results	72

4.3.1	Flexibility in the Loop Region	74
4.3.2	Dynamics of Hinge and Hairpin Regions Define Open and Closed States of the High-Affinity Binding Pocket	75
4.4	Conclusion	81
5	SUMMARY AND FUTURE DIRECTIONS	83
	APPENDIX	89
A	FORCE FIELD PARAMETERS FOR DIMANNOSE	89
B	CONVERGENCE ANALYSIS FOR OPEN AND CLOSED STATE SIMULATIONS	98
C	COPYRIGHT	100
	REFERENCES	103

LIST OF TABLES

Table	
2.1	BAR analysis for relative binding free energies for the Glu41Ala mutation 24
2.2	BAR analysis for relative binding free energies for the Glu41Gly mutation 24
3.1	BAR analysis for relative binding free energies for the Thr57Ser mutation 43
3.2	BAR analysis for relative binding free energies for the Ser52Thr mutation 53
3.3	BAR analysis for relative binding free energies for the Glu56Asp mutation 65
4.1	Enthalpies, entropies, and K_d values for a set of cyanovirin-related proteins 81
A.1	Atomic parameters for dimannose in the format of OPLS-AA 89
A.2	Bond parameters for dimannose in the format of OPLS-AA 90
A.3	Bending parameters for dimannose in the format of OPLS-AA 91
A.4	Dihedral parameters for dimannose in the format of OPLS-AA. The last two columns are all 0.000000, for the purpose of simplicity, the last two columns are not listed. 94

LIST OF FIGURES

Figure	
1.1 Two repeats are colored with red (repeat 1) and blue (repeat 2) to the left; two domains are colored with green (domain A) and yellow (domain B) to the right.	2
1.2 (PDB ID: 1L5B) The two amino acid chains of the dimer are colored with green (green and light green) and blue (dark blue and blue). The sequence repeats from the same amino acid chain are colored with same type of color but with slightly different contrast. The two hinge regions in two chains are highlighted with orange and cyan.	3
1.3 Structures of oligosaccharide Mannose 9 with the common terminal Mannose 2 highlighted in red boxes	5
1.4 The CVN bound to two dimannose. Domain A is colored blue and Domain B is colored yellow.	6
2.1 Thermodynamic cycle to compute the binding free energy difference for the mutation E41A(G)	16
2.2 (a) χ_1 of Glu 41 is in the Trans conformation centered around 180° and the side chain of Glu41 is able to establish strong hydrogen bonding with 2'OH of the non-reducing end of dimannose.(b) χ_1 of Glu41 is in the gauche conformation centered around 60° and its side chain is unable to establish hydrogen bonds with 2'OH of the non-reducing end of dimannose. In both figures, the protein backbone is represented with a large green arrow. . .	19
2.3 Hydrogen bond distance between carboxylate oxygens of side chain of glutamate and 2'OH of non-reducing end of dimannose using the AMBER11 software with the ff99SB force field for proteins and the GLYCAM06 force field for dimannose.	21
2.4 Hydrogen bond distance between carboxylate oxygens of side chain of glutamate and 2'OH of non-reducing end of dimannose using the Gromacs software suite coupled to the OPLS-AA force field for protein and the OPLS-Carbohydrates force field for dimannose	22

2.5	Three-step approach for thermodynamic integration simulations for Glu41Ala in presence of dimannose, trial I (black circles), trial II (red squares) and trial III (green diamonds). Discharging of Glutamate (a), LJ transformation (b) and discharging of Alanine (c).	27
2.6	Three-step approach for thermodynamic integration simulations for Glu41Ala in absence of dimannose, trial I (black circles), trial II (red squares) and trial III (green diamonds). Discharging of Glutamate (a), LJ transformation (b) and discharging of Alanine (c).	28
2.7	Three-step approach for Thermodynamic Integration simulations for Glu41Gly in the presence of dimannose, trial I (black circles), trial II (red squares) and trial III (green diamonds). LJ transformation (a) and discharging of Glycine (b).	29
2.8	Three-step approach for thermodynamic integration simulations for Glu41Gly in the absence of Dimannose, trial I (black circles), trial II (red squares) and trial III (green diamonds). LJ transformation (a) and discharging of Glycine (b).	30
3.1	Key residues of wt-CVN in domain B for binding to dimannose.	36
3.2	(a) Thr57 forming hydrogen bonds with vicinal residues including dimannose and Asn42 in the presence of dimannose. (b) Discharged Thr57 side chain resulting in a broken hydrogen bonding network.	41
3.3	Hydrogen bond length between Thr57 and dimannose as a function of λ when discharging the side chain of Thr57.	42
3.4	Hydrogen bond between Ser57 and Asn42.	44
3.5	(a) Binding pocket with discharged Ser57 in the absence of dimannose (b) Binding pocket with discharged Thr57 in the absence of dimannose. The green boxes highlight the side chain of the residue 57, and the Asn42 is represented by ball and stick model.	45
3.6	Three-step approach for thermodynamic integration simulations for Thr57Ser in presence of dimannose. Trial I (black circles), trial II (red squares). Discharging of Threonine (a), LJ transformation (b) and discharging of Serine (c).	46

3.7	Three-step approach for thermodynamic integration simulations for Thr57Ser in absence of dimannose. Trial I (black circles) and trial II (red squares). Discharging of Threonine (a), LJ transformation (b) and discharging of Serine (c).	47
3.8	Hydrogen bond formed between Thr52 and dimannose	49
3.9	Replica exchange rate in the temperature space for each λ starting at 300 K when discharging Thr52 in the presence of dimannose. Replica exchange rates starting at other temperatures are not shown for clarity and the exchanges are sufficient. The lowest temperature is 300.00 K and the highest is 328.87 K. There are 10 more temperatures (302.55, 305.15, 307.73, 310.31, 312.91, 315.54, 318.17, 320.82, 323.48, and 326.15 K) in between these two.	50
3.10	Three-step approach for thermodynamic integration simulations for Ser52Thr in presence of dimannose. Discharging of Serine (a), LJ transformation (b) and discharging of Threonine (c) without REMD (black circles) and with REMD (red squares).	51
3.11	Three-step approach for thermodynamic integration simulations for Ser52Thr in absence of dimannose. Discharging of Serine (a), LJ transformation (b) and discharging of Threonine (c).	52
3.12	$\partial V/\partial\lambda$ for $\lambda = 1.0$ when discharging Thr52 in the presence of dimannose.	53
3.13	Two conformations of fully discharged Thr52 in the presence of dimannose. (a) the methyl group pointing down (b) the methyl group pointing up. . .	54
3.14	Three different conformations for Gln41.	56
3.15	Discharging of Val43 for mutation Val43Ala in the absence of dimannose. The $\langle\partial V/\partial\lambda\rangle$ is not converged because of the peak at λ equals 0.1. . . .	57
3.16	$\partial V/\partial\lambda$ for $\lambda = 0.10$ when discharging Val43 in the absence of dimannose.	58
3.17	Two conformations of partially discharged Val43 ($\lambda = 0.10$) in the absence of dimannose. (a) the β -hydrogen pointing down corresponds to $\partial V/\partial\lambda = -4$ kcal/mol (b) the β -hydrogen pointing up corresponds to $\partial V/\partial\lambda = -3$ kcal/mol.	59

3.18	An illustration of the interaction between residue 56 and dimannose. (a) residue 56 is Glu and the carboxyl oxygens are far away from dimannose (b) residue 56 is Asp and the carboxyl oxygens are close to dimannose.	61
3.19	The hydrogen bond length between residue 56 and dimannose over a 20 ns NPT simulation. The distance is measured between the carboxyl oxygen of residue 56 and 1'OH of dimannose. The bond lengths between oxygen atoms of Glu56 and dimannose are colored with black and brown; whereas the bond lengths between the oxygen atoms in Asp56 and dimannose are colored with blue and cyan. A running average of 1 ns was conducted for clarity. The average distance between carboxyl group of Glu56 and dimannose is 7 Å , whereas it is 5 Å for Asp56.	62
3.20	Three-step approach for thermodynamic integration simulations of the Glu56Asp mutation in presence of dimannose. Discharging of Glutamate (a), LJ transformation (b) and discharging of Aspartate (c).	63
3.21	Three-step approach for thermodynamic integration simulations of the Glu56Asp mutation in absence of dimannose. Discharging of Glutamate (a), LJ transformation (b) and discharging of Aspartate (c).	64
4.1	(A) NMR structure of cyanovirin bound to two molecules of dimannose (PDB entry 1IIY). (B) Crystal structure (PDB entry 2RDK) of P51G-m4-CVN where dimannose is colored blue. The hinge and the hairpin regions involved in the opening and closing of the high-affinity binding site in domain B are colored red.	70
4.2	(A) Wire frame representation of the P51G-m4-CVN-binding site in domain B with bound dimannose (PDB entry 2RDK). (B) Apo form of wt-CVN from molecular dynamics simulations in explicit solvent. (C) Apo form of P51G-m4-CVN from simulations in explicit solvent. Dimannose is superposed for visualization purposes in panels B and C. In solution, because of rigidity in the hinge region, wt-CVN (B) preserves the separation between the backbone carbonyl oxygen of Asn53 (red sphere) and the amide nitrogen of Asn42 (blue sphere), rendering the pocket open for binding. There is no clash with the hydroxyl of the ligand (red sphere). Instead, the Pro51Gly mutation enhances the flexibility at the hinge region, and a close interaction between the backbone amide nitrogen of Asn42 and the backbone carbonyl oxygen of Asn53 closes the pocket and blocks binding as a clash with the hydroxyl oxygen (red sphere) from the glycan would occur (C).	73

4.3	Alignment of eight trajectory-averaged structures color-coded on the basis of the root-mean-square fluctuation (rmsf) with respect to the average structure of each individual amino acid (four structures at 300 K and four structures at 330 K) in the case of (A) wt-CVN and (B) the P51G- m4-CVN mutant. At both temperatures, the average structure of the hinge region is almost unchanged in the case of wt-CVN. The opposite is true in the case of the mutant where the hinge takes on different conformations. We notice that the hairpin in the vicinity of Asn42 appears also to be more mobile at high temperatures in the case of the mutant. The lowest and highest rmsf values were 0.38 and 1.63 Å , respectively.	76
4.4	Probability of the distance between the carbonyl oxygen of Asn53 and the amide hydrogen of Asn42 (D_{pocket}) in the case of wt-CVN and the P51G-m4-CVN mutant at 300 and 330 K from 760 ns MD simulations. Three conformations with D_{pocket} close to 5, 3.5, and 2.5 Å are detected. On the right-hand side, time-averaged (over a period of at least 10 consecutive nanoseconds) Asn53 and Asn42 structures consistent with each of the three conformations (peaks in the graph) are depicted alongside idealized representations of hinge behavior (light blue) and hairpin behavior (light green).	78
4.5	Probability distribution as a function of the ψ dihedral angle of residue 53 (in the hinge region) and φ dihedral angle of residue 42 (in the hairpin region) in the case of the P51G-m4-CVN mutant at 330 K. We see from this plot that a deactivated hairpin is consistent with two different conformations of residue 53 (open pocket at about 40° and closed pocket at about 140°). Instead, an activated hairpin appears to favor the conformation where Asn53 and Asn42 are in the proximity of each other. In this case, both residues contribute to a closed pocket and the open conformation of Asn53 is rarely observed.	79
B.1	Convergence analysis of distribution in figure 4.4 of the main text. This same plot is computed at 50, 100, 150 and 200 ns. We see that all features are captured at 50 ns however it is only at times longer than 100 ns that differences between graphs become insignificant for practical analysis. A plot of cumulative differences is also provided.	99
C.1	Copyright permission from The American Chemical Society for the article by S. K. Ramadugu, Z. Li, H. K. Kashyap, and C. J. Margulis, <i>Biochemistry</i> , 2014, 53(9), 1477-1484.	101
C.2	Copyright permission from The American Chemical Society for the article by Z. Li, A. Bolia, J. D. Maxwell, A. A. Bobkov, G. Ghirlanda, S. B. Ozkan, and C. J. Margulis, <i>Biochemistry</i> , 2015, 54(46), 6951-6960.	102

CHAPTER 1 INTRODUCTION

Cyanovirin-N (CVN), a stable potent virucidal protein against human immunodeficiency virus (HIV), was first discovered from the extract of the cyanobacterium *Nostoc ellipsosporum* in 1997.[28] The 11-kDa protein consists of 101 amino acids with a significant extent of internal sequence duplication including 13 conserved and 16 identical amino acid residues. [28, 48] The two domains (domain A and domain B) in CVN are similar in terms of both the sequences and structure. Figure 1.1[17] shows how one can define two different domains as well as two different repeats. Notice that domains and repeats do not coincide in sequence. Each of the two domains contains aminoacids belonging to both repeats. Inspection of Figure 1.1 reveals that there is no terminus in domain B. Both the N-terminus and C-terminus are in domain A. This results in a domain A that is more flexible than domain B and may be associated with an asymmetry in the binding affinity of the two domains for their corresponding substrate.[19, 67] Interestingly, it is found that under physiological conditions a small fraction of the protein can be found as a domain-swapped dimer.[26, 98] As shown in Figure 1.2, each of the two amino acid chains appears in all four domains of the dimer.

Research by Boyd *et. al.*[28] has discovered that CVN is active against HIV because of its interactions with the viral surface envelop glycoprotein. The binding of glycoprotein gp120 of HIV to CD4, which leads to conformational changes of gp120 that facilitate the subsequent binding of gp41 to chemokine receptors, plays an im-

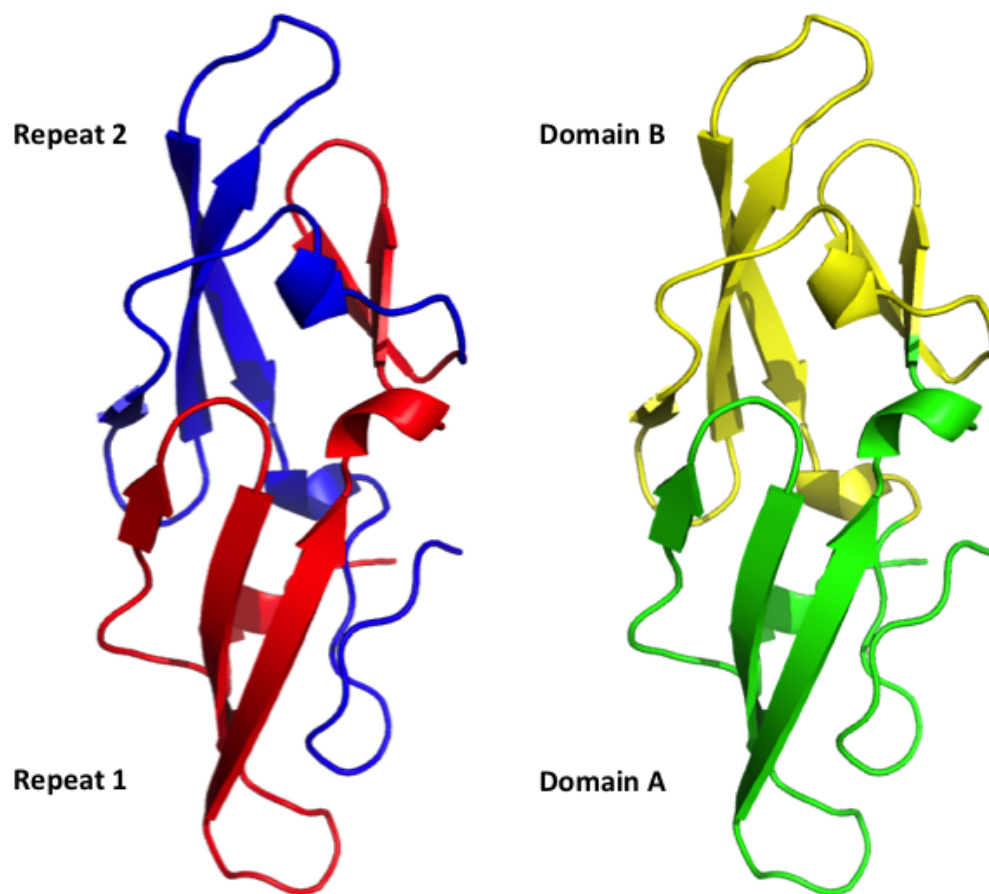


Figure 1.1: Two repeats are colored with red (repeat 1) and blue (repeat 2) to the left; two domains are colored with green (domain A) and yellow (domain B) to the right.

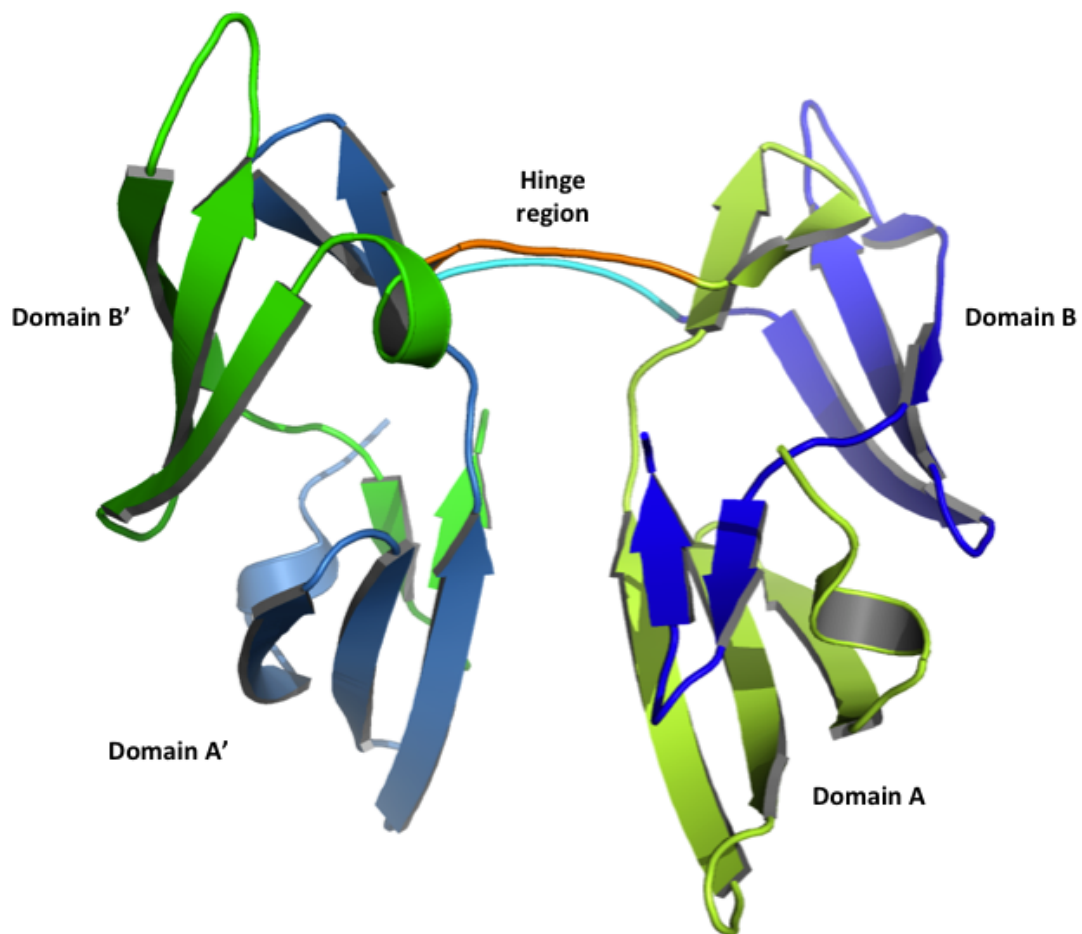


Figure 1.2: (PDB ID: 1L5B) The two amino acid chains of the dimer are colored with green (green and light green) and blue (dark blue and blue). The sequence repeats from the same amino acid chain are colored with same type of color but with slightly different contrast. The two hinge regions in two chains are highlighted with orange and cyan.

portant role in the entry of the virus. [32] The high binding affinity of CVN against gp120 is crucial for its potent antiviral activity. This is because the binding to virion-associated gp120 will impair the binding of virion-associated gp120 to cell-associated CD4, which is essential to the fusion of viral and cell membranes.[75, 38, 37, 77] In addition to HIV, CVN also inhibits the infectivity of Ebola virus because of its ability to interact with surface envelope glycoproteins.[10]

In the case of gp120 on the surface of HIV, the 24 N-linked oligosaccharides, including 11 high mannose or hybrid type, account for half of its molecular mass.[47, 66] The carbohydrates, especially $\text{Man}_9\text{GlcNAc}_2$ (Figure 1.3), can be recognized by CVN; $\text{Man}\alpha 1\rightarrow 2\text{Man}\alpha$ (dimannose, Figure 1.3), the common terminal of high mannose oligosaccharides Man_9 , is essential for high affinity binding.[19] As shown in Figure 1.4, CVN contains two carbohydrate binding sites with different binding affinities located in each domain at the end of the protein. The high binding affinity site is located in domain B with 10-fold higher affinity compared to its counterpart in domain A.[19, 67] Original work by Bewley and Otero-Quintero[19] reported that the binding dissociation constants (K_d) of domain A and domain B to dimannose for wt-CVN determined by isothermal titration calorimetry were about $1.47 \mu\text{M}$ and $0.14 \mu\text{M}$, respectively. More recent work[67] suggests instead that the K_d of domain A and domain B are $410 \mu\text{M}$ and $16 \mu\text{M}$, respectively. Whereas these two experimental results are different, they both highlight that domain B binds dimannose more strongly than domain A does.

The binding affinity of CVN for dimannose is key to the antiviral activity.

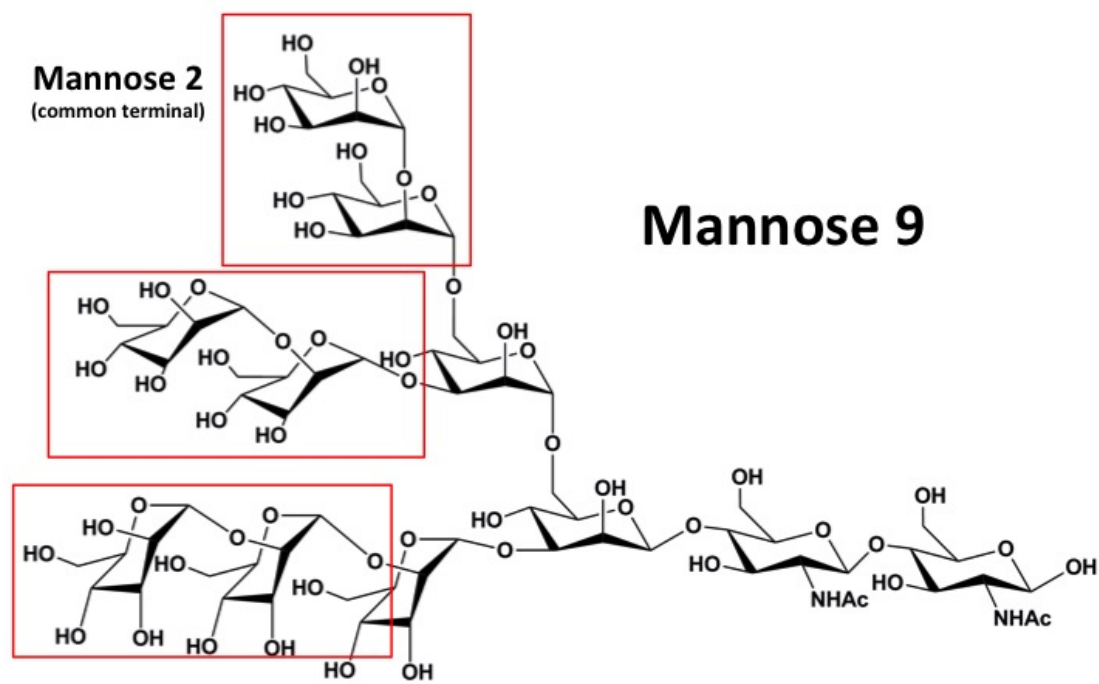


Figure 1.3: Structures of oligosaccharide Mannose 9 with the common terminal Mannose 2 highlighted in red boxes

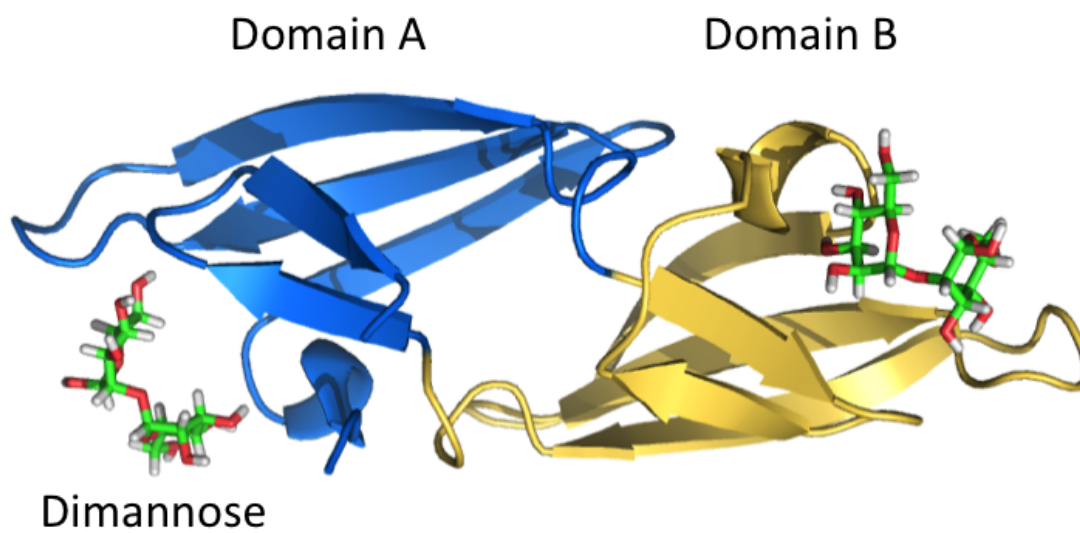


Figure 1.4: The CVN bound to two dimannose. Domain A is colored blue and Domain B is colored yellow.

Part of the work in this thesis is associated with the study of different mutations with the goal of possibly enhancing this binding affinity. In chapter 2 we first discuss a charged residue (Glu41) that at the high affinity binding site prior experimental studies[15, 42, 41] and computational studies[70, 44, 43] on wt-CVN have highlighted as likely important. In particular, chapter 2 addresses whether the presence of charge at the high affinity binding site, which is associated strong interactions with the glycan, is also fundamentally important for the free energetics of binding. In other words, does a large binding energy also imply a large binding free energy? To answer this question we use an alchemical mutation approach. This same approach is applied in chapter 3 where we discuss several other possible mutations. In both of these chapters, we work with the P51G-m4-CVN[42] mutant instead of wild type. There are several reasons for this, including that this protein favors the monomeric form and that only the high affinity binding site is active.

Of the different mutations we have tried, chapter 3 will highlight that Thr57Ser has the potential to increase binding affinity to the glycan. The system appears to prefer a Ser57 with an uncharged hydroxyl group instead of normal hydroxyl group. This is even though a prior study[96] has shown that the hydroxyl group preserves the hydrogen bonding network in the binding pocket. This chapter also discusses four other single mutations including Glu41Gln, Val43Ala, Glu56Asp and Ser52Thr. These were chosen because the locations either interact directly with the ligand or can potentially affect the interaction of other residues with the ligand. In performing these studies, we encountered important difficulties with sampling of conformers along the

chosen alchemical paths. These may very likely impact the accuracy of computed free energies but nonetheless were useful to better understand issues related to structural aspects of the high affinity binding site. In particular the 52 location associated with our Ser52Thr mutation study is in the hinge region, which chapter 4 will highlight as complex and flexible section of the protein important both to the binding affinity of wild type as well as the transformation to domain-swapped dimer.

In chapter 4 we discuss the effect that the very commonly used Pro51Gly mutation has on the flexibility of the hinge region and the effect of this on structural aspects of the high affinity binding pocket. In particular we will argue that the mutation is detrimental to binding. Comparison between our findings and experiments are discussed.

CHAPTER 2 ROLE PLAYED BY GLU41 OF P51G-M4-CVN

2.1 Introduction

The structure of wild-type CVN (wt-CVN) was first obtained using NMR spectroscopy and was determined to be in the monomeric form.[17] CVN has two domains that are pseudosymmetric. Domain A spans residues 1-38 and 90-101, while residues 39-89 form domain B.[17] More than ten different NMR or X-ray published structures of CVN and its mutants exist, some of them in the monomeric state and some others in a domain-swapped dimeric state.[7, 14, 16, 58, 68, 73, 98, 25, 26, 42, 41]. A significant number of important computational studies addressing different aspects of these systems have also been published.[70, 44, 95, 96, 43] It is believed[41, 6] that conversion of monomer to domain-swapped dimer occurs due to strain caused by the proline residue at 51 position (Pro51). Recent computational findings support this idea.[43]

Residue 51 is not the only one that determines the form of the protein. Whereas the mutation Pro51Gly stabilizes the monomeric state as well as a set of different possible oligomeric states,[62] the mutation Ser52Pro also in the hinge region appears to render it exclusively in domain-swapped dimeric form.[49] Each of the two domains in CVN can bind Man α 1-2Man which is the terminal part of the

*A portion of this chapter is reproduced from Biochemistry, vol. 53(9), page 1477, year 2014 by S. K. Ramadugu, Z. Li, H. K. Kashyap, and C. J. Margulis. DOI: 10.1021/bi4014159. Copyright - Appendix C.1

high mannose oligosaccharides (Man_9). Bewley *et. al.* have shown that dimannose binds to domain B with high affinity and to domain A with low affinity .[19] In a later article, Bewley *et. al.* mapped the binding site residues that appear to contribute the most to the binding at the high-affinity site (domain B) and at the low-affinity site (domain A). At the high-affinity site, Bewley identifies Glu41, Ser52, Asn53, Glu56, Thr57, Lys74, Thr75, Arg76 and Gln78 as key residues for binding. These residues interact either via hydrogen bonds and/or electrostatic interactions.[15] Key residues in the low-affinity site are identified as Lys3, Gln6, Thr7, Glu23 and Thr25.[15]

Molecular dynamics (MD) simulations of dimannose bound to the low- and high-affinity sites of wild type CVN have shown that, at the high affinity site, Glu41 and Arg76 appear to be important for binding due to their strong electrostatic interactions with $Man\alpha 1-2Man\alpha$ [70]. Margulis also predicted a gating or trapping mechanism in which Arg76 caps dimannose. The role of Arg76 in conformational gating was later confirmed by Fromme *et. al.* in their X-ray crystal studies as they found two distinct distributions of Arg76 in the closed and open conformations with reference to dimannose at the binding pocket.[41] Important studies on the binding and discrimination of trisaccharides at the high and low affinity binding sites are also available for wild type CVN.[18]

Fujimoto *et. al.* have used MD simulations coupled with a molecular mechanics Poisson-Boltzmann surface area (MM/PBSA) approach to study the binding affinity of $Man\alpha 1-2Man\alpha$ and $Man\alpha 1-2Man\alpha 1-2Man$ at the high and low affinity binding sites of wild type CVN. Using this MD/MM/PBSA approach they computed

the difference in binding affinities[44] at the two protein sites and found in accordance to experiment[19, 18] that binding to the high affinity was significantly more important than that at the low affinity site.

Botos *et. al.* have shown in their crystal structural studies[27] that oligomannoses, Man_9 and Man_6 bind to the low affinity site of CVN as a domain-swapped dimer. Sandstrom *et. al.* applied saturation transfer difference NMR spectroscopy in the solution phase to show that both high and low affinity sites of domain swapped dimer bind di- and trisaccharides. Saturation transfer was observed for H2, H3 and H4 of the non-reducing end of dimannose. [84] Similar observations were made by Shenoy *et. al.* using a synthetic analogue of Man_9 and Man_8 [86].

To determine whether the high-affinity site or the low affinity site are required for anti-viral activity, Barrientos *et. al.* engineered a new set of mutants, CVN^{mutDA} , CVN^{mutDB} and P51G-CVN.[9] Their conclusions were that an intact high-affinity site (domain B) was required for significant binding but domain A was important for cross linking oligosaccharides on the surface of viral glycoproteins. Later, it was demonstrated that the existence of any two binding domains, irrespective of their identities, was sufficient to retain antiviral activity.[42, 69]

Fromme *et. al.* engineered yet a new mutant, P51G-m4-CVN.[42] The purpose for this was to enforce the monomeric structure (P51G mutation) and to abolish the activity of the low affinity site (m4). The four mutations at the low affinity site in this case are K3N, T7A, E23I and N93A. The structure of P51G-m4-CVN was solved using X-ray crystallography. As expected, the protein was found to be in the

monomer form and binds dimannose only at the high-affinity site.[42] Vorontsov *et. al.* carried out the first computational studies[95, 96] on the P51G-m4-CVN mutant coupled to dimannose and a set of di-deoxy mannose ligands. Their MM/PBSA study shows that 3'OH and 4'OH of dimannose establish hydrogen-bonds with high-affinity site residues which are abolished in the case of the di-deoxy dimannose analogues. This results in much lower affinity of the di-deoxy dimannose analogues. In the first study, key residues at the high-affinity were identified as Asn42, Asp44, Ser52, Asn53, Thr57, Lys74 and Thr75 [95] and in the free energy calculation using MM/(GB)PBSA approach, Asn42 and Thr57 were highlighted as important for preserving hydrogen bonding network.[96] Interestingly Arg76, but more importantly Glu41 perceived as key to binding in experimental studies[15, 42, 41] and computational studies[70, 44, 43] on wild type CVN did not appear to be so in these studies.[95, 96] The authors proposed a possible force field dependence on this result.[95]

In this chapter, we first evaluated the suitability of two popular force field combinations (Amber ff99SB[52] for proteins coupled with Glycam06[59] for carbohydrates and OPLS-AA for proteins[55, 56] and carbohydrates [34]) for the simulation of interactions between CVN and dimannose. In particular we focus on the interactions between Glu41 and 2'OH of the sugar. With this information we then proceed to alchemically mutate residue Glu41 to Ala and Gly. This is done in order to probe if the strong Coulombic interactions between Glu41 and the glycan can be associated or not with a significant contribution to the free energy of binding.[15, 42, 41, 70, 44, 43, 95, 96] Ala and Gly are chosen because they are un-

charged and they permit a discussion not only of the effect of Coulomb interactions but also of dispersion interactions as a function of residue size.

2.2 Methods

2.2.1 System Preparation

The coordinates of the P51G-m4-CVN protein developed in the Ghirlanda lab[41] (PDB ID 2RDK) bound to Man α 1-2Man were downloaded from the protein data bank site www.pdb.org. In the present study two mutants of P51G-m4-CVN, namely E41A-P51G-m4-CVN and E41G-P51G-m4-CVN were generated by deleting the additional atoms in the Glu41 residue and renaming this residue to Ala41 or Gly41. P51G-m4-CVN, E41A-P51G-m4-CVN, and E41G-P51G-m4-CVN in complex with dimannose were solvated in a TIP3P [54] water box containing 15103 solvent molecules resulting in a cubic box of dimensions 78 X 78 X 78 Å³. The OPLS-AA force field for proteins[55, 56] combined with the OPLS force field for carbohydrates [34] was used to simulate the protein-sugar complexes using Gromacs version 4.5.x [50]. Because the OPLS-AA for carbohydrates provides parameters only for hexopyranoses, small modifications to the partial charges in the glycosidic bond region were necessary to yield the disaccharide neutral. All dimannose force field parameters are listed in Appendix A. P51G-m4-CVN has a charge of -3 electron units. Therefore 3 K^+ ions were added to maintain electroneutrality of the system. For the mutants E41A-P51G-m4-CVN and E41G-P51G-m4-CVN, 3 K^+ were also added to keep the number of counterions the same across our thermodynamic cycle.

Initially, all the protein-sugar complexes were energy minimized using a steepest-descent protocol for 5000 steps. After the initial 3000 steps, the potential energy remained nearly constant for the rest of the minimization steps. Following minimization, 100 ps were run in the NVT (constant number of particles N , volume V , and temperature $T=300$ K) ensemble followed by at least 5 ns of further equilibration in the NPT ensemble (constant number of particles N , pressure $P=1$ bar, and temperature $T=300$ K).

All subsequent simulations were also carried out at 300K and used the particle-mesh Ewald (PME)[35] protocol with a real space cutoff of 10 Å grid size of 1 Å and PME order of 6 to properly describe the electrostatic interactions. From the last 1 ns of the NPT equilibration, three snapshots separated by at least 250 ps were saved as the initial coordinates for free energy calculations.

Simulations in the absence of dimannose use the same force field parameters, simulation protocols, and include the same number of water molecules and counterions.

Whereas all our free energy calculations are carried out using the OPLS-AA force field, in order to address issues related to the force field dependence of certain results, we have also run some simulations using the Amber ff99SB[52] force field coupled with Glycam06[59] as implemented in the Amber code.[31] P51G-m4-CVN was solvated in a 68 X 64 X 71 Å³ box by 10012 SPC water molecules and three Na⁺ counterions. For equilibrium processes, we conducted a 2500 step steepest descent energy minimization, a 2500 step conjugate gradient energy minimization, and a 20

ps NVT simulation at 300 K using the Langevin piston method (collision frequency 1 ps^{-1}) sequentially. Following the equilibration, there was a 4 ns production run in NPT ensemble. Langevin piston with collision frequency of 1 ps^{-1} was used to maintain the temperature of the system at 300 K, and isotropic position scaling was applied to maintain the pressure at 1 bar. Default PME parameters were used to account for electrostatic interactions.

2.2.2 Free Energy Simulations

The objective of our free energy calculations is to computationally estimate the difference in binding free energy between dimannose bound to P51G-m4-CVN and dimannose bound to each of the two mutants.

We sought the quantity $\Delta G_3 - \Delta G_4$ in Figure 2.1 which is equal to $\Delta G_1 - \Delta G_2$. Whereas thermodynamically either one of these two free energy subtractions is equivalent, $\Delta G_1 - \Delta G_2$ is computationally more tractable because it does not involve the removal of the ligand. $\Delta G_1 - \Delta G_2$ requires the alchemical mutation of E41 to A41 or G41 in the presence (ΔG_1) and in the absence (ΔG_2) of dimannose.

ΔG_1 and ΔG_2 were further split into three processes, namely the Coulombic discharging of E41, the transformation of the Lennard-Jones interactions of E41 into those of A41 or G41, and the charging of A41 or G41. We conducted the mutation in three steps because studies[1, 82] show that turning off charges and Lennard-Jones interactions at the same time can lead to simulation instabilities. In our calculations, we computed the free energy of charging of A41 and G41 as the negative of the free

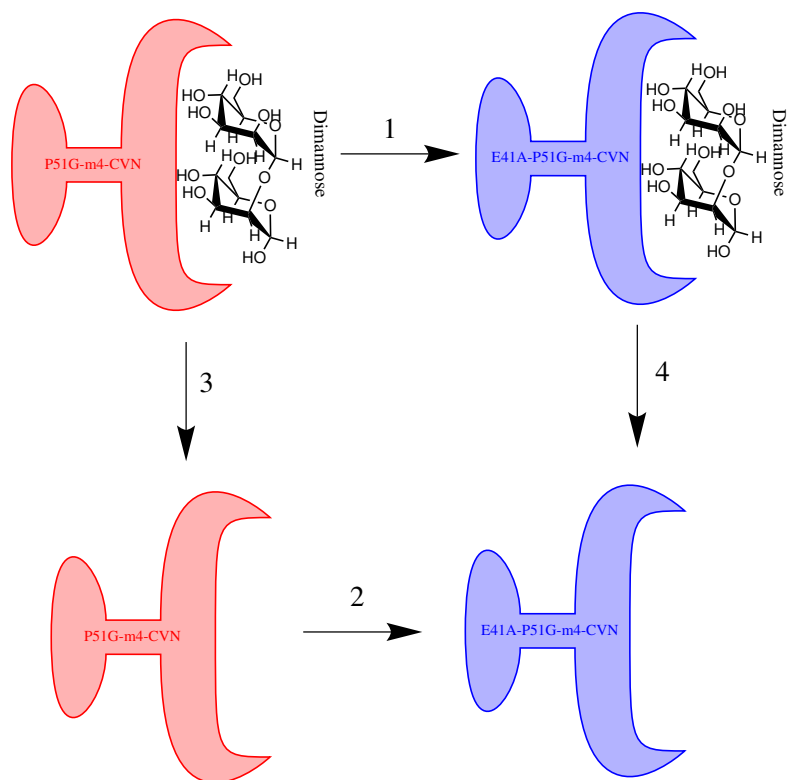


Figure 2.1: Thermodynamic cycle to compute the binding free energy difference for the mutation E41A(G)

energy of discharging.

We utilized the single topology approach [81] in which simulations with $\lambda = 0$ utilize the force field parameters of the glutamate side chain and in which simulations with $\lambda = 1$ utilize the force field parameters of alanine or glycine. The extra atoms from glutamate are converted into dummy atoms as a series of simulations transform λ from 0 to 1. For the Lennard-Jones (LJ) portion of the transformation we employed the soft-core potential approach [13, 89] as coded in Gromacs with parameters $\alpha = 0.5$, $n=1$ and $m=1$. Such settings have been previously shown to produce less error [87]

and are default values in GROMACS version 4.5.x. For a more in depth description of the implementation of the soft core potential see the GROMACS manual[93] and references therein.

Because in the case of the Coulombic discharge $\langle \partial V / \partial \lambda \rangle$, the quantity to be integrated to obtain free energy changes is fairly smooth[87, 74], we used only 11 equally spaced λ values from 0 to 1. Instead, to reduce integration noise, for the Lennard-Jones transformation we split the calculation into three parts. In the λ interval from 0 to 0.1 and 0.9 to 1 we run calculations with λ increments of 0.02. Instead, we used 0.1 increments in the intermediate region between $\lambda=0.1$ and $\lambda=0.9$.

For each λ production run the protocol was as follows; first we performed a minimization of 5000 steps followed by equilibration of 100 ps in the NVT ensemble, then we further equilibrated in the NPT ensemble for another 200 ps to finally generate a production run of 2 ns in the NPT ensemble.

Three repetitions with different initial conditions of the three-step approach in the presence and absence of dimannose resulted in a total of 0.85 μ s of combined simulation for the E41A and E41G mutations.

For the analysis of our free energy data and the estimation of errors, we utilized the g_bar program in which the Bennett Acceptance Ratio (BAR) method was implemented [11] as coded in Gromacs 4.5.x. This provides independent estimations of the error for each of the three steps of each of the three independent trials. However, we found that in some cases errors from the BAR method were smaller than the differences in average free energy values across trials. Therefore, having three trials

was very important to accurately represent our results.

2.3 Results and Discussion

In prior NMR studies, Bewley and coworkers found that Glu41 appeared to strongly hydrogen bond to the second hydroxyl group (2'OH) of the first dimannose ring[15]. This is also the case in the analysis of the P51G-m4-CVN mutant structure obtained by Fromme *et. al.* [41]

Careful scrutiny of all published monomeric and domain-swapped dimeric forms of CVN and its mutants shows that the χ_1 dihedral angle of Glu41 has two populations. In monomeric structures of CVN, whether in presence or absence of dimannose, χ_1 is reported to be in the trans conformation (with a value in the vicinity of 180°). [17, 15, 42, 41] Interestingly, in almost all domain-swapped dimeric structures χ_1 is in the gauche configuration with angle range from 45° to 75° . [33, 7, 26, 27]

This is not a minor detail because the trans conformation puts Glu41 in intimate contact with the sugar, whereas the gauche conformation puts the carboxylate part of the side chain away from the sugar unable to participate in hydrogen bonding (see Figure 2.2).

As shown in Figures 2.3 and 2.4, in the presence of dimannose, the Amber ff99SB[52] force field appears to favor the gauche conformation of χ_1 , whereas the OPLS-AA force field appears to favor the trans conformation of χ_1 .

This force field difference plays an important role during molecular dynamics simulations. Figure 2.3 shows distances between oxygen atoms in the side chain of

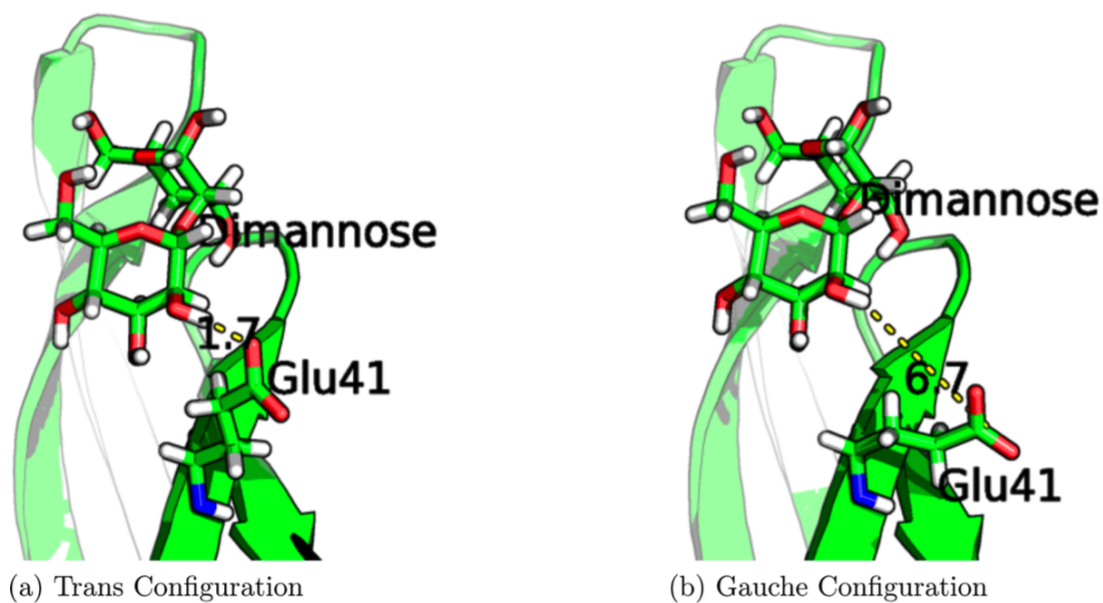


Figure 2.2: (a) χ_1 of Glu 41 is in the Trans conformation centered around 180° and the side chain of Glu41 is able to establish strong hydrogen bonding with 2'OH of the non-reducing end of dimannose.(b) χ_1 of Glu41 is in the gauche conformation centered around 60° and its side chain is unable to establish hydrogen bonds with 2'OH of the non-reducing end of dimannose. In both figures, the protein backbone is represented with a large green arrow.

Glu41 and 2' hydroxyl of the non-reducing end of dimannose as function of time during MD simulation when using ff99SB coupled with the Glycam06 force field for sugars. It is clear that except in a few instances no oxygen comes in hydrogen bonding contact with 2'OH. This is in contrast to what one finds in a simulation based on the OPLS-AA force field where all the time there is hydrogen bonding contact. This can be seen in Figure 2.4. This supports the idea put forth by Vorontsov and coworkers about the force field dependence of this result.[95]

Because in the monomeric form of CVN and its mutants $\chi 1$ of Glu41 has been shown experimentally to be almost exclusively in the trans conformation and because the results presented in Figure 2.3 and 2.4 indicate that of the two force fields probed only OPLS-AA force field[55, 56] reproduced this important experimental result, we have chosen to use only this force field to perform a set of detailed thermodynamic cycles to unravel the difference in free energy of binding between dimannose in complex with P51G-m4-CVN as compared to E41A as well as E41G mutants.

We used the BAR approach in order to estimate the relative binding free energy difference of dimannose when P51G-m4-CVN is mutated at position 41 to Ala or Gly. Whereas these numbers can be experimentally obtained, our goal is to also develop important intuition regarding the relevance of structurally derived information such as hydrogen bonding to the ligand in the bound state. Furthermore, this approach will allow us to dissect free energetic information into different interaction components such as charge and dispersion forces. The relevance of this type of analysis has been highlighted repeatedly since the early work by Karplus and collaborators.[24]

Hydrogen Bonds Between Glutamate and Dimannose AMBER99SB & GLYCAM06

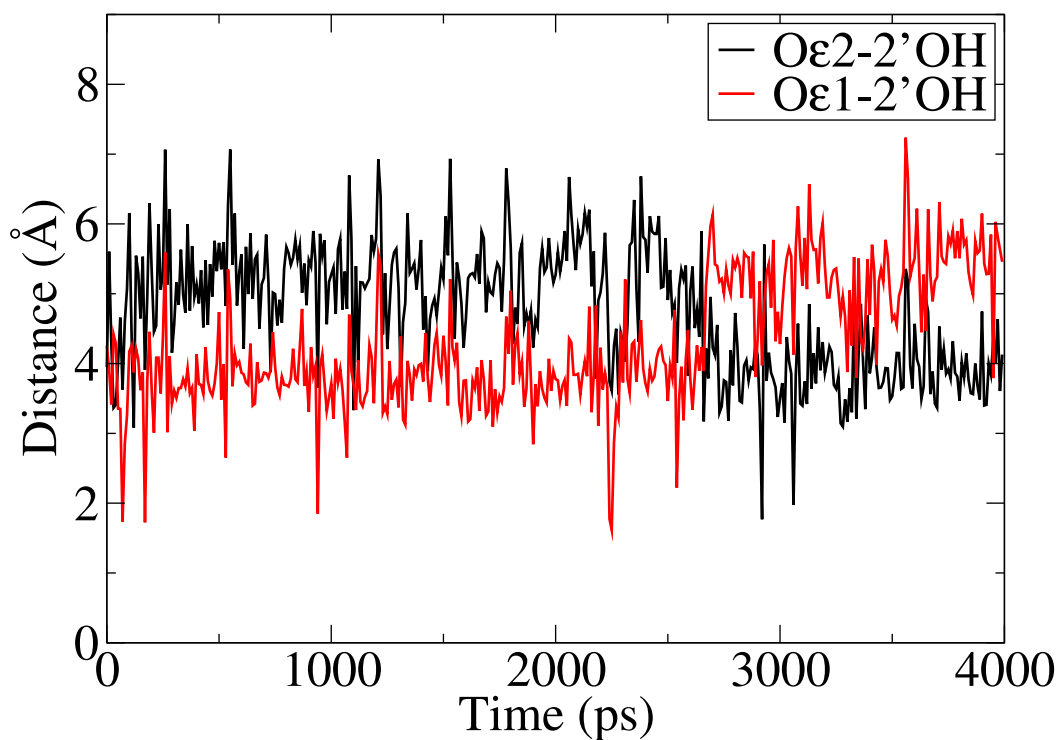


Figure 2.3: Hydrogen bond distance between carboxylate oxygens of side chain of glutamate and 2'OH of non-reducing end of dimannose using the AMBER11 software with the ff99SB force field for proteins and the GLYCAM06 force field for dimannose.

Hydrogen Bonds Between Glutamate and Dimannose OPLS-AA & OPLS-Carbohydrates

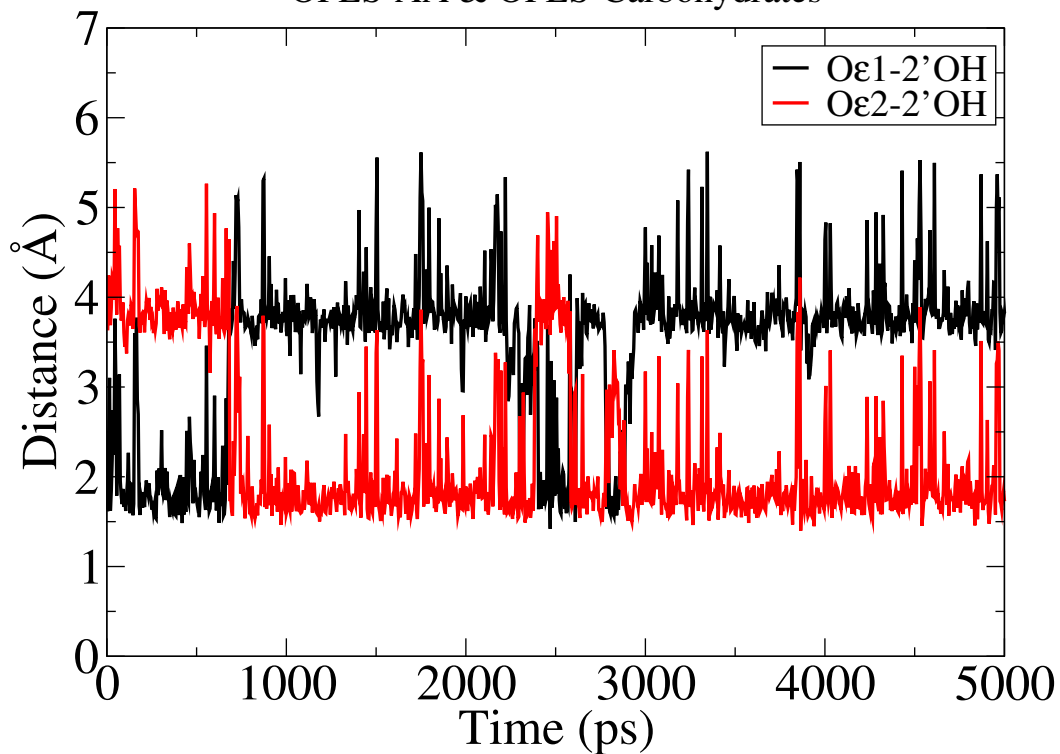


Figure 2.4: Hydrogen bond distance between carboxylate oxygens of side chain of glutamate and 2'OH of non-reducing end of dimannose using the Gromacs software suite coupled to the OPLS-AA force field for protein and the OPLS-Carbohydrates force field for dimannose

We performed three trials for this mutational analysis as stated in the Methods section, and as shown in Figure 2.1, each trial has three steps, each of the steps in presence of dimannose and absence of dimannose. Further each step has λ ranging from 0 to 1. For the E41A mutation, we have 82 simulations for each trial and each λ simulation runs for 2 ns resulting in 492 ns for all the trials. Because the Coulombic discharge of Glu41 is a common step in the thermodynamic cycle of both studied mutations, for the E41G mutation, we only needed 60 extra simulations for each trial, resulting in a total of 360 ns. Tables 2.1 and 2.2 summarize the results obtained for each step in the presence and absence of dimannose for both mutations. The dimannose is noted as Man_2 in the Tables 2.1 and 2.2; in the case of data, the error values in the parentheses are those obtained from the \bar{g} -bar utility available in the Gromacs package version 4.5.x which uses the Bennett Acceptance Ratio Method[11] to obtain the free energies. The errors reported for average ΔG are either derived from the \bar{g} -bar analysis or the standard deviation of the three trials, whichever is larger.

Table 2.1: BAR analysis for relative binding free energies for the Glu41Ala mutation

	Discharging of Glu41 (kcal/mol)	LJ Transformation (kcal/mol)	Negative of Discharging of Ala41 (kcal/mol)	Total ΔG (kcal/mol)
I in presence Man_2	177.15 (0.18)	-2.35 (0.17)	-65.77 (0.03)	
II in presence Man_2	177.60 (0.17)	-2.83 (0.19)	-65.71 (0.05)	
III in presence Man_2	177.28 (0.08)	-2.72 (0.10)	-65.70 (0.03)	
AVG ΔG	177.34 [0.23]	-2.64 [0.25]	-65.73 [0.04]	108.98 [0.34]
I in absence Man_2	175.24 (0.17)	-1.14 (0.08)	-65.23 (0.06)	
II in absence Man_2	175.22 (0.16)	-0.98 (0.15)	-65.38 (0.08)	
III in absence Man_2	175.22 (0.27)	-0.81 (0.11)	-65.24 (0.06)	
AVG ΔG	175.23 [0.21]	-0.98 [0.17]	-65.29 [0.08]	108.96 [0.28]
$\Delta \Delta G$	2.12 [0.31]	-1.66 [0.30]	-0.44 [0.09]	0.02 [0.44]

Table 2.2: BAR analysis for relative binding free energies for the Glu41Gly mutation

	Discharging of Glu41 (kcal/mol)	LJ Transformation (kcal/mol)	Negative of Discharging of Gly41 (kcal/mol)	Total ΔG (kcal/mol)
I in presence Man_2	177.15 (0.18)	-12.90 (0.13)	-64.97 (0.04)	
II in presence Man_2	177.60 (0.17)	-12.91 (0.27)	-65.03 (0.04)	
III in presence Man_2	177.28 (0.08)	-13.24 (0.15)	-64.93 (0.07)	
AVG ΔG	177.34 [0.23]	-13.02 [0.22]	-64.98 [0.05]	99.35 [0.32]
I in absence Man_2	175.24 (0.17)	-10.98 (0.29)	-64.33 (0.03)	
II in absence Man_2	175.22 (0.16)	-10.88 (0.21)	-64.24 (0.01)	
III in absence Man_2	175.22 (0.27)	-10.83 (0.31)	-64.32 (0.02)	
AVG ΔG	175.23 [0.21]	-10.90 [0.27]	-64.29 [0.05]	100.04 [0.35]
$\Delta \Delta G$	2.12 [0.31]	-2.12 [0.35]	-0.68 [0.07]	-0.69 [0.47]

One could predict, and this is indeed the case from simulation results shown in Tables 2.1 and 2.2, that a very large free energetic penalty is incurred upon Coulombic discharging of Glu41, a surface charged residue in intimate contact with dimannose. This penalty, which is on the order of 177 kcal/mol, is the result of interactions lost

with the sugar and the environment. The reader is cautioned against directly comparing the absolute discharging free energy in the case of Ala41 or Gly41 against that of Glu41 while using PME. Because the overall charge (protein plus ions) of our system is different in the case of P51G-m4-CVN and the two mutants, such direct comparison is problematic. There is no problem, however, when using a thermodynamic cycle because the same overall charge exists in each case for simulations in the presence and absence of ligand. With this caveat in mind, it is still clear that the discharge of Ala or Gly in the presence of dimannose results in a penalty that is much smaller than required to discharge Glu. It is therefore quite reasonable to expect NMR and x-ray experiments as well as simulations to display strong hydrogen bonding between Glu41 and the 2OH' of dimannose non-reducing end. However, what is often missed in the type of qualitative analysis that focuses only on interactions in the ligand-bound state is that the relative binding free energy is not only indicative of the strength of interactions in the ligand-bound state but also of the free energetics of the ligand-free transformation. What we mean by this is that whereas the charge transformation portion of ΔG_1 (the free energy of discharging Glu41 in the presence of dimannose) is a large positive number so is the discharging portion of ΔG_2 (the same process but in the absence of ligand). The discharging of Glu41 is very unfavorable as this is a surface charged residue exposed either to the ligand or in the absence of ligand to the solvent. Whereas Coulomb interactions between Glu41 and 2OH' can indeed be very strong in the bound state[70], so are the interactions of Glu41 and its solvent environment in the absence of ligand, and this is the cause of large cancellations when

computing $\Delta \Delta G$.

It is therefore very important that experimental hydrogen bonding contacts between dimannose and the binding site be considered relevant for free energetics only after careful scrutiny of the same type of interactions in the absence of ligand.

Our study predicts that the difference in free energy penalties for the discharge Glu41 in the presence and absence of dimannose is on the order of 2.1 kcal/mol. To be more specific, whereas it costs about 177 kcal/mol to discharge Glu41 in the presence of the sugar, it only costs about 175 kcal/mol to do this in the absence of ligand (see table 2.1 and Figures 2.5a and 2.6a).

Furthermore, the overall Coulomb portion of $(\Delta G_1 - \Delta G_2)$ is on the order of 1.7 kcal/mol in the case of the E41A mutation and 1.4 kcal/mol in the case of the E41G mutation (see tables 2.1 and 2.2). This is because the difference in discharging free energies for Ala41 or Gly41 in the presence and in the absence of sugar are on the order of 0.4 and 0.7 kcal/mol, respectively. Because the Coulombic $\Delta \Delta G$ for the discharging of Glu41 in the presence and absence of dimannose is significantly larger than the same free energy difference in the case of Ala and Gly, we conclude that Glu41 stabilizes the ligand from a Coulomb perspective.

What one learns from this portion of the analysis is that in the case of the mutation of the charged surface amino acid Glu41 to Ala41 or to Gly41, the Coulombic portion of the relative free energy of binding arises mostly due to the extra stabilization of Glu41 in the presence of the disaccharide as compared to Glu41 in the presence of solvent at the binding site. This is demonstrated in Figures 2.5a, 2.5c , 2.6a, 2.6c,

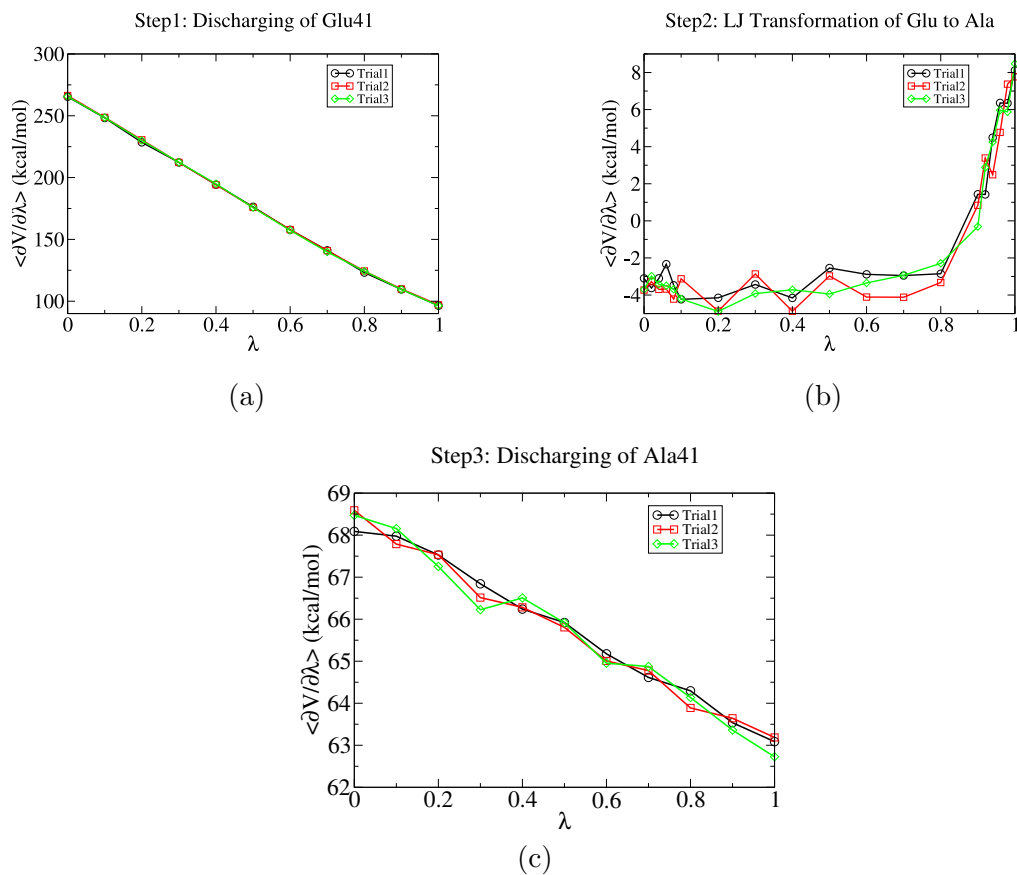


Figure 2.5: Three-step approach for thermodynamic integration simulations for Glu41Ala in presence of dimannose, trial I (black circles), trial II (red squares) and trial III (green diamonds). Discharging of Glutamate (a), LJ transformation (b) and discharging of Alanine (c).

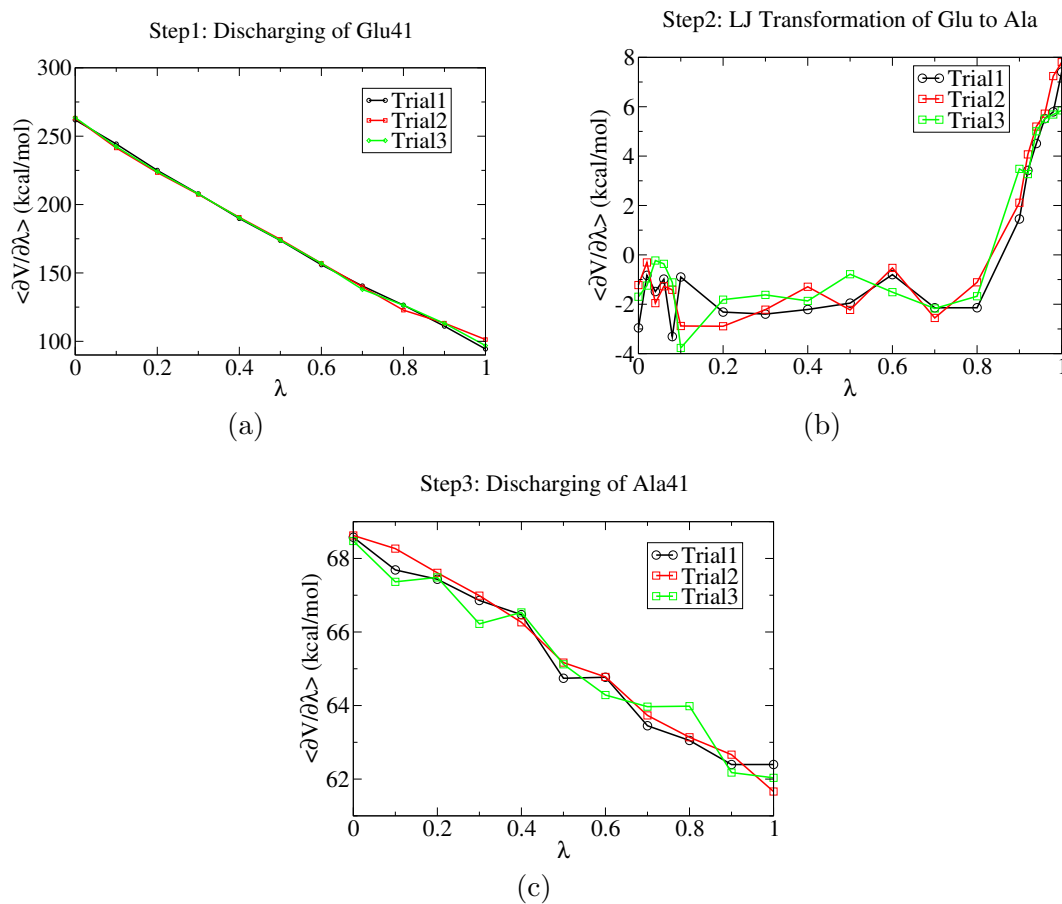


Figure 2.6: Three-step approach for thermodynamic integration simulations for Glu41Ala in absence of dimannose, trial I (black circles), trial II (red squares) and trial III (green diamonds). Discharging of Glutamate (a), LJ transformation (b) and discharging of Alanine (c).

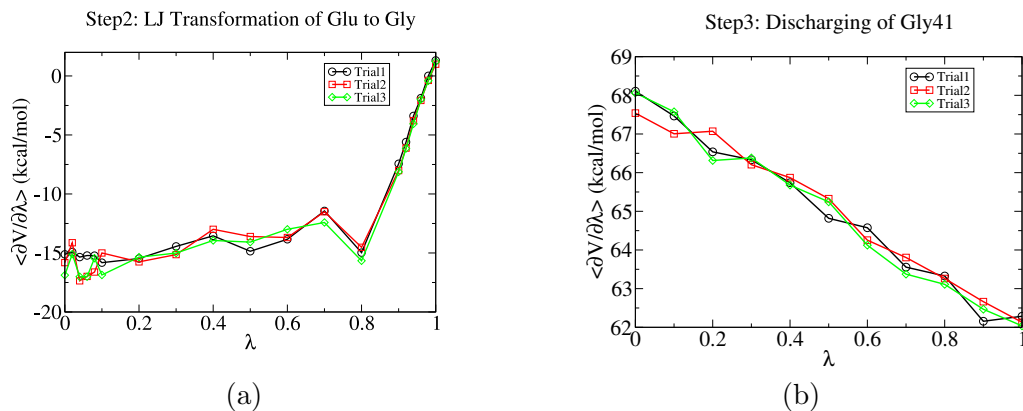


Figure 2.7: Three-step approach for Thermodynamic Integration simulations for Glu41Gly in the presence of dimannose, trial I (black circles), trial II (red squares) and trial III (green diamonds). LJ transformation (a) and discharging of Glycine (b).

2.7b, and 2.8b as well as in Tables 2.1 and 2.2. If 1.7 kcal/mol was the binding free energy difference between the protein and dimannose when Glu41 is mutated to Ala or Gly, this would be quite significant. However further cancellations occur due to dispersion interactions stabilization. A Gly side chain is smaller than an Ala side chain which in turn is smaller than a Glu side chain. In the particular case of binding to dimannose, the mutation of Glu41 into smaller residues appears to be free energetically favorable with respect to dispersion interactions. We find that from a dispersion interaction perspective Gly is better than Ala which in turns is better than Glu. This can be appreciated from Figures 2.5b, 2.6b, 2.7a and 2.8a and Tables 2.1 and 2.2. In the case of the E41A mutation, the dispersion contributions significantly diminish the

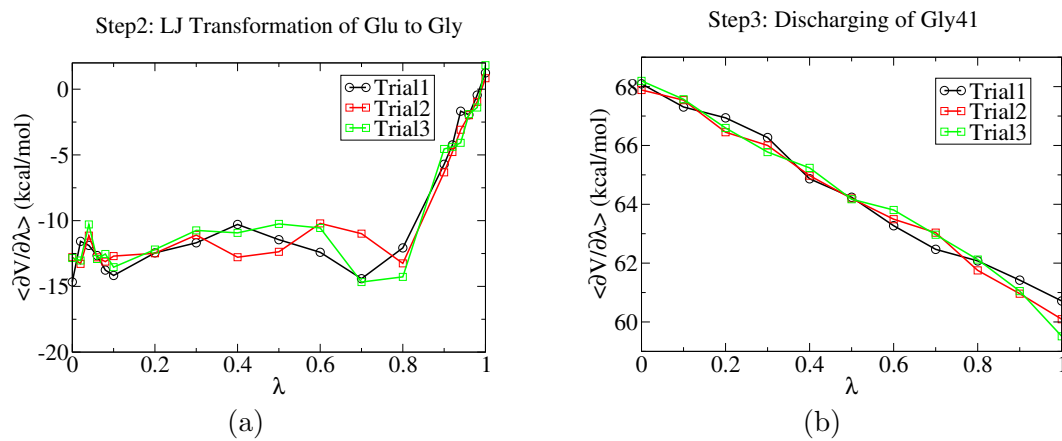


Figure 2.8: Three-step approach for thermodynamic integration simulations for Glu41Gly in the absence of Dimannose, trial I (black circles), trial II (red squares) and trial III (green diamonds). LJ transformation (a) and discharging of Glycine (b).

energetic cost of the mutation and make the mutation E41G slightly favorable. This result could be seen as somewhat unexpected based purely on structural experimental conjectures or in light of previous molecular dynamics simulations[70] showing very large electrostatic interaction between Glu41 in wild type and the disaccharide but it is not. This is because for the interpretation of free energetics one should also consider the interactions with solvent in the ligand-free state.

Recently, the Ghirlanda group has studied the binding affinity of these mutations experimentally.[22] Along with the recent experimental data for the binding affinity of P51G-m4-CVN[67], the free energy changes of E41A and E41G of P51G-m4-CVN determined by experiments are 0.78 and 0.59 kcal/mol at room temperature, respectively. These results indicated that E41G binds tighter than E41A, which is in agreement with our simulation results. Even though the trend in our computational results is correct, the free energy for mutations E41A and E41G are 0.02 and -0.69 kcal/mol respectively indicating a shift of about 1 kcal/mol in the computational results as compared to experiments. This renders the result for E41G negative as opposed to the experimental value which is positive. There are many factors that can contribute to this energy shift including the accuracy of the force field. In general, differences on the order of 1 kcal/mol in these type of calculations are expected and considered acceptable.

2.4 Conclusions

This chapter describes a detailed analysis of the importance of Glu41 to the relative binding free energy of dimannose to P51G-m4-CVN and two mutants at position 41. Several interesting findings arise from this work. First, from a purely technical perspective, there appears to be some force field dependence on the most likely conformation of the χ_1 angle of the side chain of Glu41 that impacts whether this residue is in close proximity of 2'OH of the non-reducing end of dimannose. Experimental evidence appears to indicate that, at least in the monomeric form, χ_1 should be in most cases in the trans conformation. Second, we find that although it is true that Glu41 strongly hydrogen bonds to 2'OH of the non-reducing end of dimannose, this does not necessarily result in a very large relative free energy of binding penalty for the mutation of Glu41 into Ala41 or Gly41. Coulomb interactions of Glu41 in the presence of dimannose are very strong, but so are they in the absence of ligand. Replacing the Coulombic interactions between dimannose and a charged Glu41 costs on the order of 1.7 kcal/mol in the case of E41A and 1.4 kcal/mol in the case of E41G. However, these numbers are greatly diminished by favorable dispersion free energy changes. Briefly, the cost of losing hydrogen bond interactions between the ligand and Glu41 is compensated by better dispersion interactions with Ala or Gly.

In more general terms, we learn that it is perhaps unwise to assume that what appears to be a strong hydrogen bond during molecular dynamics simulations, NMR experiments or X-ray crystallography will necessarily translate into high importance

interactions for relative free energetics of binding.

CHAPTER 3

COMPUTATIONALLY DESIGNED MUTATIONS: AN ATTEMPT AT IMPROVING THE BINDING AFFINITY OF CVN TO DIMANNOSE

3.1 Introduction

Computer-aided drug design techniques have become more sound because of the improved efficiency and speed of MD simulation packages,[50, 83, 30, 29] the development of force field, [55, 56, 34, 94, 60] and the accuracy of algorithms for free energy estimation.[2, 53, 64, 65, 88, 97] Compared to less expensive but also less accurate implicit solvent MD simulations,[40] and coarse-grained MD simulations,[71] all-atom MD studies generate high-resolution trajectories where all components are represented explicitly.[61]

Our interest is in CVN, which has potent activity against HIV because of a significant binding affinity towards $\text{Man}\alpha 1 \rightarrow 2\text{Man}\alpha$ moieties on the viral surface. [37, 38, 77, 19] In order to improve the antiviral activity of CVN, one reasonable approach is to increase its binding affinity against dimannose. Other approaches may involve multivalency. As shown in Figure 3.1, experiments have identify Glu41, Ser52, Asn53, Glu56, Thr57, Lys74, Thr75, Arg76, and Gln78 as important residues for binding.[15] Therefore, these residues constitute a good starting pool for investigation, as in binding they appear to directly interact with the ligand. We were also initially interested in vicinal residues such as Val43 which, although may not directly interact in binding, may have indirect structural influence with adjacent residues. In order to better focus on the high affinity domain, we worked with the P51G-

m4-CVN[42] mutant. Many mutations are possible considering the aforementioned aminoacids. However carrying out all of these was prohibitive and we therefore focused on the Glu56Asp, Ser52Thr, and Thr57Ser mutations. Exploratory studies were also carried out in the case of Val43Ala, Glu41Gln and a couple of dual mutations but not all of these were fully completed. These exploratory studies revealed significant complexity in the sampling of phase space particularly for discharged intermediate states. For example in the case of Val43Ala the discharge of Val43 could not be easily converged. The reduced charge intermediate has two very different possible conformations that cannot be properly sampled within the two nanoseconds allocated to each intermediate simulation. Whereas a 2 ns run may not seem like a large computational effort, a single trial of the full thermodynamic cycle requires an effort two orders of magnitude larger. I expand on this point later in this chapter. From an exploratory MD study of domain B in the presence of dimannose in which Glu41 was replaced by Gln we also learned that the new residue could sample three different conformers. Based on this I chose to abandon this particular mutation as our protocol would fail at sampling such complex system. In the case of the dual mutation Ser52Thr-Thr57Ser, a single full thermodynamic cycle was carried out but the outcome did not match unpublished data from our collaborators and it is therefore not included in this chapter. Whereas it would be interesting to look at the simulations in this chapter in the context of experiments, at the time of writing this thesis, data collected by our collaborators in the group of Ghirlanda at Arizona State University remain unpublished and will therefore not be introduced. I nonetheless

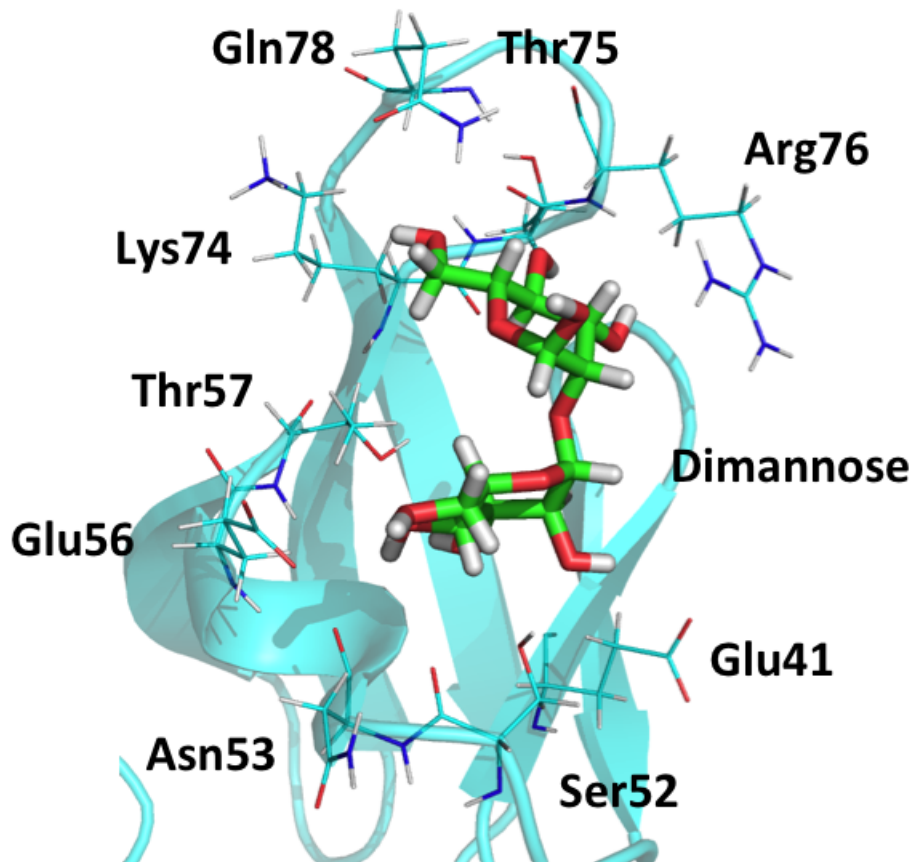


Figure 3.1: Key residues of wt-CVN in domain B for binding to dimannose.

believe that the discussion based only on my computational results is self contained and reveals important insights that can be useful for possible future studies.

3.2 Methods

3.2.1 System Preparation and Free Energy Calculation

The methods in this chapter are similar to those described in subsection 2.2. The coordinates of the mutations in the presence and absence of dimannose, including Glu56Asp, Ser52Thr and Thr57Ser, were generated from the P51G-m4-CVN bound

state using Pymol.[36] The mutants were centered in a box of dimensions 78 X 78 X 78 Å³ along with 15103 TIP3P[54] explicit water molecules. The OPLS-AA force field for proteins [55, 56] and OPLS force field for carbohydrates[34] with modifications (see Appendix A) were used in simulations with the software Gromacs version 4.5.5.[50] As none of the mutations had a net charge change, 3 K^+ ions were used to counter balance the protein charge.

The systems both in the presence and absence of dimannose were first minimized for 5000 steps using the steepest-descent algorithm followed by a 100 ps NVT run and a 5 ns NPT equilibration at 300 K. In the case of the end state for the E41Q mutation an extra 15 ns were run to better understand the conformational variability of Gln. Long-range electrostatic interactions were calculated using the PME protocol[35] (cutoff 10 Å , grid size 1 Å , and order of 6).

A thermodynamic cycle was used to estimate the binding free energy change as previously described in subsection 2.2 using the single topology approach.[81] The same λ distribution was used because the plots of $\langle \partial V / \partial \lambda \rangle$ appeared appropriately converged. There were 11 λ Coulombic discharge steps allocated evenly from 0 to 1. For the LJ transformation there were 6 evenly distributed λ values from 0 to 0.1, 7 evenly distributed λ values between 0.2 to 0.8, and 6 evenly distributed λ values from 0.9 to 1 respectively. In order to minimize the number of transformations required and to attempt an optimal convergence, the Coulombic discharging pathway and LJ transformation pathway were taken to be different compared to the approach in chapter 2. The current approach only discharges the side chain of the target residue;

whereas in chapter 2 we discharged the entire residue including the backbone of that residue. In terms of the LJ transformation, we assigned the dihedrals of the dummy particles to zero for Thr57Ser and Ser52Thr mutations, however, in chapter 2 and the case of Glu56Asp mutation, the dihedrals of dummy particles were kept as those in the original system. Because we worked with a full thermodynamic cycle, the final free energy change is still comparable but the energy change for each step is not comparable due to the different strategies. Moreover, the discharging energies in this chapter should be smaller than that in chapter 2 because the backbones were kept intact. If a mutation required changing a smaller amino acid into a bigger one, we avoided doing so by conducting the reverse LJ transformation. As an example, in the case of Ser52Thr, although the final state is Thr52, we instead conducted the LJ transformation from Thr52 to Ser. Because the free energy is a state function, the free energy of transforming Ser52 to Thr should be the negative of transforming Thr52 to Ser. We avoid growing new atoms in the system and conduct reverse LJ transformation because growing new atoms may lead to clashes due to the overlap of atoms.

For each λ simulation, the equilibration processes included a 5000-step energy minimization, a 100 ps NVT run, and a 200 ps NPT run. After equilibration, a 2 ns NPT production simulation was then generated. Because the goal in this chapter was exploratory and not the actual computation of highly accurate free energy changes we only conducted two trials of the thermodynamic cycle in the case of Thr57Ser and one for all other systems. The Bennett Acceptance Ratio method [11] and thermodynamic

integration[90, 91] were utilized to analyze the data.

3.2.2 Replica Exchange Molecular Dynamics

In cases, when conformations were harder to sample, such as at position 52, the more powerful and more expensive Replica Exchange Molecular Dynamics (REMD) technique was used.[92] One particularly problematic issue that we wanted to surmount was associated with the conformational sampling during Coulombic discharge of Thr52 in the presence of dimannose.

REMD enhances the sampling of a system at a given temperature by allowing exchanges of configurations with replicas sampled at higher temperature. We used 12 temperature replicas for each λ during the Coulombic discharging process. The highest temperature was chosen to be about 330 K because it is known[7] that around this temperature the likelihood of large conformational changes resulting in the formation of domain swapped dimer is significant in the case of CVN. Even though this is much less likely for the P51G-m4-CVN mutant, we did not want to introduce large conformational fluctuations in our study. The temperatures for replica exchange were then chosen to be 300.00, 302.55, 305.15, 307.73, 310.31, 312.91, 315.54, 318.17, 320.82, 323.48, 326.15, and 328.87 K. These temperatures were generated by the on-line temperature generator for REMD-simulations[79] based on the size of the system and exchange probability between replicas, which was set to 0.01. To see how such exchanges take place during simulation please see Figure 3.9.

3.3 Results

3.3.1 Thr57Ser

We first discuss the Thr57Ser mutation for which sampling issues are minor compared to other cases. In chapter 2 we noticed that a shorter side chain results in a free energy benefit from dispersion interactions. If this observation was to hold in the current case, the Thr57Ser mutation should be beneficial in terms of binding to dimannose.

The results from calculations are shown in Figure 3.6 and Figure 3.7. A summary is also provided in Table 3.1. One can see from Figure 3.6a and Figure 3.6c that there is a sharp drop when λ gets close to 1. This sharp drop is likely related to the loss of hydrogen bonding ability of the hydroxyl function group of the side chain of residue 57. This discharged hydroxyl group breaks the hydrogen bonding network, which was reported to be mainly preserved by Thr57.[96] When λ equals 1, the partial charge of the side chain of residue 57 was set to zero, and the electrostatic interactions between the side chain of residue 57 and vicinal residues including dimannose was completely abolished. As can be seen in Figure 3.2, the charge in Thr57 is very important to maintain the hydrogen bonding network. Figure 3.3 shows a plot of hydrogen bond distance between Thr57 and dimannose (as depicted in Figure 3.2) against λ . The drop in free energy in Figure 3.6a matches the increase of the H-bond length in Figures 3.3. The same type of behavior is observed in the case of Ser57 in the presence of sugar. We therefore conclude that the fast free energy drop in the last few λ points can be ascribed to a penalty for the disruption of the hydrogen bond

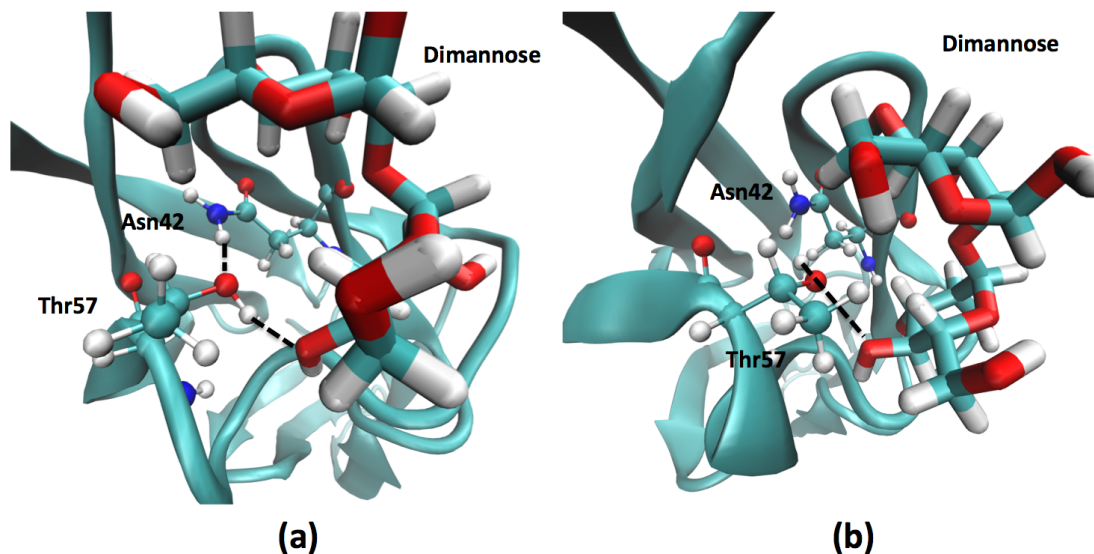


Figure 3.2: (a) Thr57 forming hydrogen bonds with vicinal residues including dimannose and Asn42 in the presence of dimannose. (b) Discharged Thr57 side chain resulting in a broken hydrogen bonding network.

network.

As shown in Table 3.1, the negative of Coulombic discharging Ser57 in the absence of dimannose is a positive quantity. In other words, the partial charge of the side chain of Ser57 is not preferred by the system in the absence of dimannose. Interestingly, different from Ser57, the energy for discharging Thr57 is positive and one can therefore conclude that the partial charge of Thr57 stabilizes the system. Because the only structural difference between Thr and Ser is the β -methyl group, the flipping sign in the free energy for discharging of Thr57 and Ser57 requires more analysis and discussion. One plausible explanation has to do with interactions between the binding pocket and the explicit solvent as shown in Figure 3.4. In Figure 3.5b, when there

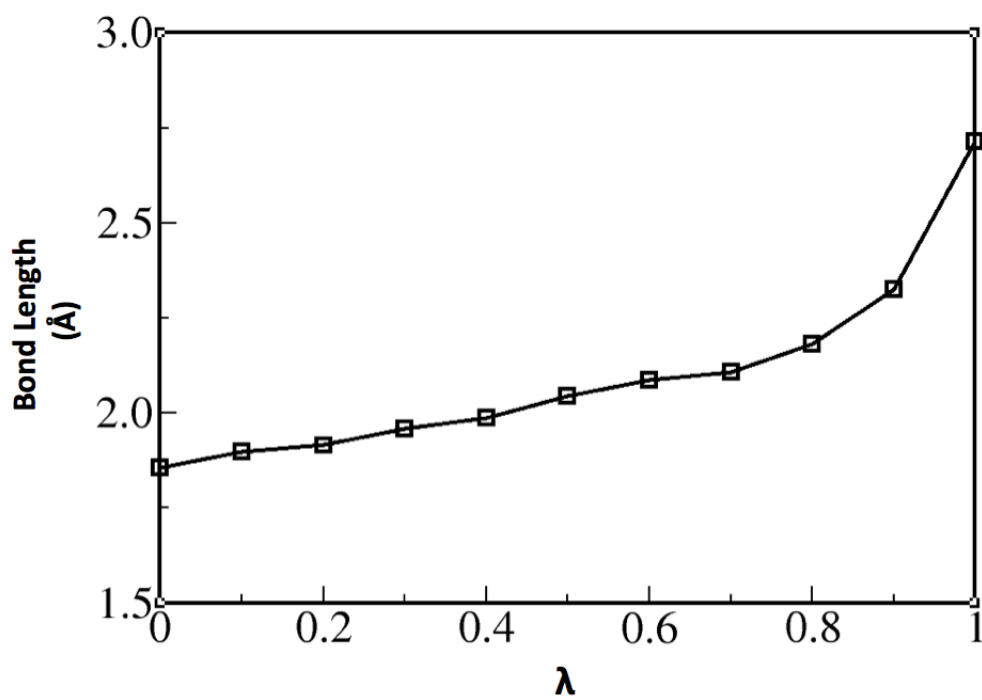


Figure 3.3: Hydrogen bond length between Thr57 and dimannose as a function of λ when discharging the side chain of Thr57.

is a methyl group, the hydroxyl group will always be close to the internal portion of the binding pocket and block water molecules from forming hydrogen bonds with neighboring residues regardless of the partial charge of the hydroxyl group. Instead, in the case of Ser57 shown in Figure 3.5a, when partial charge is diminished, the hydrogen bonds of Ser57 break followed by water molecules pushing the uncharged hydroxyl group away and forming hydrogen bonds with the binding pocket. Therefore, in the case of Ser57, the system prefers Serine without partial charges; whereas in the case of Thr57, the partial charge of Threonine is energetically preferable. As one can see in Table 3.1, the dispersion contributions diminish the energetic cost of the mutation, which is the same as what we observed in the mutation of Glu41Gly and Glu41Ala in chapter 2.

Table 3.1: BAR analysis for relative binding free energies for the Thr57Ser mutation

	Discharging of Thr57 (kcal/mol)	LJ Transformation (kcal/mol)	Negative of Discharging of Ser57 (kcal/mol)	Total ΔG (kcal/mol)
I in presence Man_2	7.00 (0.05)	-1.69 (0.15)	-3.79 (0.03)	
II in presence Man_2	7.04 (0.06)	-2.09 (0.23)	-3.78 (0.07)	
AVG ΔG	7.02 [0.08]	-1.89 [0.27]	-3.78 [0.08]	1.35 [0.29]
I in absence Man_2	3.11 (0.05)	0.09 (0.15)	1.42 (0.22)	
II in absence Man_2	3.06 (0.06)	0.24 (0.09)	1.64 (0.24)	
AVG ΔG	3.08 [0.08]	0.16 [0.17]	1.53 [0.33]	4.77 [0.38]
$\Delta \Delta G$	3.94 [0.11]	-2.05 [0.32]	-5.31 [0.34]	-3.42 [0.48]

The $\langle \partial V / \partial \lambda \rangle$ results are shown in Figure 3.6 and 3.7, and the data is summarized in Table 3.1. As one can see from Table 3.1, the final predicted free energy change for the mutation Thr57Ser is -3.42 kcal/mol which implies a 300 fold improve-

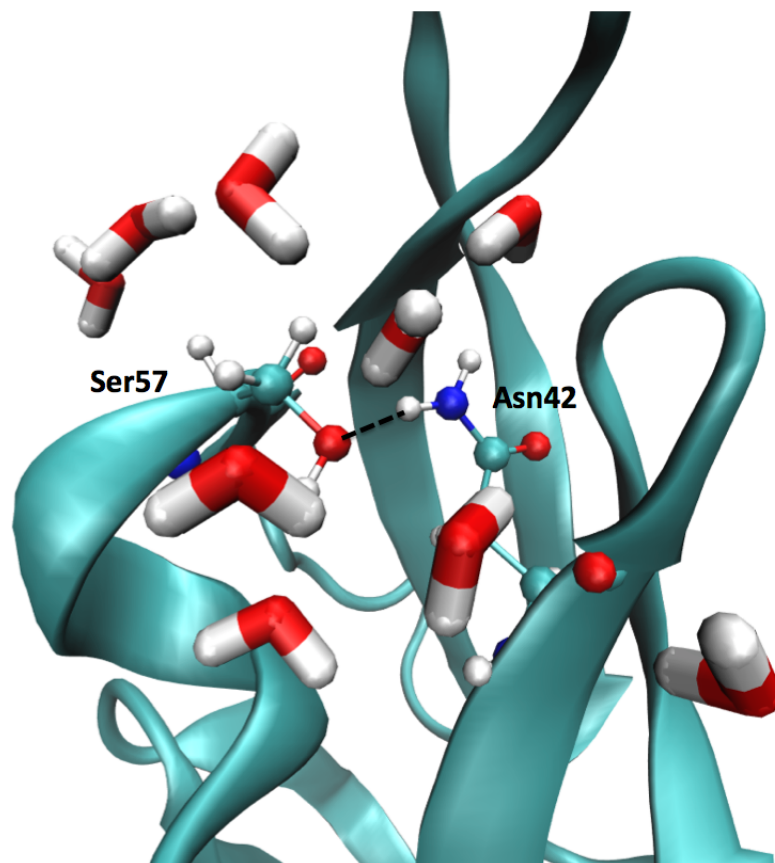


Figure 3.4: Hydrogen bond between Ser57 and Asn42.

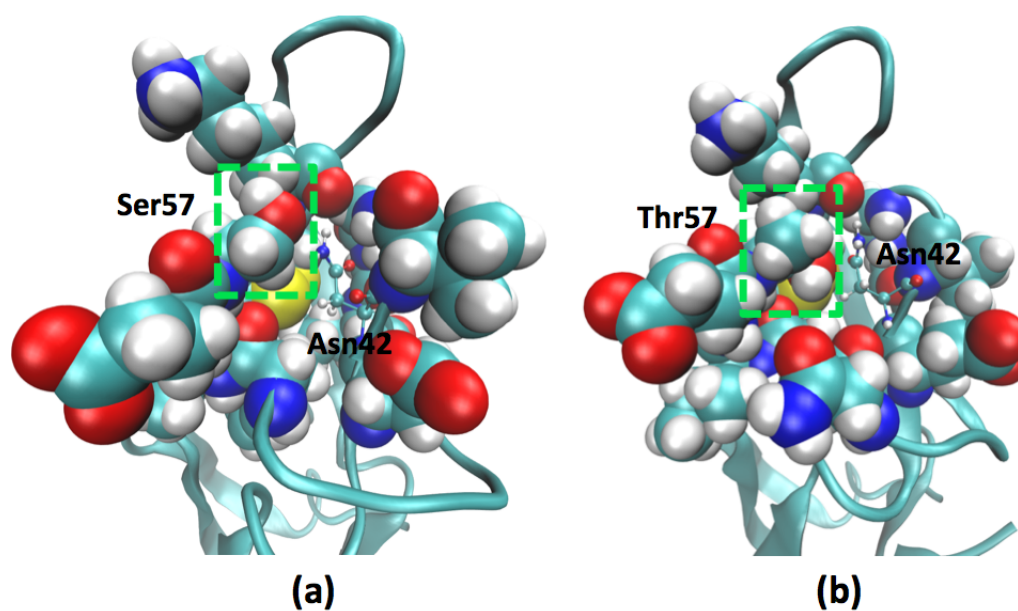


Figure 3.5: (a) Binding pocket with discharged Ser57 in the absence of dimannose
(b) Binding pocket with discharged Thr57 in the absence of dimannose. The green boxes highlight the side chain of the residue 57, and the Asn42 is represented by ball and stick model.

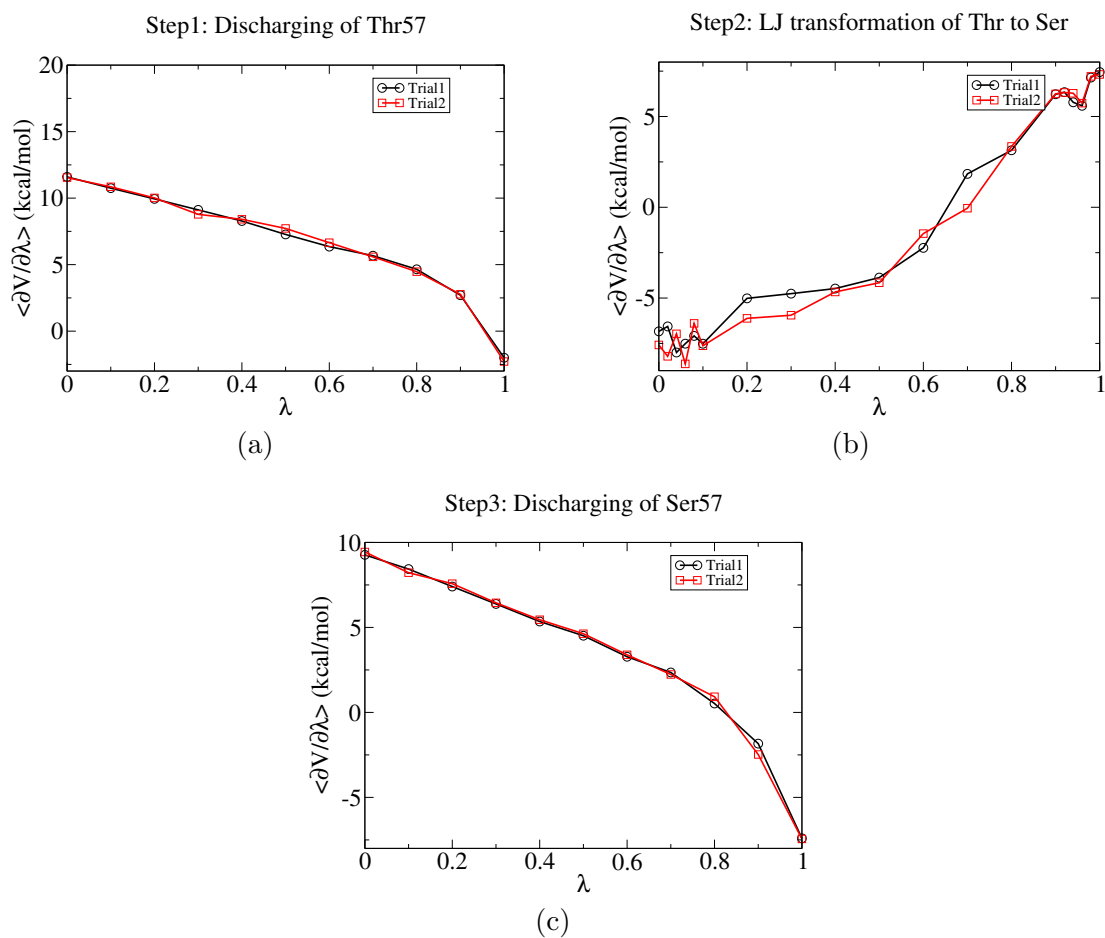


Figure 3.6: Three-step approach for thermodynamic integration simulations for Thr57Ser in presence of dimannose. Trial I (black circles), trial II (red squares). Discharging of Threonine (a), LJ transformation (b) and discharging of Serine (c).

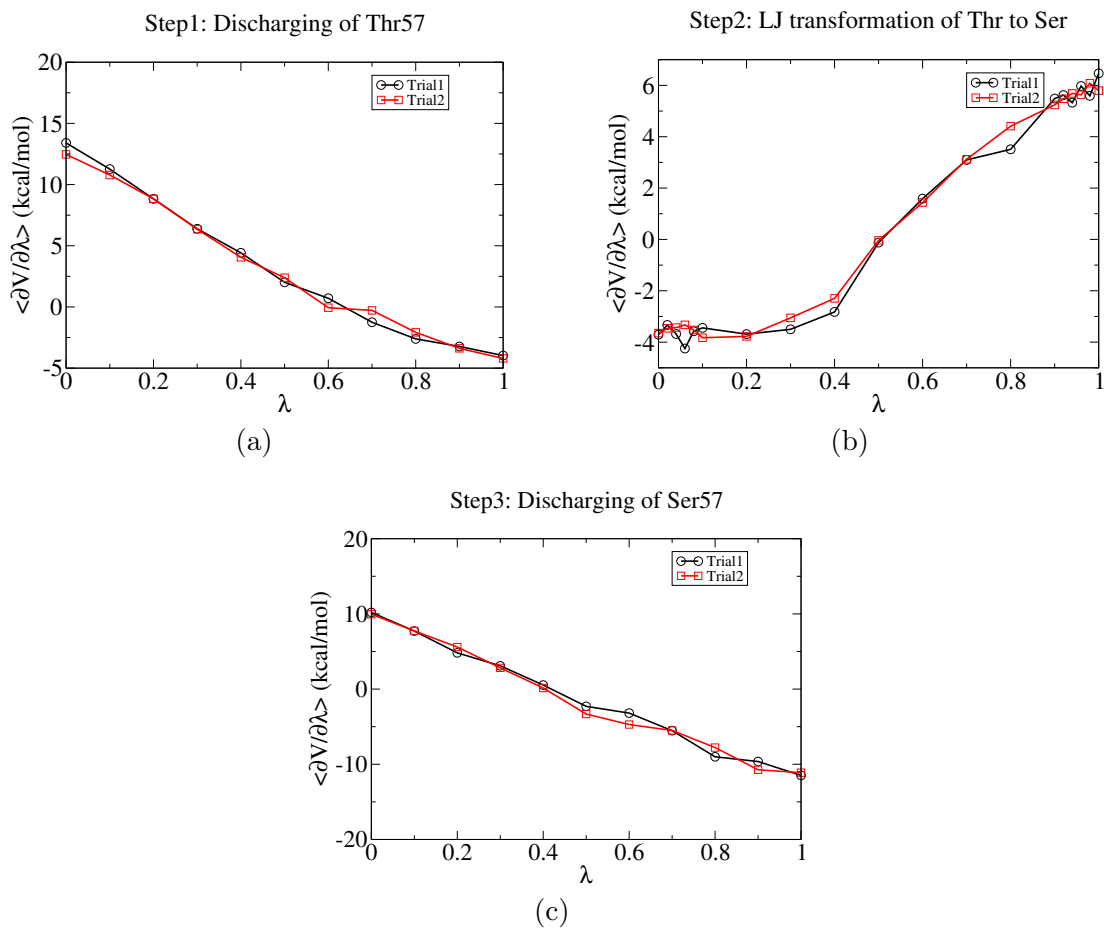


Figure 3.7: Three-step approach for thermodynamic integration simulations for Thr57Ser in absence of dimannose. Trial I (black circles) and trial II (red squares). Discharging of Threonine (a), LJ transformation (b) and discharging of Serine (c).

ment in the binding affinity of P51G-m4-CVN against dimannose. Even if this large enhancement in binding ability resulting from a single mutation may be questionable, the trend is certainly encouraging.

3.3.2 Ser52Thr

We mutated Ser52 to Thr. The thought was that because the backbone carbonyl oxygen of residue 52 forms hydrogen bond with dimannose, the interaction between solvent molecules and the hydrophobic methyl group of Thr, which points away from the binding pocket, may push the carbonyl oxygen towards the binding pocket, and therefore enhance hydrogen bonds between the carbonyl oxygen and dimannose. (See Figure 3.8) From the $\langle \partial V / \partial \lambda \rangle$ graph depicted in Figure 3.10 using the simple MD approach we conclude that convergence of Ser52Thr free energy calculation is questionable and requires improvement. This is because in the discharged state, Thr52 has two different conformations. When discharging Thr52 in the presence of dimannose (Figure 3.10c), the simple MD protocol fails to converge but results from the REMD procedure appear to be significantly improved. As one can appreciate from Figure 3.9, sufficient exchange between replicas is observed.

The sampling problem is believed to be relevant to conformational changes of the mutated amino acid in this case. The discharge of Thr52 in the presence of diamnrose ($\lambda = 1$), results in the sampling of two different values for $\partial V / \partial \lambda$. This can be seen from Figure 3.12. The value of -5 kcal/mol of $\partial V / \partial \lambda$ corresponds to the methyl group pointing down in Figure 3.13a, and the value of -10 kcal/mol

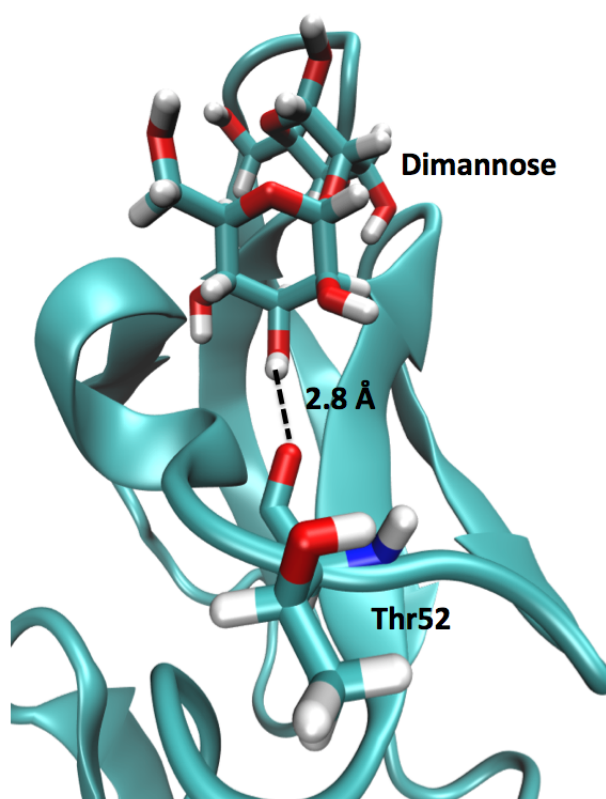


Figure 3.8: Hydrogen bond formed between Thr52 and dimannose

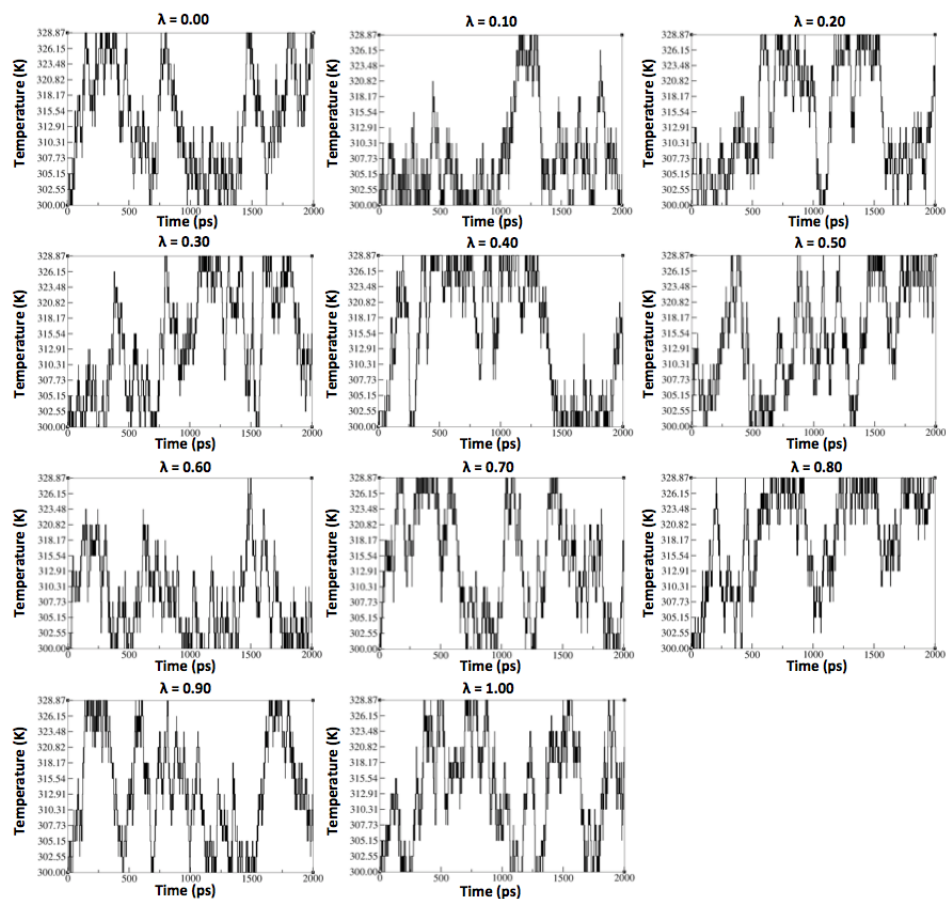


Figure 3.9: Replica exchange rate in the temperature space for each λ starting at 300 K when discharging Thr52 in the presence of dimannose. Replica exchange rates starting at other temperatures are not shown for clarity and the exchanges are sufficient. The lowest temperature is 300.00 K and the highest is 328.87 K. There are 10 more temperatures (302.55, 305.15, 307.73, 310.31, 312.91, 315.54, 318.17, 320.82, 323.48, and 326.15 K) in between these two.

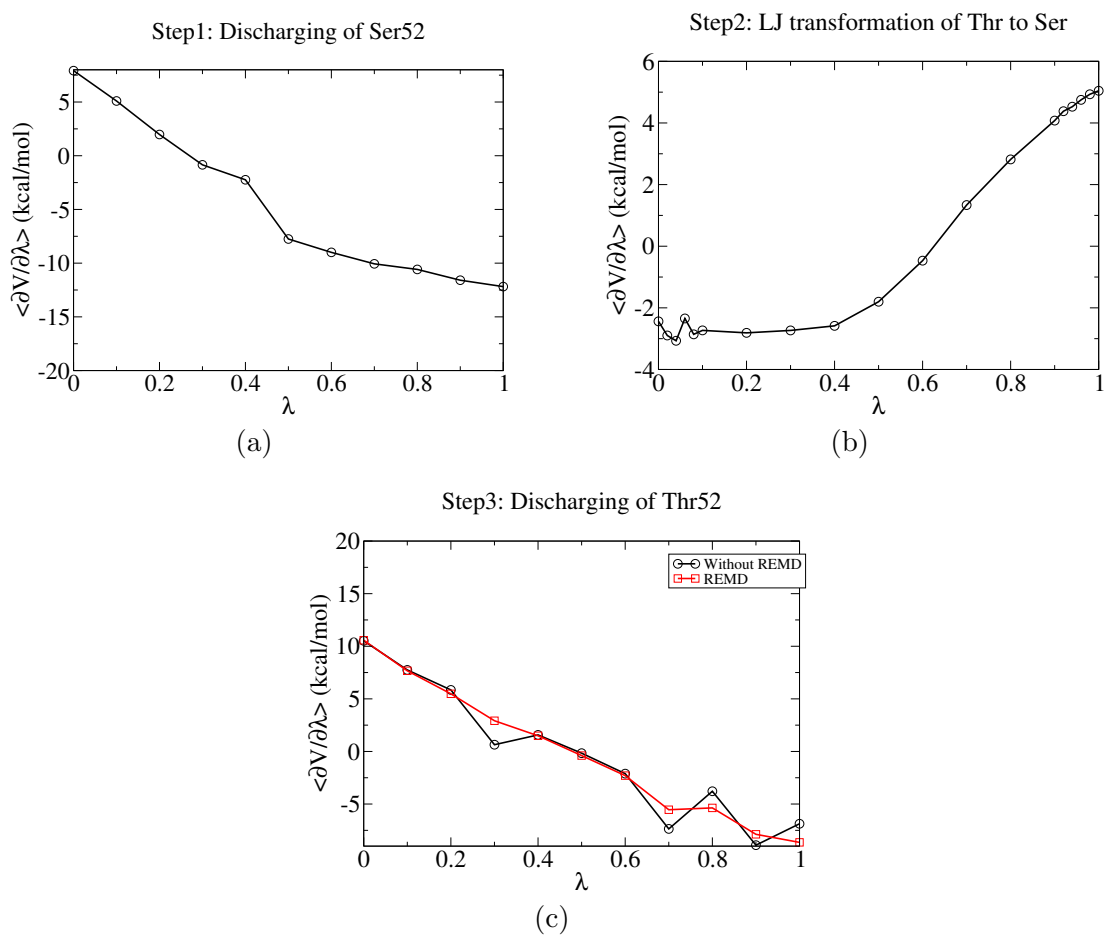


Figure 3.10: Three-step approach for thermodynamic integration simulations for Ser52Thr in presence of dimannose. Discharging of Serine (a), LJ transformation (b) and discharging of Threonine (c) without REMD (black circles) and with REMD (red squares).

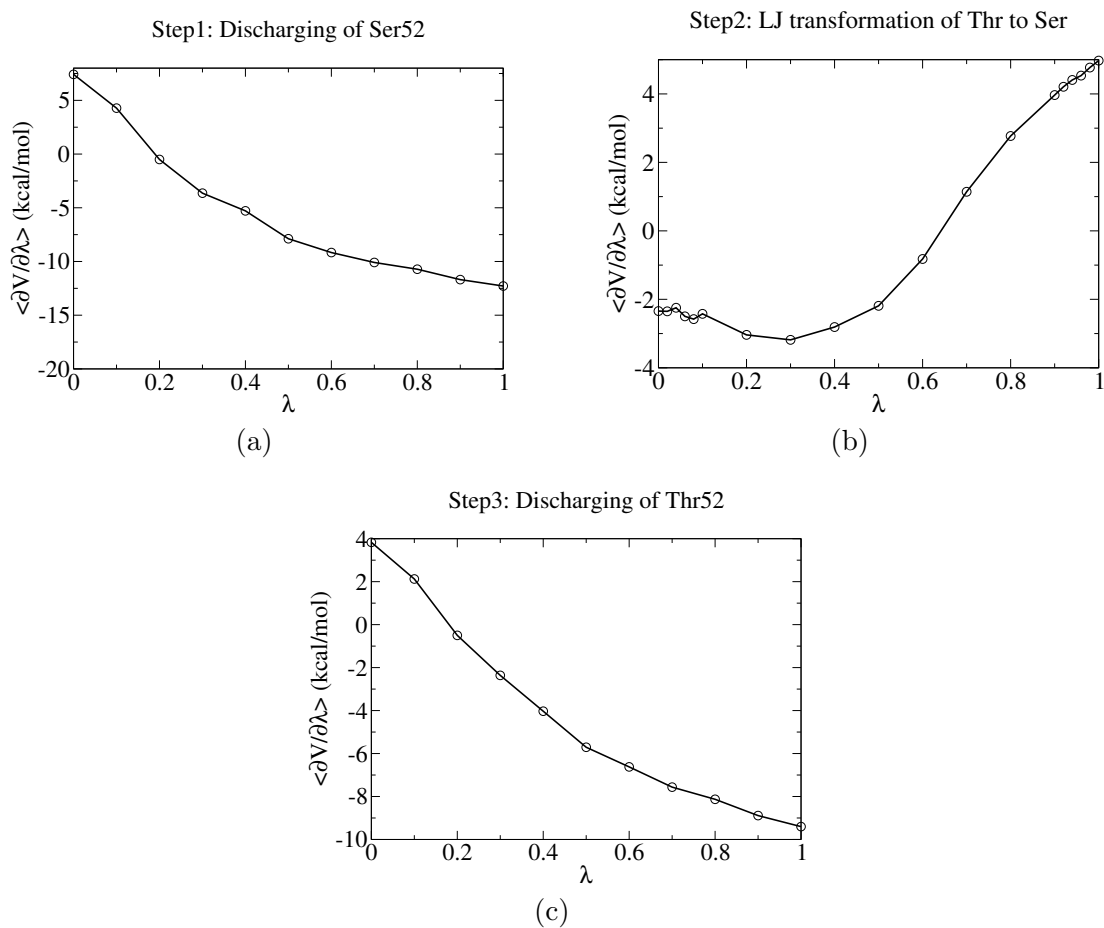


Figure 3.11: Three-step approach for thermodynamic integration simulations for Ser52Thr in absence of dimannose. Discharging of Serine (a), LJ transformation (b) and discharging of Threonine (c).

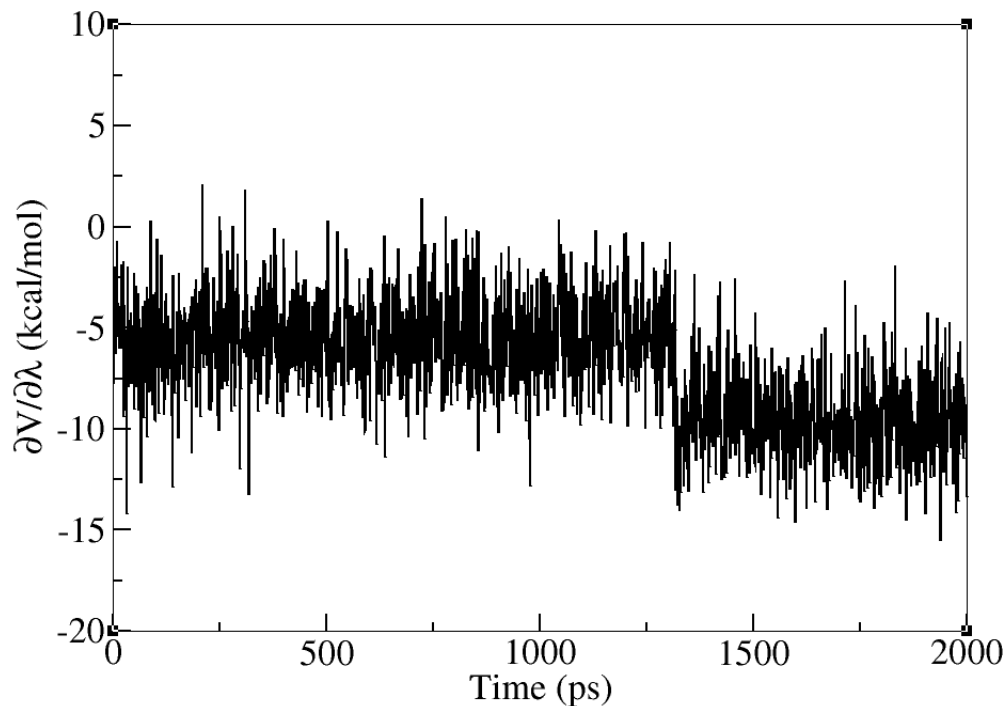


Figure 3.12: $\partial V/\partial\lambda$ for $\lambda = 1.0$ when discharging Thr52 in the presence of dimannose.

corresponds to the same methyl group pointing up as shown in Figure 3.13b.

Table 3.2: BAR analysis for relative binding free energies for the Ser52Thr mutation

	Discharging of Ser52 (kcal/mol)	Negative of LJ Transformation (kcal/mol)	Negative of Discharging of Thr52 (kcal/mol)	Total ΔG (kcal/mol)
I in presence Man_2	-4.73 (0.17)	0.38 (0.05)	0.31 (0.19)	-4.04 [0.26]
I in absence Man_2	-5.34 (0.28)	0.54 (0.01)	4.46 (0.04)	-0.34 [0.28]
$\Delta \Delta G$	0.61 [0.33]	-0.16 [0.05]	-4.15 [0.19]	-3.70 [0.38]

As we can see from Table 3.2, the computed free energy for the Ser52Thr mutation using the REMD protocol was -3.7 kcal/mol. Even though the REMD

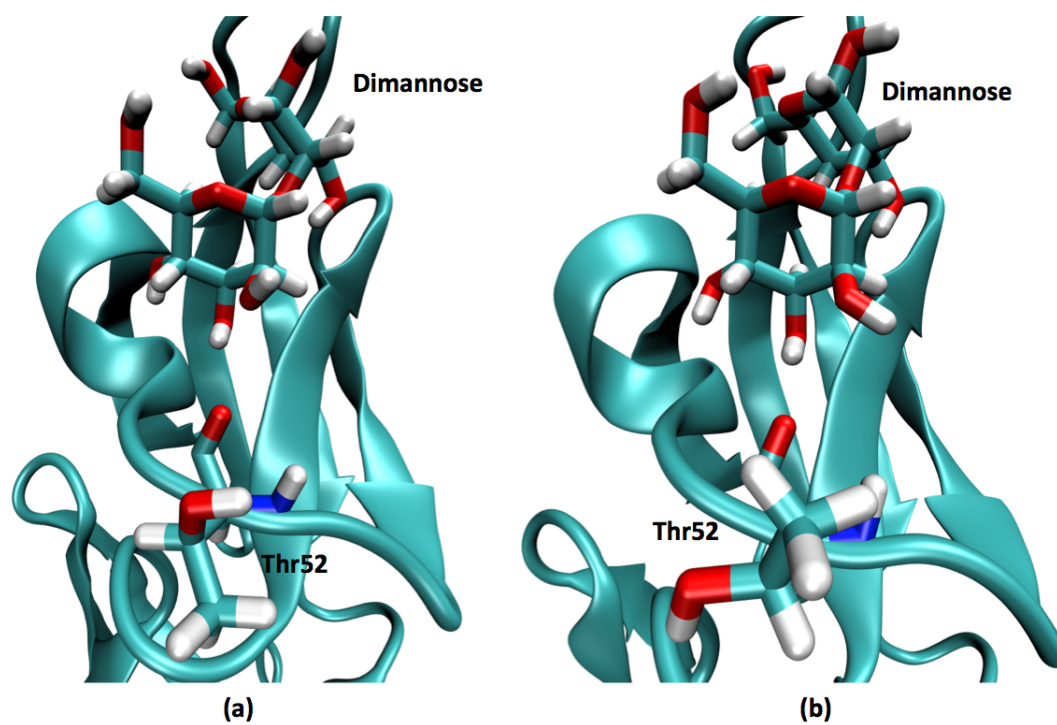


Figure 3.13: Two conformations of fully discharged Thr52 in the presence of dimannose. (a) the methyl group pointing down (b) the methyl group pointing up.

procedure improves sampling significantly, we will learn from chapter 4 that position 52 is in the very mobile hinge region. Simulations of 2 ns in duration even under the REMD protocol are highly unlikely to properly capture all necessary fluctuation in this region for the P51G-m4-CVN mutant. In fact, Ser52 has been linked to the structural stability of CVN. The S52P renders the protein uniquely in the dimeric form.[7]

3.3.3 Glu41Gln

Glu41 is a residue we already discussed in chapter 2. However, in chapter 2, we mutated Glu41 to Ala and Gly, which shorten the residue and abolish the charge on the side chain. In this chapter, we explored the change of position 41 to Gln. This change maintains the length of the side chain. However, a simple MD study of 20 ns for domain B with Gln in position 41 in the presence of dimannose revealed that there are at least three different conformations for this residue as shown in Figure 3.14. The first two conformations would be equivalent if the residue was Glu41, but are very different in the case of Gln41 as carbonyl oxygen is very different from amide nitrogen. In addition, a new conformation (c) was found for Gln41 when the amide group is pointing away from the dimannose. In this last conformation it is possible for the amide hydrogen to form hydrogen bonds with the backbone carbonyl oxygen of Trp49. The existence of these three conformations precludes the accurate calculation of free energies using the current protocol.

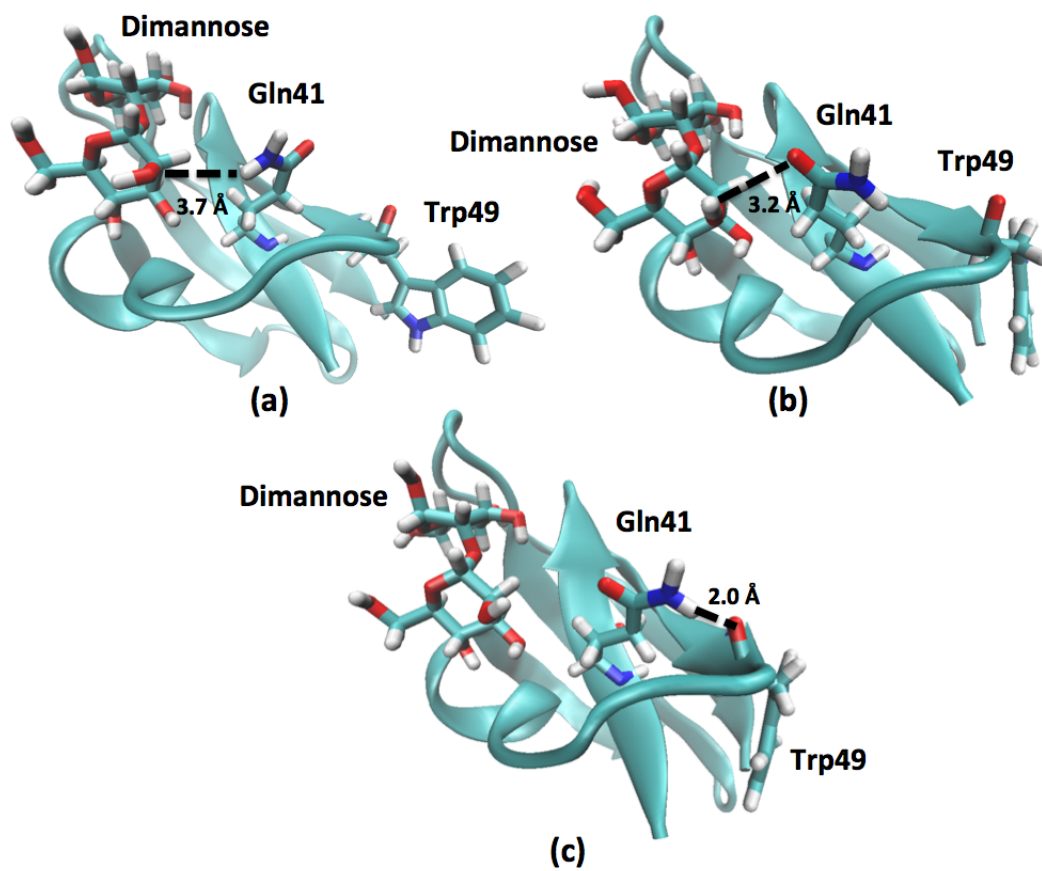


Figure 3.14: Three different conformations for Gln41.

Step1: Discharging of Val43

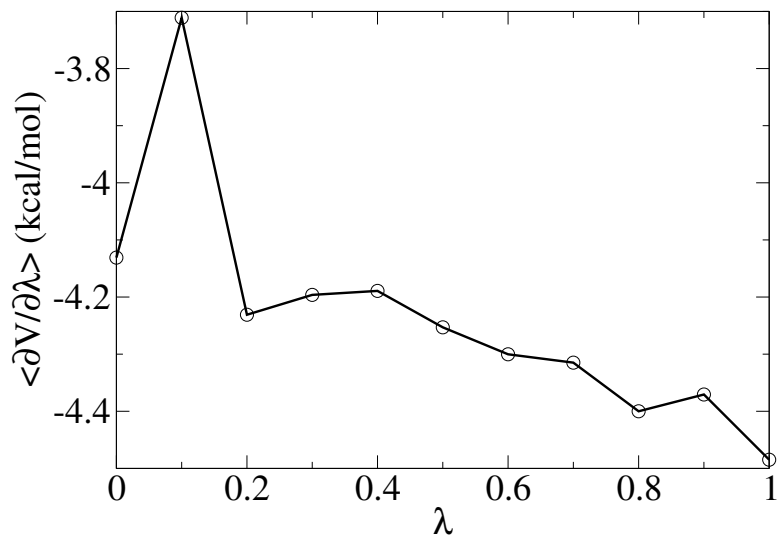


Figure 3.15: Discharging of Val43 for mutation Val43Ala in the absence of dimannose.

The $\langle \partial V / \partial \lambda \rangle$ is not converged because of the peak at λ equals 0.1.

3.3.4 Val43Ala

Val43 is a residue that dwells in the binding pocket but does not form hydrogen bonds with dimannose because of its hydrophobic nature. However, the position it holds is very important because it is in the vicinity of Asn42 and Glu41 that do interact directly with the sugar. The idea was to convert it to Ala, which is shorter than Val and does not introduce new functional groups. However, when discharging Val43 in the absence of dimannose, we encountered sampling problems as can be observed from the $\langle \partial V / \partial \lambda \rangle$ curve in Figure 3.15. Similar to what happened in the case of Ser52Thr, there are two different values for $\partial V / \partial \lambda$ when λ equals 0.10 as shown in Figure 3.16. The -4 kcal/mol and -3 kcal/mol values of $\partial V / \partial \lambda$ correspond

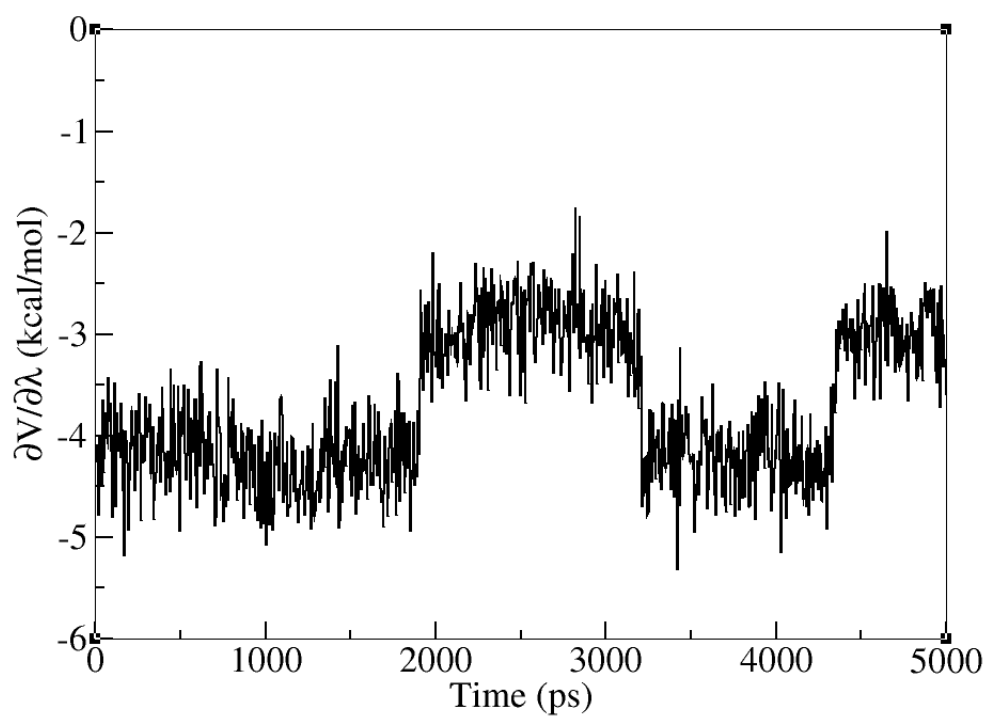


Figure 3.16: $\partial V/\partial\lambda$ for $\lambda = 0.10$ when discharging Val43 in the absence of dimannose.

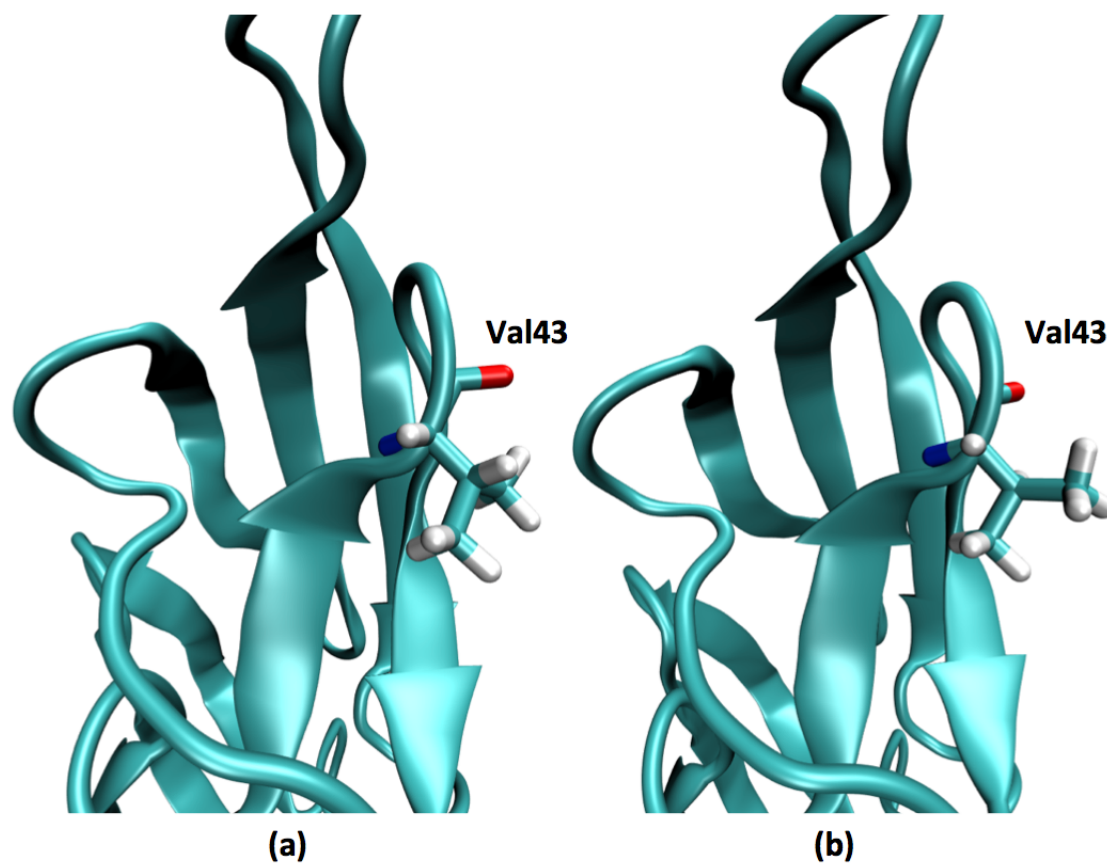


Figure 3.17: Two conformations of partially discharged Val43 ($\lambda = 0.10$) in the absence of dimannose. (a) the β -hydrogen pointing down corresponds to $\partial V/\partial\lambda = -4$ kcal/mol (b) the β -hydrogen pointing up corresponds to $\partial V/\partial\lambda = -3$ kcal/mol.

to the two conformations of Val43 depicted in Figures 3.17 a and b, respectively. This mutation was therefore not pursued any further.

3.3.5 Glu56Asp

If one can generate a mutation that improves the enthalpy of interaction in the bound state but not between water and the binding pocket in the absence of ligand this would be beneficial. This was the thought behind the proposed Glu56Asp mutation. As one can see from Figure 3.18 a and b, the carboxyl functional group of Asp56 is closer to dimannose compared to that of Glu56 and the expectation was that this would enhance this interaction. The expectation was also that both functional groups would interact with water in a similar manner as water is small and should have access to both types of residues. Therefore we expected an improvement in the interaction with ligand but not necessarily with water in the absence of ligand.

Figure 3.19 clearly demonstrates that the average distance between carboxyl groups in Glu56 and dimannose is 7 Å , whereas it is 5 Å in the case of Asp56 which should result in stronger interactions in the case of the latter. The three-step free energy calculation was then conducted to estimated the overall free energy for this mutation.

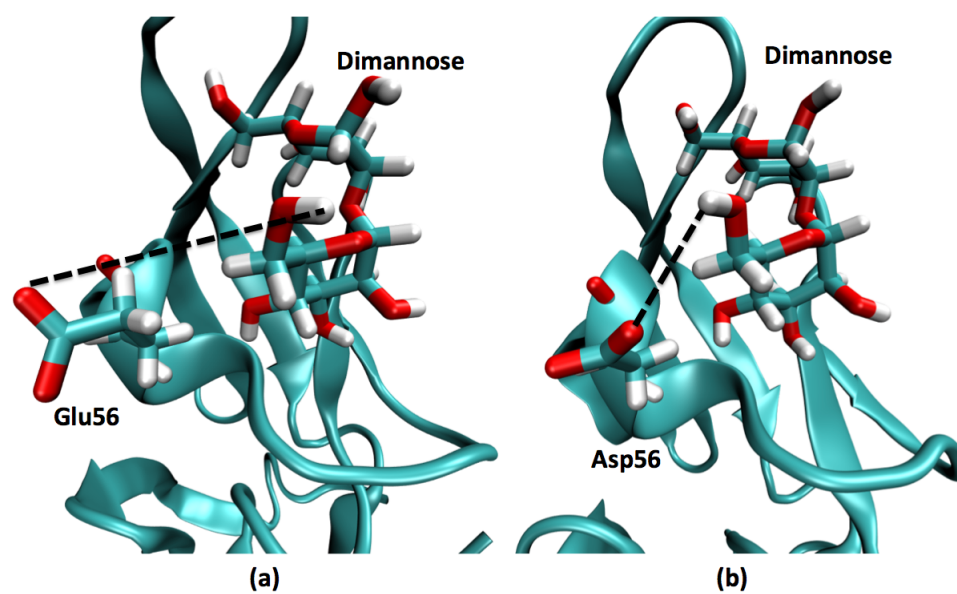


Figure 3.18: An illustration of the interaction between residue 56 and dimannose. (a) residue 56 is Glu and the carboxyl oxygens are far away from dimannose (b) residue 56 is Asp and the carboxyl oxygens are close to dimannose.

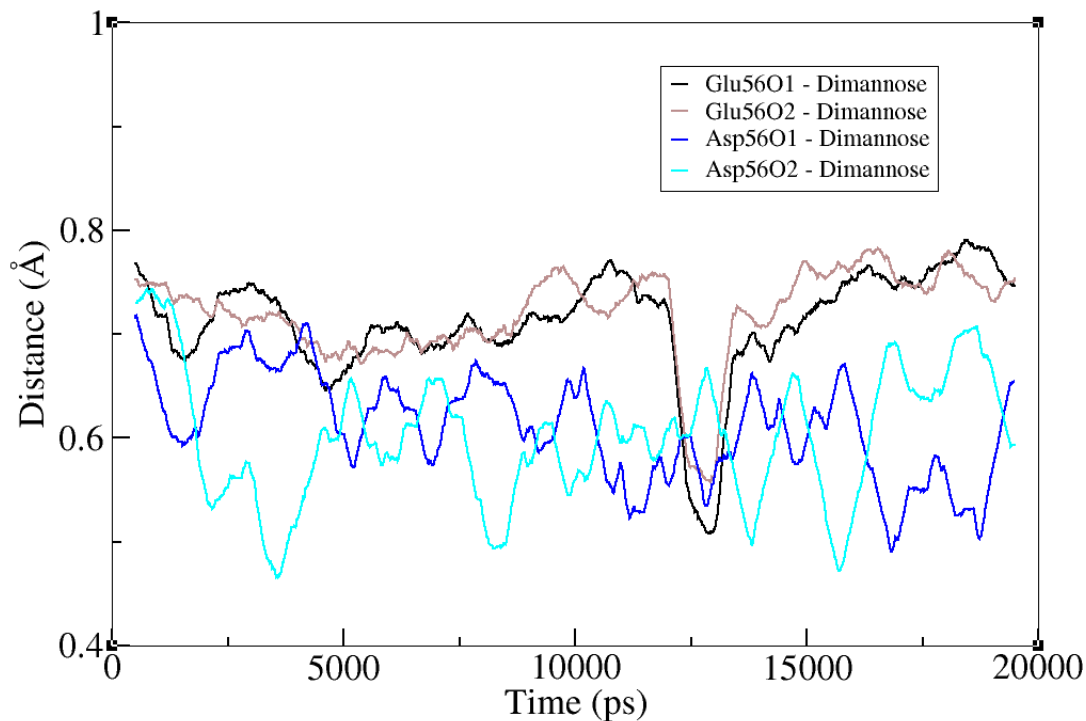


Figure 3.19: The hydrogen bond length between residue 56 and dimannose over a 20 ns NPT simulation. The distance is measured between the carboxyl oxygen of residue 56 and 1'OH of dimannose. The bond lengths between oxygen atoms of Glu56 and dimannose are colored with black and brown; whereas the bond lengths between the oxygen atoms in Asp56 and dimannose are colored with blue and cyan. A running average of 1 ns was conducted for clarity. The average distance between carboxyl group of Glu56 and dimannose is 7 Å , whereas it is 5 Å for Asp56.

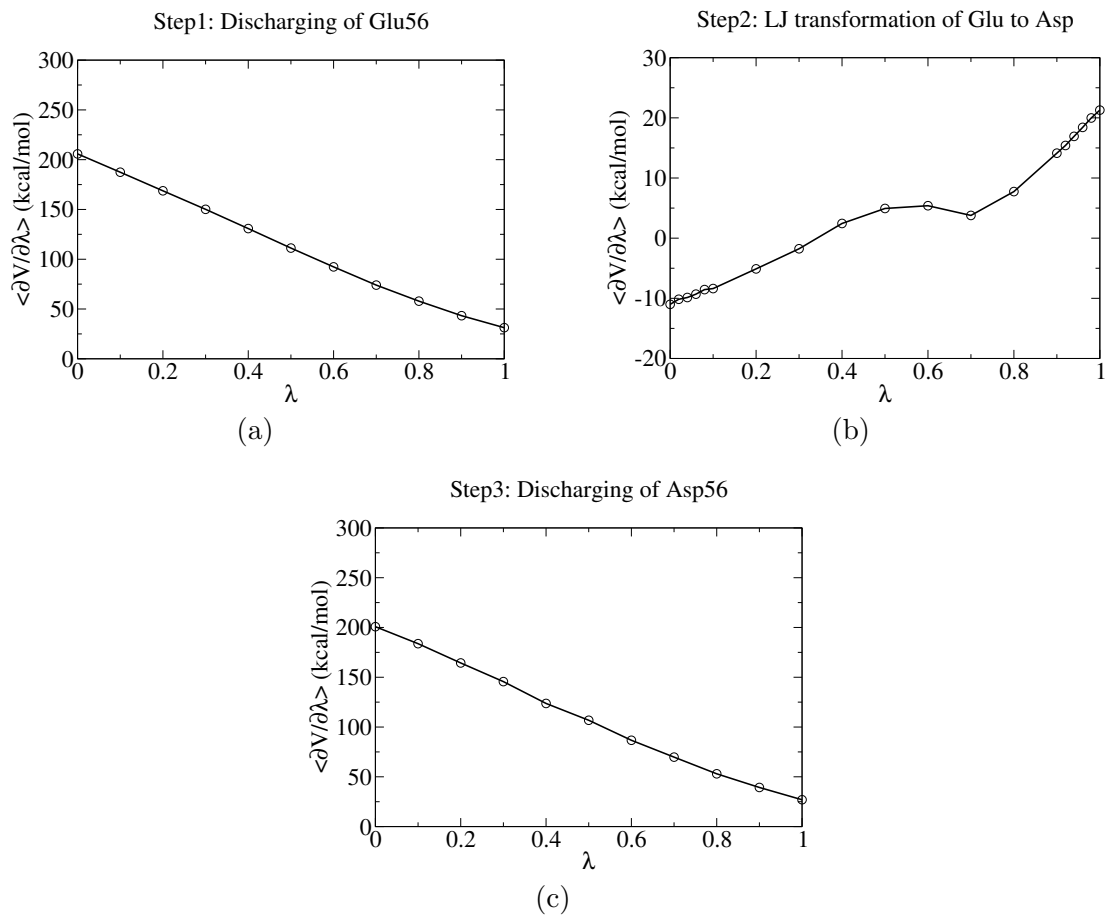


Figure 3.20: Three-step approach for thermodynamic integration simulations of the Glu56Asp mutation in presence of dimannose. Discharging of Glutamate (a), LJ transformation (b) and discharging of Aspartate (c).

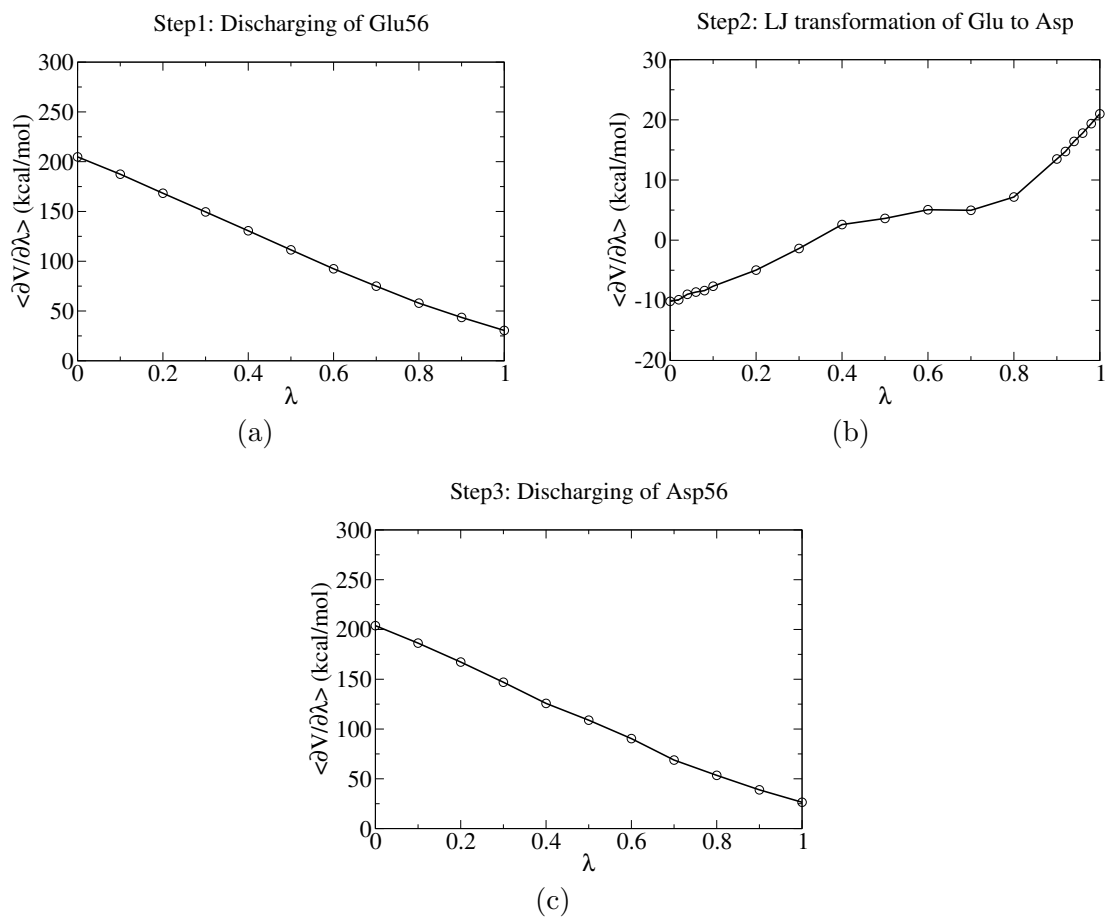


Figure 3.21: Three-step approach for thermodynamic integration simulations of the Glu56Asp mutation in absence of dimannose. Discharging of Glutamate (a), LJ transformation (b) and discharging of Aspartate (c).

Table 3.3: BAR analysis for relative binding free energies for the Glu56Asp mutation

	Discharging of Glu56 (kcal/mol)	LJ Transformation (kcal/mol)	Negative of Discharging of Asp56 (kcal/mol)	Total ΔG (kcal/mol)
I in presence <i>Man</i> ₂	113.39 (0.08)	2.81 (0.07)	-108.64 (0.22)	7.56 [0.24]
I in absence <i>Man</i> ₂	113.33 (0.04)	2.76 (0.07)	-110.15 (0.21)	5.94 [0.22]
$\Delta \Delta G$	0.06 [0.09]	0.05 [0.10]	1.51 [0.30]	1.62 [0.32]

The $\langle \partial V / \partial \lambda \rangle$ curves are shown in Figure 3.20 and 3.21. The free energy estimated by the BAR method has been summarized in Table 3.3. From this table, one can see that the free energy for discharging Glu56 both in the presence and absence of glycan are very similar as the difference is only 0.06 kcal/mol; in addition, the LJ transformation also makes a small contribution (0.05 kcal/mol) to the overall free energy of binding. This is expected as the carboxyl group of Glu 56 is fairly far away from the ligand as was described above. According to our expectation, Asp56 should instead favor the binding of ligand since it is closer (see Figure 3.19). However, the $\Delta \Delta G$ of mutating Glu56 to Asp56 is 1.6 kcal/mol, which is positive. This indicates that the mutation Glu56Asp is detrimental. The positive $\Delta \Delta G$ is a result of a large free energy penalty for the discharge of Asp56 in the absence of dimannose when compared to that in its presence.

3.4 Conclusion

In this chapter, we discussed several single mutations, including Thr57Ser, Ser52Thr and Glu56Asp. In the case of Thr57Ser, the binding ability was improved

by the fact that the partial charge of Ser57 is not preferred when dimannose is absent. Instead, discharging the partial charge of the Thr57 in the absence of dimannose causes a free energy penalty. For Ser52Thr, the partial charge of Thr52 played an important role by contributing -4.15 kcal/mol to the overall $\Delta \Delta G$. However, while estimating the free energy for the Ser52Thr mutation we encountered important sampling problems. This was particularly the case for the discharge of Thr52 in the presence of dimannose. Although, the Replica Exchange method improved sampling there are still likely important shortcomings as this residue is in the flexible hinge region. The mutation Glu56Asp is detrimental to binding as opposed to our expectation. The free energy penalties for discharging Glu56 are similar both in the presence and absence of dimannose. The free energy change caused by dispersion interaction appears to be negligible for the mutation. However, the partial charge of Asp56 appears to be favored by water to a larger extent than by dimannose, which results in an unfavorable mutation.

The studies of several other possible changes at the binding site were also discussed in the context of conformational sampling. The conclusion is that significantly more costly sampling procedures will be required if these are to be properly converged.

CHAPTER 4

OPEN AND CLOSED BINDING POCKET IN DOMAIN B AFFECTING THE BINDING ABILITY OF CVN TO DIMANNOSE

4.1 Introduction

As shown by screening of a glycan array,[80] CVN is highly specific as it exclusively binds oligomannosides containing a minimal dimannose unit ($\text{Man}\alpha 1 \rightarrow 2\text{Man}\alpha$) with affinity in the low nanomolar range. CVN comprises two quasi-symmetric domains, A (residues 1-38 and 90-101) and B (residues 39-89), each defining a carbohydrate binding pocket (Figure 4.1A). Despite a low degree of sequence homology of only 30 %, the two domains are strikingly similar, possessing three interstrand loops as well as five β -sheet elements. The protein can form a domain-swapped dimer by extension of a hinge region centered around Pro51.[98, 7] The dimeric form of CVN was initially found in the crystal structure[98]; it is a metastable form and converts to a monomer by partial unfolding.[7] However, its presence has been implicated in the mechanism of action of CVN because it allows multivalent binding to gp120.[69, 57]

Whereas both domains bind oligomannose glycans,[75, 38, 25, 10, 23, 5, 85, 42, 15, 72, 44, 19, 27, 70, 95, 41] the higher-affinity domain B has dissociation constants in the low micromolar range.[69, 57, 42, 72, 19, 41, 73, 18, 86] A number of structural studies of the wild type and mutants both in dimannose and apo

*A portion of this chapter is reproduced from Biochemistry, vol. 54(46), page 6951, year 2015 by Z. Li, A. Bolia, J. D. Maxwell, A. A. Bobkov, G. Ghirlanda, S. B. Ozkan, and C. J. Margulis. DOI: 10.1021/acs.biochem.5b00635. Copyright - Appendix C.2. Part of the content has also appeared in the thesis of another coauthor. A. Bolia, Modeling Protein Ligand Interactions Using Multi-Scale Computational Approaches, Arizona State University, 2015.

forms[80, 42, 15, 19, 27, 41, 86, 63] have significantly contributed to our understanding of glycan recognition. The structural data have been complemented by computational studies that helped dissect the origins of the unusual specificity of CVN for its glycan target. MD simulations and MM/PBSA (molecular mechanics/Poisson-Boltzmann surface area) analysis were used to assess the relative binding free energies for the two binding sites of CVN. These studies highlighted the role of backbone hydrogen bond interactions[80, 44, 70, 95, 43, 96] and conformational gating of the binding site by Arg76.[70] In addition, MD simulations provided two possible ways of relieving the clash between the amide protons of Ser52 and Asn53 observed in the NMR structures (PDB entries 2EZM and 1IIY): a crank-shaft motion that involves the backbone of Ser52 and Asn53 or a cis peptide bond isomerization between Pro51 and Ser52. The latter was observed only once during a 3 μ s molecular dynamics simulation.[43] Individual contributions to binding affinity by residues in the pocket of domain B were predicted by BP-Dock analysis and validated experimentally.[22] The Ozkan and Atilgan groups have recently computationally explored the role that conformational dynamics may play in protein binding[3, 4, 45, 20, 21, 46] and have found intriguing connections between motions, sometimes far removed from the binding site, that modulate interactions at the ligand-binding location. In the case of CVN and its mutants, this is important because the hinge region in the vicinity of Pro51 is directly associated with the protein's ability to access the domain-swapped dimer form. While a Ser52Pro mutation will result in almost exclusively dimeric protein, Pro51Gly locks CVN in the monomeric form.[7] This mutation also results in

increased folding stability of the monomeric form.[7] Thus, the rigidity in the hinge region provided by proline facilitates the formation of the dimer, whereas more flexible residues at position 51 favor the monomeric form. Because the hinge region is a constituent part of the high-affinity binding site in domain B (see Figure 4.1B), it is reasonable to question what structural effects dynamical flexibility or lack thereof may have on binding. To better address this problem, we focus on the comparison between wt-CVN and P51G-m4-CVN,[42] which contains the Pro51Gly mutation as well as four others that abolish dimannose binding at domain A.

In addition to the hinge region, we found that another key dynamic element at the high-affinity binding site is the hairpin where Asn42 is located (see Figure 4.1B). Our computational results indicate that Asn42, at the hairpin, in combination with Asn53, at the hinge region, is critical for the opening and closing of the binding site at domain B, which in turn is crucial for glycan binding. On the basis of our results, the following question arises: Is there a backbone structural flexibility origin to the higher or lower binding affinity for dimannose and, possibly, other more complex glycans by wild-type CVN as compared to mutants that has been so far overlooked? We conjecture that the rigidity in the hinge region provided by Pro51 is absent when this residue is mutated to glycine, and this accomplishes two very important goals. (1) It favors the formation of the dimer, therefore conferring a higher degree of multivalency, and (2) in the case of the monomeric form, it enhances the binding affinity at the high-affinity site; these advantages are absent when this residue is mutated to glycine.

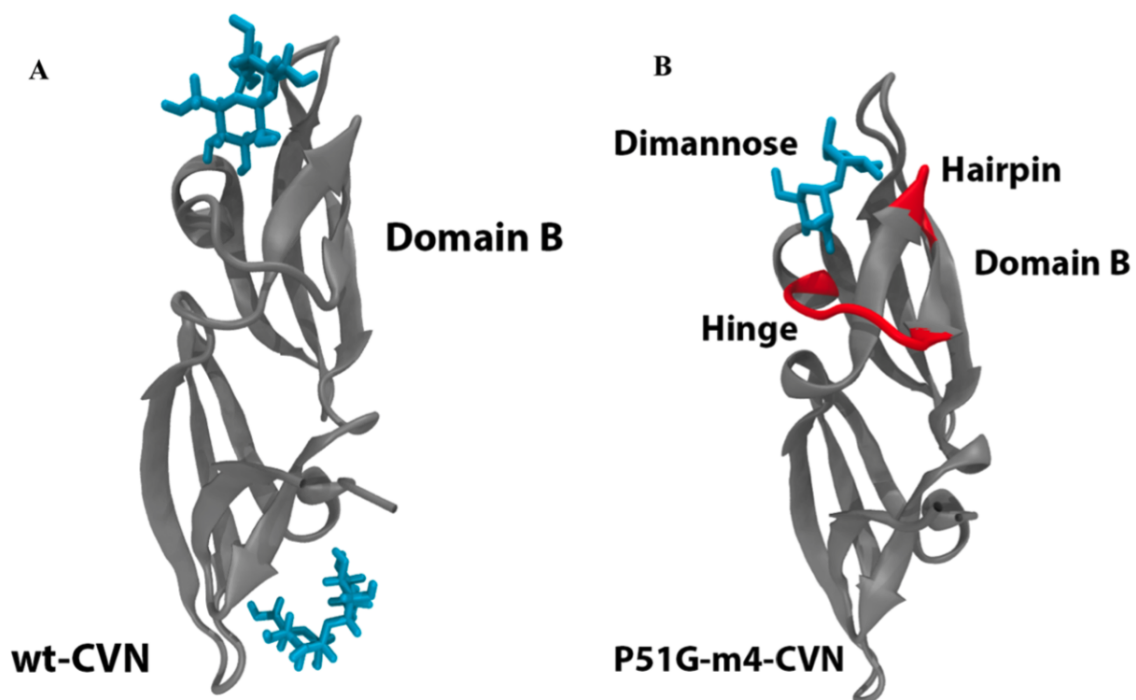


Figure 4.1: (A) NMR structure of cyanovirin bound to two molecules of dimannose (PDB entry 1IIY). (B) Crystal structure (PDB entry 2RDK) of P51G-m4-CVN where dimannose is colored blue. The hinge and the hairpin regions involved in the opening and closing of the high-affinity binding site in domain B are colored red.

4.2 Methods

To improve our understanding of the role of flexibility at the high-affinity binding site of wt-CVN and the P51G-m4-CVN mutant at 300 and 330 K, four long production molecular dynamics simulations of 200 ns each were conducted in the isothermal and isobaric (NPT) ensemble using the OPLS-AA force field.[55, 56] This resulted in an overall run time of 800 ns for each system and for each temperature. Protein structures were initially derived from NMR studies in the case of wild-type CVN (PDB entry 2EZM) and crystallography in the case of the P51G-m4-CVN (PDB entry 2Z21) mutant. These PDB structures were first energy minimized in the presence of 15103 TIP3P water molecules[54] and three K^+ counterions that were introduced to preserve electroneutrality. Subsequently, the two systems were equilibrated using the Berendsen method[12] for 100 ps in the constant volume, temperature, and number of particles ensemble (NVT) at the two respective temperatures. Following this, we conducted further equilibration in the constant number of particles, pressure, and temperature ensemble (NPT) for 200 ps using the Berendsen method. In all cases, periodic boundary conditions and the particle mesh Ewald[35] method were deployed with parameters (order = 4, Ewald_rtol = 10^{-5} , and Fourier spacing = 0.12) as coded in Gromacs version 4.5.5.[50, 83] The real part of the electrostatic interactions and dispersion interactions were cut off at 1.2 nm. NPT production trajectories used the Nose-Hoover[76, 39, 51] and Parrinello-Rahman[78] temperature and pressure coupling methods with frequency parameters of 0.5 and 5 ps, respectively. The first 10 ns of our 200 ns runs were used as further equilibration and not

considered in our data analysis. We found that the amide clashes observed in the starting NMR structures were relieved after at most 3.1 ns (in most cases in a few picoseconds) through a crank-shaft motion that involves a concerted rotation of the φ dihedral angle of Ser52 and the ψ dihedral angle of Asn53 similar to that observed by Fujimoto and Green.[43]

In addition to the work presented in this chapter, our collaborators from Arizona State University in the Ozkan group and the Ghirlanda group have carried out bioinformatics studies as well as isothermal titration calorimetry (ITC) experiments that support our findings. The experimental results will be discussed later in this chapter in the context of our simulations and the bioinformatics findings can be found in a recently published paper.[67]

4.3 Results

We will attempt to make the case that stiffness of the hinge region controls glycan binding by modulating the structure of the binding pocket with subtle backbone conformational changes strongly correlating with observed binding affinity. We will show that stiffness of the hinge region promotes an open domain B pocket whereas mutations that render the hinge more flexible promote changes in the conformation of Asn53 that hinder binding. Direct interaction between Asn53 and Asn42 in the hinge and hairpin regions depicted in Figure 4.1B results in an obstruction that should severely impede binding. We qualitatively represent this concept in Figure 4.2, where scenarios corresponding to the open and closed conformation of the binding site in

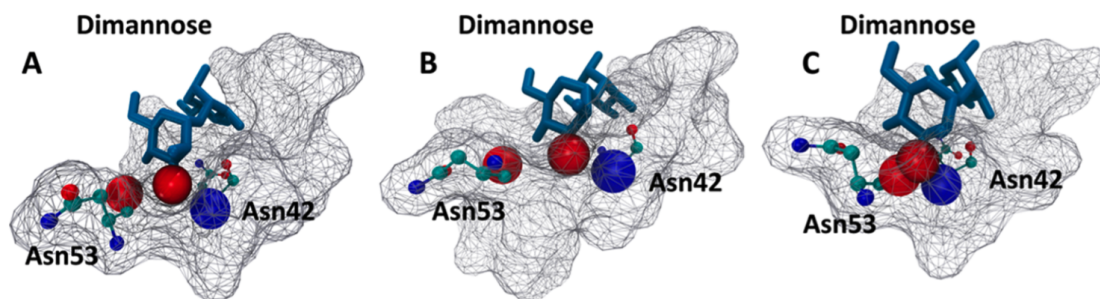


Figure 4.2: (A) Wire frame representation of the P51G-m4-CVN-binding site in domain B with bound dimannose (PDB entry 2RDK). (B) Apo form of wt-CVN from molecular dynamics simulations in explicit solvent. (C) Apo form of P51G-m4-CVN from simulations in explicit solvent. Dimannose is superposed for visualization purposes in panels B and C. In solution, because of rigidity in the hinge region, wt-CVN (B) preserves the separation between the backbone carbonyl oxygen of Asn53 (red sphere) and the amide nitrogen of Asn42 (blue sphere), rendering the pocket open for binding. There is no clash with the hydroxyl of the ligand (red sphere). Instead, the Pro51Gly mutation enhances the flexibility at the hinge region, and a close interaction between the backbone amide nitrogen of Asn42 and the backbone carbonyl oxygen of Asn53 closes the pocket and blocks binding as a clash with the hydroxyl oxygen (red sphere) from the glycan would occur (C).

domain B of wt-CVN and the P51G-m4-CVN mutant are displayed.

Figure 4.2A shows in a wire frame representation the experimental configuration of dimannose in the binding pocket of P51G-m4-CVN; depicted as balls and sticks are residues Asn53 and Asn42. Highlighted as van der Waals spheres are the amide nitrogen of Asn42 and the carbonyl oxygen of Asn53 as well as a relevant hydroxyl oxygen from the ligand. Panels B and C of Figure 4.2 show structures of the apo form of wt-CVN and P51G-m4-CVN from molecular dynamics simulations where the ligand has been superposed for the sake of visual clarity. In cases A and B, the binding pocket is open, but in case C, one can see from the clash of van de Waals spheres that interactions between Asn53 and Asn42 are inconsistent with glycan binding. We predict that the close conformation depicted in Figure 4.2C is ubiquitous when the Pro51Gly mutation is present, particularly at higher temperatures.

4.3.1 Flexibility in the Loop Region

Panels A and B of Figure 4.3 show a comparison of alignments at two different temperatures (four at 300 K and another four at 330 K) of wild-type CVN and the mutant, respectively. The eight trajectory-averaged structures are shown color-coded to quantify individual residue root-mean-square fluctuations. It is clear from Figure 4.3A that the average pocket structure is almost identical across simulations in the case of wt-CVN at 300 K and at 330 K. This is in contrast to the case of the mutant in Figure 4.3B where the average structure changes more pronouncedly across the same temperature trajectories and across different temperatures. In other words, the

hinge region of wt-CVN is significantly more rigid. Resilience to changes in the hinge region in the case of the wild type can also be expected upon perturbations induced by forces associated with approaching ligands. On the other hand, the Pro51Gly mutation provides extra flexibility to the region, and our analysis discussed below shows that this flexibility favors conformational changes in the vicinity of Asn53 that are detrimental for binding.

Another interesting aspect of this analysis is that the mobility of the hairpin in the vicinity of Asn42 can significantly increase at higher temperatures. In our trajectories, we see this most prominently in the case of the mutant. We will show that the combined effect of a flexible hinge and a motionally activated hairpin often leads to interactions between Asn53 (in the hinge) and Asn42 (in the hairpin) that should hinder binding as depicted in Figure 4.2C.

4.3.2 Dynamics of Hinge and Hairpin Regions Define Open and Closed States of the High-Affinity Binding Pocket

It is instructive to define the distance between the carbonyl oxygen in Asn53 and the amide hydrogen in Asn42 as an order parameter that we will call D_{pocket} . When D_{pocket} is smaller than a certain distance, the binding pocket is closed and does not support glycan binding. On the other hand, when D_{pocket} is larger, the binding pocket is open and binding is allowed. A combination of data from all our simulations was used to generate statistics for the probability of observing different values of D_{pocket} at different temperatures for wt-CVN and the mutant. This is shown

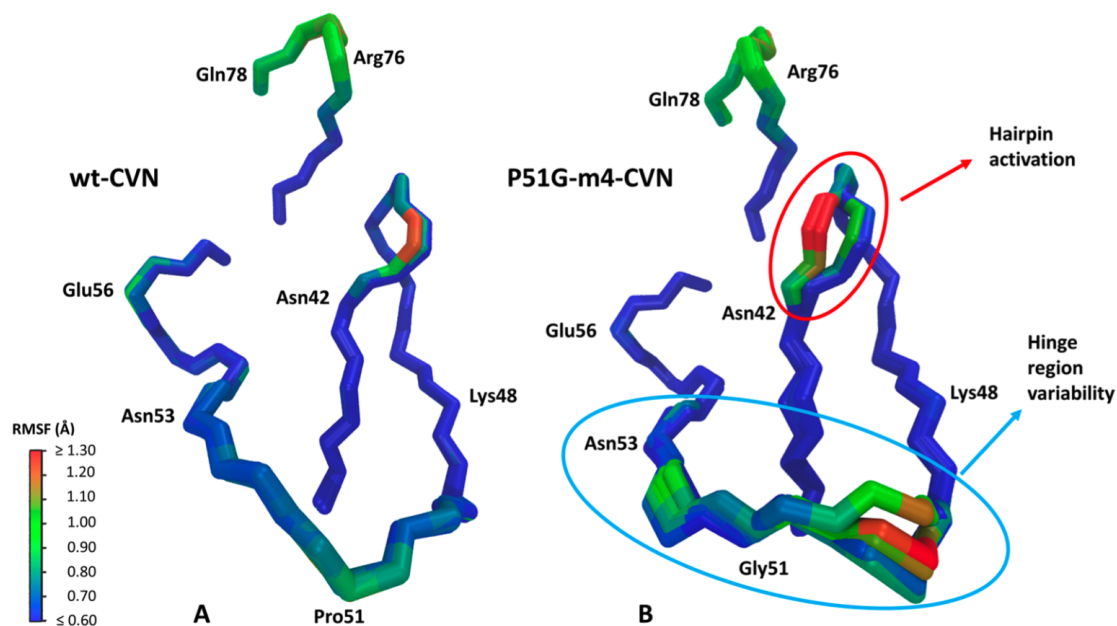


Figure 4.3: Alignment of eight trajectory-averaged structures color-coded on the basis of the root-mean-square fluctuation (rmsf) with respect to the average structure of each individual amino acid (four structures at 300 K and four structures at 330 K) in the case of (A) wt-CVN and (B) the P51G- m4-CVN mutant. At both temperatures, the average structure of the hinge region is almost unchanged in the case of wt-CVN. The opposite is true in the case of the mutant where the hinge takes on different conformations. We notice that the hairpin in the vicinity of Asn42 appears also to be more mobile at high temperatures in the case of the mutant. The lowest and highest rmsf values were 0.38 and 1.63 Å , respectively.

in Figure 4.4. The convergence of this histogram is analyzed as shown in Appendix B Figure B.1.

In all cases, three clear regions can be defined from the maxima in Figure 4.4: one at a D_{pocket} value slightly above 4 Å one centered slightly above 3.5 Å and one between 2 and 3 Å . When D_{pocket} is at a value of ≥ 4 Å (peaks to the right), one can consider the pocket to be open. A significant fraction of the population of the intermediate peak is characteristic of a situation in which Asn53 blocks the pocket but the hairpin where Asn42 is located is not motionally activated and does not block the pocket. Instead, peaks close to the left in Figure 4.4 are typical of the case in which the hairpin is activated (mainly at residue 43) and close interactions between Asn53 and Asn42 can be detected. Dynamical changes giving rise to these different peaks occur on a time scale on the order of ≥ 50 ns; therefore, it is possible that prohibitively expensive simulations on the order of many microseconds may be needed to guarantee the most accurate convergence of these distribution functions. However, analysis over our extensive 760 ns simulations at the two temperatures shows that the probability (sum of frequencies in Figure 4.4) of finding the pocket open for binding ($D_{pocket} \geq 4$ Å) decreases in the following order: wt-CVN at 330 K > wt-CVN at 300 K > P51G-m4-CVN at 300 K > P51G-m4-CVN at 330 K. We see that at higher temperatures little change is observed in the case of wild-type CVN; however, the mutant has significant probability in the region around a D_{pocket} of 2.5 Å that is the fully blocked configuration where changes in the overall hinge shape as well as coupling between Asn53 and Asn42 should make binding highly

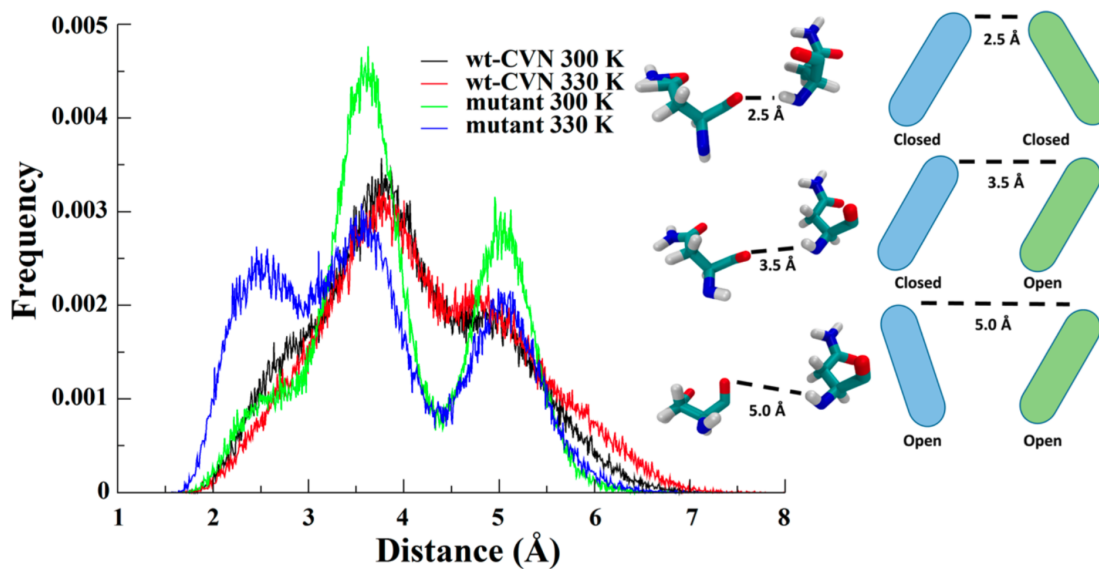


Figure 4.4: Probability of the distance between the carbonyl oxygen of Asn53 and the amide hydrogen of Asn42 (D_{pocket}) in the case of wt-CVN and the P51G-m4-CVN mutant at 300 and 330 K from 760 ns MD simulations. Three conformations with D_{pocket} close to 5, 3.5, and 2.5 Å are detected. On the right-hand side, time-averaged (over a period of at least 10 consecutive nanoseconds) Asn53 and Asn42 structures consistent with each of the three conformations (peaks in the graph) are depicted alongside idealized representations of hinge behavior (light blue) and hairpin behavior (light green).

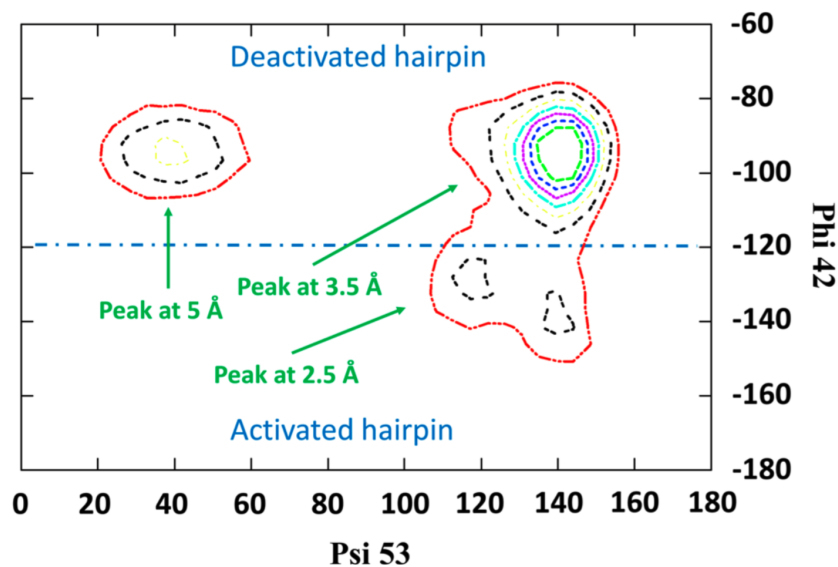


Figure 4.5: Probability distribution as a function of the ψ dihedral angle of residue 53 (in the hinge region) and φ dihedral angle of residue 42 (in the hairpin region) in the case of the P51G-m4-CVN mutant at 330 K. We see from this plot that a deactivated hairpin is consistent with two different conformations of residue 53 (open pocket at about 40° and closed pocket at about 140°). Instead, an activated hairpin appears to favor the conformation where Asn53 and Asn42 are in the proximity of each other. In this case, both residues contribute to a closed pocket and the open conformation of Asn53 is rarely observed.

unlikely. One can conclude from Figure 4.4 that in the monomeric form, rigidity in the hinge region protects the structure of the high-affinity binding site of wt-CVN and that the Pro51Gly mutation should be detrimental to binding as can be appreciated pictorially from Figure 4.2C. However, this additional interaction may explain the increased folding stability observed experimentally for CVN variants containing the Pro51Gly mutation. [7, 42]

In the case of the mutant at room temperature, blocking of the binding site is mostly accomplished by Asn53. At higher temperatures, we also see important participation by Asn42 in the hairpin region. It is reasonable to ask whether these motions are correlated. Figure 4.5 attempts to address these questions for the P51G-m4-CVN mutant at 330 K.

We first notice that when the hairpin is not activated, there are two possible conformers corresponding to two different orientations of the ψ angle of residue 53 (top part of Figure 4.5). One of these has a maximum at 40° and the other a maximum at 140° . The first corresponds to the fully open pocket (peak to the right in Figure 4.4) where binding is most likely, and the other corresponds to the situation in which Asn53 blocks the pocket (intermediate peak in Figure 4.4). When the hairpin is activated, our 760 ns simulations appear to indicate that conformations at 40° become of low probability (bottom portion of Figure 4.5). In other words, an activated hairpin at residue 42 tends to correlate with a blocking Asn53. The region where both residues 53 and 42 block the pocket is seen on the bottom right-hand side of Figure 4.5.

Table 4.1: Enthalpies, entropies, and K_d values for a set of cyanovirin-related proteins

protein	enthalpy (ΔH) (kcal/mol)	entropy ($T\Delta S$) (kcal/mol)	K_d (μM)
wt	-12.5 ± 0.3	-6.0 ± 0.1	16 ± 1 ; site A, 410 ± 20
P51G	-7.74 ± 0.04	-2.87 ± 0.01	213 ± 9 (two sites)
m4 (m)	-6.9 ± 0.1	0.0631 ± 0.001	7.2 ± 0.5
m4 (dimer)	1.12 ± 0.05	7.1 ± 0.3	40 ± 4
P51G-m4	-11 ± 1	-5.9 ± 0.5	141 ± 9

Results presented here are supported by bioinformatics studies reported in a recent paper[67] as well as by experimental isothermal calorimetry measurements presented in Table 4.1 from the Ghirlanda group.[67] Analysis of these results reveals that proteins containing a native proline at position 51 in the hinge region, e.g., wt-CVN and m4, display binding constants (K_d) in the low micromolar range at the high-affinity site, identified as domain B. In sharp contrast, the presence of glycine at position 51 results in a dramatic loss of affinity, with binding constant of approximately $212.8 \mu M$ at this site.[67]

4.4 Conclusion

The computational studies presented here have provided significant evidence that the Pro51Gly mutation enhances the flexibility of the hinge region in a way that is detrimental to glycan binding. These findings are fully supported by ITC measurements[67] on a set of CVN-related proteins with and without the mutation conducted by our collaborators from Arizona State University. Asn53 is a main culprit as motion of the hinge region results in this residue blocking the high-affinity binding

site. Instead, the rigidity of the hinge region due to Pro51 in the case of wt-CVN alters the position of Asn53, preventing it from blocking the binding site. Such rigidity maintains the binding pocket in an open conformation a large fraction of the time. The hairpin including Asn42 can become activated and also contribute to blocking the binding pocket. This is seen most prominently in the case of the mutant at higher temperatures. Notably, a full survey of published structures of CVN and engineered mutants reveals that almost in all cases when glycan binding is observed the distance between Asn53 and Asn42 is larger than 4 Å, indicating an open pocket, and in cases when binding is abolished the distance between these residues is small, suggesting a closed pocket blocking binding.[7, 72, 63, 17, 8] We think that this mechanism of temporarily blocking the binding site by Asn53 or a symmetrically located residue in domain A may be pervasive. The fraction of the time this reversible blocking occurs should be statistically correlated with the higher or lower binding affinity of wt-CVN and its analogues.

CHAPTER 5 SUMMARY AND FUTURE DIRECTIONS

The focus of this thesis has been on CVN and related mutants because of their potent antiviral activity against HIV. A primary goal was to understand what controls affinity and flexibility at the high affinity pocket in CVN and its mutants. Mutations that can potentially enhance affinity may affect flexibility as well, and flexibility is directly connected with major changes in the protein structure that lead to the formation of a dimeric form.

The first question we attempted to address was if the strong Coulombic interactions with Glu41, a charged residue in direct contact with the glycan ligand, would translate into a fundamentally important contribution to the free energy of binding. Our results from free energy alchemical mutations that changed Glu41 into Ala and Gly appear to contradict this hypothesis as the free energy changes were found to be 0.02 kcal/mol and -0.69 kcal/mol respectively. From these initial studies we learned that strong interactions at the binding site do not always necessarily translate into strong contributions to the free energy of binding. This is because one always must consider the interactions that may occur between protein surface residues in the absence of ligand and the solvent as well as other neighboring residues. The binding sites of CVN are at the protein surface and water has access to these when the glycan is absent. Mutating a residue that strongly interacts with the ligand also significantly impacts interactions with the solvent in the absence of ligand. These two effects compensate in ways that are not easily predictable without expensive calculations or

experiments. Even though experimental results follow a similar trend (a mutation to Ala is worse than a mutation to Gly), they do not fully support the finding that Glu41 is unimportant when the free energy of binding is considered. Both mutations are unfavorable (0.78 and 0.59 kcal/mol for Glu41Ala and Glu41Gly respectively). This discrepancy between simulations and experiment can be ascribed to force field inaccuracies and other cumulative errors. Discrepancies on the order of 1 kcal/mol are considered to be within the accuracy of the alchemical method.

Following this initial study, several other residues were considered as possible targets for mutations that could enhance binding affinity. Whereas many of these studies were unsuccessful for sampling reasons, each taught us about different aspects of protein-sugar interactions, issues of conformational variability and local protein flexibility. These studies focused not on wild type CVN but instead on the P51G-m4-CVN mutant. The rationale behind this choice was that the mutant favors the monomeric form of the protein and that the low affinity binding site is abolished. In these studies the Thr57Ser mutation emerged as a potentially good candidate for enhancing the free energy of binding. A possible explanation for this appears to be associated with an asymmetry in the behavior of the side chain of the two residues in the absence of ligand. The methyl group of Thr57 sterically hinders access of water molecules no matter if partial charges are on or off during the alchemical cycle. Instead, the lack of a methyl group allows for reorientation of the OH group and interactions with the solvent appear to be different when partial charges are on or off during the alchemical cycle.

We originally thought that Ser52Thr could be a good candidate for mutation. The idea was that interactions between water and the methyl group of residue 52 could push the backbone carbonyl oxygen closer to dimannose to form stronger hydrogen bonds. However, we encountered problems sampling Thr52 as this residue appears to have two distinct conformations that must be thoroughly sampled particularly in the presence of dimannose. Although the Replica Exchange method significantly helped with sampling, the fact that position 52 is at the hinge region and is key to locking the protein either in the monomeric or dimeric form is problematic. The hinge region is associated with protein conformational motions that occur on a time scale orders or magnitude larger than we are able to sample. Ultimately only experiments can tell if such mutation is favorable as fully atomistic computational studies are inadequate to sample large scale motions in this region of significant flexibility.

Val at position 43 has two conformations with the β -hydrogen pointing in two opposite directions. These two conformations lead to two different values for $\partial V/\partial\lambda$ causing poor convergence of the $\langle\partial V/\partial\lambda\rangle$ curve. Similarly an exploratory study for Gln at position 41 revealed several different conformations for this residue particularly in its interaction with the backbone of residue 49. A general lesson learned from all of these alchemical mutation studies is that the most problematic portion is the behavior of aminoacids upon discharge. This is because upon discharge these could adopt many conformations that are unfavorable for the same residue in realistic situations. The occurrence of many conformations with similar probability resulted in severe sampling problems.

A particularly interesting study was the Glu56Asp mutation. Free energy analysis based on the three steps method shows that discharging of the partial charge of Glu56 is independent of whether it is done in the presence or absence of dimannose as $\Delta \Delta G$ for discharging is 0.06 kcal/mol. Additionally, the $\Delta \Delta G$ for dispersion was found to be 0.05 kcal/mol. In other words, the success of this change completely hinges on the ability of Asp to be a better residue in relation to the ligand than it would be in its absence and not on any property of the original Glu56. The overall free energy is therefore mainly determined by the alchemical step involving the discharging of Asp56 in the presence and absence of sugar. Because a hydrogen bond distance analysis indicated that the carboxyl group of Asp56 should be closer to dimannose, we expected this mutation to be favorable. However, we found the opposite result to be correct. In very simple terms, both water and the ligand prefer a charged Asp56, but water likes it more. In spite of the fact that this particular attempt at alchemically mutating position 56 was not successful, it opened the opportunity for future studies. This is because one only needs to worry about finding a replacement for Glu56 that prefers the sugar over the solvent. The portion of the cycle associated with the disappearance of Glu is unimportant and this simplifies matters significantly.

As discussed previously, most of our mutation studies were done on the P51G-m4-CVN mutant. It is important to realize that whereas the Pro51Gly mutation may be important in the stabilization of the monomeric form of the protein, it may not be innocent in the context of the free energetics of binding. In our studies, the flexibility of P51G-m4-CVN was compared to that of wt-CVN. Our simulations and bioinfor-

matic studies from the Ozkan group show that the hinge region of wt-CVN resists the mobility associated with an increase of temperature, whereas the hinge region of P51G-m4-CVN fluctuates significantly more when temperature is increased. Until recently the effect of this mutation on binding was not well understood. Our simulations alongside with the experimental ITC results from the Ghirlanda laboratory in collaboration with Andrey Bobkov at the Sanford Burnham Medical Research Institute as well as bioinformatics work in the Ozkan lab found that the Pro51Gly change impedes binding by enhancing the occlusion of the high affinity binding pocket. This occlusion is intimately related to the behavior of residue 53 in the flexible hinge region and its interaction with residue 42 in a hairpin on the opposite side of the binding pocket that can get activated upon a temperature increase.

Many aspects of this protein and its mutants require further work. To enhance the binding affinity of the protein for dimannose, multi-mutation studies should be undertaken. However, such studies are expensive and their accuracy may be questionable. Furthermore, alchemical studies only probe local changes but multi-mutations may result in folding issues or important protein conformational changes. Research on the folding mechanism of CVN is also important as it will provide a more nuanced understanding of the hinge region linked to the formation of domain swapped dimer. It may be that focusing on enhancing the binding affinity at domain B of CVN is not necessarily the best route to better overall antiviral activity. Perhaps a more fruitful approach has to do with enhancing the multivalency of engineered protein constructs. The issue of multivalency should be better understood because it is well

appreciated that it plays an important role in antiviral activity. The exact mechanism of how CVN inhibits HIV infectivity is still elusive but it is now widely accepted that multivalency is very important.

APPENDIX A
FORCE FIELD PARAMETERS FOR DIMANNOSE

The parameters of dimannose in the format of OPLS-AA are listed below. The O8 is the oxygen we have modified based on the OPLS force field for carbohydrates[34] due to the fact that the OPLS-AA for carbohydrates provides parameters only for hexopyranoses.

Table A.1: Atomic parameters for dimannose in the format of OPLS-AA

Name	Charge	Sigma	Epsilon
C1	0.365	0.350	0.276
C2	0.205	0.350	0.276
C3	0.205	0.350	0.276
C4	0.205	0.350	0.276
C5	0.170	0.350	0.276
C6	0.145	0.350	0.276
O7	-0.700	0.307	0.711
O8	-0.465	0.290	0.586
O9	-0.700	0.307	0.711
O10	-0.700	0.307	0.711
O11	-0.400	0.290	0.586
O12	-0.683	0.312	0.711
H13	0.100	0.250	0.126
H14	0.060	0.250	0.126
H15	0.060	0.250	0.126
H16	0.060	0.250	0.126
H17	0.030	0.250	0.126
H18	0.060	0.250	0.126
H19	0.060	0.250	0.126
HO20	0.435	0.000	0.000
HO21	0.435	0.000	0.000
HO22	0.435	0.000	0.000
HO23	0.418	0.000	0.000

Continued on Next Page...

*A portion of this chapter is reproduced from Biochemistry, vol. 53(9), page 1477, year 2014 by S. K. Ramadugu, Z. Li, H. K. Kashyap, and C. J. Margulis. DOI: 10.1021/bi4014159. Copyright - Appendix C.1

Table A.1 – Continued

Name	Charge	Sigma	Epsilon
C24	0.300	0.350	0.276
C25	0.205	0.350	0.276
C26	0.205	0.350	0.276
C27	0.205	0.350	0.276
C28	0.170	0.350	0.276
C29	0.145	0.350	0.276
O30	-0.700	0.307	0.711
O31	-0.700	0.307	0.711
O32	-0.700	0.307	0.711
O33	-0.400	0.290	0.586
O34	-0.683	0.312	0.711
H35	0.100	0.250	0.126
H36	0.060	0.250	0.126
H37	0.060	0.250	0.126
H38	0.060	0.250	0.126
H39	0.030	0.250	0.126
H40	0.060	0.250	0.126
H41	0.060	0.250	0.126
HO42	0.435	0.000	0.000
HO43	0.435	0.000	0.000
HO44	0.435	0.000	0.000
HO45	0.418	0.000	0.000

Table A.2: Bond parameters for dimannose in the format of OPLS-AA

Atom i	Atom j	Func	b0	kb
C1	C2	1	0.1529	224262.4
C1	O7	1	0.1380	267776.0
C1	O11	1	0.1380	267776.0
C1	H13	1	0.1090	284512.0
C2	C3	1	0.1529	224262.4
C2	O8	1	0.1410	267776.0
C2	H14	1	0.1090	284512.0
C3	C4	1	0.1529	224262.4
C3	O9	1	0.1410	267776.0
C3	H15	1	0.1090	284512.0
C4	C5	1	0.1529	224262.4
C4	O10	1	0.1410	267776.0
C4	H16	1	0.1090	284512.0
C5	C6	1	0.1529	224262.4
C5	O11	1	0.1410	267776.0

Continued on Next Page...

Table A.2 – Continued

Atom i	Atom j	Func	b0	kb
C5	H17	1	0.1090	284512.0
C6	O12	1	0.1410	267776.0
C6	H18	1	0.1090	284512.0
C6	H19	1	0.1090	284512.0
O7	HO20	1	0.0945	462750.4
O8	C24	1	0.1380	267776.0
O9	HO21	1	0.0945	462750.4
O10	HO22	1	0.0945	462750.4
O11	HO23	1	0.0945	462750.4
C24	C25	1	0.1529	224262.4
C24	O33	1	0.1380	267776.0
C24	H35	1	0.1090	284512.0
C25	C26	1	0.1529	224262.4
C25	O30	1	0.1410	267776.0
C25	H36	1	0.1090	284512.0
C26	C27	1	0.1529	224262.4
C26	O31	1	0.1410	267776.0
C26	H37	1	0.1090	284512.0
C27	C28	1	0.1529	224262.4
C27	O32	1	0.1410	267776.0
C27	H38	1	0.1090	284512.0
C28	C29	1	0.1529	224262.4
C28	O33	1	0.1410	267776.0
C28	H39	1	0.1090	284512.0
C29	O34	1	0.1410	267776.0
C29	H40	1	0.1090	284512.0
C29	H41	1	0.1090	284512.0
O30	HO42	1	0.0945	462750.4
O31	HO43	1	0.0945	462750.4
O32	HO44	1	0.0945	462750.4
O33	HO45	1	0.0945	462750.4

Table A.3: Bending parameters for dimannose in the format of OPLS-AA

Atom i	Atom j	Atom k	Func	Angle	Const
C2	C1	O7	1	109.50	418.40
C2	C1	O11	1	109.50	418.40
C2	C1	H13	1	110.70	313.80
O7	C1	O11	1	111.55	774.88
O7	C1	H13	1	109.50	292.88
O11	C1	H13	1	109.50	292.88

Continued on Next Page...

Table A.3 – Continued

Atom i	Atom j	Atom k	Func	Angle	Const
C1	C2	C3	1	112.70	488.27
C1	C2	O8	1	109.50	418.40
C1	C2	H14	1	110.70	313.80
C3	C2	O8	1	109.50	418.40
C3	C2	H14	1	110.70	313.80
O8	C2	H14	1	109.50	292.88
C2	C3	C4	1	112.70	488.27
C2	C3	O9	1	109.50	418.40
C2	C3	H15	1	110.70	313.80
C4	C3	O9	1	109.50	418.40
C4	C3	H15	1	110.70	313.80
O9	C3	H15	1	109.50	292.88
C3	C4	C5	1	112.70	488.27
C3	C4	O10	1	109.50	418.40
C3	C4	H16	1	110.70	313.80
C5	C4	O10	1	109.50	418.40
C5	C4	H16	1	110.70	313.80
O10	C4	H16	1	109.50	292.88
C4	C5	C6	1	112.70	488.27
C4	C5	O11	1	109.50	418.40
C4	C5	H17	1	110.70	313.80
C6	C5	O11	1	109.50	418.40
C6	C5	H17	1	110.70	313.80
O11	C5	H17	1	109.50	292.88
C5	C6	O12	1	109.50	418.40
C5	C6	H18	1	110.70	313.80
C5	C6	H19	1	110.70	313.80
O12	C6	H18	1	109.50	292.88
O12	C6	H19	1	109.50	292.88
H18	C6	H19	1	107.80	276.14
C1	O7	HO20	1	108.50	460.24
C2	O8	C24	1	109.50	502.08
C3	O9	HO21	1	108.50	460.24
C4	O10	HO22	1	108.50	460.24
C1	O11	C5	1	109.50	502.08
C6	O12	HO23	1	108.50	460.24
O8	C24	C25	1	109.50	418.40
O8	C24	O33	1	111.55	774.88
O8	C24	H35	1	109.50	292.88
C25	C24	O33	1	109.50	418.40
C25	C24	H35	1	110.70	313.80

Continued on Next Page...

Table A.3 – Continued

Atom i	Atom j	Atom k	Func	Angle	Const
O33	C24	H35	1	109.50	292.88
C24	C25	C26	1	112.70	488.27
C24	C25	O30	1	109.50	418.40
C24	C25	H36	1	110.70	313.80
C26	C25	O30	1	109.50	418.40
C26	C25	H36	1	110.70	313.80
O30	C25	H36	1	109.50	292.88
C25	C26	C27	1	112.70	488.27
C25	C26	O31	1	109.50	418.40
C25	C26	H37	1	110.70	313.80
C27	C26	O31	1	109.50	418.40
C27	C26	H37	1	110.70	313.80
O31	C26	H37	1	109.50	292.88
C26	C27	C28	1	112.70	488.27
C26	C27	O32	1	109.50	418.40
C26	C27	H38	1	110.70	313.80
C28	C27	O32	1	109.50	418.40
C28	C27	H38	1	110.70	313.80
O32	C27	H38	1	109.50	292.88
C27	C28	C29	1	112.70	488.27
C27	C28	O33	1	109.50	418.40
C27	C28	H39	1	110.70	313.80
C29	C28	O33	1	109.50	418.40
C29	C28	H39	1	110.70	313.80
O33	C28	H39	1	109.50	292.88
C28	C29	O34	1	109.50	418.40
C28	C29	H40	1	110.70	313.80
C28	C29	H41	1	110.70	313.80
O34	C29	H40	1	109.50	292.88
O34	C29	H41	1	109.50	292.88
H40	C29	H41	1	107.80	276.14
C25	O30	HO42	1	108.50	460.24
C26	O31	HO43	1	108.50	460.24
C27	O32	HO44	1	108.50	460.24
C24	O33	C28	1	109.50	502.08
C29	O34	HO45	1	108.50	460.24

Table A.4: Dihedral parameters for dimannose in the format of OPLS-AA. The last two columns are all 0.000000, for the purpose of simplicity, the last two columns are not listed.

i	j	k	l	Func	c1	c2	c3	c4
O7	C1	C2	C3	3	-2.794912	2.794912	0.000000	0.000000
O7	C1	C2	O8	3	9.035348	-9.035348	0.000000	0.000000
O7	C1	C2	H14	3	0.979056	2.937168	0.000000	-3.916224
O11	C1	C2	C3	3	-2.794912	2.794912	0.000000	0.000000
O11	C1	C2	O8	3	9.035348	-9.035348	0.000000	0.000000
O11	C1	C2	H14	3	0.979056	2.937168	0.000000	-3.916224
H13	C1	C2	C3	3	0.765672	2.297016	0.000000	-3.062688
H13	C1	C2	O8	3	0.979056	2.937168	0.000000	-3.916224
H13	C1	C2	H14	3	0.665256	1.995768	0.000000	-2.661024
C2	C1	O7	HO20	3	-4.322072	0.845168	12.062472	-8.585568
O11	C1	O7	HO20	3	-10.179672	2.648472	7.556304	-0.025104
H13	C1	O7	HO20	3	0.941400	2.824200	0.000000	-3.765600
C2	C1	O11	C5	3	1.715440	2.845120	1.046000	-5.606560
O7	C1	O11	C5	3	-6.458004	0.809604	5.681872	-0.033472
H13	C1	O11	C5	3	1.589920	4.769760	0.000000	-6.359680
C1	C2	C3	C4	3	3.566860	-1.889076	0.656888	-2.334672
C1	C2	C3	O9	3	-2.794912	2.794912	0.000000	0.000000
C1	C2	C3	H15	3	0.765672	2.297016	0.000000	-3.062688
O8	C2	C3	C4	3	-2.794912	2.794912	0.000000	0.000000
O8	C2	C3	O9	3	9.035348	-9.035348	0.000000	0.000000
O8	C2	C3	H15	3	0.979056	2.937168	0.000000	-3.916224
H14	C2	C3	C4	3	0.765672	2.297016	0.000000	-3.062688
H14	C2	C3	O9	3	0.979056	2.937168	0.000000	-3.916224
H14	C2	C3	H15	3	0.665256	1.995768	0.000000	-2.661024
C1	C2	O8	C24	3	1.715440	2.845120	1.046000	-5.606560
C3	C2	O8	C24	3	1.715440	2.845120	1.046000	-5.606560
H14	C2	O8	C24	3	1.589920	4.769760	0.000000	-6.359680
C2	C3	C4	C5	3	3.566860	-1.889076	0.656888	-2.334672
C2	C3	C4	O10	3	-2.794912	2.794912	0.000000	0.000000
C2	C3	C4	H16	3	0.765672	2.297016	0.000000	-3.062688
O9	C3	C4	C5	3	-2.794912	2.794912	0.000000	0.000000
O9	C3	C4	O10	3	18.966072	-18.966072	0.000000	0.000000
O9	C3	C4	H16	3	0.979056	2.937168	0.000000	-3.916224
H15	C3	C4	C5	3	0.765672	2.297016	0.000000	-3.062688
H15	C3	C4	O10	3	0.979056	2.937168	0.000000	-3.916224
H15	C3	C4	H16	3	0.665256	1.995768	0.000000	-2.661024
C2	C3	O9	HO21	3	-4.322072	0.845168	12.062472	-8.585568
C4	C3	O9	HO21	3	-4.322072	0.845168	12.062472	-8.585568

Continued on Next Page...

Table A.4 – Continued

i	j	k	l	Func	c1	c2	c3	c4
H15	C3	O9	HO21	3	0.941400	2.824200	0.000000	-3.765600
C3	C4	C5	C6	3	3.566860	-1.889076	0.656888	-2.334672
C3	C4	C5	O11	3	-2.794912	2.794912	0.000000	0.000000
C3	C4	C5	H17	3	0.765672	2.297016	0.000000	-3.062688
O10	C4	C5	C6	3	-2.794912	2.794912	0.000000	0.000000
O10	C4	C5	O11	3	9.035348	-9.035348	0.000000	0.000000
O10	C4	C5	H17	3	0.979056	2.937168	0.000000	-3.916224
H16	C4	C5	C6	3	0.765672	2.297016	0.000000	-3.062688
H16	C4	C5	O11	3	0.979056	2.937168	0.000000	-3.916224
H16	C4	C5	H17	3	0.665256	1.995768	0.000000	-2.661024
C3	C4	O10	HO22	3	-4.322072	0.845168	12.062472	-8.585568
C5	C4	O10	HO22	3	-4.322072	0.845168	12.062472	-8.585568
H16	C4	O10	HO22	3	0.941400	2.824200	0.000000	-3.765600
C4	C5	C6	O12	3	-2.794912	2.794912	0.000000	0.000000
C4	C5	C6	H18	3	0.765672	2.297016	0.000000	-3.062688
C4	C5	C6	H19	3	0.765672	2.297016	0.000000	-3.062688
O11	C5	C6	O12	3	9.035348	-9.035348	0.000000	0.000000
O11	C5	C6	H18	3	0.979056	2.937168	0.000000	-3.916224
O11	C5	C6	H19	3	0.979056	2.937168	0.000000	-3.916224
H17	C5	C6	O12	3	0.979056	2.937168	0.000000	-3.916224
H17	C5	C6	H18	3	0.665256	1.995768	0.000000	-2.661024
H17	C5	C6	H19	3	0.665256	1.995768	0.000000	-2.661024
C4	C5	O11	C1	3	1.715440	2.845120	1.046000	-5.606560
C6	C5	O11	C1	3	1.715440	2.845120	1.046000	-5.606560
H17	C5	O11	C1	3	1.589920	4.769760	0.000000	-6.359680
C5	C6	O12	HO23	3	-4.322072	0.845168	12.062472	-8.585568
H18	C6	O12	HO23	3	0.941400	2.824200	0.000000	-3.765600
H19	C6	O12	HO23	3	0.941400	2.824200	0.000000	-3.765600
C2	O8	C24	C25	3	1.715440	2.845120	1.046000	-5.606560
C2	O8	C24	O33	3	-6.458004	0.809604	5.681872	-0.033472
C2	O8	C24	H35	3	1.589920	4.769760	0.000000	-6.359680
O8	C24	C25	C26	3	-2.794912	2.794912	0.000000	0.000000
O8	C24	C25	O30	3	9.035348	-9.035348	0.000000	0.000000
O8	C24	C25	H36	3	0.979056	2.937168	0.000000	-3.916224
O33	C24	C25	C26	3	-2.794912	2.794912	0.000000	0.000000
O33	C24	C25	O30	3	9.035348	-9.035348	0.000000	0.000000
O33	C24	C25	H36	3	0.979056	2.937168	0.000000	-3.916224
H35	C24	C25	C26	3	0.765672	2.297016	0.000000	-3.062688
H35	C24	C25	O30	3	0.979056	2.937168	0.000000	-3.916224
H35	C24	C25	H36	3	0.665256	1.995768	0.000000	-2.661024
O8	C24	O33	C28	3	-6.458004	0.809604	5.681872	-0.033472

Continued on Next Page...

Table A.4 – Continued

i	j	k	l	Func	c1	c2	c3	c4
C25	C24	O33	C28	3	1.715440	2.845120	1.046000	-5.606560
H35	C24	O33	C28	3	1.589920	4.769760	0.000000	-6.359680
C24	C25	C26	C27	3	3.566860	-1.889076	0.656888	-2.334672
C24	C25	C26	O31	3	-2.794912	2.794912	0.000000	0.000000
C24	C25	C26	H37	3	0.765672	2.297016	0.000000	-3.062688
O30	C25	C26	C27	3	-2.794912	2.794912	0.000000	0.000000
O30	C25	C26	O31	3	18.966072	-18.966072	0.000000	0.000000
O30	C25	C26	H37	3	0.979056	2.937168	0.000000	-3.916224
H36	C25	C26	C27	3	0.765672	2.297016	0.000000	-3.062688
H36	C25	C26	O31	3	0.979056	2.937168	0.000000	-3.916224
H36	C25	C26	H37	3	0.665256	1.995768	0.000000	-2.661024
C24	C25	O30	HO42	3	-4.322072	0.845168	12.062472	-8.585568
C26	C25	O30	HO42	3	-4.322072	0.845168	12.062472	-8.585568
H36	C25	O30	HO42	3	0.941400	2.824200	0.000000	-3.765600
C25	C26	C27	C28	3	3.566860	-1.889076	0.656888	-2.334672
C25	C26	C27	O32	3	-2.794912	2.794912	0.000000	0.000000
C25	C26	C27	H38	3	0.765672	2.297016	0.000000	-3.062688
O31	C26	C27	C28	3	-2.794912	2.794912	0.000000	0.000000
O31	C26	C27	O32	3	18.966072	-18.966072	0.000000	0.000000
O31	C26	C27	H38	3	0.979056	2.937168	0.000000	-3.916224
H37	C26	C27	C28	3	0.765672	2.297016	0.000000	-3.062688
H37	C26	C27	O32	3	0.979056	2.937168	0.000000	-3.916224
H37	C26	C27	H38	3	0.665256	1.995768	0.000000	-2.661024
C25	C26	O31	HO43	3	-4.322072	0.845168	12.062472	-8.585568
C27	C26	O31	HO43	3	-4.322072	0.845168	12.062472	-8.585568
H37	C26	O31	HO43	3	0.941400	2.824200	0.000000	-3.765600
C26	C27	C28	C29	3	3.566860	-1.889076	0.656888	-2.334672
C26	C27	C28	O33	3	-2.794912	2.794912	0.000000	0.000000
C26	C27	C28	H39	3	0.765672	2.297016	0.000000	-3.062688
O32	C27	C28	C29	3	-2.794912	2.794912	0.000000	0.000000
O32	C27	C28	O33	3	9.035348	-9.035348	0.000000	0.000000
O32	C27	C28	H39	3	0.979056	2.937168	0.000000	-3.916224
H38	C27	C28	C29	3	0.765672	2.297016	0.000000	-3.062688
H38	C27	C28	O33	3	0.979056	2.937168	0.000000	-3.916224
H38	C27	C28	H39	3	0.665256	1.995768	0.000000	-2.661024
C26	C27	O32	HO44	3	-4.322072	0.845168	12.062472	-8.585568
C28	C27	O32	HO44	3	-4.322072	0.845168	12.062472	-8.585568
H38	C27	O32	HO44	3	0.941400	2.824200	0.000000	-3.765600
C27	C28	C29	O34	3	-2.794912	2.794912	0.000000	0.000000
C27	C28	C29	H40	3	0.765672	2.297016	0.000000	-3.062688
C27	C28	C29	H41	3	0.765672	2.297016	0.000000	-3.062688

Continued on Next Page...

Table A.4 – Continued

i	j	k	l	Func	c1	c2	c3	c4
O33	C28	C29	O34	3	9.035348	-9.035348	0.000000	0.000000
O33	C28	C29	H40	3	0.979056	2.937168	0.000000	-3.916224
O33	C28	C29	H41	3	0.979056	2.937168	0.000000	-3.916224
H39	C28	C29	O34	3	0.979056	2.937168	0.000000	-3.916224
H39	C28	C29	H40	3	0.665256	1.995768	0.000000	-2.661024
H39	C28	C29	H41	3	0.665256	1.995768	0.000000	-2.661024
C27	C28	O33	C24	3	1.715440	2.845120	1.046000	-5.606560
C29	C28	O33	C24	3	1.715440	2.845120	1.046000	-5.606560
H39	C28	O33	C24	3	1.589920	4.769760	0.000000	-6.359680
C28	C29	O34	HO45	3	-4.322072	0.845168	12.062472	-8.585568
H40	C29	O34	HO45	3	0.941400	2.824200	0.000000	-3.765600
H41	C29	O34	HO45	3	0.941400	2.824200	0.000000	-3.765600

APPENDIX B CONVERGENCE ANALYSIS FOR OPEN AND CLOSED STATE SIMULATIONS

Convergence analysis of the distribution function in Figure 4.4. We analyzed the data in 50 ns, 100 ns, 150 ns, and 200 ns of each molecular dynamics simulation (see Figure B.1) discussed in Chapter 4 in order to determine the change of distribution along simulation time.

*A portion of this chapter is reproduced from *Biochemistry*, vol. 54(46), page 6951, year 2015 by Z. Li, A. Bolia, J. D. Maxwell, A. A. Bobkov, G. Ghirlanda, S. B. Ozkan, and C. J. Margulis. DOI: 10.1021/acs.biochem.5b00635. Copyright - Appendix C.2

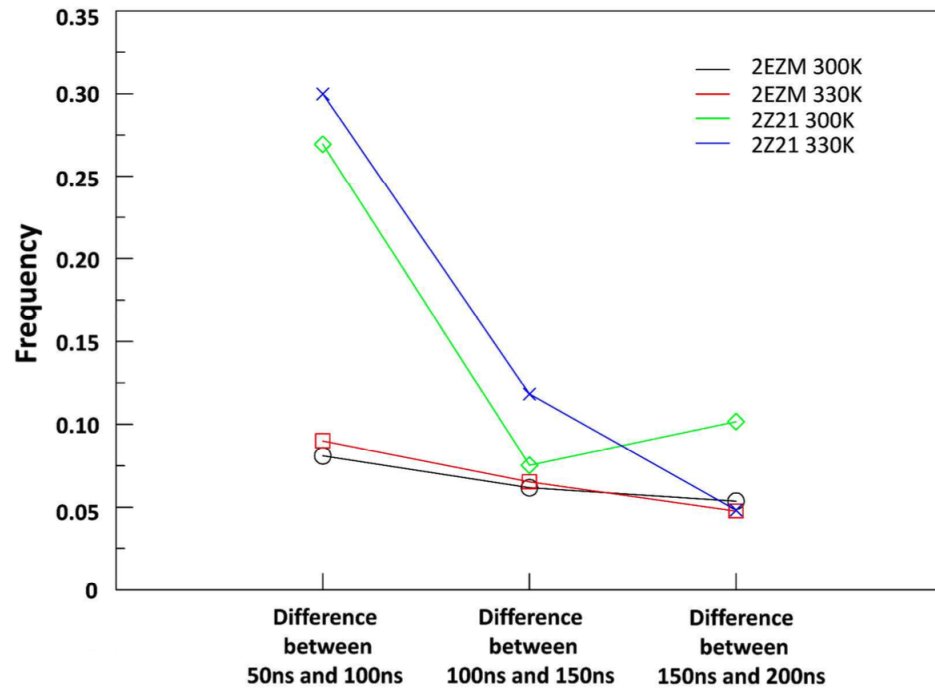
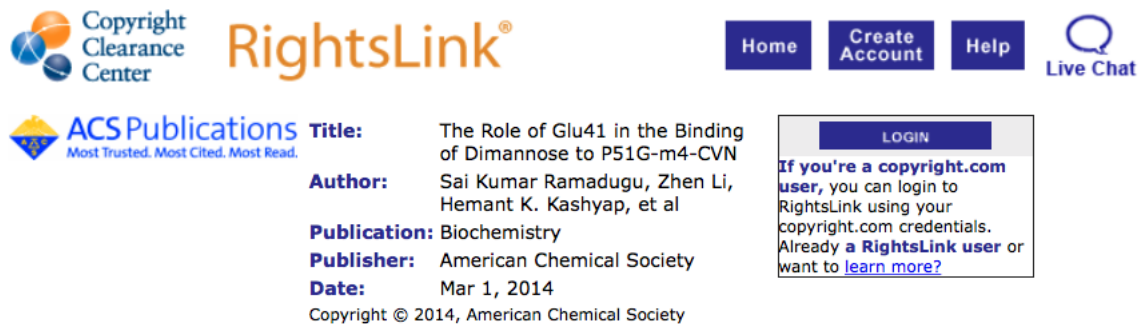


Figure B.1: Convergence analysis of distribution in figure 4.4 of the main text. This same plot is computed at 50, 100, 150 and 200 ns. We see that all features are captured at 50 ns however it is only at times longer than 100 ns that differences between graphs become insignificant for practical analysis. A plot of cumulative differences is also provided.

APPENDIX C COPYRIGHT

Reprinted with permission from Ramadugu, S. K., Li, Z., Kashyap, H. K., Margulis, C. J. (2014). The Role of Glu41 in the Binding of Dimannose to P51G-m4-CVN. *Biochemistry*, 53(9), 1477-1484. Copyright 2014, American Chemical Society.

Reprinted with permission from Li, Z., Bolia, A., Maxwell, J. D., Bobkov, A. A., Ghirlanda, G., Ozkan, S. B., Margulis, C. J. (2015). A Rigid Hinge Region Is Necessary for High-Affinity Binding of Dimannose to Cyanovirin and Associated Constructs. *Biochemistry*, 54(46), 6951-6960. Copyright 2015, American Chemical Society.



Copyright Clearance Center **RightsLink** Home Create Account Help Live Chat

ACS Publications **Title:** The Role of Glu41 in the Binding of Dimannose to P51G-m4-CVN
Author: Sai Kumar Ramadugu, Zhen Li, Hemant K. Kashyap, et al
Publication: Biochemistry
Publisher: American Chemical Society
Date: Mar 1, 2014
 Copyright © 2014, American Chemical Society

LOGIN
 If you're a **copyright.com** user, you can login to RightsLink using your copyright.com credentials. Already a **RightsLink** user or want to [learn more?](#)

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

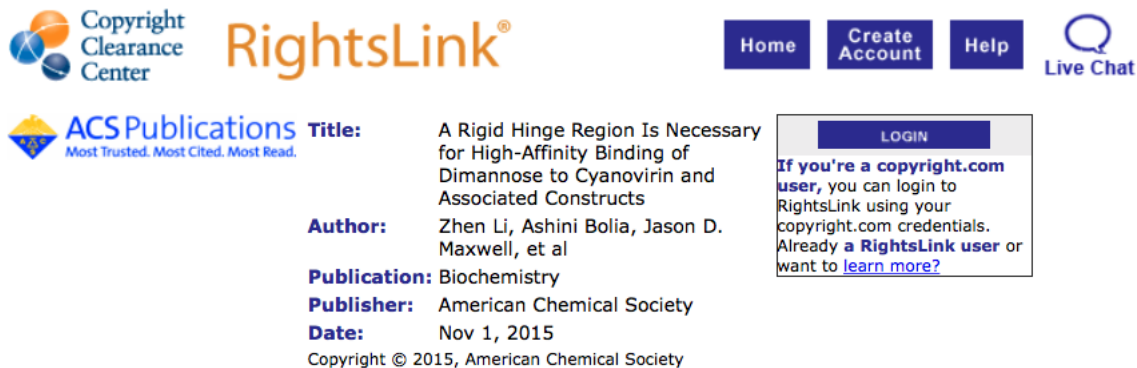
- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from (COMPLETE REFERENCE CITATION). Copyright (YEAR) American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your request. No additional uses are granted (such as derivative works or other editions). For any other uses, please submit a new request.

BACK

CLOSE WINDOW

Copyright © 2016 Copyright Clearance Center, Inc. All Rights Reserved. [Privacy statement](#). [Terms and Conditions](#).
 Comments? We would like to hear from you. E-mail us at customercare@copyright.com

Figure C.1: Copyright permission from The American Chemical Society for the article by S. K. Ramadugu, Z. Li, H. K. Kashyap, and C. J. Margulis, *Biochemistry*, 2014, 53(9), 1477-1484.



Copyright Clearance Center RightsLink® Home Create Account Help Live Chat

ACS Publications Most Trusted. Most Cited. Most Read.

Title: A Rigid Hinge Region Is Necessary for High-Affinity Binding of Dimannose to Cyanovirin and Associated Constructs

Author: Zhen Li, Ashini Bolia, Jason D. Maxwell, et al

Publication: Biochemistry

Publisher: American Chemical Society

Date: Nov 1, 2015

Copyright © 2015, American Chemical Society

LOGIN

If you're a **copyright.com** user, you can login to RightsLink using your copyright.com credentials. Already a **RightsLink** user or want to [learn more?](#)

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from (COMPLETE REFERENCE CITATION). Copyright (YEAR) American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your request. No additional uses are granted (such as derivative works or other editions). For any other uses, please submit a new request.

BACK

CLOSE WINDOW

Copyright © 2016 Copyright Clearance Center, Inc. All Rights Reserved. [Privacy statement](#). [Terms and Conditions](#).
Comments? We would like to hear from you. E-mail us at customer@copyright.com

Figure C.2: Copyright permission from The American Chemical Society for the article by Z. Li, A. Bolia, J. D. Maxwell, A. A. Bobkov, G. Ghirlanda, S. B. Ozkan, and C. J. Margulis, *Biochemistry*, 2015, 54(46), 6951-6960.

REFERENCES

- [1] Jamshed Anwar and David M Heyes. Robust and accurate method for free-energy calculation of charged molecular systems. *The Journal of chemical physics*, 122(22):224117, 2005.
- [2] Johan Åqvist, Carmen Medina, and Jan-Erik Samuelsson. A new method for predicting binding affinity in computer-aided drug design. *Protein engineering*, 7(3):385–391, 1994.
- [3] C Atilgan, ZN Gerek, SB Ozkan, and AR Atilgan. Manipulation of conformational change in proteins by single-residue perturbations. *Biophysical Journal*, 99(3):933–943, 2010.
- [4] Canan Atilgan and Ali Rana Atilgan. Perturbation-response scanning reveals ligand entry-exit mechanisms of ferric binding protein. *PLoS Comput Biol*, 5(10):e1000544, 2009.
- [5] Jan Balzarini. Targeting the glycans of glycoproteins: a novel paradigm for antiviral therapy. *Nature Reviews Microbiology*, 5(8):583–597, 2007.
- [6] Laura G Barrientos, Fátima Lasala, Rafael Delgado, Anthony Sanchez, and Angela M Gronenborn. Flipping the switch from monomeric to dimeric cv-n has little effect on antiviral activity. *Structure*, 12(10):1799–1807, 2004.
- [7] Laura G Barrientos, John M Louis, Istvan Botos, Toshiyuki Mori, Zhaozhong Han, Barry R O’Keefe, Michael R Boyd, Alexander Wlodawer, and Angela M Gronenborn. The domain-swapped dimer of cyanovirin-n is in a metastable folded state: reconciliation of x-ray and nmr structures. *Structure*, 10(5):673–686, 2002.
- [8] Laura G Barrientos, John M Louis, Daniel M Ratner, Peter H Seeberger, and Angela M Gronenborn. Solution structure of a circular-permuted variant of the potent hiv-inactivating protein cyanovirin-n: structural basis for protein stability and oligosaccharide interaction. *Journal of molecular biology*, 325(1):211–223, 2003.
- [9] Laura G Barrientos, Elena Matei, Fátima Lasala, Rafael Delgado, and Angela M Gronenborn. Dissecting carbohydrate–cyanovirin-n binding by structure-guided mutagenesis: functional implications for viral entry inhibition. *Protein Engineering Design and Selection*, 19(12):525–535, 2006.

- [10] Laura G Barrientos, Barry R OKeefe, Mike Bray, Anthony Sanchez, Angela M Gronenborn, and Michael R Boyd. Cyanovirin-n binds to the viral surface glycoprotein, gp 1, 2 and inhibits infectivity of ebola virus. *Antiviral research*, 58(1):47–56, 2003.
- [11] Charles H Bennett. Efficient estimation of free energy differences from monte carlo data. *Journal of Computational Physics*, 22(2):245–268, 1976.
- [12] Herman JC Berendsen, JPM van Postma, Wilfred F van Gunsteren, ARHJ Di-Nola, and JR Haak. Molecular dynamics with coupling to an external bath. *The Journal of chemical physics*, 81(8):3684–3690, 1984.
- [13] Thomas C Beutler, Alan E Mark, René C van Schaik, Paul R Gerber, and Wilfred F van Gunsteren. Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations. *Chemical physics letters*, 222(6):529–539, 1994.
- [14] Carole A Bewley. Rapid validation of the overall structure of an internal domain-swapped mutant of the anti-hiv protein cyanovirin-n using residual dipolar couplings. *Journal of the American Chemical Society*, 123(5):1014–1015, 2001.
- [15] Carole A Bewley. Solution structure of a cyanovirin-n: $\text{Man}\alpha 1\text{-}2\text{man}\alpha$ complex: structural basis for high-affinity carbohydrate-mediated binding to gp120. *Structure*, 9(10):931–940, 2001.
- [16] Carole A Bewley and G Marius Clore. Determination of the relative orientation of the two halves of the domain-swapped dimer of cyanovirin-n in solution using dipolar couplings and rigid body minimization. *Journal of the American Chemical Society*, 122(25):6009–6016, 2000.
- [17] Carole A Bewley, Kirk R Gustafson, Michael R Boyd, David G Covell, Ad Bax, G Marius Clore, and Angela M Gronenborn. Solution structure of cyanovirin-n, a potent hiv-inactivating protein. *Nature Structural & Molecular Biology*, 5(7):571–578, 1998.
- [18] Carole A Bewley, Shigeki Kiyonaka, and Itaru Hamachi. Site-specific discrimination by cyanovirin-n for α -linked trisaccharides comprising the three arms of man 8 and man 9. *Journal of molecular biology*, 322(4):881–889, 2002.
- [19] Carole A Bewley and Sarah Otero-Quintero. The potent anti-hiv protein cyanovirin-n contains two novel carbohydrate binding sites that selectively bind to man8 d1d3 and man9 with nanomolar affinity: implications for binding to the hiv envelope protein gp120. *Journal of the American Chemical Society*, 123(17):3892–3902, 2001.

- [20] Ashini Bolia, Z Nevin Gerek, Ozlem Keskin, Sefika Banu Ozkan, and Kumlesh K Dev. The binding affinities of proteins interacting with the pdz domain of pick1. *Proteins: Structure, Function, and Bioinformatics*, 80(5):1393–1408, 2012.
- [21] Ashini Bolia, Z Nevin Gerek, and S Banu Ozkan. Bp-dock: A flexible docking scheme for exploring protein–ligand interactions based on unbound structures. *Journal of chemical information and modeling*, 54(3):913–925, 2014.
- [22] Ashini Bolia, Brian W Woodrum, Angelo Cereda, Melissa A Ruben, Xu Wang, S Banu Ozkan, and Giovanna Ghirlanda. A flexible docking scheme efficiently captures the energetics of glycan-cyanovirin binding. *Biophysical journal*, 106(5):1142–1151, 2014.
- [23] Anders J Bolmstedt, Barry R O’Keefe, Shilpa R Shenoy, James B McMahon, and Michael R Boyd. Cyanovirin-n defines a new class of antiviral agent targeting n-linked, high-mannose glycans in an oligosaccharide-specific manner. *Molecular Pharmacology*, 59(5):949–954, 2001.
- [24] Stefan Boresch and Martin Karplus. The meaning of component analysis: decomposition of the free energy in terms of specific interactions. *Journal of molecular biology*, 254(5):801–807, 1995.
- [25] I Botos and A Wlodawer. Cyanovirin-n: a sugar-binding antiviral protein with a new twist. *Cellular and Molecular Life Sciences CMLS*, 60(2):277–287, 2003.
- [26] Istvan Botos, Toshiyuki Mori, Laura K Cartner, Michael R Boyd, and Alexander Wlodawer. Domain-swapped structure of a mutant of cyanovirin-n. *Biochemical and biophysical research communications*, 294(1):184–190, 2002.
- [27] Istvan Botos, Barry R O’Keefe, Shilpa R Shenoy, Laura K Cartner, Daniel M Ratner, Peter H Seeberger, Michael R Boyd, and Alexander Wlodawer. Structures of the complexes of a potent anti-hiv protein cyanovirin-n and high mannose oligosaccharides. *Journal of Biological Chemistry*, 277(37):34336–34342, 2002.
- [28] Michael R Boyd, Kirk R Gustafson, James B McMahon, Robert H Shoemaker, Barry R O’Keefe, Toshiyuki Mori, Robert J Gulakowski, Lin Wu, Maria I Rivera, Carolyn M Laurencot, et al. Discovery of cyanovirin-n, a novel human immunodeficiency virus-inactivating protein that binds viral surface envelope glycoprotein gp120: potential applications to microbicide development. *Antimicrobial agents and chemotherapy*, 41(7):1521–1530, 1997.

- [29] Bernard R Brooks, Charles L Brooks, Alexander D MacKerell, Lennart Nilsson, Robert J Petrella, Benoît Roux, Youngdo Won, Georgios Archontis, Christian Bartels, Stefan Boresch, et al. Charmm: the biomolecular simulation program. *Journal of computational chemistry*, 30(10):1545–1614, 2009.
- [30] Bernard R Brooks, Robert E Bruccoleri, Barry D Olafson, David J States, S Swaminathan, and Martin Karplus. Charmm: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of computational chemistry*, 4(2):187–217, 1983.
- [31] David A Case, TA Darden, TE Cheatham, Carlos L Simmerling, J Wang, Robert E Duke, Ray Luo, RC Walker, W Zhang, KM Merz, et al. Amber 11. Technical report, University of California, 2010.
- [32] David C Chan and Peter S Kim. Hiv entry and its inhibition. *Cell*, 93(5):681–684, 1998.
- [33] G Marius Clore and Carole A Bewley. Using conjoined rigid body/torsion angle simulated annealing to determine the relative orientation of covalently linked protein domains from dipolar couplings. *Journal of Magnetic Resonance*, 154(2):329–335, 2002.
- [34] Wolfgang Damm, Antonio Frontera, Julian Tirado-Rives, and William L Jorgensen. Opls all-atom force field for carbohydrates. *Journal of Computational Chemistry*, 18(16):1955–1970, 1997.
- [35] Tom Darden, Darrin York, and Lee Pedersen. Particle mesh ewald: An nlog(n) method for ewald sums in large systems. *The Journal of chemical physics*, 98(12):10089–10092, 1993.
- [36] Warren L DeLano. The pymol molecular graphics system. 2002.
- [37] Barna Dey, Danica LLerner, Paolo Lusso, Michael R Boyd, John H Elder, and Edward A Berger. Multiple antiviral activities of cyanovirin-n: blocking of human immunodeficiency virus type 1 gp120 interaction with cd4 and coreceptor and inhibition of diverse enveloped viruses. *Journal of virology*, 74(10):4562–4569, 2000.
- [38] Mark T Esser, Toshiyuki Mori, Isabelle Mondor, Quentin J Sattentau, Barna Dey, Edward A Berger, Michael R Boyd, and Jeffrey D Lifson. Cyanovirin-n binds to gp120 to interfere with cd4-dependent human immunodeficiency virus type 1 virion binding, fusion, and infectivity but does not affect the cd4 binding site on gp120 or soluble cd4-induced conformational changes in gp120. *Journal of virology*, 73(5):4360–4371, 1999.

- [39] Denis J Evans and Brad Lee Holian. The nose–hoover thermostat. *The Journal of chemical physics*, 83(8):4069–4074, 1985.
- [40] Michael Feig and Charles L Brooks. Recent advances in the development and application of implicit solvent models in biomolecule simulations. *Current opinion in structural biology*, 14(2):217–224, 2004.
- [41] Raimund Fromme, Zivile Katiliene, Petra Fromme, and Giovanna Ghirlanda. Conformational gating of dimannose binding to the antiviral protein cyanovirin revealed from the crystal structure at 1.35 Å resolution. *Protein Science*, 17(5):939–944, 2008.
- [42] Raimund Fromme, Zivile Katiliene, Barbara Giomarelli, Federica Bogani, James Mc Mahon, Toshiyuki Mori, Petra Fromme, and Giovanna Ghirlanda. A monovalent mutant of cyanovirin-n provides insight into the role of multiple interactions with gp120 for antiviral activity. *Biochemistry*, 46(32):9199–9207, 2007.
- [43] Yukiji K Fujimoto and David F Green. Carbohydrate recognition by the antiviral lectin cyanovirin-n. *Journal of the American Chemical Society*, 134(48):19639–19651, 2012.
- [44] Yukiji K Fujimoto, Ryan N TerBush, Vadim Patsalo, and David F Green. Computational models explain the oligosaccharide specificity of cyanovirin-n. *Protein Science*, 17(11), 2008.
- [45] Z Nevin Gerek and S Banu Ozkan. A flexible docking scheme to explore the binding selectivity of pdz domains. *Protein Science*, 19(5):914–928, 2010.
- [46] Z Nevin Gerek and S Banu Ozkan. Change in allosteric network affects binding affinities of pdz domains: analysis through perturbation response scanning. *PLoS Comput Biol*, 7(10):e1002154, 2011.
- [47] Hildegard Geyer, Christel Holschbach, Gerhard Hunsmann, and Josef Schneider. Carbohydrates of human immunodeficiency virus. structures of oligosaccharides linked to the envelope glycoprotein 120. *Journal of Biological Chemistry*, 263(24):11760–11767, 1988.
- [48] Kirk R Gustafson, Raymond C Sowder, Louis E Henderson, John H Cardellina, James B McMahan, Umamaheswari Rajamani, Lewis K Pannell, and Michael R Boyd. Isolation, primary sequence determination, and disulfide bond structure of cyanovirin-n, an anti-hiv (human immunodeficiency virus) protein from the cyanobacterium *Nostoc ellipsosporum*. *Biochemical and biophysical research communications*, 238(1):223–228, 1997.

- [49] Zhaozhong Han, Changyun Xiong, Toshiyuki Mori, and Michael R Boyd. Discovery of a stable dimeric mutant of cyanovirin-n (cv-n) from a t7 phage-displayed cv-n mutant library. *Biochemical and biophysical research communications*, 292(4):1036–1043, 2002.
- [50] Berk Hess, Carsten Kutzner, David Van Der Spoel, and Erik Lindahl. Gromacs 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of chemical theory and computation*, 4(3):435–447, 2008.
- [51] William G Hoover. Canonical dynamics: equilibrium phase-space distributions. *Physical Review A*, 31(3):1695, 1985.
- [52] Viktor Hornak, Robert Abel, Asim Okur, Bentley Strockbine, Adrian Roitberg, and Carlos Simmerling. Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins: Structure, Function, and Bioinformatics*, 65(3):712–725, 2006.
- [53] Shuanghong Huo, Junmei Wang, Piotr Cieplak, Peter A Kollman, and Irwin D Kuntz. Molecular dynamics and free energy analyses of cathepsin d-inhibitor interactions: insight into structure-based ligand design. *Journal of medicinal chemistry*, 45(7):1412–1419, 2002.
- [54] William L Jorgensen, Jayaraman Chandrasekhar, Jeffry D Madura, Roger W Impey, and Michael L Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics*, 79(2):926–935, 1983.
- [55] William L Jorgensen, David S Maxwell, and Julian Tirado-Rives. Development and testing of the opls all-atom force field on conformational energetics and properties of organic liquids. *Journal of the American Chemical Society*, 118(45):11225–11236, 1996.
- [56] George A Kaminski, Richard A Friesner, Julian Tirado-Rives, and William L Jorgensen. Evaluation and reparametrization of the opls-aa force field for proteins via comparison with accurate quantum chemical calculations on peptides. *The Journal of Physical Chemistry B*, 105(28):6474–6487, 2001.
- [57] Jennifer R Keeffe, Priyanthi NP Gnanapragasam, Sarah K Gillespie, John Yong, Pamela J Bjorkman, and Stephen L Mayo. Designed oligomers of cyanovirin-n show enhanced hiv neutralization. *Proceedings of the National Academy of Sciences*, 108(34):14079–14084, 2011.
- [58] Brendan S Kelley, Leng Chee Chang, and Carole A Bewley. Engineering an obligate domain-swapped dimer of cyanovirin-n with enhanced anti-hiv activity. *Journal of the American Chemical Society*, 124(13):3210–3211, 2002.

- [59] Karl N Kirschner, Austin B Yongye, Sarah M Tschampel, Jorge González-Outeiriño, Charlisa R Daniels, B Lachele Foley, and Robert J Woods. Glycam06: a generalizable biomolecular force field. carbohydrates. *Journal of computational chemistry*, 29(4):622–655, 2008.
- [60] Jeffery B Klauda, Richard M Venable, J Alfredo Freites, Joseph W OConnor, Douglas J Tobias, Carlos Mondragon-Ramirez, Igor Vorobyov, Alexander D MacKerell Jr, and Richard W Pastor. Update of the charmm all-atom additive force field for lipids: validation on six lipid types. *The journal of physical chemistry B*, 114(23):7830–7843, 2010.
- [61] John L Klepeis, Kresten Lindorff-Larsen, Ron O Dror, and David E Shaw. Long-timescale molecular dynamics simulations of protein structure and function. *Current opinion in structural biology*, 19(2):120–127, 2009.
- [62] Leonardus MI Koharudin, Lin Liu, and Angela M Gronenborn. Different 3d domain-swapped oligomeric cyanovirin-n structures suggest trapped folding intermediates. *Proceedings of the National Academy of Sciences*, 110(19):7702–7707, 2013.
- [63] Leonardus MI Koharudin, Arturo R Viscomi, Jun-Goo Jee, Simone Ottonello, and Angela M Gronenborn. The evolutionarily conserved family of cyanovirin-n homologs: structures and carbohydrate specificity. *Structure*, 16(4):570–584, 2008.
- [64] Peter A Kollman, Irina Massova, Carolina Reyes, Bernd Kuhn, Shuanghong Huo, Lillian Chong, Matthew Lee, Taisung Lee, Yong Duan, Wei Wang, et al. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Accounts of chemical research*, 33(12):889–897, 2000.
- [65] Bernd Kuhn and Peter A Kollman. Binding of a diverse set of ligands to avidin and streptavidin: an accurate quantitative prediction of their relative affinities by a combination of molecular mechanics and continuum solvent models. *Journal of medicinal chemistry*, 43(20):3786–3791, 2000.
- [66] Cordelia K Leonard, Michael W Spellman, Lavon Riddle, Reed J Harris, James N Thomas, and TJ Gregory. Assignment of intrachain disulfide bonds and characterization of potential glycosylation sites of the type 1 recombinant human immunodeficiency virus envelope glycoprotein (gp120) expressed in chinese hamster ovary cells. *Journal of Biological Chemistry*, 265(18):10373–10382, 1990.

- [67] Zhen Li, Ashini Bolia, Jason D Maxwell, Andrey A Bobkov, Giovanna Ghirlanda, S Banu Ozkan, and Claudio J Margulis. A rigid hinge region is necessary for high-affinity binding of dimannose to cyanovirin and associated constructs. *Biochemistry*, 54(46):6951–6960, 2015.
- [68] Lin Liu, In-Ja L. Byeon, Ivet Bahar, and Angela M. Gronenborn. Domain swapping proceeds via complete unfolding: A 19f- and 1h-nmr study of the cyanovirin-n protein. *Journal of the American Chemical Society*, 134(9):4229–4235, 2012. doi: 10.1021/ja210118w.
- [69] Yinan Liu, Jacob R Carroll, Lindsey A Holt, James McMahon, Barbara Giomarelli, and Giovanna Ghirlanda. Multivalent interactions with gp120 are required for the anti-hiv activity of cyanovirin. *Peptide Science*, 92(3):194–200, 2009.
- [70] Claudio. J. Margulis. Computational study of the dynamics of mannose disaccharides free in solution and bound to the potent anti-hiv virucidal protein cyanovirin. *The Journal of Physical Chemistry B*, 109(8):3639–3647, 2005.
- [71] Siewert J Marrink, Alex H De Vries, and Alan E Mark. Coarse grained model for semiquantitative lipid simulations. *The Journal of Physical Chemistry B*, 108(2):750–760, 2004.
- [72] Elena Matei, William Furey, and Angela M Gronenborn. Solution and crystal structures of a sugar binding site mutant of cyanovirin-n: no evidence of domain swapping. *Structure*, 16(8):1183–1194, 2008.
- [73] Elena Matei, Andrew Zheng, William Furey, Jeremy Rose, Christopher Aiken, and Angela M Gronenborn. Anti-hiv activity of defective cyanovirin-n mutants is restored by dimerization. *Journal of Biological Chemistry*, 285(17):13057–13065, 2010.
- [74] David L Mobley, Élise Dumont, John D Chodera, and Ken A Dill. Comparison of charge models for fixed-charge force fields: small-molecule hydration free energies in explicit solvent. *The Journal of Physical Chemistry B*, 111(9):2242–2254, 2007.
- [75] Toshiyuki Mori and Michael R Boyd. Cyanovirin-n, a potent human immunodeficiency virus-inactivating protein, blocks both cd4-dependent and cd4-independent binding of soluble gp120 (sgp120) to target cells, inhibits scd4-induced binding of sgp120 to cell-associated cxcr4, and dissociates bound sgp120 from target cells. *Antimicrobial agents and chemotherapy*, 45(3):664–672, 2001.
- [76] Shūichi Nosé. A molecular dynamics method for simulations in the canonical ensemble. *Molecular physics*, 52(2):255–268, 1984.

- [77] Barry R O’Keefe, Shilpa R Shenoy, Dong Xie, Wentao Zhang, Jeffrey M Muschik, Michael J Currens, Irwin Chaiken, and Michael R Boyd. Analysis of the interaction between the hiv-inactivating protein cyanovirin-n and soluble forms of the envelope glycoproteins gp120 and gp41. *Molecular pharmacology*, 58(5):982–992, 2000.
- [78] Michele Parrinello and Aneesur Rahman. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied physics*, 52(12):7182–7190, 1981.
- [79] Alexandra Patriksson and David van der Spoel. A temperature predictor for parallel tempering simulations. *Physical Chemistry Chemical Physics*, 10(15):2073–2077, 2008.
- [80] Vadim Patsalo, Daniel P Raleigh, and David F Green. Rational and computational design of stabilized variants of cyanovirin-n that retain affinity and specificity for glycan ligands. *Biochemistry*, 50(49):10698–10712, 2011.
- [81] David A Pearlman. A comparison of alternative approaches to free energy calculations. *The Journal of Physical Chemistry*, 98(5):1487–1493, 1994.
- [82] Jed W Pitera and Wilfred F van Gunsteren. A comparison of non-bonded scaling approaches for free energy calculations. *Molecular Simulation*, 28(1-2):45–65, 2002.
- [83] Sander Pronk, Szilárd Páll, Roland Schulz, Per Larsson, Pär Bjelkmar, Rossen Apostolov, Michael R Shirts, Jeremy C Smith, Peter M Kasson, David van der Spoel, et al. Gromacs 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, 29(7):845–854, 2013.
- [84] Corine Sandström, Olivier Berteau, Emiliano Gemma, Stefan Oscarson, Lennart Kenne, and Angela M Gronenborn. Atomic mapping of the interactions between the antiviral agent cyanovirin-n and oligomannosides by saturation-transfer difference nmr. *Biochemistry*, 43(44):13926–13931, 2004.
- [85] Christopher N Scanlan, John Offer, Nicole Zitzmann, and Raymond A Dwek. Exploiting the defensive sugars of hiv-1 for drug and vaccine design. *Nature*, 446(7139):1038–1045, 2007.
- [86] Shilpa R Shenoy, Laura G Barrientos, Daniel M Ratner, Barry R O’Keefe, Peter H Seeberger, Angela M Gronenborn, and Michael R Boyd. Multisite and multivalent binding between cyanovirin-n and branched oligomannosides: calorimetric and nmr characterization. *Chemistry & biology*, 9(10):1109–1118, 2002.

- [87] Michael R Shirts and Vijay S Pande. Solvation free energies of amino acid side chain analogs for common molecular mechanics water models. *The Journal of chemical physics*, 122(13):134508, 2005.
- [88] Richard H Smith, William L Jorgensen, Julian Tirado-Rives, Michelle L Lamb, Paul AJ Janssen, Christopher J Michejda, and Marilyn B Kroeger Smith. Prediction of binding affinities for tibo inhibitors of hiv-1 reverse transcriptase using monte carlo simulations in a linear response method. *Journal of medicinal chemistry*, 41(26):5272–5286, 1998.
- [89] Thomas Steinbrecher, David L Mobley, and David A Case. Nonlinear scaling schemes for lennard-jones interactions in free energy calculations. *The Journal of chemical physics*, 127(21):214108, 2007.
- [90] TP Straatsma and HJC Berendsen. Free energy of ionic hydration: Analysis of a thermodynamic integration technique to evaluate free energy differences by molecular dynamics simulations. *The Journal of chemical physics*, 89(9):5876–5886, 1988.
- [91] TP Straatsma and JA McCammon. Multiconfiguration thermodynamic integration. *The Journal of chemical physics*, 95(2):1175–1188, 1991.
- [92] Yuji Sugita and Yuko Okamoto. Replica-exchange molecular dynamics method for protein folding. *Chemical physics letters*, 314(1):141–151, 1999.
- [93] D Van Der Spoel, E Lindahl, B Hess, AR van Buuren, E Apol, PJ Meulenhoff, DP Tieleman, ALTM Sijbers, KA Feenstra, R van Drunen, et al. Gromacs user manual version 4.5. 4, 2010. *GROMACS-Groningen Machine for Chemical Simulations*.
- [94] Kenno Vanommeslaeghe, Elizabeth Hatcher, Chayan Acharya, Sibsankar Kundu, Shijun Zhong, Jihyun Shim, Eva Darian, Olgun Guvench, P Lopes, Igor Vorobyov, et al. Charmm general force field: A force field for drug-like molecules compatible with the charmm all-atom additive biological force fields. *Journal of computational chemistry*, 31(4):671–690, 2010.
- [95] Ivan I Vorontsov and Osamu Miyashita. Solution and crystal molecular dynamics simulation study of m4-cyanovirin-n mutants complexed with di-mannose. *Biophysical journal*, 97(9):2532–2540, 2009.
- [96] Ivan I Vorontsov and Osamu Miyashita. Crystal molecular dynamics simulations to speed up mm/pb (gb) sa evaluation of binding free energies of di-mannose deoxy analogs with p51g-m4-cyanovirin-n. *Journal of computational chemistry*, 32(6):1043–1053, 2011.

- [97] Wei Wang, Oreola Donini, Carolina M Reyes, and Peter A Kollman. Biomolecular simulations: recent developments in force fields, simulations of enzyme catalysis, protein-ligand, protein-protein, and protein-nucleic acid noncovalent interactions. *Annual review of biophysics and biomolecular structure*, 30(1):211–243, 2001.
- [98] Fan Yang, Carole A Bewley, John M Louis, Kirk R Gustafson, Michael R Boyd, Angela M Gronenborn, G Marius Clore, and Alexander Wlodawer. Crystal structure of cyanovirin-n, a potent hiv-inactivating protein, shows unexpected domain swapping. *Journal of molecular biology*, 288(3):403–412, 1999.