

Practical Applications of Molecular Modeling Pertaining to Oxidative Damage and Disease

William Joseph Allen

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in
partial fulfillment of the requirements for the degree of

Doctor of Philosophy
In
Biochemistry

David R. Bevan, Chair
Richard F. Helm
Jianyong Li
Edward J. Smith
James M. Tanko

March 23, 2011
Blacksburg, VA

Keywords: molecular modeling, monoamine oxidase B, p53 C-terminal domain, B2 suppressor
of RNA silencing, *Arachis Ara h 2* protein, lipid bilayer analysis

Copyright 2011 by William Joseph Allen

Practical Applications of Molecular Modeling Pertaining to Oxidative Damage and Disease

William Joseph Allen

ABSTRACT

Molecular modeling is a term referring to the study of proteins, nucleic acids, lipids, and other bio-, macro-, or small molecules at the atomistic level using a combination of computational methods, physico-chemical principles, and mathematical functions. It can be generally subdivided into two areas: molecular mechanics, which is the treatment of atoms and bonds as Newtonian particles and springs, and quantum mechanics, which models electronic behaviors using the Schrödinger equation and wavefunctions. Each technique is a powerful tool that, when used alone or in combination with wet lab experiments, can yield useful results, the products of which have broad applications in studying human disease models, oxidative damage, and other biomolecular processes that are otherwise not easily observed by experiment alone. Within this document, we study seven different such systems. This includes the mode of inhibitor binding to the enzyme monoamine oxidase B, the active site mechanism of that same enzyme, the dynamics of the unstructured p53 C-terminal domain in complex with globular, structured proteins, the process of the viral protein B2 unbinding from double-stranded RNA, and a focus on the dynamics of a variable loop in the antigenic peanut protein Ara h 2. In addition to those conventional molecular modeling studies, several of which were done in tandem with wet lab experiment, we also discuss the validation of charges and charge group parameters for small molecules used in molecular mechanics, and the development of software for the analysis of lipid bilayer systems in molecular mechanics simulations. As computational resources continue to evolve, and as more structural information becomes available, these methods are becoming an integral part of the study of biomolecules in the context of disease.

*These works are dedicated to
my loving wife, Krista.*

Acknowledgments

First, I owe my deepest gratitude to my advisor, David Bevan. Performing the projects described herein and writing this dissertation would not have been possible without his continuing support and guidance. I sincerely thank him for providing me the opportunity to pursue graduate research in his lab, and for creating an environment in which I was able to grow and excel as a researcher. I am also very grateful for the rest of my committee members, including Rich Helm, Jianyong Li, Ed Smith, and James Tanko. Through meaningful discussions, they have each helped me to refine and expand upon my projects, resulting in better work, better publications, and greater scientific impact. I thank each of them for their patience and for their belief in my abilities.

Many other members of the faculty at Virginia Tech have been instrumental to my growth and success as a graduate student. Erin Dolan and Glenda Gillaspay have been incredibly helpful and supportive as mentors, especially in teaching and other academic pursuits. Florian Schubot, E. M. Gregory, and several other faculty have been fantastic teachers and have made my coursework both interesting and valuable. Sue Duncan, Tim Long, Carla Finkielstein, and Daniel Capelluto have each supported part of my work financially and through their intellectual contributions. The rest of the faculty and staff of the Department of Biochemistry have created a friendly and inviting atmosphere in which it was a pleasure to work. For each of these things, I am exceedingly grateful.

I extend my thanks to my fellow lab mates, Justin Lemkul, Nikki Lewis, and Somaye Badiyan. They have each been exceptionally constructive towards the improvement of my work through lab meetings, informal discussions, e-mail exchanges, critical proof-reads, general advice, and the list goes on. I thank Annie Aigster and my other fellow MILES graduate students for encouragement and camaraderie through the IGERT program. I also thank Michelle Grimm, Sheena Friend, and the other co-authors on these interdisciplinary projects for finding an interest in molecular modeling and believing that through me its products could supplement their own research.

Finally, but most of all, I thank my family and friends. Daniel Ragheb and Mike Wiley have been great friends to me these past four and a half years, and it is through their mutual encouragement and support that we have all succeeded. Lastly, my wife, Krista, has been incredibly patient and supportive throughout my graduate career. These works are dedicated to her because she is both my motivation and my reason to work hard and to try my best in everything I do.

Attribution

Many of the works presented in this document are the product of collaboration in terms of both the research and the writing. The contributing coauthors (other than those listed on the title page of this document) include Professor Emeritus Neal Castagnoli, Jr., graduate student Michelle L. Grimm, and undergraduate Meghan Finn from the Department of Chemistry at Virginia Tech; Assistant Professors Daniel G. S. Capelluto and Carla V. Finkielstein, graduate student Sheena A. Friend, and Professor Khidir W. Hilu from the Department of Biological Sciences at Virginia Tech; Assistant Professors Zachary N. Adelman and Kevin M. Myles, and graduate student Michael R. Wiley from the Entomology Department at Virginia Tech; and graduate student Justin A. Lemkul from the Department of Biochemistry at Virginia Tech. Within each chapter, specific contributions to the research and writing made by these authors are described in detail. All other works presented herein are my own. Copyrights for published articles are cited within individual chapters.

Table of Contents

Chapter 1: Introduction to Molecular Modeling	1
Molecular Mechanics.....	1
Quantum Mechanics.....	4
Practical Applications.....	6
References.....	8
Chapter 2: Steered Molecular Dynamics Simulations Reveal Important Mechanisms in Reversible Monoamine Oxidase B Inhibition	9
Attribution.....	9
Abstract.....	9
Introduction.....	10
Materials and Methods.....	12
File Preparation.....	12
Conventional MD Simulations.....	14
Steered MD Simulations.....	15
Results.....	16
Molecular Dynamics of MAO B.....	16
Steered Molecular Dynamics of MAO B Inhibitors.....	24
Discussion.....	32
Acknowledgments.....	34
Supplemental Information.....	35
Inhibitor and Cofactor Parameters.....	35
Tabulated Δ PMF and Peak Force Values.....	43
Additional RMSF Plots.....	44
Linear Regression Plots.....	45
References.....	48

Chapter 3: Reaction of Benzophenone Triplet with Aliphatic Amines. What a Potent Neurotoxin Can Tell Us About the Reaction Mechanism 52

Attribution.....	52
Abstract.....	52
Introduction.....	53
Materials and Methods.....	58
Results and Discussion.....	60
Calculations.....	60
Laser Flash Photolysis.....	65
Conclusions.....	68
Acknowledgments.....	69
Supplemental Information.....	70
Concentration Profiles for MPTP and N-cyclopropyl MPTP.....	70
References.....	73

Chapter 4: Modeling the Relationship Between the p53 C-Terminal Domain and Its Binding Partners Using Molecular Dynamics 77

Attribution.....	77
Abstract.....	77
Introduction.....	78
Computational Methods.....	80
PDB File Preparation.....	80
Molecular Dynamics Simulations.....	82
Analysis.....	83
Results and Discussion.....	86
1DT7.....	86
1MA3.....	94
1H26.....	101
1JSP.....	107
1XQH.....	114
Conclusions.....	119
Acknowledgments.....	120

Supplemental Information.....	121
Modified Lysine Parameters.....	121
Supplemental Tables for Energy Contributions.....	123
References.....	126
Chapter 5: Using Modeling to Predict Attenuated Mutants of the Viral Protein B2, A Potent Suppressor of RNA Silencing	133
Attribution.....	133
Abstract.....	133
Introduction.....	134
Materials and Methods.....	136
Computational.....	136
Experimental.....	137
Results.....	138
Energy Profiles for NoV B2-dsRNA Interaction.....	138
Steered MD Simulations of NoV B2-dsRNA Complex.....	143
Self-Directed NoV RNA1 F49 Mutants in Cells.....	146
Discussion.....	147
References.....	150
Chapter 6: Modeling Loop Dynamics of Antigenic Peanut Protein Orthologs from the Genus <i>Arachis</i>	153
Attribution.....	153
Abstract.....	153
Introduction.....	154
Methods.....	154
Results and Discussion.....	157
Part I: RMSD Analysis.....	157
Part II: Radius of Gyration (R_g)	160
Part III: RMSF Analysis.....	163
Part IV: Clustering Analysis.....	166
Part V: DSSP Analysis.....	169

Part VI: Solvent-Accessible Surface Area.....	174
Conclusions.....	176
References.....	178
Chapter 7: Practical Considerations for Building GROMOS-Compatible Small-Molecule Topologies	180
Attribution.....	180
Abstract.....	180
Introduction.....	181
Methods.....	183
Condensed-Phase and Gas-Phase-Systems.....	183
Interfacial Amino Acid Systems.....	184
UDP-Galactopyranose Mutase.....	188
Results and Discussion.....	190
Condensed-Phase and Gas-Phase-Systems.....	190
Interfacial Amino Acid Systems.....	193
UDP-Galactopyranose Mutase.....	202
Strategies for Charge Assignment in GROMOS Topologies.....	206
Conclusions.....	211
Acknowledgments.....	212
Supplemental Information.....	213
References.....	220
Chapter 8: GridMAT-MD: A Grid-Based Membrane Analysis Tool for Use With Molecular Dynamics	226
Attribution.....	226
Abstract.....	226
Introduction.....	227
Description.....	228
Processing the Coordinate File.....	229
Calculating Thickness.....	230
Calculating Area Per Lipid Headgroup.....	233

Validation Simulations.....	234
Methods.....	234
Results.....	236
Summary.....	238
Acknowledgments.....	239
References.....	240
 Chapter 9: Conclusions	 242
 Appendix	 244
Appendix A: Sample <i>.mdp</i> Files.....	244
Appendix B: Perl Scripts.....	256

List of Figures

Chapter 2

2.1	The rmsd of the protein backbone, residues 1-485.....	17
2.2	MAO B fluctuation profile for bilayer and solvent forms.....	18
2.3	Overview of MAO B homodimer embedded in a lipid bilayer.....	20
2.4	Comparison of the area of active site channel opening when MAO B is in a bilayer (a) vs MAO B in a box of solvent (b).....	21
2.5	Partial density plot for important lipid and protein components along the z -axis.....	22
2.6	Bilayer thickness plots.....	23
2.7	Δ PMF plots for inhibitor unbinding.....	25
2.8	Average Δ PMF values for pulling inhibitors from the protein active site.....	28
2.9	Summated and averaged energies of interaction between specific inhibitors and MAO B when pulling in a bilayer.....	29
2.10	Sum of binding energy components for each inhibitor.....	30
2.11	View of the MAO B active site during inhibitor C17 (green) egress at 0.5 nm displacement (top) and 1.5 nm displacement (bottom).....	32
S2.1	All six RMSF plots comparing degree of fluctuation per MAO B residue in solvent vs. bilayer.....	44
S2.2	Linear regression plots and R^2 values correlating the ΔG_{dissoc} to peak force for inhibitor unbinding in both media.....	45
S2.3	Linear regression plots and R^2 values correlating the ΔG_{dissoc} to Δ PMF for inhibitor unbinding in both media.....	46
S2.4	Linear regression plots and R^2 values correlating the ΔG_{dissoc} to the sum of interaction energies during steered MD.....	47

Chapter 3

3.1	Reaction coordinate diagram for the ring opening of $\mathbf{12}^{*+}$ calculated at the UHF/6-31G* level of theory.....	62
-----	--	----

3.2	Reaction coordinate diagram for the ring opening of 13^{•+} calculated at the UHF/6-31G* level of theory.....	62
3.3	Reaction coordinate diagram for the ring opening of 14^{•+} calculated at the UHF/6-31G* level of theory.....	63
3.4	Depiction of how variation on C-C bond length of 14^{•+} is accompanied by simultaneous rotation of the <i>N</i> -methyl group.....	64
3.5	Reaction coordinate diagram for the ring opening of 15^{•+} calculated at the UHF/6-31G* level of theory.....	64
3.6	Transient absorption spectra of (i) MPTP [1] in CH ₃ CN and, (ii) <i>N</i> -cyclopropylphenyl MPTP [8] in CH ₃ CN.....	66
3.7	Transient absorption spectra of (i) MPTP [1] in benzene, (ii) MPTP [1] in CH ₃ CN in the presence of LiOCL ₄	66
3.8	Transient absorption spectra of (i) <i>N</i> -cyclopropylphenyl MPTP [8] in benzene, (ii) <i>N</i> -cyclopropylphenyl MPTP [8] in CH ₃ CN in the presence of LiOCL ₄	67
S3.1	Concentration profile for the reaction of benzophenone triplet (³ BP) with MPTP [1] in benzene.....	70
S3.2	Concentration profile for the reaction of benzophenone triplet (³ BP) with MPTP [1] in acetonitrile.....	70
S3.3	Concentration profile for the reaction of benzophenone triplet (³ BP) with MPTP [1] in acetonitrile in the presence of 0.5M LiOCL ₄	71
S3.4	Concentration profile for the reaction of benzophenone triplet (³ BP) with <i>N</i> -cyclopropyl MPTP [2] in benzene.....	71
S3.5	Concentration profile for the reaction of benzophenone triplet (³ BP) with <i>N</i> -cyclopropyl MPTP [2] in acetonitrile.....	72
S3.6	Concentration profile for the reaction of benzophenone triplet (³ BP) with <i>N</i> -cyclopropyl MPTP [2] in acetonitrile in the presence of 0.5M LiOCL ₄	72

Chapter 4

4.1	NMR and simulation subspace overlap for 1DT7.....	84
4.2	NMR and simulation subspace overlap for 1JSP.....	85
4.3	DSSP analysis for 1DT7.....	87
4.4	RMSD and RMSF analyses for 1DT7.....	88
4.5	Surface rendering and backbone trace of the p53 CTD (colored in red) bound to cartoon representation of one S100B β-subunit (colored in gray).....	89
4.6	PC analysis for 1DT7.....	90
4.7	Potential energy of interaction between the p53 CTD and S100B(ββ) by residue.....	91

4.8	Hydrophobic SASA of the p53 CTD binding site on each apo-binding partner.....	92
4.9	Average number of hydrogen bonds formed between the main chain of each p53 CTD residue and the receptor, S100B($\beta\beta$).....	93
4.10	DSSP analysis for 1MA3.....	95
4.11	RMSD and RMSF analyses for 1MA3.....	96
4.12	PC analysis for 1MA3.....	97
4.13	Cartoon representation of the p53 CTD (colored in red), bound to Sir2.....	99
4.14	Average number of hydrogen bonds formed between the main chain of each p53 CTD residue and Sir2.....	99
4.15	Potential energy of interaction between the p53 CTD and Sir2 by residue.....	100
4.16	DSSP analysis for 1H26.....	101
4.17	RMSD and RMSF analysis for 1H26.....	103
4.18	PC analysis for 1H26.....	104
4.19	Surface rendering and cartoon of the p53 CTD (colored in red).....	105
4.20	Potential energy of interaction between the p53 CTD and cyclin A by residue.....	106
4.21	DSSP analysis for 1JSP.....	108
4.22	RMSD and RMSF analyses for 1JSP.....	109
4.23	Surface rendering and cartoon of the p53 CTD (colored in red).....	110
4.24	PC analysis for 1JSP.....	111
4.25	Potential energy of interaction between the p53 CTD and the CBP bromodomain by residue.....	113
4.26	DSSP analysis for 1XQH.....	114
4.27	RMSD and RMSF analyses for 1XQH.....	115
4.28	Surface rendering and cartoon of the p53 CTD (colored in red).....	116
4.29	PC analysis for 1XQH.....	117
4.30	Potential energy of interaction between the p53 CTD and Set9 by residue.....	119
S4.1	Atom numbering scheme for acetylated lysine (AcLys), methylated lysine (MeLys), and unmodified lysine (Lys).....	122

Chapter 5

5.1	Model of NoV B2 WT protein in complex with dsRNA.....	139
5.2	Energy profile for NoV B2 WT broken down by individual residue.....	140

5.3	NoV F49 mutants change B2 sequence without affecting the protein A/B1 amino acid sequence.....	141
5.4	Energy <i>difference</i> profiles for NoV B2 mutants.....	142
5.5	Force profiles for steered MD simulations.....	144
5.6	Average peak force on spring from pulling simulations.....	145
5.7	NoV F49 B2 mutants affect replication in S2 cells.....	147

Chapter 6

6.1	Sequence alignment of different Ara h 2 proteins used in this study.....	155
6.2	RMSD values measuring backbone deviation.....	159
6.3	Radius of gyration, or ‘foldedness’ of protein backbone.....	162
6.4	RMSF analysis of bat2-1 across all 60 ns of simulation (black line), during the last 40 ns of simulation (red line) and during the last 20 ns of simulation (green line).....	163
6.5	RMSF values measuring sidechain fluctuation for the last 20 ns of simulation.....	165
6.6	Initial conformation of <i>bat2</i> protein.....	166
6.7	Top members of each of the 4 largest <i>bat2</i> clusters.....	168
6.8	Secondary structure content of <i>bat2</i> as a function of time for each replicate.....	170
6.9	Secondary structure content of <i>dar</i> as a function of time for each replicate.....	170
6.10	Secondary structure content of <i>dur</i> as a function of time for each replicate.....	171
6.11	Secondary structure content of <i>dur2</i> as a function of time for each replicate.....	171
6.12	Secondary structure content of <i>glab</i> as a function of time for each replicate.....	172
6.13	Secondary structure content of <i>ipa</i> as a function of time for each replicate.....	172
6.14	Secondary structure content of <i>mace</i> as a function of time for each replicate.....	173
6.15	Secondary structure content of <i>tris2</i> as a function of time for each replicate.....	173
6.16	Secondary structure content of <i>hypo</i> as a function of time for each replicate.....	174
6.17	The SASA of the disordered loop between $\alpha 2$ and $\alpha 3$ for the last 20 ns of each replicate simulation.....	175
6.18	The (n)SASA of the disordered loop between $\alpha 2$ and $\alpha 3$ for the last 20 ns of each replicate simulation.....	176

Chapter 7

7.1	Comparison of GROMOS96 43A1 parameters and PRODRG charges and charge groups for hexane (top), ethanol (middle) and <i>p</i> -cresol (bottom).....	184
-----	---	-----

7.2	Initial position of lysine in the cyclohexane-water interfacial system.....	186
7.3	Comparison of GROMOS96 43A1 parameters and PRODRG charges and charge groups for alanine (top), asparagine (middle) and aspartate (bottom).....	187
7.4	Comparison of GROMOS96 43A1 parameters and PRODRG charges and charge groups for isoleucine (top), lysine (middle) and serine (bottom).....	188
7.5	Comparison of GROMOS96 43A1 parameters and PRODRG charges and charge groups for FAD.....	189
7.6	RDF for different atom pairs in (A) ethanol and (B) <i>p</i> -cresol.....	192
7.7	Overlays of amino acid positions generated by GROMOS96 43A1 parameters.....	195
7.8	Overlays of amino acid positions generated by PRODRG parameters.....	196
7.9	(A) Electrostatic and (B) Lennard-Jones interaction energies between the amino acids (listed on the x-axis) and water.....	198
7.10	RDF for SPC oxygen (O_w) atoms with respect to lysine sidechain atoms using (A) GROMOS 43A1 and (B) PRODRG parameters.....	199
7.11	(A) Backbone and (B) sidechain hydrogen bonds to water.....	201
7.12	Alignment of FAD structures, with the reference position (the position of FAD following equilibration) for each simulation shown in gray.....	203
7.13	RMSF of the UGM backbone.....	205

Chapter 8

8.1	Flowchart of GridMAT-MD logic structure.....	229
8.2	Diagram illustrating GridMAT-MD decision structure in choosing protein atoms.....	232
8.3	Schematic diagram of the extraction of thickness from a lipid bilayer.....	233
8.4	Snapshots from the KALP ₁₅ -DPPC trajectory.....	236
8.5	The KALP ₁₅ -DPPC system, at 50 ns, in a traditional rendering.....	237
8.6	Representation of area per lipid headgroup in the top leaflet, from the final configuration of the system.....	238

Chapter 9

9.1	Number of citations containing the keyword “molecular dynamics” arranged by year.....	243
-----	---	-----

List of Schemes

Chapter 2

2.1	Structures of MAO B inhibitors used in this study.....	13
-----	--	----

Chapter 3

3.1	Proposed pathways for MAO oxidations.....	54
3.2	The MAO B catalyzed bioactivation of MPTP [1] to MPDP ⁺ [2 ⁺] and MPP ⁺ [3 ⁺].....	55
3.3	Reaction of benzophenone triplet (³ BP) with aromatic amines.....	55
3.4	Structure of compound 1 •.....	56
3.5	Structure of compounds 1 , 7 , and 8	56
3.6	Ring opening of cyclopropylcarbinyl neutral radical.....	57
3.7	Ring opening of <i>N</i> -(2-phenylcyclopropyl) aminyl radical cations.....	58
3.8	Structures of compounds 11-15	61
3.9	Resonance stabilization of the distonic radical cation derived from 8 • ⁺ and 14 • ⁺	65
3.10	Reaction pathways for ³ BP with MPTP 1) SET, and 2) HAT.....	67

List of Tables

Chapter 2

2.1	Summary of inhibitor structures arranged by K_i	13
2.2	Position of important groups along z -coordinate	22
S2.1	Parameters for 1,4-diphenyl-2-butene (1PB).....	35
S2.2	Parameters for N-methyl-1(R)-aminoindan (RM1).....	36
S2.3	Parameters for 1H-indole-2,3-dione (isatin, ISN).....	37
S2.4	Parameters for farnesol (FOH).....	38
S2.5	Parameters for (S)-(+)-2-[4-(fluorobenzyloxy-benzylamino) propionamide] (SAG)...	39
S2.6	Parameters for 7-[(3-chlorobenzyl)oxy]-2-oxo-2H-chromene-4-carbaldehyde (C17)...	40
S2.7	Parameters for 7-[(3-chlorobenzyl)oxy]-4-[(methylamino)methyl]-2H-chromen-2-one (C18).....	41
S2.8	Parameters for flavin adenine dinucleotide (FAD).....	42
S2.9	Average Δ PMF and peak force values for steered MD simulations.....	43

Chapter 3

3.1	Rate constants for the reaction of 3 BP with 1 and 2	68
-----	---	----

Chapter 4

4.1	Radius of gyration (R_g).....	85
4.2	Secondary structure content.....	85
S4.1	Acetylated lysine parameters.....	121
S4.2	Methylated lysine parameters.....	121
S4.3	Original lysine parameters from ffG53a6.rtp (listed as [LYSH]).....	122
S4.4	Energy contributions for the interaction between the p53 CTD and S100B($\beta\beta$).....	123
S4.5	Energy contributions for the interaction between the p53 CTD and Sir2.....	123
S4.6	Energy contributions for the interaction between the p53 CTD and cyclin A.....	124

S4.7	Energy contributions for the interaction between the p53 CTD and the CBP bromodomain.....	124
S4.8	Energy contributions for the interaction between the p53 CTD and Set9.....	125

Chapter 6

6.1	System-specific simulation conditions.....	157
6.2	Cluster results.....	167
6.3	Helix locations in starting structures (as determined by PyMOL).....	169
6.4	Loop location and length for each protein.....	175

Chapter 7

7.1	Density and ΔH_{vap} for small molecules.....	191
7.2	Charges calculated by Antechamber semi-empirical methods.....	209
7.3	Charges calculated by methods available in Spartan '04.....	210
S7.1	Charges calculated by Antechamber semi-empirical methods (full).....	214
S7.2	Charges calculated by methods available in Spartan '04 (full).....	217

List of Abbreviations

AGO2	Argonaute-2
AM1	Austin model 1
APL	area per lipid
ATP	adenosine triphosphate
BHK	baby hamster kidney
³ BP	benzophenone triplet
cAMP	cyclic adenosine monophosphate
CBP	CREB-binding protein
CREB	cAMP response element-binding
CTD	C-terminal domain
DET	dissociative electron transfer
DFT	density functional theory
DPPE	di-palmitoyl-phosphatidyl-choline
dsRNA	double-stranded RNA
DSSP	define of secondary structure of proteins
DTBPO	di- <i>t</i> -butoxyl peroxide
EC-ESI-MS	electrochemistry coupled to electrospray mass spectrometry
EPR	electron paramagnetic resonance
ESP	electrostatic charge
FAD	flavin adenine dinucleotide
FHV	flock house virus
FMN	flavin mononucleotide
GAFF	General AMBER force field
ΔG_{dissoc}	Gibbs free energy of dissociation
GridMAT-MD	grid-based membrane analysis tool for use with molecular dynamics
HAT	hydrogen atom transfer
HF	Hartree-Fock

ΔH_{vap}	heat of vaporization
LC-MS	liquid chromatography—mass spectrometry
LFP	laser flash photolysis
LINCS	linear constraint solver
MAO	monoamine oxidase
MD	molecular dynamics
MM	molecular mechanics
MP2	second-order Møller-Plesset
MPTP	1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine
NAD	nicotinamide-adenine dinucleotide
NMR	nuclear magnetic resonance
NoV	Nodamura virus
NPT	constant <u>N</u> umber of atoms, <u>P</u> ressure, and <u>T</u> emperature
(n)SASA	normalized solvent-accessible surface area
nt	nucleotides
NVT	constant <u>N</u> umber of atoms, <u>V</u> olume, and <u>T</u> emperature
ORF	open reading frame
P-LINCS	parallel linear constraint solver
PC	principal components
pCDK2	Phosphorylated cyclin-dependant protein kinase 2
PDB	Protein Data Bank
PME	particle-mesh Ewald
Δ PMF	potential of mean force
POPC	palmitoyl-oleoyl-phosphatidyl-choline
POPE	palmitoyl-oleoyl-phosphatidyl-ethanolamine
PTM	post-translational modification
QM	quantum mechanics
RDF	radial distribution function
RISC	RNA silencing complex
R_g	radius of gyration
RMSD	root mean square deviation
RMSF	root mean square fluctuation

RNAi	RNA interference
SASA	solvent-accessible surface area
SET	single electron transfer
siRNA	small interfering RNA
SPC	simple point charge
SRS	suppressor of RNA silencing
SSIP	solvent separated ion pair
TAD	transactivation domain
^t BuO*	<i>t</i> -butoxyl radical
UDP	uridine diphosphate
UHF	unrestricted Hartree-Fock
UGM	UDP-galactopyranose mutase
UMP2	unrestricted second-order Møller-Plesset
UMP4STDQ	unrestricted fourth-order Møller-Plesset with single, double, triple, and quadruple excitations
VDW	van der Waals
WT	wild-type

Chapter 1

Introduction to Molecular Modeling

The advent and development of new technologies can change how researchers approach a scientific question. One such tool that is changing how we approach questions in the life sciences is *Molecular Modeling*. Modeling, for short, is a method that uses structural information of biomolecules along with computational methods, physico-chemical principles, and mathematical functions to answer scientific questions from a unique perspective. Modeling is becoming much more widely used in part because of better availability of structural data in databanks such as the Research Collaboratory for Structural Biology (RCSB) (1), and also because of an increasing standard in computing power; many of the world's fastest supercomputers have moved past teraflop capacity to the *petaflop* scale.

Through modeling we can study nano-scale atomistic phenomena and interactions, things that are oftentimes quite difficult to study by wet-lab experimentation. Modeling has important applications in drug design, studying diseases of protein folding or aggregation, elucidating reaction mechanisms including those that are related to oxidative stress, and many other biophysical purposes. Pursuing this kind of research in tandem with wet-lab experiment can save researchers valuable time and money in the long run. Within this document, we will discuss projects that apply two specific sub-disciplines of molecular modeling – molecular mechanics and quantum mechanics, and their practical applications in certain disease models.

Molecular Mechanics

The goal of molecular mechanics (MM) is to find the potential energy minimum of a molecule by (*i*) calculating the potential energy of interaction between pairs of interacting atoms,

then (ii) systematically updating the position of each atom in small increments. This process is repeated until subsequent atomic position updates only result in a positive change in potential energy, at which point it can be said that a potential energy minimum has been reached. There are several important underlying principals that make MM possible. First, MM presumes that atoms and bonds act as spherical particles (of defined radius) connected by springs. Second, potential energies of interaction imparted on single atoms from multiple sources are additive. Third, because electronic motion is much slower than atomic motion, individual electrons need not be modeled; rather each atom is assigned a charge state. In pure MM, bonds are never broken or formed, and the charge state of nuclei cannot change. This final assumption is called the Born-Oppenheimer approximation.

Molecular dynamics (MD) simulation is a technique that uses the underlying principles of MM to solve Newton's equations of motion for systems of interacting atoms. (It is this technique and its applications that form the bulk of the experiments described in this document). Several software packages have been developed to perform MD simulations, and the exact equations each one used may vary slightly between different distributions. Below, we will be discussing the theory used by the widely accepted open-source package GROMACS, which was employed for all of the MD simulation described in this document (2). To begin, we examine the different factors that contribute to the potential energy of interaction. There are three bonded terms – bond stretching (2-atom interactions), bond angles (3-atom interactions), and dihedral angles (4-atom interactions). The sum of these terms make up the bonded interactions:

$$V_{bonded} = \frac{1}{2} \sum_{bonds} k_b (b - b_0)^2 + \frac{1}{2} \sum_{angles} k_\theta (\theta - \theta_0)^2 + \frac{1}{2} \sum_{dihedrals} k_\phi [1 + \cos(n\phi - \phi_s)]$$

To summarize this function, the bonded potential energy (V_{bonded}) of a given atom is the sum of all of the three types of bonded interactions. In the first term, bond stretching, k_b is a harmonic spring constant, b is the measured bond length, and b_0 is the equilibrium bond length – a value that is unique for different pairs of covalently bound atoms. In the second term, bond angles, k_θ is a harmonic constant, θ is the measured bond angle, and θ_0 is the equilibrium bond angle – another value that is unique for different triads of atoms. Finally, in the third term, dihedral angles, k_ϕ is another harmonic constraint, and the $(n\phi - \phi_s)$ term is the difference between the

measured dihedral angle and the equilibrium dihedral angle. In practice, these terms are measured and summed simultaneously for each atom in a system of N interacting atoms.

In addition to the bonded terms, there are two non-bonded terms; the Lennard-Jones potential measures the van der Waals interactions between atomic radii, and the Coulomb potential measures the electrostatic interactions between atomic charges. The sum of the non-bonded interactions takes the form:

$$V_{nonbonded} = \sum_{\substack{\text{Lennard} \\ \text{Jones}}} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \sum_{\text{Coulomb}} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}$$

To summarize this equation, the non-bonded potential energy of a given atom is the sum of two terms. The first term, the Lennard-Jones potential, employs unique ϵ and σ values for each possible pair of atoms, i and j . The ϵ value is the minimum potential energy of interaction for the two interacting species at the ideal radius of separation, and the σ value is the distance at which the potential of interaction is zero. The final variable, r_{ij} is the radius of separation between atoms i and j . In the second term, the Coulomb potential, the energy of interaction between two atoms (i and j) depends on the charges of each atom (q_i and q_j), the permittivity of free space, ϵ_0 (a constant), and the radius of separation of the two charged atoms, r_{ij} . The sum of the bonded terms and non-bonded terms from each source is the total potential energy of interaction (V_{total}) of a given atom:

$$V_{total} = \sum V_{bonded} + \sum V_{nonbonded}$$

The negative derivative of potential energy with respect to position is equal to the force that is exerted on that atom. Forces on each atom are computed using a potential function of the form:

$$\mathbf{F}_i = -\frac{\partial V_i}{\partial \mathbf{r}_i}$$

Where the force (F_i) is computed as the derivative of potential energies of an atom i (V_i) with respect to the position of that atom (r_i). Finally, this information is used in Newton's second law, which can be described as:

$$\frac{\partial r_i^2}{\partial^2 t} = F_i m_i$$

In this form of the equation, we can determine how the position of an atom i (r_i) changes with respect to time (t), in response to forces derived from potentials of interaction (F_i) and taking into account the mass of the atom (m_i). This equation is a simple rearrangement of the widely-recognized $F=ma$.

In MD simulation, forces and position updates are traditionally computed on the femtosecond time scale. The computation is repeated millions or even billions of times to gather an ensemble of structural conformations that are thermodynamically or statistically favorable for the system in question. In practice, there are many other factors to consider when performing MD simulations, the breadth of which will not be covered here. Included in the references cited below are excellent resources for further reading, especially (3).

Quantum Mechanics

In MM we treat electrons as uniform clouds rather than modeling them individually because, as the Born-Oppenheimer approximation states, electronic motion occurs on a much faster scale than does atomic motion. In quantum mechanics (QM), the approximation is reversed. We assume that atomic motion is much slower than electronic motion, so when modeling individual electrons, the kinetics of atoms is ignored. QM calculations are an appropriate computational method when the question at hand is related to electronic behavior. For example, QM can be used to (i) perform energy calculations of molecules, (ii) determine equilibrium or transition state molecular geometries, (iii) predict IR, NMR, or UV/Vis spectra, or (iv) analyze energy profiles along a reaction coordinate including the breaking or formation of a bond, or the rotation about a bond or dihedral.

The most widely used formulation in QM models is the Schrödinger equation. It appears as follows:

$$\hat{H}\Psi = E\Psi$$

Where H is what is known as a Hamiltonian operator, or a function that represents the combined kinetic and potential energy of an electron. The symbol Ψ is a wavefunction, describing the coordinates of the electrons in question, and E is the total energy of the system. It is said that for systems comprising more than one electron, it is impossible to solve the Schrödinger equation exactly. Previously, Heisenberg has shown that for multiple electron systems, we cannot know both the exact location (x) and momentum (p) of any given electron simultaneously. In fact, the more we know about one, the less we know about the other. This is known as the Heisenberg uncertainty principal:

$$\Delta x \Delta p \geq \frac{h}{4\pi}$$

In this equation, h is Planck's constant. Therefore, in order to study systems of multiple electrons using QM, approximations must be made. The first approximation has been described previously – atomic motions are significantly slow and they may be neglected (the Born-Oppenheimer approximation). Second, electrons move independently from one another within discrete energy functions about a central point at a certain probability. This concept is understood from chemistry as s , p , d , and f , orbitals. Finally, the Linear Combination of Atomic Orbitals (LCAO) approximation states that multiple-electron models are equivalent to the sum of their single-electron model components.

In practice, performing QM calculations requires that the researcher choose a *model chemistry*. A model chemistry is made up of two components; a theoretical model and a basis set. A theoretical model describes the structure of electrons, and each model exists at a fine balance between accuracy and speed. For example, lower level theoretical models, such as Austin Model 1 (AM1) make more approximations in favor of speedy calculations at the expense of accuracy. Higher level theoretical models, such as Hartree-Fock (HF) or Møller-Plesset (MP), make fewer

approximations in favor of more accurate calculations but at the expense of speed. Theoretical models can be further divided into those that are defined purely by theory (“*ab initio*”), those that are defined by experimental observation (“empirical”), and those with roots in both theory and experiment (“semi-empirical”). Basis sets are sets of mathematical functions that describe the shape and populational probability of an electronic orbital. Similarly to theoretical models, basis sets with fewer approximations are generally accepted as more accurate, but are much slower in computation.

These principals form the foundation of QM theory, and it is difficult to describe them in further detail while maintaining brevity. The references cited below are excellent resources for more information about QM, especially (4).

Practical Applications

Within this document we discuss seven unique applications of the techniques described above. The commonality among all of the projects is that they have some sort of medical application, particularly in disease models or in systems where oxidative damage occurs. Chapter 2 deals with the enzyme monoamine oxidase B in a lipid bilayer. We discuss the interactions between the enzyme and the bilayer, as well as the dynamics of inhibitor unbinding. The following chapter, Chapter 3, also concerns monoamine oxidase, but in the context of the active site mechanism. We use QM techniques to study the barrier to ring-opening of certain cypromine analogues with the hope that we will be able to distinguish more specifics about the reaction mechanism. Chapter 4 is a study of the unstructured p53 C-terminal domain in complex with other protein-binding partners. Chapter 5 is a study of the viral suppressor of RNA silencing, B2, in complex with double-stranded RNA. In Chapter 6, we examine the dynamics of an epitope unique to peanut proteins that may be linked to allergenicity. The final two chapters steer slightly away from conventional molecular modeling and are more focused on method development and validation. In Chapter 7, we test the validity of small molecule parameters assigned by the widely used PRODRG server using a series of MM methods, and we propose a solution to derive better parameters using QM methods. Finally, in Chapter 8 we discuss a program that we

developed to analyze lipid bilayers in molecular dynamics simulations. Following Chapter 8, we will present a brief conclusion.

Because each of the projects in this document is so distinct, there is no comprehensive literature review for the whole document. Rather, in the introduction of each chapter the appropriate literature will be reviewed and cited. Each chapter is formatted as a unique manuscript, complete with an independent references section.

References

- (1) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235-242.
- (2) van der Spoel, D.; Lindahl, E.; Hess, B.; van Buuren, A. R.; Apol, E.; Meulenhoff, P. J.; Tieleman, D. P.; Sijbers, A. L. T. M.; Feenstra, K. A.; van Drunen, R.; Berendsen, H. J. C. *Gromacs User Manual version 4.5*; www.gromacs.org, **2010**.
- (3) *Molecular Modeling and Simulation: An Interdisciplinary Guide*; Schlick, T.; Interdisciplinary Applied Mathematics, Vol. 22; Springer: New York, **2002**.
- (4) Hehre, W. J. *A guide to Molecular Mechanics and Quantum Chemical Calculations*, 1st ed.; Wavefunction, Inc.: Irvine, CA, **2003**.

Chapter 2

Steered Molecular Dynamics Simulations Reveal Important Mechanisms in Reversible Monoamine Oxidase B Inhibition

William J. Allen and David R. Bevan

Submitted to Biochemistry.

Attribution

Of the work discussed within this chapter, Allen performed all of the simulations, analyses, and wrote and prepared the manuscript. Bevan significantly contributed to the editing and review of the manuscript, as well as provided guidance in the research.

Abstract

Monoamine oxidase B (MAO B) catalyzes the degradation of certain monoamine neurotransmitters, including phenylethylamine, benzylamine, and dopamine. The oxidative process results in decreased dopamine availability in neuronal cells as well as increased production of hydrogen peroxide, a potent pro-oxidant species. For these reasons, MAO B is widely considered a drug target for neurodegenerative diseases, including Parkinson's disease. In order to design more specific inhibitors for this monotopic bilayer enzyme, it is important to determine what factors govern the inhibitor binding process, including the role of the lipid bilayer, the active site loop, and several key residues within the binding pocket. We performed molecular dynamics simulations of MAO B embedded in a lipid bilayer and free in solution. The results of the simulations suggest that the bilayer controls the degree of fluctuation in two key

loops that form the greater part of the active site entrance (residues 85-110 and 155-165), a level of control that extends to the interaction of the protein with an inhibitor. We also simulated MAO B in complex with seven different reversible inhibitors that represent a wide spectrum of size, functional groups, and K_i values. We used steered molecular dynamics to simulate the release of these compounds from the active site in the presence and absence of the lipid bilayer. The results identify several specific residues and interactions that are important to inhibitor binding, and thus should be exploited when designing new inhibitors.

Introduction

There are two isoforms of the enzyme monoamine oxidase (MAO): A and B. The latter (designated MAO B) regulates synaptic levels of certain amine neurotransmitters including phenylethylamine, benzylamine, and dopamine by oxidative deamination, a process that also results in the production of hydrogen peroxide (1). MAO B expression increases with age, causing a decreased availability of dopamine and an increased degree of oxidative damage to neuronal cells by hydrogen peroxide – a pathway that, in combination with other factors, can lead to neurodegeneration and Parkinson's disease (2). Treatment strategies have included supplementation with levodopa, a dopamine precursor, as a method to replenish depleted dopamine levels in patients showing parkinsonian-type symptoms. However, this treatment is only a temporary fix that does nothing to slow or reverse the progression of neuronal cell death (3). For this reason, a wealth of research has been devoted to the development of effective MAO B inhibitors to be used as anti-neurodegenerative therapeutics (3).

The irreversible MAO B inhibitors selegiline (1-deprenyl) (4) and rasagiline (5) are currently used as therapies to slow the progression, but not prevent Parkinson's disease. Additional studies have suggested that the effects elicited by selegiline and rasagiline are not solely due to MAO B inhibition, but instead a multitude of different pathways (6). Newer reversible inhibitors, including safinamide and coumarin analogues, show some promise for their high specificity to the MAO B isoform (7), but are not yet approved as therapeutics. As the general body of information grows, it is prudent to continue to design new MAO B inhibitors to be used as neuroprotectant therapeutics with higher specificities and fewer deleterious side effects. A

detailed understanding of the specific molecular interactions that govern MAO B inhibition is essential to this process.

MAO B is a monotopic membrane protein embedded in the outer mitochondrial membrane. The 520-amino acid protein forms a homodimer with two helical tails that extend across the bilayer (8). The bulk of the solvent-exposed part of the protein is divided into a region that covalently binds a flavin adenine dinucleotide (FAD) cofactor (9), and a region that houses the active site. The active site is dipartite, divided into a catalytic cavity that is proximal to the isoalloxazine ring of the FAD, and an entrance cavity that is positioned closer to the interface of the bilayer. Their combined volume is approximately 700 \AA^3 (10). The crystal structure of MAO B was first solved in 2001 to 3 \AA resolution (8). There are now (as of March, 2011) 36 unique crystal structures available in the Protein Data Bank (PDB, www.pdb.org) (11) of MAO B and MAO B mutants in complex with different reversible and irreversible inhibitors, at as high as 1.6 \AA resolution (7). The wealth of available structural data makes this enzyme an ideal target for rational structure-based drug design.

Molecular dynamics (MD) is a useful computational tool that can offer insight into specific molecular interactions between protein and inhibitor at the atomic level – beyond what is learned from a static crystal structure. For example, crystal structures revealed that residues Ile 199 (12, 13) and Tyr 326 (14) are important for ligand specificity. The “aromatic cage” formed by Tyr 398 and Tyr 435 plays a role in coordinating ligands in the active site to the FAD molecule (15). In addition, the “active site loop”, formed by residues 99-112 could be important in allowing ligand access to the active site (8, 16). Aside from these static-structure derived observations and predictions, one can uncover more about specific molecular processes and interactions using MD simulations. For example, Apostolov *et al.* recently demonstrated that the bilayer itself plays a role in influencing the dynamics of the MAO A isoform, especially with regards to the entrance to the active site (17).

Here we have performed MD simulations of several different inhibitor-MAO B complexes in an attempt to better understand the nature and the mechanism of inhibition of this isoform. We sought to identify specific residues that contribute importantly to the binding of MAO B inhibitors, as well as identify behaviors, especially pertaining to the active site and active site entrance, which might also influence inhibitor binding. The resulting dynamic model of inhibitor

binding and interaction within the active site is important in the design of novel anti-neurodegenerative therapeutics.

Materials and Methods

File Preparation

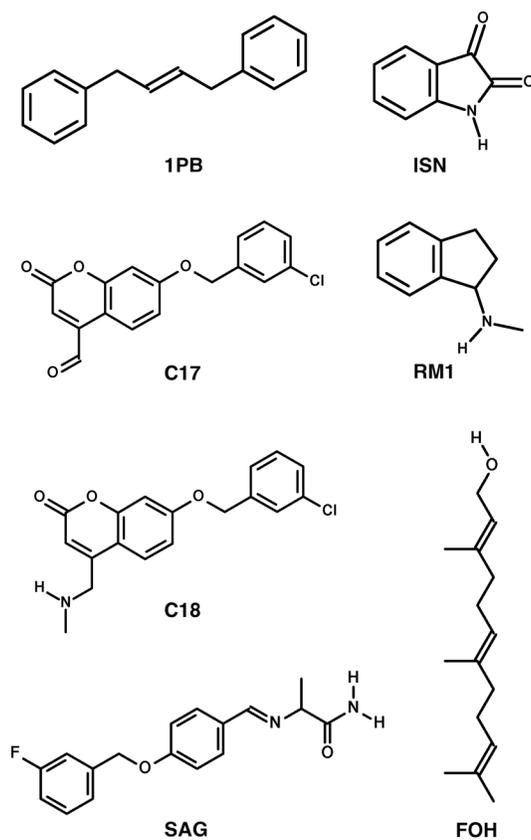
Seven crystal structures were downloaded from the PDB: 1OJ9, 1OJA (18), 2BK3 (13), 2C67 (19), 2V5Z, 2V60, and 2V61 (7). Each crystal structure consists of two protein chains (chains A and B) comprising the MAO B homodimer. Each chain is covalently bound to an FAD cofactor and contains a reversible inhibitor in the active site unique to the crystal structure. Important chemical information and inhibitor structures are summarized in Table 2.1 and Scheme 2.1. In all cases, chain A includes residues Asn 3 to Ile 501 (Ile 501 is missing atoms), and chain B includes residues Asn 3 to Ile 496. The full-length homodimer contains 520 residues per chain; missing C-terminal residues were modeled in LEaP of the AmberTools 1.4 suite of programs (20), and dihedral angles characteristic of an α -helix were imposed to correctly reflect the trans-membrane nature of the residues. The two missing N-terminal residues were not modeled into the structure.

The protein topology was based on the united-atom GROMOS96 43A1 force field. The FAD topology was assembled from the existing adenosine triphosphate (ATP) and flavin mononucleotide (FMN) topologies in that force field. Inhibitor topologies were generated by the PRODRG server (21). We previously showed that charges and charge groups generated by the PRODRG server are inconsistent with the GROMOS96 43A1 force field (22), and accordingly the charges were refined either by analogy from existing groups in the GROMOS96 43A1 force field, or by using the Hartree-Fock level of theory with the 6-31G** basis set in Titan 1.0.8 (Wavefunction, Inc., Irvine, CA). The FAD and small molecule topologies are included as Supplemental Information (Table S2.1-S2.8).

Table 2.1. Summary of inhibitor structures arranged by K_i .

PDB code	Res. (Å)	Inhibitor	Abbrev.	K_i (μM)	ΔG_{dissoc} (kJ/mol) ^e
1OJ9	2.3	1,4-diphenyl-2-butene	1PB	35 ^a	26.4
2C67	1.7	N-methyl-1(R)-aminoindan	RM1	17 ^b	28.3
1OJA	1.7	1H-indole-2,3-dione (isatin)	ISN	3 ^a	32.8
2BK3	1.8	farnesol	FOH	2.3 ^c	33.5
2V5Z	1.6	(S)-(+)-2-[4-(fluorobenzyloxy-benzylamino) propionamide]	SAG	0.45±0.13 ^d	37.7
2V60	2.0	7-[(3-chlorobenzyl)oxy]-2-oxo-2H-chromene-4-carbaldehyde	C17	0.40±0.02 ^d	38.0
2V61	1.7	7-[(3-chlorobenzyl)oxy]-4-[(methylamino)methyl]-2H-chromen-2-one	C18	0.10±0.02 ^d	41.5

^a Ref. 18; ^b Ref. 19; ^c Ref. 13; ^d Ref. 7; ^e calculated by $\Delta G_{\text{dissoc}} = -RT \ln(K_i)$

**Scheme 2.1.** Structures of MAO B inhibitors used in this study.

Lipid parameters (23) and the lipid bilayer starting structure (24) were obtained from the Biocomputing Group at the University of Calgary (<http://moose.bio.ucalgary.ca/>). The starting structure contained 128 pre-equilibrated palmitoyl-oleoyl-phosphatidylcholine (POPC) lipids arranged in a periodic cell. The system was multiplied into a 2x3 array along the x - y plane, for a total of 768 lipid molecules, in order to create a sufficiently large foundation for the simulations. An in-house script was used to randomly select 329 lipids, and then mutate the choline group to ethanolamine, thereby creating palmitoyl-oleoyl-phosphatidylethanolamine (POPE) molecules. The final lipid bilayer contained 439 POPC lipids and 329 POPE lipids, a 4:3 ratio that mimics the composition of the *in vivo* outer-mitochondrial membrane (25) and a previous simulation (16).

Conventional MD Simulations

All simulations were performed using GROMACS 4.0.7 (26). A twin-range cutoff was used to calculate short-range non-bonded interactions, with electrostatic interactions truncated at 0.8 nm and van der Waals interactions truncated at 1.4 nm. The particle mesh Ewald (PME) method (27, 28) was used to evaluate long-range electrostatic interactions with fourth-order spline interpolation and a 0.12 nm Fourier grid spacing. All bond lengths were constrained using a parallelized linear constraint solver (P-LINCS) (29, 30), and simulations were performed with a 2 fs timestep. The simple point charge (SPC) water model was used (31). Isochoric-isothermal (NVT) simulations employed the Berendsen thermostat (32) with a temperature coupling constant (τ_T) of 0.1 ps. Isothermal-isobaric (NPT) simulations employed the Nosé-Hoover thermostat (33, 34) with a temperature coupling constant of 0.1 ps and a semi-isotropic Parrinello-Rahman barostat (35, 36) with a temperature coupling constant (τ_p) of 1.0 ps and a compressibility of $4.5e^{-5}$ bar⁻¹ applied separately in the x - y and z directions, a common practice for membrane simulations (37).

The bilayer system was minimized with a steepest descents integrator until the maximum force on any atom was less than 1000 kJ mol⁻¹ nm⁻¹. The system was then simulated under an NVT ensemble at 100 K for 200 ps. Following equilibration around the target temperature, the system was annealed to 310 K linearly over 900 ps under an NPT ensemble. Finally, a

production simulation followed under the same conditions for 20 ns, with this final conformation of the bilayer simulations being used as the starting platform for protein insertion.

The seven crystal structures of MAO B described above were inserted into the bilayer using the “INFLATEGRO” methodology described by Kandt *et al.* (37). The apo-form of the crystal structure originating from PDB 1OJA (1.70 Å resolution) provided an eighth simulation condition that was simulated in triplicate. Minimization, equilibration, and production MD were performed in the same way as described for the bilayer. The protein backbone atoms were restrained until the production MD. These eight simulation conditions were then repeated in the absence of the bilayer, with the protein surrounded by bulk solvent (the apo-form was also simulated in triplicate). Following minimization, proteins simulated in a box of water were equilibrated at 310 K under an NVT ensemble for 200 ps, followed by 1 ns of NPT equilibration at the same temperature, and finally production MD as before. In the simulations of protein in bulk solvent, the pressure-coupling algorithm was applied isotropically and weak restraints were imposed on the backbone atoms of the helical portion of the protein tail (residues 488-520). (Previous simulations in our lab demonstrated that performing MD with the trans-membrane tails unrestrained in a box of solvent results in unnatural fluctuations between the protein subunits that consistently crashed the simulations, data not shown). The tails were retained even in the absence of the lipid bilayer in case their proximity to the active site entrance was shown to be important.

Steered MD Simulations

Following production MD of inhibitor-bound systems, non-equilibrium pulling was performed independently on each reversible inhibitor in the direction of active site egress. In each case, the pulling simulation was performed in quintuplicate. The starting configurations and velocities for the pulling simulations were taken at 200 ps intervals. All constant-velocity steered MD simulations were implemented with a spring force constant of 1000 kJ mol⁻¹ nm⁻² and at a speed of 0.01 nm ps⁻¹.

Analyses were performed using utilities available in the GROMACS suite of programs or by scripts written in-house. Three-dimensional images were rendered with PyMol (38). Chemical structures were created in PubChem Sketcher (<http://pubchem.ncbi.nlm.nih.gov/>) and two-dimensional line art was generated with Grace (39). Bilayer thickness calculations and area per

lipid (APL) were measured by the GridMAT-MD software (40) and plotted with Xmatrix (<http://www.matpack.de>).

Results

The simulation results will be discussed as follows. First, we discuss the triplicate, conventional MD simulations of MAO B in the apo-form. Thereafter, we will discuss the steered MD simulations of MAO B with inhibitors bound. Systems comprising MAO B in a bilayer will henceforth be referred to as MAO B-bilayer; systems of MAO B in bulk solvent will be referred to as MAO B-solvent.

Molecular Dynamics of MAO B

To obtain sufficient sampling, we simulated each replicate of MAO B-bilayer and MAO B-solvent under production MD conditions for 50 ns. When measured as a function of time, the root mean square deviation (rmsd) of a protein is a good indicator of the convergence in protein structure change over the course of a simulation. The MAO B trans-membrane hydrophobic tails (residues 485-520) were restrained in the solvent simulations and therefore not considered for the rmsd calculations, nor were they considered for the rmsd calculations of the bilayer simulations for the sake of consistency. The backbone rmsd of the MAO B-bilayer replicates equilibrated to between 0.19 nm and 0.21 nm after approximately 20 ns of simulation (Figure 2.1). Similarly, the backbone rmsd of each MAO B-solvent replicate equilibrated just slightly higher, between 0.21 nm and 0.24 nm, in the same time frame. The average rmsd over the last 30 ns of each set of three replicates was 0.19 ± 0.01 nm for MAO B-bilayer and 0.22 ± 0.01 nm for MAO B-solvent. The relative stability of each structure after the first 20 ns provides assurance that the final 30 ns of simulation are suitable for analysis.

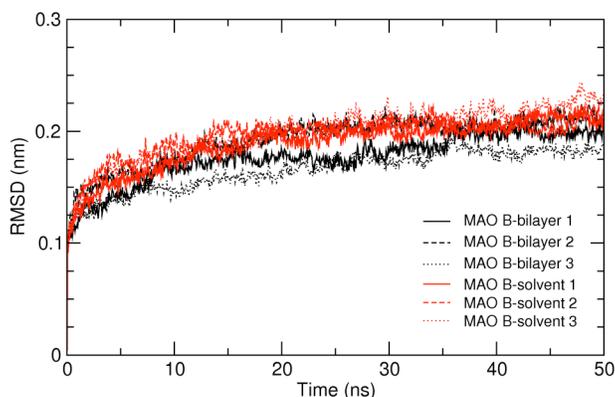


Figure 2.1. The rmsd of the protein backbone, residues 1-485. The rmsd value is averaged over the two monomers in each replicate simulation.

As rmsd is a global measurement of protein motion, root mean square fluctuation (RMSF) is a local measurement that provides a higher resolution detail of residue fluctuations. Each of the triplicate simulations contained the full homodimeric form of MAO B, thus providing six individual monomers for analysis. During the last 30 ns of simulation, the differences in fluctuation pattern between MAO B-bilayer and MAO B-solvent in two particular regions of the enzyme demanded further investigation. These regions were residues 85-110, which includes most of the active site loop (residues 99-112), and residues 155-165. A representative RMSF plot is shown in Figure 2.2a (the other RMSF plots are included as Supplemental Information at the end of this chapter in Figure S2.1). On average, the degree of fluctuation exhibited by these two regions was larger for MAO B-solvent than it was for MAO B-bilayer. The fluctuation from residues 85-110 was 0.10 ± 0.03 nm when in a bilayer and 0.16 ± 0.06 nm when in solvent, a statistically significant increase as determined by a Student's t-test ($p < 0.05$). The fluctuation from residues 155-165 was 0.081 ± 0.019 nm when in a bilayer and 0.101 ± 0.036 nm when in solvent, also statistically significant by the same test ($p < 0.05$). No other areas in the protein were consistently higher or lower in fluctuation. This observation is particularly significant because these two regions form the entrance to the active site cavity (Figure 2.2b), indicating that the bilayer exerts a degree of control over the fluctuations of these loops and subsequently the interactions that may occur in these regions.

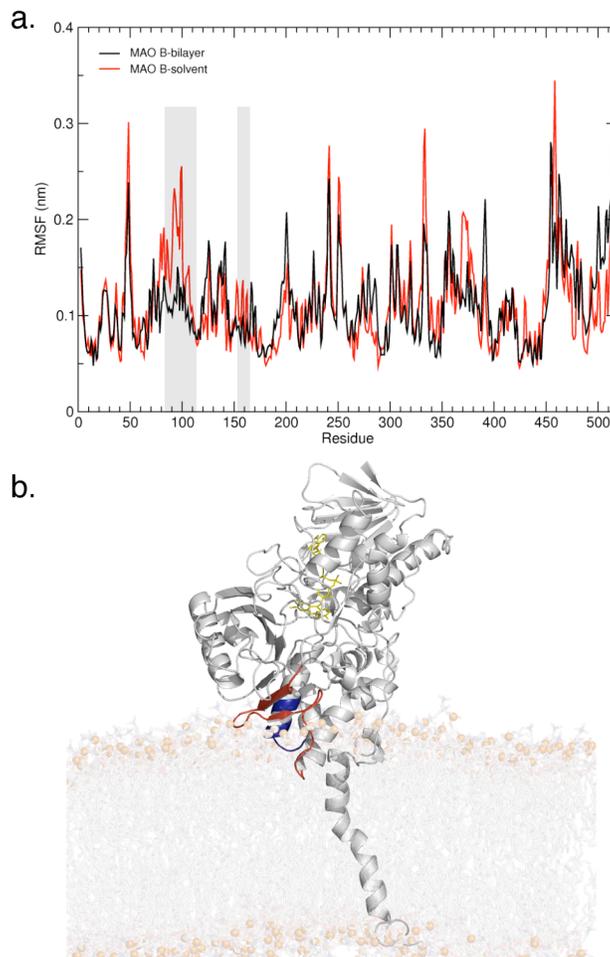


Figure 2.2. MAO B fluctuation profile for bilayer and solvent forms. A representative RMSF plot is shown in panel (a), comparing degree of fluctuation per MAO B residue in solvent vs. bilayer. An example snapshot is shown in panel (b); residues 85-110 are highlighted red and residues 155-165 are highlighted blue. The bilayer was made transparent and the phosphorus atoms are represented as orange spheres. The FAD cofactor is shown as yellow sticks. One MAO B subunit and water atoms were removed for simplicity.

As these two loop regions comprise the ring-like ‘mouth’ of the active site, it is convenient to then define the size of the active site entrance by measuring the best-fit ellipsoidal surface area formed by opposing residues in the loops. Thus, the center-of-mass distance between Lys 93 and Trp 107 was treated as one axis of the ellipse, whereas the center-of-mass distance between Tyr 97 and Ser 160 was treated as the second axis. This concept is illustrated in Figure 2.3. It stands to reason that a larger ellipsoidal surface area would be reflective of a more open conformation, while a smaller surface area would indicate a more closed conformation. The following equation was used to calculate the surface area of the active site entrance opening:

$$SurfaceArea = \pi \left[\frac{1}{2} (Lys93 - Trp107) \right] \left[\frac{1}{2} (Tyr97 - Ser160) \right]$$

The residue names in the equation refer to the center of mass position of those residues. The resulting surface area of each active site entrance is plotted with respect to time in Figure 2.4. In the MAO B-bilayer simulations, five of the six replicates converged to an active site entrance opening between 1.00 nm and 1.50 nm. One replicate of MAO B-bilayer, however, demonstrated a behavior that could be described as the opening of the active site entrance. The size of the opening in replicate 1, chain B, increased to more than 2.25 nm², in the presence of the bilayer. We found that MAO B-solvent typically maintained an active site entrance surface area between approximately 0.75 nm² to 1.50 nm² during the simulations. Further, the fluctuation in this value was greater than it was for MAO B-bilayer, reflecting the previous observation that MAO B in the absence of the bilayer exhibited a higher RMSF in these regions. These data provide further evidence that the lipid bilayer regulates behaviors that occur near the entrance to the active site.

To further explore how the lipid bilayer modulates the fluctuation of these loops, and to understand the mechanism by which the active site mouth opens and closes, we examined more closely the interaction between the protein and the bilayer. Qualitative analysis of the MD simulations indicated that the protein active site entrance that opened in the presence of the bilayer did so because it dipped down below the lipid head groups towards the hydrophobic fatty acid tails. In doing so, native interactions between residues in the loops were disrupted and new interactions with the hydrophobic tails were formed. Figure 2.3 compares two active site entrances – that of a closed conformation (Figure 2.3c) and that of an open conformation (Figure 2.3d). From the snapshot of the open conformation we can see that the Tyr 97 moved up and away from the bilayer so the plane of the aromatic ring was parallel to the surface of the bilayer, and the Trp 107 moved further down into the bilayer, forming hydrophobic contacts with the lipid tail. In the closed conformation, these two residues formed pi-stacking interactions with one another. In addition, the Lys 93 and Ser 160 residues snorkeled into the lipid headgroups in the closed conformation, whereas the same residues were further into the aqueous phase in the open conformation.

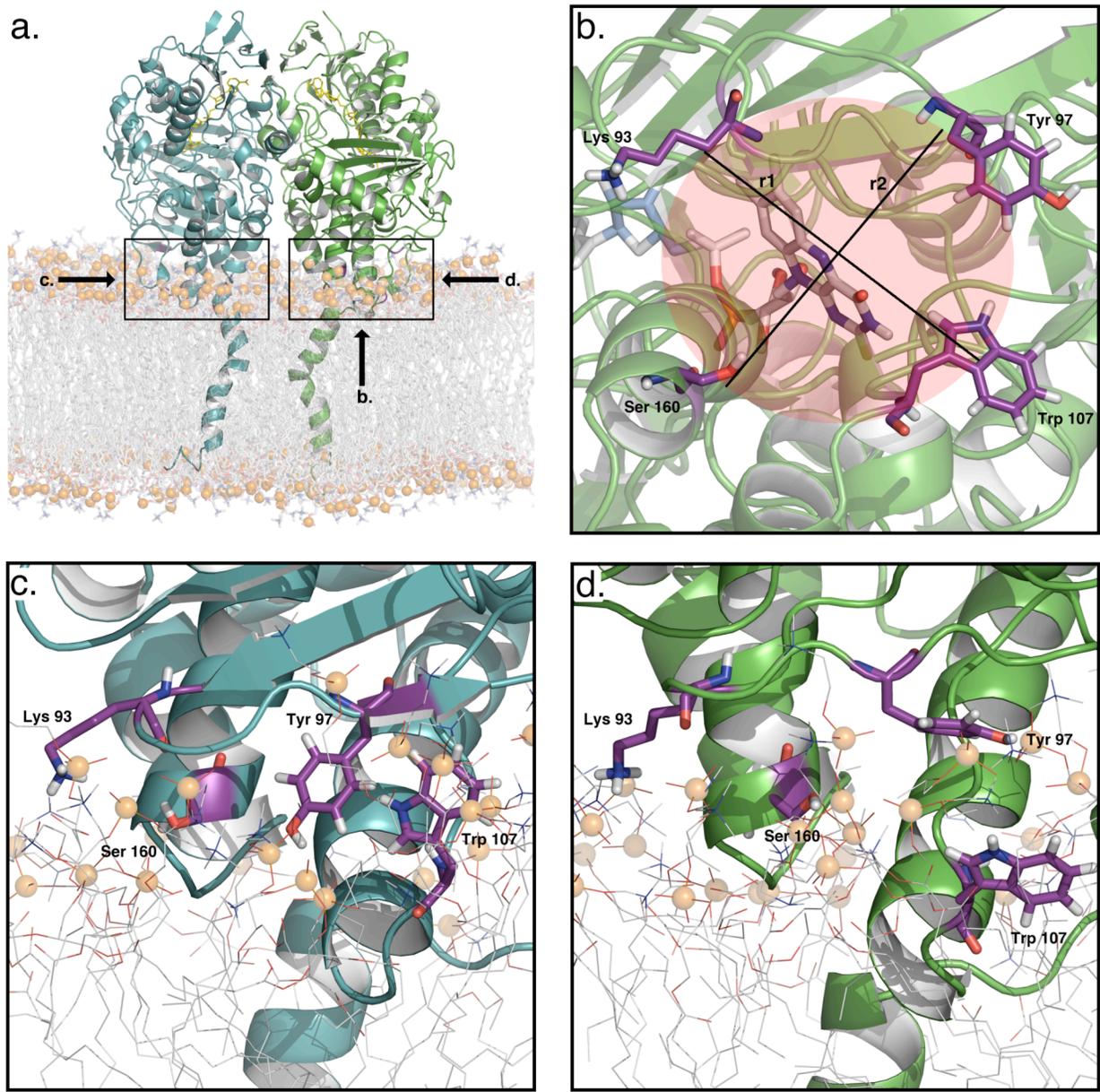


Figure 2.3. Overview of MAO B homodimer embedded in a lipid bilayer. Each arrow indicates the direction from which the subsequent panels are viewed. (b) A view into the active site with opposing residues shown. Ellipse approximates active site entrance opening area. (c), (d) Active site opening residues shown in a closed (c) and open (d) conformation.

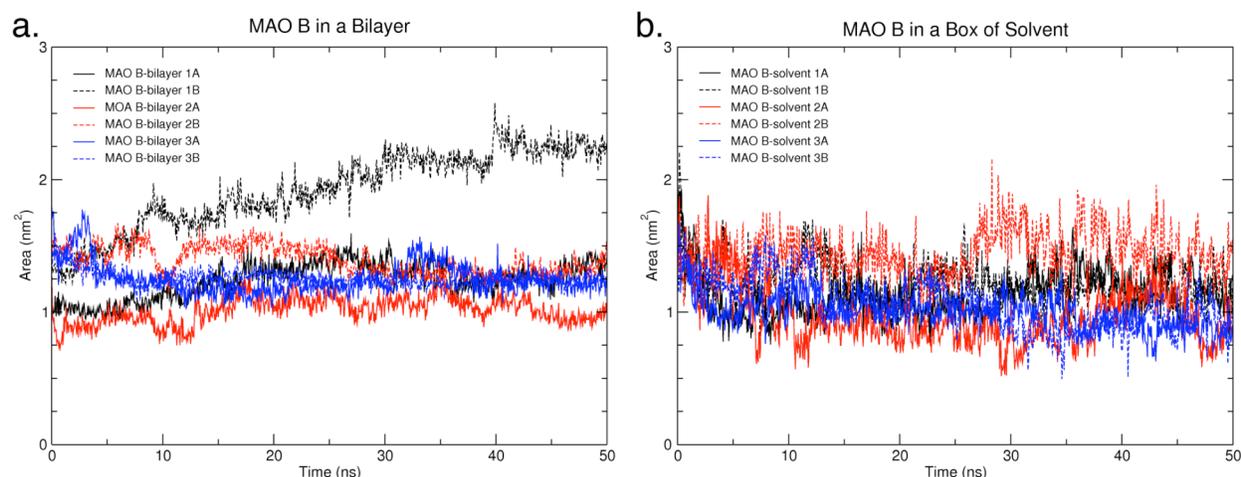


Figure 2.4. Comparison of the area of the active site channel opening when MAO B is in a bilayer (a) vs MAO B in a box of solvent (b).

A measurement of partial densities as a function of the bilayer normal established a quantifiable and time-averaged comparison between closed and open conformations. The partial densities of each lipid component, the bulk of the protein, and the active site entrance of each monomeric subunit were measured (Figure 2.5). The key observation lies in the positioning and shape of the active site entrance in the open conformation. Its position was shifted along the z -axis, further into the bilayer, creating a bimodal peak. Because the active site entrance of the other subunit within the dimer remained closed, one possible explanation is that the enzyme tilted with respect to the plane of the bilayer. Also, because all of the active site entrances in the MAO B-solvent replicate simulations remained in the closed conformation, we can hypothesize that it is the formation of new hydrophobic interactions that facilitated the active site entrance-opening event. Table 2.2 shows the average height of the center-of-mass of the active site entrance opening with respect to the average position of the phospholipid headgroups. In each replicate, the average position of one of the active site mouths of the dimer dipped down with respect to the phospholipid heads, from 0.12 nm – 0.16 nm. The other chain in the same replicate stayed above the phospholipid head group, near 0.32 nm – 0.33 nm. In a single replicate, the active site entrance residues that dropped below the headgroups shifted into the open configuration, suggesting that interactions within this hydrophobic region control the size of the active site entrance. We note that the distance of the active site mouth to the bilayer surface did not always result in an observable opening of the active site (Figure 2.4a). This is likely due to

the short time frame of our simulations relative to the apparently slow period of active site channel opening and closing, and thus it would be improbable to observe this behavior in all replicates without applying non-physiological techniques such as locally enhanced sampling, or greatly extending the length of the simulations.

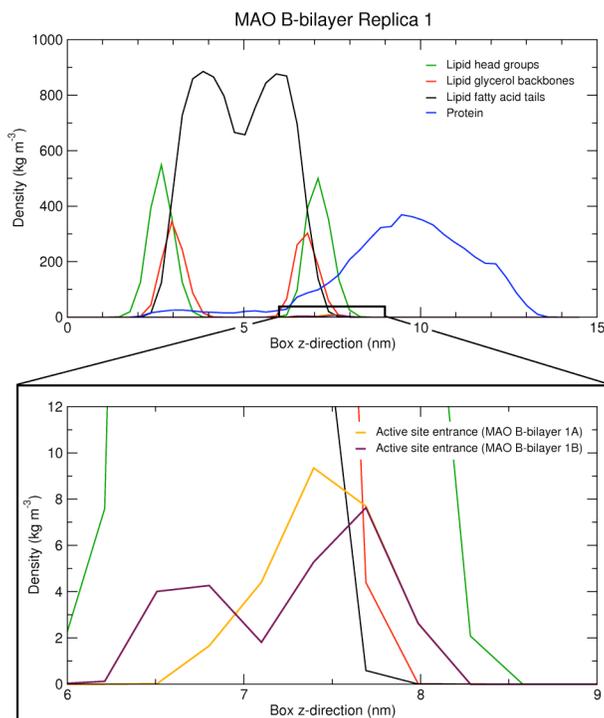


Figure 2.5. Partial density plot for important lipid and protein components along the z -axis. Insert is a close-up of the active site entrance opening residues. The residues considered for this measurement are Lys 93, Tyr 97, Trp 107, and Ser 160.

Table 2.2. Position of important groups along z -coordinate.

Replica	Lipid Head (nm)	Glycerol (nm)	Entrance to Active Site A (nm)	Entrance to Active Site B (nm)
1	0	-0.35	+0.33	+0.16
2	0	-0.25	+0.33	+0.12
3	0	-0.35	+0.13	+0.32

In addition to the effects elicited by the bilayer on the active site entrance, the enzyme itself also changed the local structure and dynamics of the bilayer such that it may be more conducive to inhibitor or substrate trafficking. Foremost, over the last 30 ns of simulation, the average

bilayer thickness was notably decreased around MAO B when compared to a control bilayer (Figure 2.6). The control bilayer ranged in thickness (phosphorus atom – phosphorus atom) from about 4.0 nm – 4.7 nm, as most of the bilayer maintained 4.3 nm thickness. In the presence of MAO B, however, the same measure of thickness increased in range to about 3.7 nm – 4.8 nm, and the occurrence of extreme thickening or thinning was more frequent. This observation is best illustrated by the fact that there were two large pockets in close proximity to MAO B that exhibited a large amount of thinning. This observation is consistent for 5 of the 6 monomeric subunits. One monomer in the third replicate seemed to have no major effect on the thickness of the bilayer that immediately surrounds it, suggesting that the degree of thinning may not always be evident.

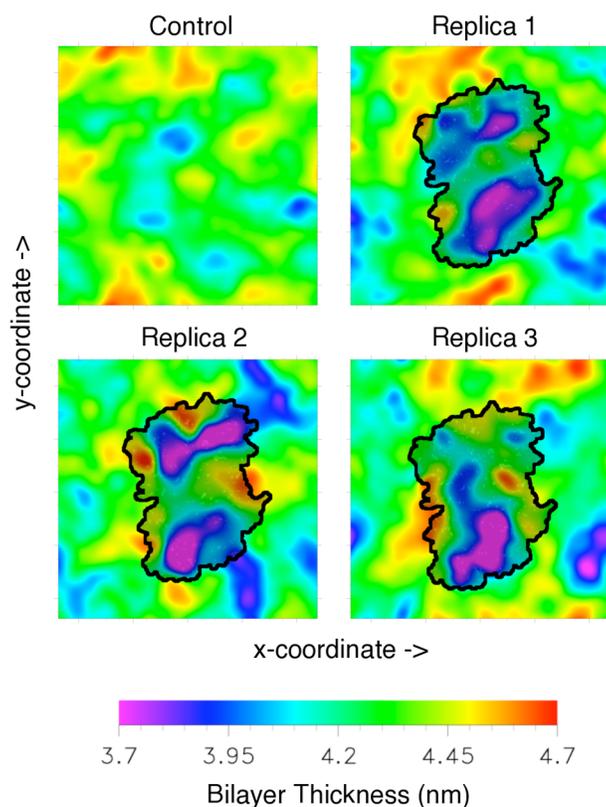


Figure 2.6. Bilayer thickness plots. The top left plot is a top-down perspective thickness plot of a control bilayer. The remaining three plots are the thicknesses of the replicate simulations. The position of the MAO B homodimer is overlaid as a black, transparent surface.

The presence of MAO B also influenced the area per lipid headgroup (APL) of the bilayer systems. The control bilayer maintained an average APL of $53.8 \pm 0.2 \text{ \AA}^2$ per leaflet during

simulation. Once MAO B was inserted, the average APL in the top leaflet dropped to $50.2 \pm 0.6 \text{ \AA}^2$ ($p < 0.05$) and the average APL in the bottom leaflet dropped to $51.7 \pm 0.5 \text{ \AA}^2$ ($p < 0.05$), as averaged over the last 30 ns of simulation.

Steered Molecular Dynamics of MAO B Inhibitors

In addition to the simulations of apo-MAO B embedded in a bilayer and in bulk solvent, we performed several simulations of MAO B in complex with various reversible inhibitors. Herein, the inhibitors will be referred to by a three-letter identifier, which they were assigned in the PDB (Summarized in Table 2.1 and Scheme 2.1). Following a preliminary 20 ns production run of these inhibitor-bound systems, we saved full atomic coordinates and velocities in 200 ps increments. These snapshots served as starting points for non-equilibrium steered MD simulations, as described in the Methods. In this study we employed constant velocity pulling. Briefly, the inhibitor was tethered to a dummy atom (an atom having neither mass nor charge) by a classical spring. The dummy atom traveled along a pre-specified vector at a constant velocity. As tension in the spring increased, the inhibitor was steered away from its original position, along the same general path as the vector. The force in the spring was recorded with respect to time or displacement of the inhibitor. We performed five such steered simulations for each inhibitor in each condition (bilayer and solvent), for a total of 70 simulations.

During a steered simulation, the force required to pull each inhibitor was plotted as a function of displacement from the original position, or distance traveled from the binding site, for each simulation (Figure 2.7). By integrating each curve we obtained the work performed on the system for each simulation. Jarzynski's equality (41) equates the average work performed on replicate ensembles to the potential of mean force (ΔPMF) for a non-equilibrium process, or the overall change in free energy between two states along a reaction coordinate:

$$\exp(-\Delta PMF / k_B T) = \langle \exp(-W / k_B T) \rangle$$

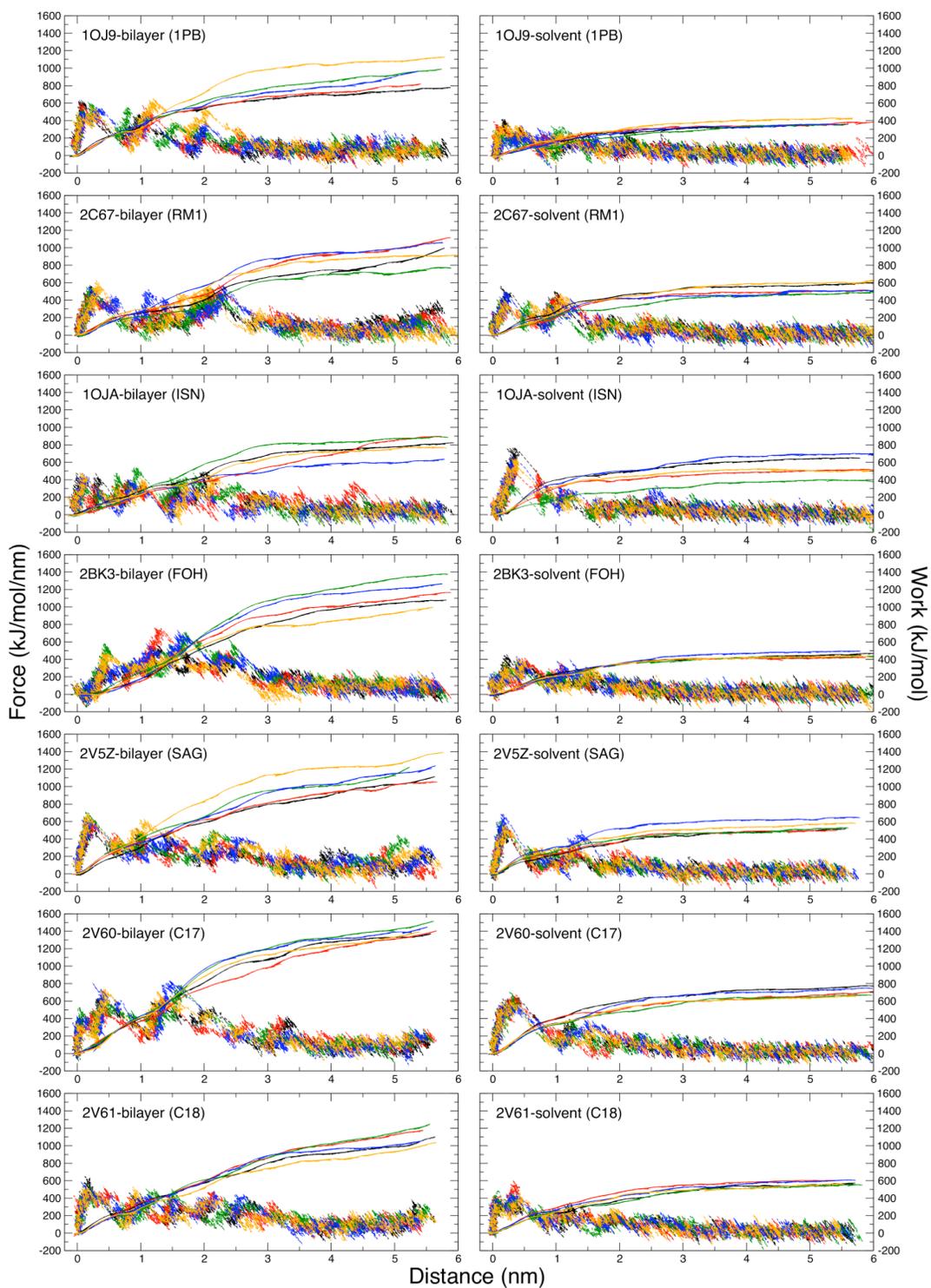


Figure 2.7. Δ PMF plots for inhibitor unbinding. The plots in the left column are bilayer simulations, the right column are solvent. The five replicates are colored differently (black = 1, red = 2, green = 3, blue = 4, orange = 5). The smooth curves are integrals of the irregular curves, corresponding to work, displayed on the right axis.

In Jarzynski's equality, W is the work performed on a system, k_B is the Boltzmann constant, and T is temperature in Kelvin. The angled brackets signify an average over ensembles. We considered the beginning of the reaction coordinate to be the MAO B-inhibitor complex, and the end of the reaction coordinate to be the state in which the distance of separation from any inhibitor atom and any protein atom was greater than or equal to 0.7 nm.

From these data it is convenient to make two important comparisons: the pulling of different inhibitors into the same media (bilayer, for example), and the pulling of the same inhibitor into two different media (bilayer and solvent) (Figure 2.7 and Table S2.9). For example, pulling the weak inhibitor 1PB into the bilayer required a peak force of $606 \pm 36 \text{ kJ mol}^{-1} \text{ nm}^{-1}$, whereas pulling the strong inhibitor C17 into the same media required a greater peak force, $730 \pm 82 \text{ kJ mol}^{-1} \text{ nm}^{-1}$, a statistically significant increase ($p < 0.05$). Conversely, the inhibitor 1PB required a peak force of $373 \pm 35 \text{ kJ mol}^{-1} \text{ nm}^{-1}$ to pull into solvent, a much smaller amount than was required to pull the same inhibitor into the bilayer. In fact, all seven of the inhibitors required a smaller peak force to pull into solvent with the exception of ISN. It is convenient to plot the Gibbs free energy of dissociation (ΔG_{dissoc}) of each inhibitor by the greatest (peak) force required to pull the inhibitor from the binding site (Figure S2.2). The best-fit linear regression of that plot had a positive slope in each case, that is, the greatest amount of force required to pull an inhibitor from the binding site increased with increasing strength of the inhibitor. However, the correlation coefficient values were quite low; 0.04 for pulling into the bilayer, and 0.27 for pulling into solvent. Although there is no strong correlation between the strength of each inhibitor and the peak force required to pull, the general increase in force required between solvent and bilayer is apparent. At first glance it seems as if pulling into different media itself would cause the discrepancy, however, we find that the peak force typically occurs while the inhibitor is still deep within the binding pocket, before it encounters solvent or bilayer. Therefore, it stands to reason that the medium itself may allosterically regulate the dynamics of residues in the binding pocket.

The ΔPMF of pulling 1PB, the weakest inhibitor, into the membrane was $770 \pm 150 \text{ kJ mol}^{-1}$, as averaged over five simulations (Figure 2.7 and Table S2.9). Conversely, the ΔPMF of pulling C17, a very strong inhibitor, into the membrane was $1201 \pm 94 \text{ kJ mol}^{-1}$, a statistically significant increase ($p < 0.05$). The two ΔPMF values of the weak inhibitor RM1 and of ISN fell below 800 kJ mol^{-1} . The ΔPMF of ISN was particularly low, which could be attributed to its diminutive size compared to the other inhibitors. The two ΔPMF values of two of the strongest inhibitors, FOH

and SAG, were above 1000 kJ mol^{-1} and 950 kJ mol^{-1} , respectively. The inhibitor C18 seemed to be the most bothersome outlier, with a ΔPMF of $823 \pm 72 \text{ kJ mol}^{-1}$. It is difficult to determine the exact reason for the discrepancy, but it must be attributed to interactions between the secondary amide group and surrounding residues, for this group is the only difference between C18 and the structure of C17. In Figure 2.8, the average ΔPMF values required to pull each inhibitor into solvent or bilayer are arranged by the strength of the inhibitor. The correlation between ΔG_{dissoc} and ΔPMF is slightly higher than it is for ΔG_{dissoc} and peak force, and thus is a slightly better rubric for measuring the validity of the simulations. The best-fit linear regression for pulling inhibitors of different strength out of the binding pocket and into the bilayer or solvent are both positively sloped and have correlation coefficients of 0.19 and 0.40, respectively (Figure S2.3). Although the correlation is still quite low, the increase in correlation does indicate that the total amount of work performed on a system is a better indication of inhibitor strength than is the peak amount of force applied. In addition to the slightly improved correlations, we also observed a greater separation of data between pulling into the bilayer or solvent. For every inhibitor, it required more work to pull into the bilayer than it did to pull into solvent. The ΔPMF of pulling each inhibitor into bulk solvent is drastically decreased when compared to pulling each inhibitor into the bilayer. There are two simple explanations for why such a difference could occur: a decrease in force required to pull through the media (bilayer vs. solvent), or a decrease in interaction between the protein and inhibitor because the change in environment has made the protein less conducive to inhibitor binding. To address the first explanation, when the inhibitor was pulled through just media (between roughly 350 – 550 ps in Figure 2.7), the force required to pull through the bilayer was only 40 – 70 $\text{kJ mol}^{-1} \text{ nm}^{-1}$ higher on average. It is unlikely that simply the entrance into the media resulted in a change in work as great as observed for some inhibitors (Figure 2.8). In fact, to account for a 400 kJ mol^{-1} change in work, one would need to pull through the more viscous media an extra distance of 6 – 10 nm, much further than they were pulled in the actual simulation. Thus, it is likely that the change in medium from bilayer-bound to solvent-exposed effectively decouples native protein-ligand interactions that are typically occurring. These observations attest to the importance of studying the inhibitor-protein complex in the most appropriate physiological environment.

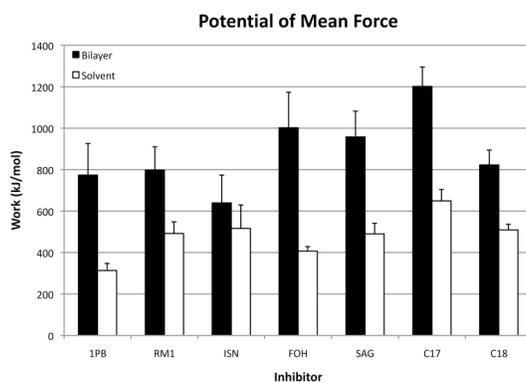


Figure 2.8. Average Δ PMF values for pulling inhibitors from the protein active site. Black bars represent inhibitor being pulled into the bilayer, white bars represent inhibitor being pulled into solvent. Inhibitors are arranged from left to right by increasing magnitude of ΔG_{dissoc} .

Quantification of average energy of interaction between each inhibitor and specific residues in the MAO B active site channel provided an excellent indication of which residues were important in inhibitor binding. For each steered MD simulation, we tabulated all MAO B residues that were within 0.7 nm of the inhibitor at any time. Then, the van der Waals and electrostatic contributions to the potential of interaction between inhibitor and individual residues were recorded. Finally, for each residue we took an average of any value that was above a 0.01 kJ mol^{-1} cut-off, and plotted those interactions that contributed the most to the protein-inhibitor interaction (Figure 2.9). Some inhibitors interacted with as many as 45 different MAO B residues during active site egress, but we plotted only the 20 strongest interactions for simplicity. The weakest inhibitor, 1PB, had a total interaction of $-176 \pm 12 \text{ kJ mol}^{-1}$ with MAO B residues. The total energy of interaction increased in magnitude for each sequentially stronger inhibitor, with the exception of the inhibitor FOH at $-277 \pm 10 \text{ kJ mol}^{-1}$, and the inhibitor C18 at $-270 \pm 6 \text{ kJ mol}^{-1}$. In Figure 2.10, we arranged the sums of the individual energy components by increasing strength of the inhibitor. Most encouragingly, the sums of the individual energy components correlated strongly with the ΔG_{dissoc} of the inhibitor (Figure S2.4). The correlation coefficient for the best-fit linear regression that describes the increasing potential of interaction (increasing in magnitude, that is) with increasing inhibitor strength is 0.67. Notably, the correlation coefficient for the same relationship, but only considering the electrostatic contributions is 0.86. The implication here is that the addition of electrostatic interactions contributes most meaningfully to the increased strength of the inhibitor.

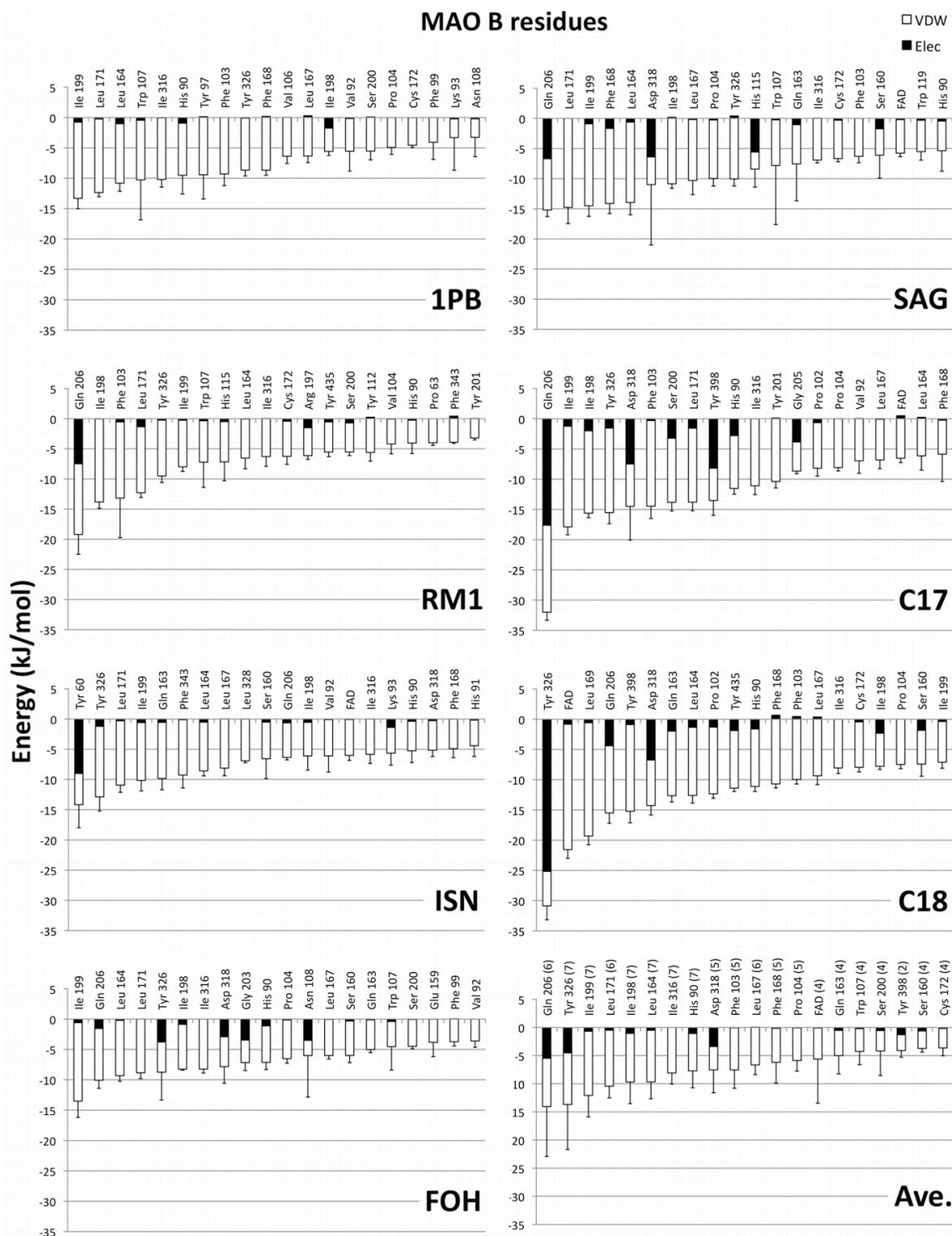


Figure 2.9. Summated and averaged energies of interaction between specific inhibitors and MAO B when pulling in a bilayer. The black component of each bar represents electrostatic interactions, and the white component represents van der Waals interactions. The final panel is an average of the other seven panels.

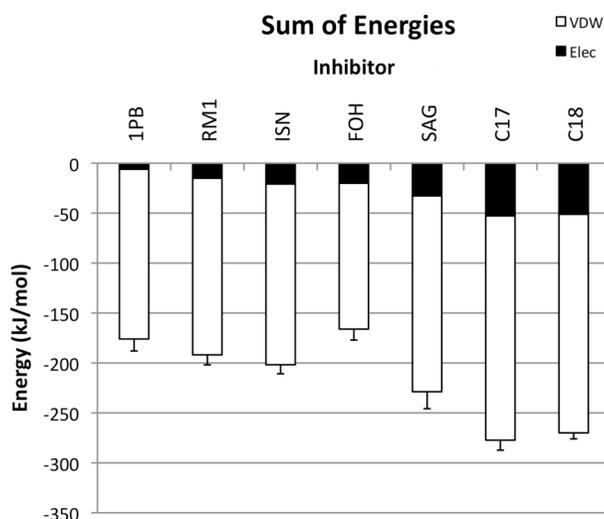


Figure 2.10. Sum of binding energy components for each inhibitor. The black part of each bar corresponds to electrostatic interactions, and the white part of each bar corresponds to van der Waals interactions. Inhibitors are arranged from left to right by increasing magnitude of ΔG_{dissoc} .

From analysis of the energies of interaction, we are also able to identify residues and specific interactions that may be important to inhibitor binding. The last panel on Figure 2.9 summarizes the average energy of interaction between MAO B residues and all inhibitors. The number in parenthesis following the inhibitor name in the final panel of Figure 2.9 indicates the frequency with which that residue appeared in the other top-20 plots. The three strongest-interacting residues are Gln 206, Tyr 326, and Ile 199 (Figure 2.11). It is evident that a strong electrostatic contact near the site of Gln 206 or Tyr 326 (both of which sit above the active site very near to the FAD cofactor) is a core component that helps account for a large ΔG_{dissoc} value. Further, a strong and large contribution to the interactions in the hydrophobic core is also essential to good inhibition. Of the top ten contributing residues, six are hydrophobic, five of which line the bottom of the active site and active site channel. Finally, His 90 and Asp 318 are also two important polar residues that contribute to the binding of the inhibitor. All of the inhibitors examined in this study bind adjacent to the FAD molecule in the back of the active site. From this position, these inhibitors are too short to form contacts with His 90 or Asp 318, which are positioned closer to the active site entrance. This indicates that the inhibitors only started interacting with these residues during the active site egress. Therefore, it would make sense to design an inhibitor that is larger and contains the appropriate group to form electrostatic contacts or hydrogen bonds with His 90 and Asp 318. Finally, we must remember that although several of

the electrostatic contacts are tempting to exploit, most of the restoring force comes from hydrophobic van der Waals type contacts, which should be the focus of any pursuit in inhibitor design.

To better characterize those specific interactions that are important to inhibitor binding, and to validate the model for inhibitor design proposed above, we briefly return to Figure 2.7. The peaks in the force curves correspond to different bound states of the inhibitors; the greater the height of the peak, the more tightly it is bound. Therefore it is useful to examine what specific interactions were occurring at the tallest peaks, and what interactions broke causing the downward slope in the curves. Presumably, at the top of the force peaks, the inhibitors should be most closely interacting with those residues identified as important in Figure 2.9. In fact, that is the case. For example, we examined the egress of C17 from the MAO B-bilayer active site. There are two notable peaks that occur during inhibitor egress, one at approximately 0.5 nm displacement, and another at 1.5 nm displacement (Figure 2.7). At 0.5 nm displacement, C17 formed hydrogen bonds with Gln 206 and His 90, pi-stacking interactions with Phe 103 and Tyr 326, and van der Waals interactions with Leu 171, Ile 198, Ile 199, and Ile 316. Therefore, the height of that first peak is a measure of the force required to overcome these interactions. Immediately past the tip of the first peak, those interactions were broken and the ligand moved further through the active site. At 1.5 nm displacement, the ligand formed electrostatic contacts with Phe 103, His 90, and Asp 318, and van der Waals contacts with Ile 316. The break in this second peak corresponded to a disruption in these specific interactions and total egress from the active site (Figure 2.11). This trend holds true for the other six inhibitors. In general, and especially in the case of the inhibitors with lower K_i values, the residues that are responsible for binding the inhibitor the most tightly are those that were identified as important by the energy profile in Figure 2.9. Increasing the strength of future inhibitors now becomes an exercise in modifying specific functional groups to optimize the strength of these interactions.

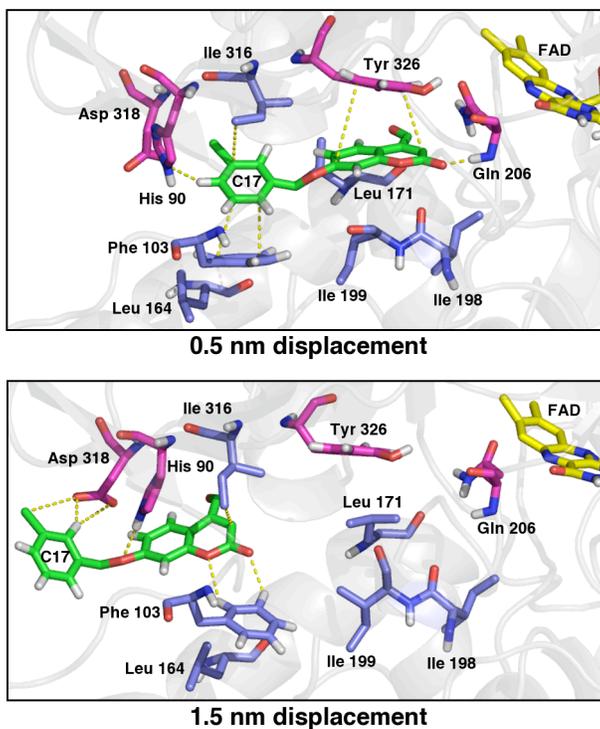


Figure 2.11. View of the MAO B active site during inhibitor C17 (green) egress at 0.5 nm displacement (top) and 1.5 nm displacement (bottom). Contacts between the inhibitor and hydrophobic (blue) or polar (red) residues under 4 nm are shown as yellow dashes.

Discussion

We performed classical MD simulations of the MAO B homodimer in a POPC-POPE mixed bilayer and in bulk solvent, followed by steered MD simulations of several reversible MAO B inhibitors also in bilayer or solvent. Taken together, these data effectively describe the mechanism of inhibition of MAO B by reversible inhibitors, and the role of the bilayer in governing the dynamics of MAO B – especially around the active site entrance.

The dynamics of two specific loops, residues 85-110 and 155-165 are closely modulated by the bilayer. Previously, it was suggested that the “active site loop”, residues 99-112, is important in inhibitor specificity (8). From our observations, it seems the modulation of the loop by the bilayer is a control mechanism in which the protein and bilayer work in a concerted effort to control ligand binding. The bilayer modulates these two loops by facilitating the transition between an open and a closed conformation. Fowler *et al.* (16) suggested that perhaps MAO B

undergoes a periodic tilting. While we did not find conclusive evidence to support or refute this hypothesis, we believe it could explain some behaviors we observed. When one active site entrance dipped down into the hydrophobic core of the bilayer, the change in the nature of the environment disrupted key interactions between residues at the mouth of the active site, facilitating the transition into the open conformation. We observed a dipping motion in these key active site residues with respect to the bilayer, and we found that the protein tended to be tilted with respect to the height of the active site entrance along the z -dimension. Previously, Apostolov *et al.* (17) examined a similar model, the monomeric MAO A isoform, and observed that the bilayer modulates the dynamics of the protein, and subsequently, ligand binding. However, they found that the same active site loop region (residues 105-115) fluctuates *more* in a bilayer, not less as we observed for MAO B. This difference could be due to subtle differences in the sequence between the MAO A and B isoforms, and may contribute to their distinguishing ligand specificities. The bilayer becomes thinner in the immediate vicinity of the protein, a phenomenon that was also observed by Fowler *et al.* (16). Here we showed that the thinning occurred especially at the protein interface with the bilayer.

The dynamics of inhibitor egress from the active site were also modulated by the bilayer. The Δ PMF of the unbinding process is a valuable measure that was weakly correlated to the strength of the inhibitor. We observed that the Δ PMF changed drastically in the presence of solvent, indicating that the bilayer maintained a level of control over the dynamics of MAO B in and around the active site. The increase in the number of contacts between MAO B and an inhibitor in the presence of a bilayer further illustrated this point. Finally, the energies of interaction, when taken together, provided key information that will be useful in designing new inhibitors. The right combination of hydrophobic and hydrophilic groups within the same inhibitor are paramount to the success of binding, and our model of the active site with important residues highlighted is key in designing such an inhibitor. Static crystal structures alone cannot fully describe the intricacies of the protein-ligand relationship. Dynamic models such as this are invaluable in identifying important interactions that may otherwise be missed.

Recently, Colizzi *et al.* used a form of steered MD to pull several inhibitors from the active site of *Plasmodium falciparum* β -hydroxyacyl-ACP dehydratase, thereby using the resulting Δ PMF values to distinguish between active and inactive inhibitors (42). Following that study, Jorgensen wrote a brief commentary in *Nature* describing the merits of steered MD simulations

and its utility in drug discovery (43). He questions whether the peak force applied during active site egress would correlate more strongly with inhibitor strength. In our study, we found that the peak force required to pull the inhibitor did not correlate more strongly with the strength of each inhibitor than did the Δ PMF value. However, the limitations of Jarzynski's equality have been well described (44). The accuracy of Δ PMF measurements will continue to increase as the number of replicates increases, and as the pulling speed and force constant in the spring decrease, each of which, in practice, increases the computing hours required to perform the simulation. That being said, despite the number of replicates we used and the pulling conditions, we observed a very strong correlation between inhibitor strength and the sum of the per-residue energies of interactions from the energy profiles (Figures 2.9-2.10). We propose that this rubric be used in tandem with force and Δ PMF measurements to further discern between inhibitors of varying strength.

In summary, the implication of these observations is quite profound. Foremost, we have demonstrated that MAO B in a bilayer does not behave similarly to MAO B in solvent, especially pertaining to the active site and active site entrance dynamics. Therefore, when studying MAO B, particularly in the context of drug design, the correct model should always be to study this enzyme in a bilayer. This suggestion extends well beyond just MAO B, and should be applied to any system studied by MD. Studying the biomolecules in the correct environment is key in obtaining accurate or meaningful results. Described herein are some important observations and implications that are useful to those taking a knowledge-based drug design approach to the design of powerful and reversible inhibitors of MAO B.

Acknowledgments

The authors thank Justin Lemkul for useful discussions regarding this project, James M. Tanko for allowing us the use of Titan, and the administrators of the Advanced Research Computing facility at Virginia Tech (<http://www.arc.vt.edu/>) for computing hours.

Supplemental Information

Inhibitor and Cofactor Parameters

Table S2.1. Parameters for 1,4-diphenyl-2-butene (1PB).

Structure	Atom	Atom Name	Charge (e)	Charge Group
	1	CH	-0.144	1
	2	HH	0.144	1
	3	CI	-0.141	2
	4	HI	0.141	2
	5	CJ	-0.151	3
	6	HJ	0.151	3
	7	CK	-0.141	4
	8	HK	0.141	4
	9	CL	-0.144	5
	10	HL	0.144	5
	11	CG	-0.004	6
	12	CA	0.001	6
	13	CB	0.003	6
	14	CC	0.003	6
	15	CD	0.001	6
	16	CI	-0.004	6
	17	CN	-0.144	7
	18	HN	0.144	7
	19	CO	-0.141	8
	20	HO	0.141	8
	21	CP	-0.151	9
	22	HP	0.151	9
	23	CQ	-0.141	10
	24	HQ	0.141	10
	25	CR	-0.144	11
	26	HR	0.144	11

Table S2.2. Parameters for N-methyl-1(R)-aminoindan (RM1).

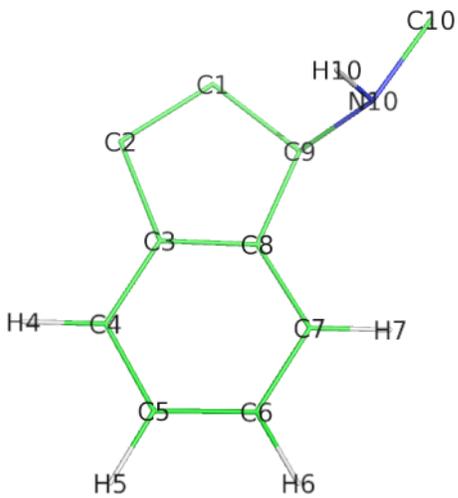
Structure	Atom	Atom Name	Charge (<i>e</i>)	Charge Group
	1	C1	-0.010	1
	2	C2	0.010	1
	3	C3	0.000	2
	4	C4	-0.140	3
	5	H4	0.140	3
	6	C5	-0.144	4
	7	H5	0.144	4
	8	C6	-0.144	5
	9	H6	0.144	5
	10	C7	-0.140	6
	11	H7	0.140	6
	12	C8	0.000	7
	13	C9	0.170	8
	14	N10	-0.620	8
	15	H10	0.270	8
	16	C10	0.180	8

Table S2.3. Parameters for 1H-indole-2,3-dione (isatin, ISN).

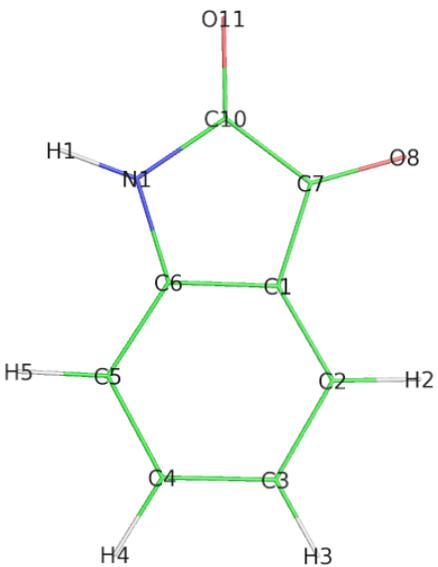
Structure	Atom	Atom Name	Charge (<i>e</i>)	Charge Group
	1	O8	-0.380	1
	2	C7	0.380	1
	3	C1	0.000	2
	4	C2	-0.100	3
	5	H2	0.100	3
	6	C3	-0.100	4
	7	H3	0.100	4
	8	C4	-0.100	5
	9	H4	0.100	5
	10	C5	-0.100	6
	11	H5	0.100	6
	12	C6	0.000	7
	13	N1	-0.050	8
	14	H1	0.050	8
	15	C10	0.380	9
	16	O11	-0.380	9

Table S2.4. Parameters for farnesol (FOH).

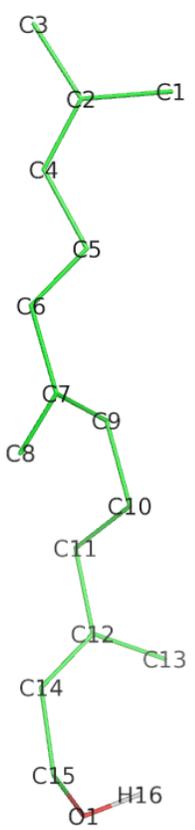
Structure	Atom	Atom Name	Charge (<i>e</i>)	Charge Group
	1	C1	0.005	1
	2	C2	0.015	1
	3	C3	0.005	1
	4	C4	-0.035	1
	5	C5	0.010	1
	6	C6	0.005	2
	7	C7	0.015	2
	8	C8	0.005	2
	9	C9	-0.035	2
	10	C10	0.010	2
	11	C11	0.010	3
	12	C12	0.030	3
	13	C13	0.010	3
	14	C14	-0.050	3
	15	C15	0.320	4
	16	O1	-0.650	4
	17	H16	0.330	4

Table S2.5. Parameters for (S)-(+)-2-[4-(fluorobenzyloxy-benzylamino) propionamide] (SAG).

Structure	Atom	Atom Name	Charge (e)	Charge Group
	1	C2	-0.190	1
	2	H2	0.190	1
	3	C3	0.420	2
	4	F3	-0.420	2
	5	C4	-0.190	3
	6	H4	0.190	3
	7	C5	-0.140	4
	8	H5	0.140	4
	9	C6	-0.144	5
	10	H6	0.144	5
	11	C1	-0.060	6
	12	C7	0.320	6
	13	O8	-0.680	6
	14	C9	0.420	6
	15	C14	-0.164	7
	16	H14	0.164	7
	17	C13	-0.115	8
	18	H13	0.115	8
	19	C10	-0.164	9
	20	H10	0.164	9
	21	C11	-0.115	10
	22	H11	0.115	10
	23	C12	-0.080	11
	24	C15	0.280	11
	25	N16	-0.400	11
	26	C17	0.150	11
	27	C18	0.050	11
	28	C19	0.720	12
	29	O20	-0.610	12
	30	N21	-0.745	12
	31	H75	0.325	12
	32	H21	0.310	12

Table S2.6. Parameters for 7-[(3-chlorobenzyl)oxy]-2-oxo-2H-chromene-4-carbaldehyde (C17).

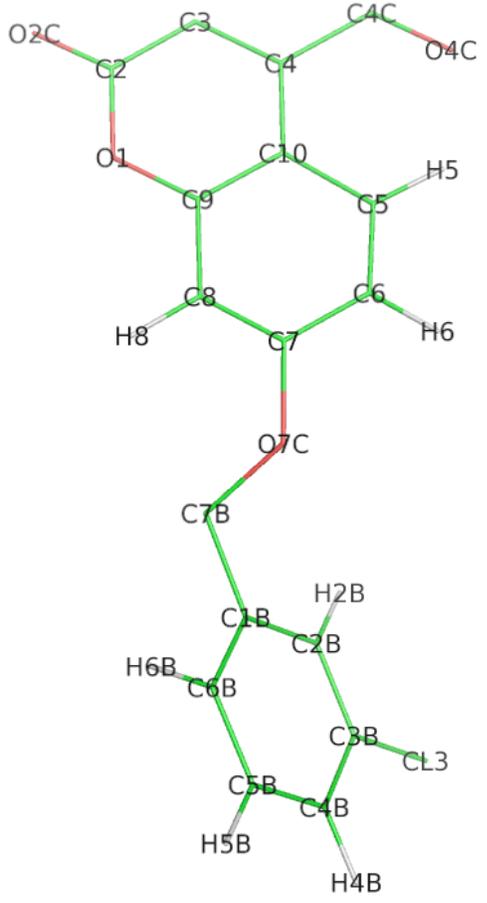
Structure	Atom	Atom Name	Charge (<i>e</i>)	Charge Group
	1	O4C	-0.500	1
	2	C4C	0.550	1
	3	C4	-0.050	1
	4	C3	-0.005	2
	5	C2	0.570	2
	6	O2C	-0.565	2
	7	O1	-0.600	3
	8	C9	0.520	3
	9	C8	-0.115	3
	10	H8	0.195	3
	11	C10	-0.100	4
	12	C5	-0.070	4
	13	H5	0.170	4
	14	C6	-0.144	5
	15	H6	0.144	5
	16	C7	0.330	6
	17	O7C	-0.645	6
	18	C7B	0.350	6
	19	C1B	-0.035	6
	20	C2B	-0.140	7
	21	H2B	0.140	7
	22	C3B	0.260	8
	23	CL3	-0.260	8
	24	C4B	-0.155	9
	25	H4B	0.155	9
	26	C5B	-0.130	10
	27	H5B	0.130	10
	28	C6B	-0.144	11
	29	H6B	0.144	11

Table S2.7. Parameters for 7-[(3-chlorobenzyl)oxy]-4-[(methylamino)methyl]-2H-chromen-2-one (C18).

Structure	Atom	Atom Name	Charge (e)	Charge Group
	1	C4N	0.195	1
	2	N4C	-0.620	1
	3	H8M	0.280	1
	4	C4C	0.205	1
	5	C4	-0.060	1
	6	C3	-0.005	2
	7	C2	0.570	2
	8	O2C	-0.565	2
	9	O1	-0.600	3
	10	C9	0.520	3
	11	C8	-0.115	3
	12	H8	0.195	3
	13	C10	-0.100	4
	14	C5	-0.070	4
	15	H5	0.170	4
	16	C6	-0.144	5
	17	H6	0.144	5
	18	C7	0.330	6
	19	O7C	-0.645	6
	20	C7B	0.350	6
	21	C1B	-0.035	6
	22	C2B	-0.140	7
	23	H2B	0.140	7
	24	C3B	0.260	8
	25	CL3	-0.260	8
	26	C4B	-0.155	9
	27	H4B	0.155	9
	28	C5B	-0.130	10
	29	H5B	0.130	10
	30	C6B	-0.144	11
	31	H6B	0.144	11

Table S2.8. Parameters for flavin adenine dinucleotide (FAD).

Structure	Atom	Atom Name	Charge (e)	Charge Group
	1	AP	0.660	1
	2	AO1	-0.635	1
	3	AO2	-0.635	1
	4	AO5*	-0.360	1
	5	AC5*	0.150	1
	6	AC4*	0.160	2
	7	AO4*	-0.360	2
	8	AC1*	0.200	2
	9	AC3*	0.150	3
	10	AO3*	-0.548	3
	11	HO3*	0.398	3
	12	AC2*	0.150	4
	13	AO2*	-0.548	4
	14	HO2*	0.398	4
	15	AN9	-0.200	5
	16	AC4	0.200	5
	17	AC8	0.360	6
	18	AN7	-0.360	6
	19	AC5	0.000	6
	20	AN1	-0.360	7
	21	AC6	0.360	7
	22	AN6	-0.830	8
	23	HN61	0.415	8
	24	HN62	0.415	8
	25	AC2	0.360	9
	26	AN3	-0.360	9
	27	N1	-0.360	10
	28	C10	0.360	10
	29	C2	0.380	11
	30	O2	-0.380	11
	31	N3	-0.280	12
	32	H3	0.280	12
	33	C4	0.380	13
	34	O4	-0.380	13
	35	C4A	0.180	14
	36	N5	-0.280	14
	37	C5A	0.100	14
	38	C6	0.000	15
	39	C7	0.000	16
	40	C7M	0.000	16
	41	C8	0.000	17
	42	C8M	0.000	17
	43	C9	0.000	18
	44	C9A	0.200	19
	45	N10	-0.200	19
	46	C1*	0.000	20
	47	C2*	0.150	21
	48	O2*	-0.548	21
	49	H2*	0.398	21
	50	C3*	0.150	22
	51	O3*	-0.548	22
	52	H3*	0.398	22
	53	C4*	0.150	23
	54	O4*	-0.548	23
	55	H4*	0.398	23
	56	C5*	0.150	24
	57	O5*	-0.360	24
	58	P	0.660	24
	59	O1P	-0.635	24
	60	O2P	-0.635	24
	61	O3P	-0.360	24

Tabulated Δ PMF and Peak Force Values

Table S2.9. Average Δ PMF and peak force values for steered MD simulations.

Inhibitor	ΔG_{dissoc} (kJ/mol)	Bilayer		Solvent	
		ΔPMF (kJ/mol)	Peak Force (kJ/mol/nm)	ΔPMF (kJ/mol)	Peak Force (kJ/mol/nm)
1PB	26.4	770 \pm 150	606 \pm 36	313 \pm 34	373 \pm 35
RM1	28.3	798 \pm 113	583 \pm 25	492 \pm 56	482 \pm 56
ISN	32.8	640 \pm 134	477 \pm 54	517 \pm 113	614 \pm 127
FOH	33.5	1002 \pm 171	646 \pm 117	407 \pm 22	419 \pm 29
SAG	37.7	958 \pm 125	644 \pm 58	490 \pm 51	583 \pm 69
C17	38.0	1201 \pm 94	730 \pm 82	649 \pm 55	651 \pm 39
C18	41.5	823 \pm 72	560 \pm 52	509 \pm 27	482 \pm 70

Additional RMSF Plots

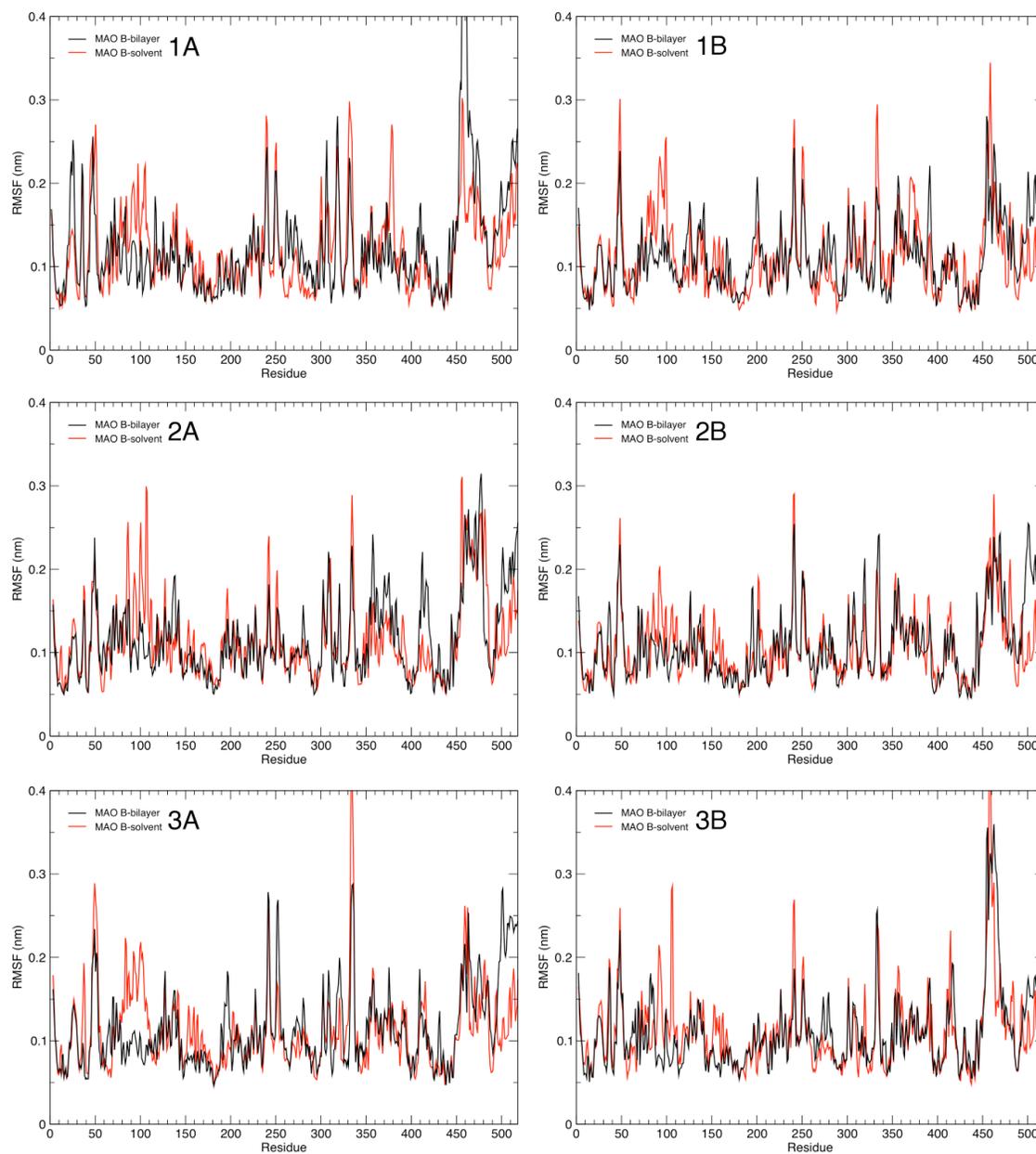


Figure S2.1. All six RMSF plots comparing degree of fluctuation per MAO B residue in solvent vs. bilayer. The identifiers (1A, 2A, 1B...etc.) refer to a replicate (1-3) and a protein chain (A or B).

Linear Regression Plots

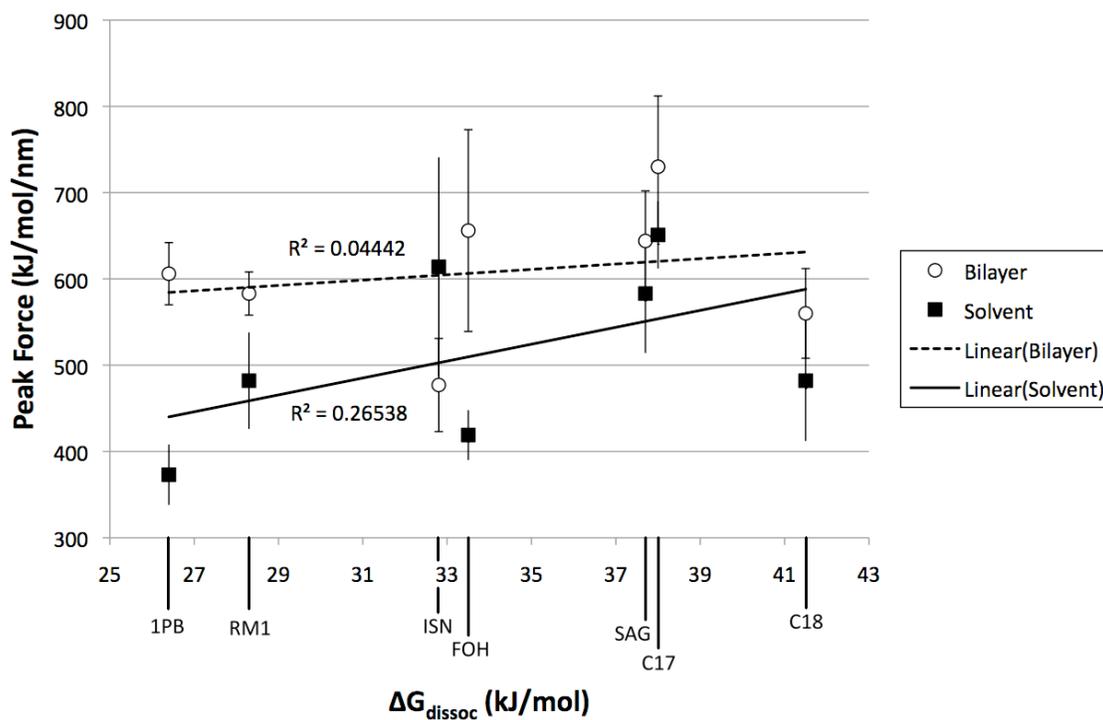


Figure S2.2. Linear regression plots and R^2 value correlating the ΔG_{dissoc} to peak force for inhibitor unbinding in both media.

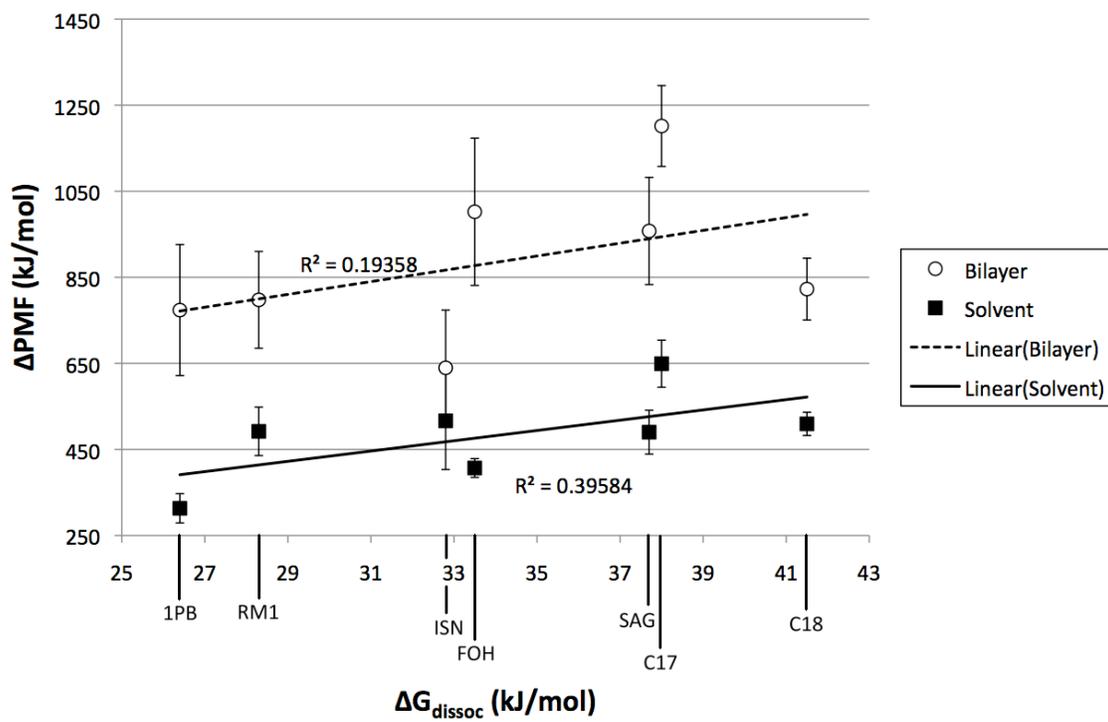


Figure S2.3. Linear regression plots and R^2 value correlating the ΔG_{dissoc} to ΔPMF for inhibitor unbinding in both media.

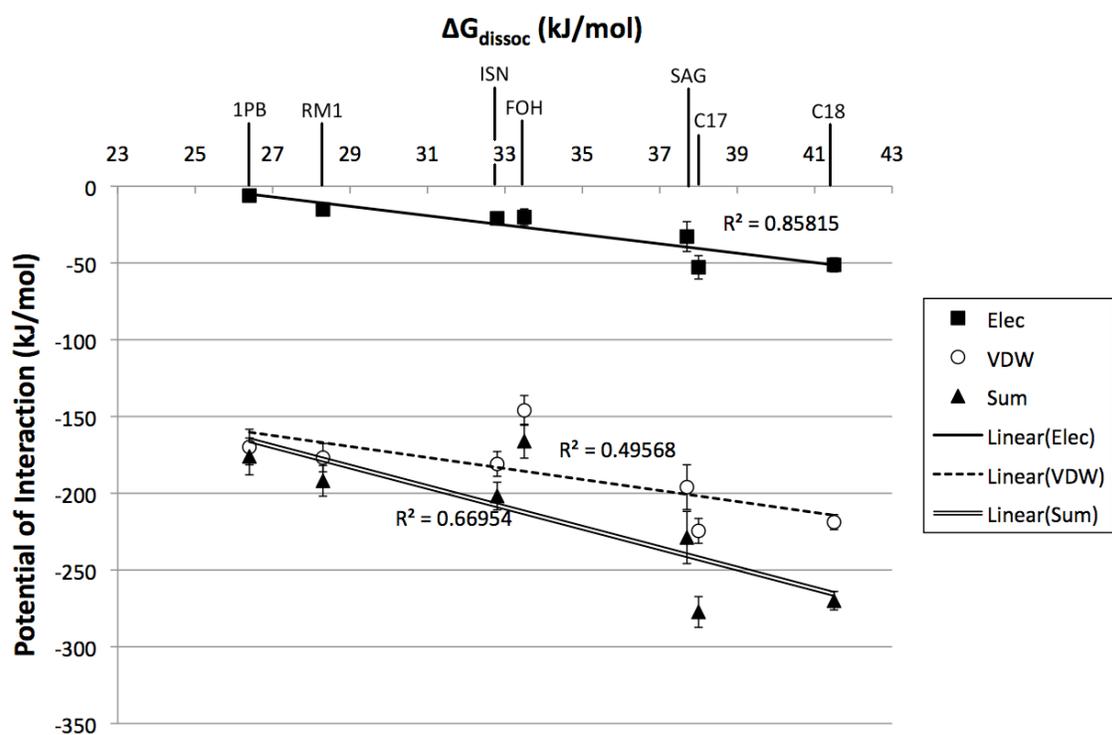


Figure S2.4. Linear regression plots and R^2 values correlating ΔG_{dissoc} to the sum of interaction energies during steered MD.

References

- (1) Shih, J. C.; Chen, K.; Ridd, M. J. Monoamine Oxidase: From Genes to Behavior. *Annu. Rev. Neurosci.* **1999**, *22*, 197-217.
- (2) Fowler, J. S.; Volkow, N. D.; Wang, G.-J.; Logan, J.; Pappas, N.; Shea, C.; Macgregor, R. Age-Related Increases in Brain Monoamine Oxidase B in Living Healthy Human Subjects. *Neurobiol. Aging* **1997**, *18*, 431-435.
- (3) Tabakman, R.; Lecht, S.; Lazarovici, P. Neuroprotection by Monoamine Oxidase B Inhibitors: A Therapeutic Strategy for Parkinson's Disease? *BioEssays* **2004**, *26*, 80-90.
- (4) Birkmayer, W.; Knoll, J.; Riederer, P.; Youdim, M. B.; Hars, V.; Marton, J. Increased Life Expectancy Resulting from Addition of L-deprenyl to Madopar Treatment in Parkinson's Disease: A Longterm Study. *J. Neural Transm.* **1985**, *64*, 113-127.
- (5) Finberg, J. P. M.; Lamensdorf, I.; Commissiong, J. W.; Youdim, M. B. H. Pharmacology and Neuroprotective Properties of Rasagiline. *J. Neural Transm.* **1996**, *48*, 95-101.
- (6) Naoi, M.; Maruyama, W. Functional Mechanism of Neuroprotection by Inhibitors of Type B Monoamine Oxidase in Parkinson's Disease. *Expert Rev. Neurother.* **2009**, *9*, 1233-1250.
- (7) Binda, C.; Wang, J.; Pisani, L.; Caccia, C.; Carotti, A.; Salvati, P.; Edmondson, D. E.; Mattevi, A. Structures of Human Monoamine Oxidase B Complexes with Selective Noncovalent Inhibitors: Safinamide and Coumarin Analogs. *J. Med. Chem.* **2007**, *50*, 5848-5852.
- (8) Binda, C.; Newton-Vinson, P.; Hubálek, F.; Edmondson, D. E.; Mattevi, A. Structure of Human Monoamine Oxidase B, a Drug Target for the Treatment of Neurological Disorders. *Nat. Struct. Biol.* **2002**, *9*, 22-26.
- (9) Edmondson, D. E.; Binda, C.; Mattevi, A. The FAD Binding Sites of Human Monoamine Oxidases A and B. *Neurotoxicology* **2004**, *25*, 63-72.
- (10) Edmondson, D. E.; Binda, C.; Mattevi, A. Structural insights into the mechanism of amine oxidation by monoamine oxidases A and B. *Arch. Biochem. Biophys.* **2007**, *464*, 269-276.

- (11) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235-242.
- (12) Tsugeno, Y.; Ito, A. A Key Amino Acid Responsible for Substrate Selectivity of Monoamine Oxidase A and B. *J. Biol. Chem.* **1997**, *272*, 14033-14036.
- (13) Hubálek, F.; Binda, C.; Khalil, A.; Li, M.; Mattevi, A.; Castagnoli, N.; Edmondson, D. E. Demonstration of Isoleucine 199 as a Structural Determinant for the Selective Inhibition of Human Monoamine Oxidase B by Specific Reversible Inhibitors. *J. Biol. Chem.* **2005**, *280*, 15761-15766.
- (14) Geha, R. M.; Rebrin, I.; Chen, K.; Shih, J. C. Substrate and Inhibitor Specificities for Human Monoamine Oxidase A and B Are Influenced by a Single Amino Acid. *J. Biol. Chem.* **2001**, *276*, 9877-9882.
- (15) Li, M.; Binda, C.; Mattevi, A.; Edmondson, D. E. Functional Role of the "Aromatic Cage" in Human Monoamine Oxidase B: Structures and Catalytic Properties of Tyr435 Mutant Proteins. *Biochemistry* **2006**, *45*, 4775-4784.
- (16) Fowler, P. W.; Balali-Mood, K.; Deol, S.; Coveney, P. V.; Sansom, M. S. P. Monotopic Enzymes and Lipid Bilayers: A Comparative Study. *Biochemistry* **2007**, *46*, 3108-3115.
- (17) Apostolov, R.; Yonezawa, Y.; Standley, D. M.; Kikugawa, G.; Takano, Y.; Nakamura, H. Membrane Attachment Facilitates Ligand Access to the Active Site in Monoamine Oxidase A. *Biochemistry* **2009**, *48*, 5864-5873.
- (18) Binda, C.; Li, M.; Hubálek, F.; Restelli, N.; Edmondson, D. E.; Mattevi, A. Insights into the Mode of Inhibition of Human Mitochondrial Monoamine Oxidase B from High-Resolution Crystal Structures. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 9750-9755.
- (19) Binda, C.; Hubálek, F.; Li, M.; Herzig, Y.; Sterling, J.; Edmondson, D. E.; Mattevi, A. Binding of Rasagiline-Related Inhibitors to Human Monoamine Oxidases: A Kinetic and Crystallographic Analysis. *J. Med. Chem.* **2005**, *48*, 8148-8154.
- (20) Case, D. A.; Darden, T. A.; Cheatham III, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Crowley, M.; Walker, R. C.; Zhang, W.; Merz, K. M.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossvery, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.;

- Cui, G.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Kollman, P. A. AMBER 10, University of California, San Francisco, **2008**.
- (21) Schüttelkopf, A. W.; van Aalten, D. M. F. PRODRG: A Tool for High-Throughput Crystallography of Protein-Ligand Complexes. *Acta Crystallogr.* **2004**, *D60*, 1355-1363.
- (22) Lemkul, J. A.; Allen, W. J.; Bevan, D. R. Practical Considerations for Building GROMOS-Compatible Small Molecule Topologies. *J. Chem. Inf. Model.* **2010**, *50*, 2221-2235.
- (23) Berger, O.; Edholm, O.; Jähnig, F. Molecular Dynamics Simulations of a Fluid Bilayer of Dipalmitoylphosphatidylcholine at Full Hydration, Constant Pressure, and Constant Temperature. *Biophys. J.* **1997**, *72*, 2002-2013.
- (24) Tieleman, D. P.; Sansom, M. S. P.; Berendsen, H. J. C. Alamethicin Helices in a Bilayer and in Solution: Molecular Dynamics Simulations. *Biophys. J.* **1999**, *76*, 40-49.
- (25) van Meer, G.; Voelker, D. R.; Feigenson, G. W. Membrane Lipids: Where They Are and How They Behave. *Nat. Rev. Mol. Cell Biol.* **2008**, *9*, 112-124.
- (26) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435-447.
- (27) Darden, T.; York, D.; Pedersen, L. G. Particle mesh Ewald: An N-log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089-10092.
- (28) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H. Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577-8593.
- (29) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997**, *18*, 1463-1472.
- (30) Hess, B. P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *J. Chem. Theory Comput.* **2007**, *4*, 116-122.
- (31) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. *Intermolecular Forces, Interaction Models for Water in Relation to Protein Hydration*, D. Reidel Publishing Co., Dordrecht, The Netherlands, **1981**.
- (32) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81*, 3684-3690.

- (33) Nosé, S. A Molecular Dynamics Method for Simulations in the Canonical Ensemble. *Mol. Phys.* **1984**, *52*, 255-268.
- (34) Hoover, W. G. Canonical dynamics: Equilibrium Phase-Space Distributions. *Phys. Rev. A* **1985**, *31*, 1695-1697.
- (35) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52*, 7182-7190.
- (36) Nosé, S.; Klein, M. L. Constant Pressure Molecular Dynamics for Molecular Systems. *Mol. Phys.* **1983**, *50*, 1055-1076.
- (37) Kandt, C.; Ash, W. L.; Tieleman, D. P. Setting Up and Running Molecular Dynamics Simulations of Membrane Proteins. *Methods* **2007**, *41*, 475-488.
- (38) DeLano, W. L. *The PyMOL User's Manual*, DeLano Scientific, Palo Alto, CA, USA, **2002**.
- (39) Turner, P. J. *Grace*, Center for Coastal and Land-Margin Research Oregon Graduate Institute of Science and Technology, Beaverton, Oregon. **2002**.
- (40) Allen, W. J.; Lemkul, J. A.; Bevan, D. R. GridMAT-MD: A Grid-Based Membrane Analysis Tool for Use with Molecular Dynamics. *J. Comput. Chem.* **2009**, *30*, 1952-1958.
- (41) Jarzynski, C. Nonequilibrium Equality for Free Energy Differences. *Phys. Rev. Lett.* **1997**, *78*, 2690-2693.
- (42) Colizzi, F.; Perozzo, R.; Scapozza, L.; Recanatini, M.; Cavalli, A. Single-Molecule Pulling Simulations Can Discern Active from Inactive Enzyme Inhibitors. *J. Am. Chem. Soc.* **2010**, *132*, 7361-7371.
- (43) Jorgensen, W. L. Drug discovery: Pulled from a protein's embrace. *Nature* **2010**, *466*, 42-43.
- (44) Baştuğ, T.; Chen, P.-C.; Patra, S. M.; Kuyucak, S. Potential of mean force calculations of ligand binding to ion channels from Jarzynski's equality and umbrella sampling. *J. Chem. Phys.* **2008**, *128*, 155104.1-155104.9.

Chapter 3

Reaction of Benzophenone Triplet with Aliphatic Amines. What a Potent Neurotoxin Can Tell Us About the Reaction Mechanism

Michelle L. Grimm, William J. Allen, Meghan Finn, Neal Castagnoli Jr., and James M. Tanko

Reproduced with permission from *Bioorganic & Medicinal Chemistry* **2011**, *19* (4), 1458-1463. Copyright 2011, Elsevier.

Attribution

Of the work discussed within this chapter, Grimm performed the laser flash photolysis of compounds **9** and **10**, and made significant contributions to the writing. Finn determined rate constants for compounds **2** and **9**, and performed the laser flash photolysis of compound **2**. Allen performed many of the calculations and contributed to the writing of the computational methods and results sections. Tanko performed further calculations and prepared much of the manuscript. Castagnoli Jr. reviewed and edited the manuscript, and provided guidance in the research.

Abstract

A photochemical model study of benzophenone triplet (^3BP) with the MAO B substrate 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP [**1**]) and two of its derivatives, 1-cyclopropyl-4-phenyl-1,2,3,6-tetrahydropyridine [**2**] and (+/-)-trans-2-phenylcyclopropyl-4-phenyl-1,2,3,6-tetrahydropyridine [**3**] were performed. Literature precedent and calculations reported herein suggest that the barrier to ring opening for aminyl radical cations derived from N-cyclopropyl derivatives of tertiary amines (such as MPTP) will be low. The laser plash

photolysis results reported herein demonstrate that pathways for the reaction of ³BP with **1**, **2**, and **3** are very similar. In each instance, disappearance of ³BP is accompanied solely by appearance of bands corresponding to the diphenylhydroxymethyl radical and neutral radical derived from MPTP and its two derivatives **2** and **3**. These results suggest that the reaction between ³BP and tertiary aliphatic amines proceed via a simple hydrogen atom transfer reaction. Additionally, these model examinations provide evidence that oxidations of N-cyclopropyl derivatives of MPTP catalyzed by monoamine oxidase B may not be consistent with a pure single electron transfer pathway.

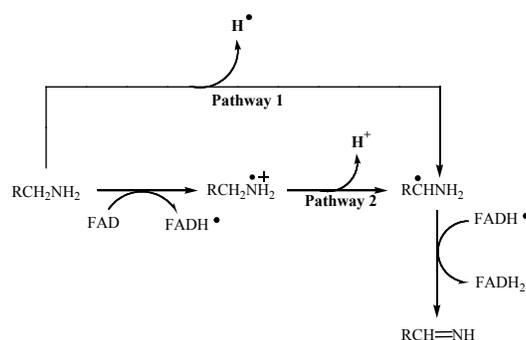
Introduction

Monoamine oxidase (MAO) A and B are flavoenzymes that catalyze the oxidation of various monoaminergic neurotransmitters, including serotonin and dopamine (1), as well as some xenobiotics (2). MAO A selectively catalyzes the oxidation of serotonin; the B isoform is the principal enzyme that catalyzes the oxidation of dopamine (3). MAO B has gained additional attention since it catalyzes the bioactivation of certain tetrahydropyridinyl derivatives to neurotoxic metabolites that mediate the degradation of dopaminergic neurons in the *substantia nigra* resulting in a parkinsonian syndrome closely resembling idiopathic Parkinson's disease (4).

Model studies have led to several proposals to account for the catalytic activity of MAO B. A polar pathway involving the addition of the aminyl substrate across the 4a–5 double bond of the covalently bound flavin adenine dinucleotide (FAD) co-factor, followed by an intramolecular redox reaction has been proposed by Mariano (5, 6) and, more recently, Edmondson (7). Two radical-based pathways also have been considered. Silverman has been a strong proponent of the single electron transfer (SET) mechanism (1, 8) and Castagnoli has presented evidence consistent with a hydrogen atom transfer (HAT) mechanism (2, 9) (Scheme 3.1).

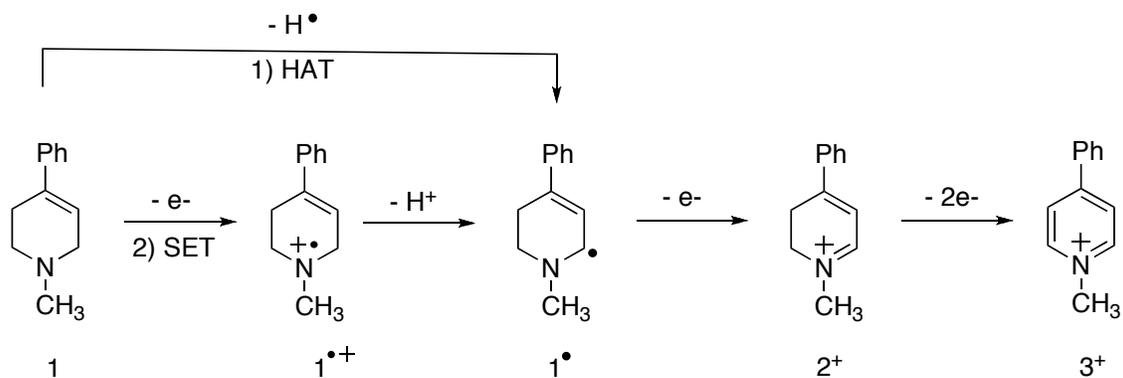
Scheme 3.1 outlines the differences between the SET and HAT proposals. Although these two pathways can account for some of the behaviour of substrates and inhibitors, it is our view that neither of these radical-based mechanisms nor the polar pathway account for all of the available experimental evidence. Since our interests are linked to the MAO-catalyzed metabolic

activation of cyclic tertiary allylamines, the discussion that follows will focus on the radical pathways since nucleophilic addition of tertiary amines across the 4a–5 double bond (the polar pathway) should be sterically prohibited. The present manuscript describes the results of our attempts to provide evidence to help distinguish the potential contributions of the SET and HAT mechanisms using photochemically-activated benzophenone as the electron or hydrogen atom acceptor and various tetrahydropyridinyl derivatives that are known MAO B substrates and/or for which independent documentation of the fate of electrochemically mediated 1-electron oxidations is available.



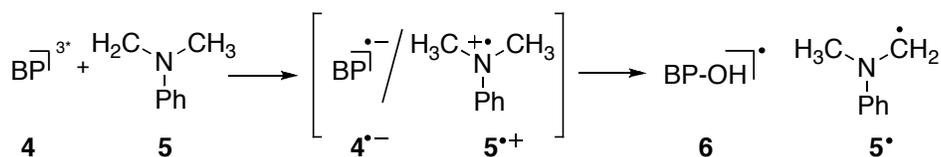
Scheme 3.1. Proposed Pathways for MAO oxidations.

Over the past 30 years, much focus has been given to 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP [**1**]; bold numbers in square brackets are specific compound abbreviations), a compound known to induce an irreversible parkinsonian disorder in humans and animal models (10, 11). Subsequent biochemical studies led to the discovery that MPTP is oxidatively metabolized in the brain to give the dihydropyridiniumyl species MPDP⁺ [**2**] that subsequently undergoes a two-electron oxidation to give the toxic pyridiniumyl product MPP⁺ [**3**] (9) (Scheme 3.2). MPP⁺ is thought to mediate the degeneration of dopaminergic neurons (4). Mechanistic details of this important bioactivation pathway remain poorly understood. What is known is that the initial 2-electron oxidation is catalyzed efficiently by MAO B.



Scheme 3.2. The MAO-B catalyzed bioactivation of MPTP [1] to MPDP⁺ [2⁺] and MPP⁺ [3⁺].

Benzophenone is a common photosensitizer; its chemistry with aromatic amines has been well characterized. Although the products of the reaction give the *appearance* of a HAT process, laser flash photolysis (LFP) experiments have shown that the reaction of benzophenone triplet [4] with aromatic amines proceeds by SET, generating a solvent separated ion pair (SSIP) consisting of benzophenone radical anion ([4^{•-}] $\lambda_{max} = 715$ nm) and an aminyl radical cation [5^{•+}] (12, 13). This caged-pair subsequently reacts via proton transfer, with the aminyl radical cation acting as an acid and the benzophenone radical anion as a base, generating diphenylmethanol radical ([6] $\lambda_{max} = 545$ nm) and the neutral aminyl radical [5[•]] (Scheme 3.3).

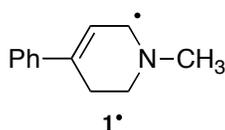


Scheme 3.3. Reaction of benzophenone triplet (³BP) with aromatic amines.

Similarly, ³BP reacts readily with aliphatic amines with a reactivity order 3° > 2° > 1° (9). Rate constants for tertiary and secondary amines are on the order of diffusion controlled, while primary amines are on the order of 10⁷ M⁻¹ s⁻¹ (14). Although the reaction is also a formal HAT process, the details of the mechanism remain ambiguous. Assuming electron transfer occurs, proton transfer between the initially produced 4^{•-}/5^{•+} caged pair may occur too rapidly for these species to be detected. Also, unlike aromatic amines, spectroscopic characterization of the initially formed R₃N^{•+} is difficult because these species absorb weakly and thus are invisible by UV. Consequently, both HAT and SET mechanisms have been suggested for the reactions of

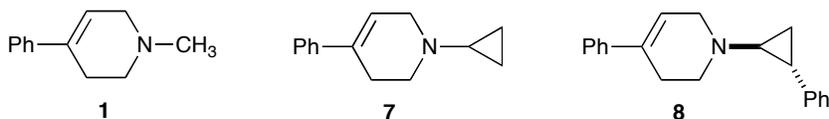
benzophenone triplet and aliphatic amines. Achieving an unambiguous assignment of the reaction mechanism has proven problematic.

Recently, we reported rate constants for the reaction of *t*-butoxyl radical (${}^t\text{BuO}\cdot$) with the potent neurotoxin MPTP (15, 16). In this study, ${}^t\text{BuO}\cdot$ was generated by the photolysis of di-*t*-butyl peroxide (DTBPO). This method was employed to measure the absolute rate constants for hydrogen atom abstractions from MPTP by ${}^t\text{BuO}\cdot$, generated by LFP, in order to gain insight into the radical-based chemistry of MPTP. Unlike the radicals derived from HAT in simple aliphatic amines (17,18), the MPTP derived radical [$\mathbf{1}\cdot$] gave rise to a transient species ($\lambda_{\text{max}} = 385 \text{ nm}$) that could be easily monitored. The strongly absorbing species was assigned to the α -allylic MPTP radical based on isotopic labeling studies. The observed rate constant for the formation of the MPTP derived radical was determined to be $2.27 \times 10^8 \text{ M}^{-1} \text{ s}^{-1}$ (15, 16).



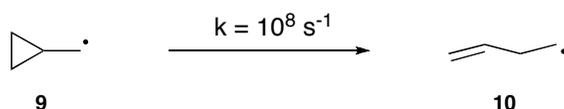
Scheme 3.4. Structure of compound $\mathbf{1}\cdot$.

In principle, hydrogen atom abstraction can occur at each of the carbons alpha to nitrogen in MPTP, as was demonstrated for hydrogen abstractions by ${}^t\text{BuO}\cdot$ (15). However, only neutral radical $\mathbf{1}\cdot$ possesses a suitable chromophore that can be readily monitored at 385 nm. Accordingly, it was envisioned that this radical would provide a unique spectroscopic handle to study the mechanism of the reaction between ${}^3\text{BP}$ and aliphatic amines. This manuscript describes the photochemical investigation of reaction of ${}^3\text{BP}$ with MPTP [$\mathbf{1}$] and two of its derivatives; 1-cyclopropyl-4-phenyl-1,2,3,6-tetrahydropyridine [*N*-cyclopropyl MPTP] ($\mathbf{7}$) and (+/-)-[*trans*-2-phenylcyclopropyl-4-phenyl-1,2,3,6-tetrahydropyridine] [*N*-cyclopropylphenyl MPTP] ($\mathbf{8}$).



Scheme 3.5. Structures of compounds $\mathbf{1}$, $\mathbf{7}$, and $\mathbf{8}$.

The *N*-cyclopropyl MPTP derivatives provide a unique way to probe the mechanism of these ³BP-mediated oxidations because in direct analogy to the cyclopropylcarbinyl → homoallyl neutral radical rearrangement (Scheme 3.6), the derived radical cations (if produced) are expected to undergo instantaneous ring opening (19). Moreover, because the barrier to this ring opening is exceedingly small (*vide infra*), this process should be competitive with deprotonation of the aminyl radical cation.



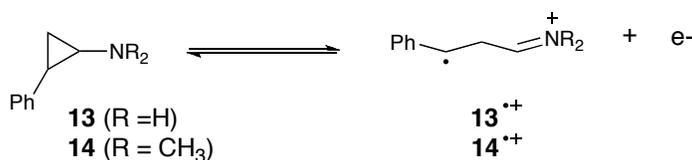
Scheme 3.6. Ring opening of the cyclopropylcarbinyl neutral radical.

Literature precedent suggests that **7**^{•+} and **8**^{•+} will have little or no barrier to ring opening. Indeed, 1-electron oxidation of **8** may occur via a dissociative electron transfer (DET) reaction, wherein electron transfer and ring opening occur simultaneously (20). Electron paramagnetic resonance (EPR) examinations of the radical cation generated from *N*-cyclopropylamine [**11**] support this statement (19). EPR experiments in a solid matrix have provided evidence for the distonic radical cation (ring opened) structure based upon hyperfine coupling constants and have shown that the aminyl radical cation ring-opens to the distonic radical cation rapidly upon electron removal (21, 22). Subsequent molecular orbital calculations at the second-order Møller-Plesset (MP2) and other levels of theory have suggested that the removal of an electron from *N*-cyclopropylamine occurs simultaneously with ring opening; the only stable structure is the distonic radical ion, which collapses to a more stable aminopropenyl ion (23, 24).

Studies on MPTP and the cyclopropyl analogs **7** and **8** by online coupling of electrochemistry and electrospray mass spectrometry (EC-ESI-MS) have been reported (25). The technique involves passing a substrate solution through a porous graphite electrode and varying the potential from 0 to 1500 mV. The concentrations of substrates and products resulting from 1-electron oxidation are then monitored via liquid chromatography-mass spectrometry (LC-MS). Experiments conducted with compound **1** showed evidence for electron transfer to generate the aminyl radical cation **1**^{•+} followed by subsequent deprotonation at the allylic position to yield **1**[•]. On the other hand, the analogous reactions with compounds **7** and **8** were consistent with rapid ring opening of the aminyl radical cations to give corresponding distonic radical cation. No

dihydropyridinium products corresponding to electron transfer followed by allylic α -carbon deprotonation were detected. These experiments were further probed by computational analysis at the unrestricted Hartree-Fock (UHF) level of theory. No global minima corresponding to the ring-closed aminyl radical cation $\mathbf{8}^{+\bullet}$ was ever obtained, indicating that the electron transfer and ring opening proceed by a concerted process.

Dinnocenzo, *et al.* (26) studied the fate of a series of structurally-related *N*-(2-phenylcyclopropyl) aminyl radical cations ($[\mathbf{11}^{+\bullet}]$ and $[\mathbf{12}^{+\bullet}]$) by nanosecond LFP. Ring opening of the radical cation was estimated to be exothermic by $-26 \text{ kcal mol}^{-1}$. It was suggested that ring opening and electron transfer were occurring in a concerted fashion and that *N*-(2-phenylcyclopropyl) aminyl radical cations have virtually no barrier for ring opening (Scheme 3.7).



Scheme 3.7. Ring opening of *N*-(2-phenylcyclopropyl) aminyl radical cations.

Literature precedent as well as higher order molecular orbital calculations suggest that the barrier for ring opening of tertiary amine radical cations such as *N*-(2-phenylcyclopropyl) derivative $[\mathbf{3}]$ to be virtually nonexistent. Due to the unique chromophore derived from $\mathbf{3}$, LFP can be used to spectroscopically observe a reactive intermediate that is generated from excitation, be it radical or radical ion. Herein we report: (1) computational analysis of ring opening reactions of relevant cyclopropyl amines, and (2) the results of model studies on the photochemically-mediated oxidation of MPTP and two of its derivatives that are designed to evaluate the HAT vs. the SET pathways.

Materials and Methods

Materials. All solvents and fine chemicals used in this study were obtained from Aldrich and used as received unless otherwise noted. Benzophenone was recrystallized in methanol prior to

use. Syntheses of MPTP and derivatives have been described previously (20, 28); these compounds were stored as their oxalate salts. For sample preparation, the oxalate salts were suspended in saturated potassium carbonate and extracted using ethyl acetate to render the free base prior to use. Solution concentrations were calculated using the mass of the free base. **Caution:** MPTP is a known neurotoxin and should be handled in a well-ventilated hood and proper personal protective equipment should be worn while working with this material. Procedures for the safe handling of MPTP have been documented (29).

Apparatus. Steady-state UV/Vis spectra were recorded on a Hewlett Packard diode array UV/Visible spectrophotometer (HP 8452A). LFP experiments were conducted using an Applied Photophysics LKS.60 spectrometer using the third harmonic of a Continuum Surelite I-10 Nd:YAG laser (4 - 6 ns pulse, 355 nm). Absorption spectra were monitored by a Hewlett Packard Infinium digital oscilloscope and analyzed with an Applied Photophysics SpectraKinetic Workstation software package (v. 4.59). Experiments were performed with a jacketed cell holder connected to a VWR Scientific Products (PolyScience) variable temperature circulating bath (model 1150-A) thermally equilibrated to 25 °C.

Laser Flash Photolysis (LFP). Sample solutions were prepared in benzene or acetonitrile and deoxygenated with argon prior to use. (Steady-state UV/Vis spectra were recorded to ensure that benzophenone was the only species absorbing at the excitation wavelength). In most LFP experiments, 1:1 concentrations of the MPTP derivatives and benzophenone (4.5 mM) were examined in acetonitrile with varying concentrations of lithium perchlorate.

Concentration profile for MPTP and benzophenone. Solutions containing varying concentration of amine and benzophenone (4 mM) were prepared. All solutions were degassed prior to use. Pseudo first order rate constants for the reaction of benzophenone triplet (³BP) with MPTP were obtained by plotting the experimentally observed rate (k_{obs}) constants of ³BP decay as a function of amine concentration [MPTP] as follows:

$$k_{obs} = k_0 + k[MPTP]$$

Transient Absorption Spectroscopy. Sample solutions containing approximately a 1:1 ratio of benzophenone and MPTP (4.5 mM) were prepared in distilled acetonitrile or benzene. Solutions were degassed prior to use. All experiments were conducted with the use of a flow cell to ensure

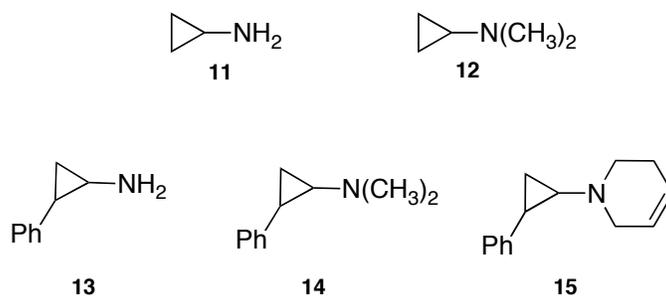
that fresh solution was used for all runs. UV/Vis spectra were recorded prior to runs to ensure benzophenone was the only absorbing species.

Calculations. Density functional theory calculations were performed using the Titan (30) molecular modeling software and/or Gaussian '03 (31). Initial calculations for all four compounds [11-15] included the structure of the radical cation in the ring-closed form transition state, and ring-opened form – each of which were isolated at the Austin Model 1 (AM1) level of theory. Structural results were imported from the AM1 level calculations, followed by geometry optimization and frequency calculations at the UHF level of theory with the 6-31G* basis set. The results (structural coordinates) from this calculation were taken to a higher energy level, and once again geometry optimization and frequency calculations were performed at the second-order unrestricted Møller-Plesset (UMP2) level of theory with the 6-31G* basis set. Finally, the energies of the structures that were determined at UMP2 were calculated once more at the fourth-order unrestricted Møller-Plesset level of theory with single, double, triple, and quadruple excitation states (UMP4SDTQ) and the 6-311G** basis set.

Results and Discussion

Calculations

The ring opening of radical cations derived from: *N*-cyclopropylamine [11], *N,N*-dimethylaminocyclopropane [12], 2-phenylcyclopropylamine [13], *N,N*-dimethyl-*N*-(2-phenylcyclopropyl)amine and [14] 1-cyclopropylphenyl-1,2,3,6-tetrahydropyridine [15] were examined computationally. As noted, calculations pertaining to the ring opening chemistry of **11**^{•+} have been previously reported (21-24), thereby allowing us to validate the computational methods before examining the chemistry of compounds **12**^{•+} through **15**^{•+}. The following are the specific issues to be addressed: Will a tertiary aminyl radical cation such as **12**^{•+} be more stable than primary radical **11**^{•+}, and thus have a barrier to ring opening? Will phenyl substitution on the cyclopropyl group lower or eliminate the barrier to ring opening, if indeed such a barrier exists?



Scheme 3.8. Structures of compounds **11-15**.

N-cyclopropylaminyl radical cation [**11**^{•+}]. Consistent with published results (21-24), an energy profile (plot of energy vs. C-C bond length) of **11**^{•+} at the UHF/6-31G* level of theory suggest that the ring-closed form of **11**^{•+} does not exist at a potential energy minimum. Attempts to minimize the energy of the cyclopropyl ring-closed form of **11**^{•+} led to a minimized structure corresponding to the distonic (ring-opened) radical cation. Similar results were obtained at the UMP2/6-31G* and UMP4SDTQ/6-311G** levels of theory/basis sets. Geometry optimization led exclusively to the ring-opened form. These results confirm that there is no barrier associated with the ring-opening of **11**^{•+}. Because these results correspond with previously published studies (21-24), it can be said with confidence that these computational methods can also be utilized to examine compounds **12**^{•+} through **15**^{•+}.

N,N-dimethyl-*N*-cyclopropylaminyl radical cation [**12**^{•+}]. Energy profiles on **12**^{•+} at the UHF/6-31G* (Figure 3.1), UMP2/6-31G*, and, UMP4SDTQ/6-31G** levels all reveal a) the ring-closed and ring-opened forms of **12**^{•+} reside at potential energy minima (no imaginary frequencies), and b) that there is small barrier to the ring opening reaction. A transition state for ring opening was successfully located and characterized by one imaginary frequency corresponding to C-C bond cleavage. In the case of UHF/6-31G*, the energy of activation for the ring opening pathway was found to be 4.7 kcal mol⁻¹. Ring opening was exothermic by ca. -7.5 kcal mol⁻¹.

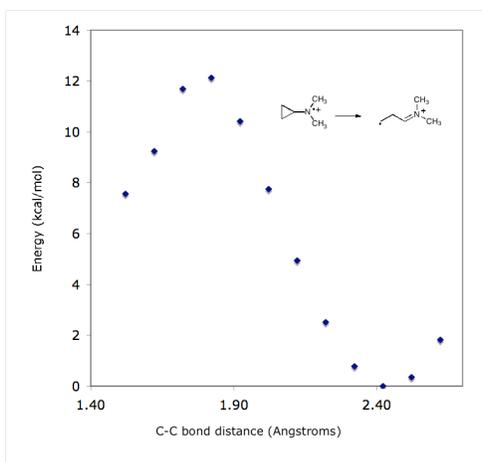


Figure 3.1. Reaction coordinate diagram for the ring opening of $12^{\bullet+}$ calculated at the UHF/6-31G* level of theory.

N-(2-phenylcyclopropyl)aminyl radical cation [$13^{\bullet+}$]. As was the case for $11^{\bullet+}$, calculations at the UHF/6-31G* (Figure 3.2) level of theory revealed that ring-opening of $13^{\bullet+}$ occurs with essentially no barrier, consistent with the results of Dinnocenzo, *et al.* (26). The interaction between C1 and C3 is purely repulsive. Similar results were obtained at UMP2/G-31G*, and UMP4STDQ/6-31G** levels of theory, further confirming that the ring closed form of $13^{\bullet+}$ does not exist at a potential energy minimum, and that no barrier to ring opening exists.

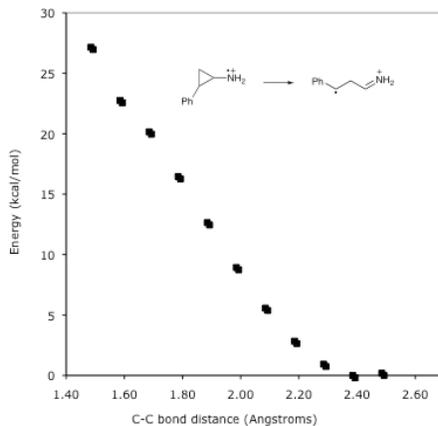


Figure 3.2. Reaction coordinate diagram for the ring opening of $13^{\bullet+}$ calculated at the UHF/6-31G* level of theory.

N,N-dimethyl-*N*-(2-phenylcyclopropyl)aminyl radical cation [$14^{\bullet+}$]. Geometry optimizations at UHF/6-31G* revealed a stable ring-opened form and a stable ring-closed form of the radical cation with no imaginary frequencies for either. ΔE° for ring opening was found to be -27.0 kcal

mol⁻¹, consistent with the results of Dinnocenzo, *et al.* (26). Curiously, an energy profile of the cyclopropyl ring opening reaction (Figure 3.3) did not reveal either a minimum for the ring-closed form of **14**^{•+} or a transition state for ring opening. Similarly, geometry optimization calculations at the UMP2 level of theory using the 6-31G* basis set also found both the ring-opened and ring-closed form of **14**^{•+}. Again, the barrier to ring opening was nonexistent. (Unfortunately, the limitations of the computational system used in this study prevented us from probing this problem further as energy profiles at UMP2/6-31G* and UMP4STDQ/6-31G** levels of theory were unsuccessful).

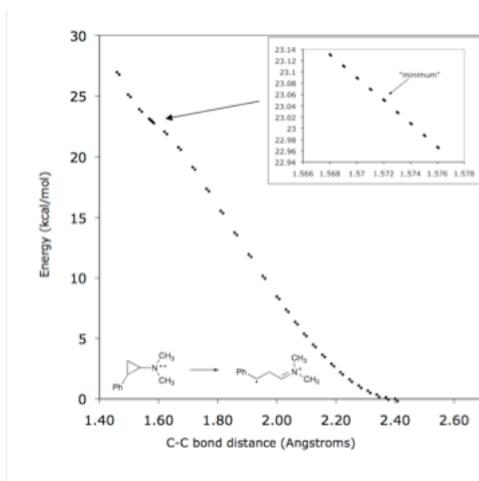


Figure 3.3. Reaction coordinate diagram for the ring opening of **14**^{•+} calculated at the UHF/6-31G* level of theory.

Careful inspection of the profile for **14**^{•+}, and varying the C-C bond length in very small increments (0.001 Å) in the vicinity of the ring-closed minimum proved informative. Variation of the C-C bond length is accompanied by a simultaneous rotation of one of the methyl groups as depicted in Figure 3.4. Thus, a simple, two-dimensional reaction coordinate diagrams such as in Figure 3.3 does not depict the full dynamics of the system, explaining why no minimum appears in this plot. All attempts to locate a transition state for ring opening failed, and as the calculations pertain to 0 K, we conclude that there is no significant barrier to ring opening of **14**^{•+} at room temperature.

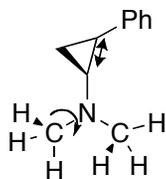


Figure 3.4. Depiction of how variation on C-C bond length of $14^{\bullet+}$ is accompanied by simultaneous rotation of the *N*-methyl group.

Because of the anomaly with rotation of the methyl groups in $14^{\bullet+}$, ring opening of cyclic aminyl radical cation $15^{\bullet+}$ was reexamined as a model for the ring opening of the radical cation generated from *trans*-2-phenylcyclopropyl-4-phenyl-1,2,3,6-tetrahydro-pyridine [8]. No minimum corresponding to a cyclopropane ring-closed structure could be located at any level of theory. The reaction coordinate diagram (Figure 3.5) is unambiguous in showing that the interaction between C1 and C3 is purely repulsive, consistent with earlier observations (20). Accordingly, it appears quite reasonable to assume that ring-opening of $8^{\bullet+}$ is barrier-free.

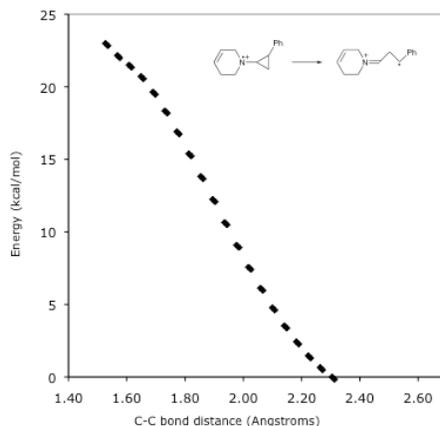
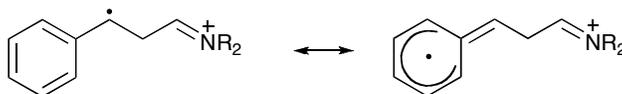


Figure 3.5. Reaction coordinate diagram for the ring opening of $15^{\bullet+}$ calculated at the UHF/6-31G* level of theory.

Comparison of $11^{\bullet+}$ and $12^{\bullet+}$. Computationally, we observed that $11^{\bullet+}$ exhibits no stable ring closed form and opens spontaneously with the transfer of an electron, consistent with earlier reports on this system. On the other hand, $12^{\bullet+}$ can lose an electron and remain a stable ring-closed radical cation. A barrier to ring-opening exists, thus a transition state exists. A chemical reason for this is that the electron donating character of the methyl groups on $12^{\bullet+}$ can stabilize the radical cation in the ring-closed form.

Comparison of 12^{•+} and 14^{•+}. Computational analysis of **12^{•+}** clearly reveals a barrier to ring opening. Conversely, ring opening of **14^{•+}** is highly exothermic ($\Delta E^\circ = -27 \text{ kcal mol}^{-1}$) with no significant barrier to ring opening, consistent with the report of Dinnocenzo and coworkers (26). One-electron oxidation of **8** and **14** likely occurs via a concerted DET reaction. This process is driven by the fact that the ring-opened form is stabilized by the formation of a benzylic radical (Scheme 3.9). For **12^{•+}**, no such stabilization exists resulting in a $4.71 \text{ kcal mol}^{-1}$ barrier for the ring-opening pathway.



Scheme 3.9. Resonance stabilization of the distonic radical cation derived from **8^{•+}** and **14^{•+}**.

Laser Flash Photolysis

The transient absorption spectra for the reaction of ³BP with MPTP [**1**] in acetonitrile are presented in Figure 3.6; virtually identical results were obtained in benzene (Figure 3.7-3.8). The key features of these spectra are that for MPTP, the data suggest that the decay of benzophenone triplet [**4^{3*}** ($\lambda_{max} = 520$)] (12) is accompanied by the formation of two new species at $\lambda_{max} = 545$ and 385 nm (12, 16). These two species are the diphenylhydroxymethyl radical [**6** (545 nm)] and MPTP_H• [**1[•]** (385 nm)], which are formed either by in-cage proton abstraction in the caged ion pair, or direct hydrogen atom abstraction (Scheme 3.10). The spectra showed no indication that the benzophenone radical anion [**4^{•-}** ($\lambda_{max} = 650 \text{ nm}$)] was formed.

In order to probe further the mechanism of this system, and to differentiate between the direct HAT and SET pathways, an analogous LFP study was conducted on the corresponding *N*-cyclopropyl MPTP derivatives **7** and **8**. In the case of MPTP, either hydrogen atom abstraction, or SET followed by in-cage deprotonation, would result in the formation of **1[•]**. However in the case of the two cyclopropyl derivatives, if SET is occurring, cyclopropyl ring opening is expected to compete with deprotonation because there is little (in the case of **7^{•+}**) or no (in the case of **8^{•+}**) barrier to ring opening (*vide supra*). Electron transfer from **8** is likely to be concerted with ring opening. This system is an ultra-sensitive probe for SET – competitive with events such as deprotonation, which may be occurring within the lifetime of a caged-radical ion pair.

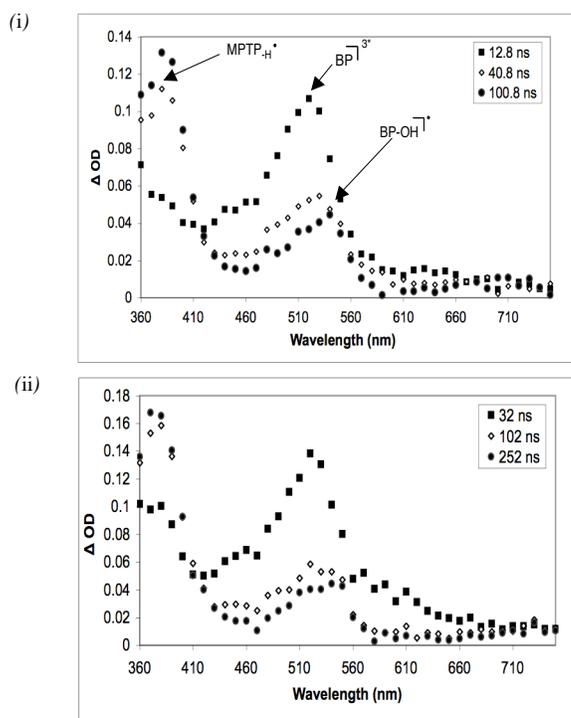


Figure 3.6. Transient absorption spectra of (i) MPTP [1] in CH₃CN and, (ii) *N*-cyclopropylphenyl MPTP [8] in CH₃CN.

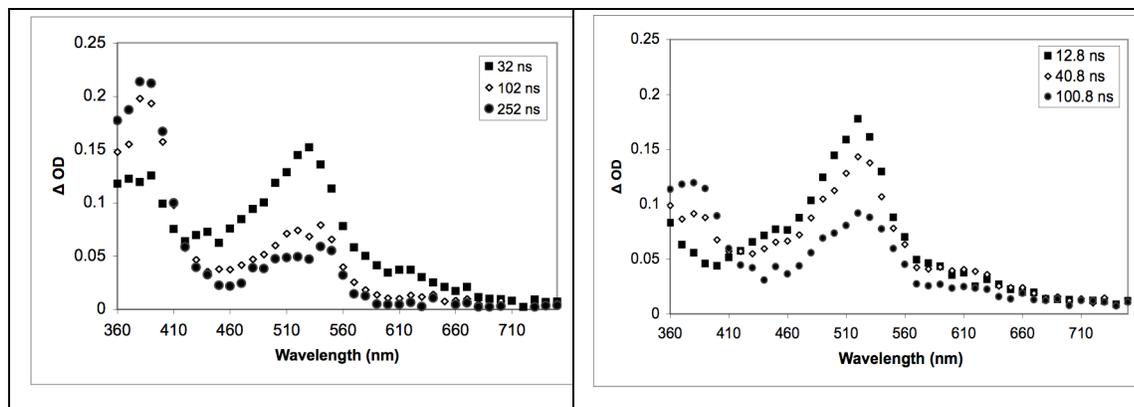


Figure 3.7. Transient absorption spectra of (i) MPTP [1] in benzene, (ii) MPTP [1] in CH₃CN in the presence of LiOCl₄.

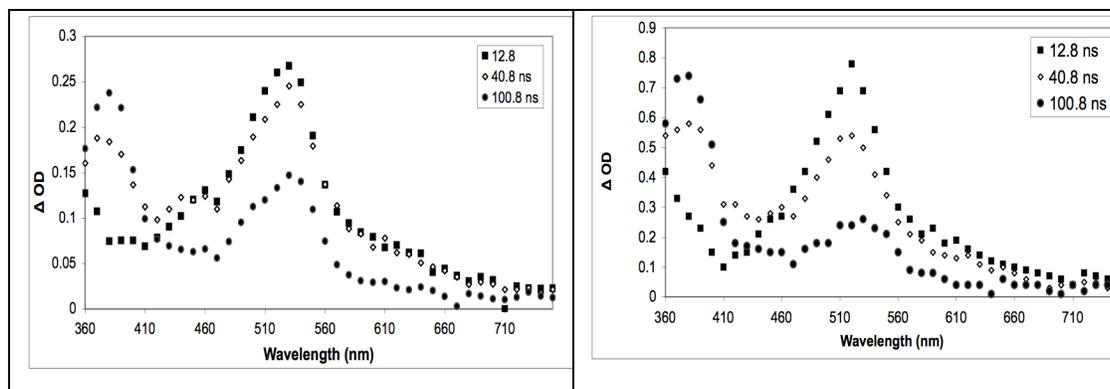
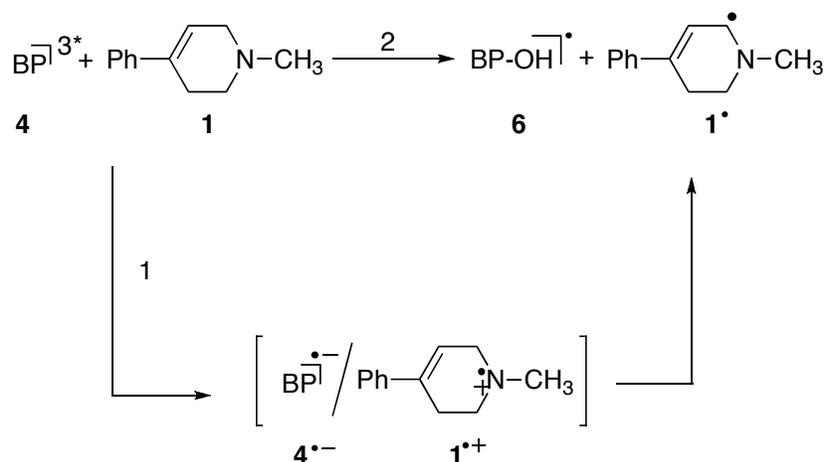


Figure 3.8. Transient absorption spectra of (i) *N*-cyclopropylphenyl MPTP [8] in benzene, (ii) *N*-cyclopropylphenyl MPTP [8] in CH_3CN in the presence of LiOCl_4 .



Scheme 3.10. Reaction pathways for ^3BP with MPTP 1) SET, and 2) HAT.

For each, the reaction of ^3BP only produced species that absorbed at 540 nm, corresponding to diphenylmethanol radical (**6**) and 380 nm corresponding to the hydrogen atom abstraction product. In examining the chemistry of these compounds, using ^3BP as a sensitizer it has been inferred that the “hydrogen atom abstraction” is the only observable process.

These experiments were also conducted in the presence of 0.5 M lithium perchlorate (Figure 3.7-3.8), which has been shown to aid in the separation of the contact ion pair to facilitate the diffusion of the contact ion pair forming a solvent separated ion pair (27). Lithium cation will complex with benzophenone radical anion, which can be monitored at $\lambda_{\text{max}} = 650$ nm on a nanosecond time regime. Examination of all three compounds showed no evidence of benzophenone radical anion in the presence of lithium perchlorate. In the case of the *N*-cyclopropyl derivatives, the MPTP derived radical absorption band [**7** $^{\cdot}$] should be diminished

and the band corresponding to the benzophenone radical anion should be visible in the absorption spectra, making the reasonable assumption that ring opening is competitive with deprotonation. In the case of compound **8**, the molecule is expected to undergo a concerted DET reaction, and therefore neither the band at 385 nm (corresponding to the MPTP derived radical [**8'**]) or band at 545 nm (corresponding to **6'**) are expected to be present. Because both of these bands *are* observed, it can be surmised that an electron transfer process is not occurring in this system. These results are fully consistent with the proposal that these oxidations of MPTP and its derivatives by benzophenone occur by a HAT process.

Rate constants for the reaction of ³BP with **1** and **7** were examined by varying the concentration of MPTP derivatives and monitoring the growth of the MPTP radical species on a nanosecond time regime as described previously (14). For both substrates, the rate constants were on the order of diffusion controlled and not significantly affected by added electrolyte (Table 3.1).

Table 3.1. Rate constants for the reaction of ³BP with **1** and **2**.

Substrate	Solvent	k _{obs} (M ⁻¹ s ⁻¹)
1	Benzene	2.12 x 10 ⁹
1	Acetonitrile	8.34 x 10 ⁹
1	Acetonitrile/ LiOCl ₄	5.83 x 10 ⁹
7	Benzene	2.19 x 10 ⁹
7	Acetonitrile	5.47 x 10 ⁹
7	Acetonitrile/LiOCl ₄	6.37 x 10 ⁹

Conclusions

Literature precedent and the calculations reported herein suggest for tertiary amines (such as these MPTP derivatives) that the barrier to ring opening of the corresponding radical cations is anticipated to be low for the *N*-cyclopropyl derivative (*e.g.*, **7**^{•+}) (21-24) and non-existent for the *N*-(2-phenylcyclopropyl) derivative (*e.g.*, **8**^{•+}) (20, 26). These systems are thus expected to be ultra-sensitive probes for SET. The LFP results reported herein demonstrate that for the reaction of ³BP with **1**, **7**, and **8**, the intermediates formed are very similar. In each instance,

disappearance of ^3BP is accompanied solely by appearance of bands corresponding to the diphenylhydroxymethyl radical and neutral radical derived from MPTP and its derivatives. In principle, this is consistent with either an apparent hydrogen abstraction from the MPTP derivative, or deprotonation within the radical ion pair formed by SET. However, the results obtained for **7**, and especially for **8** do not support an electron transfer pathway. Electron transfer from **8** is expected to lead immediately to the cyclopropyl ring opened product. If this were occurring, the observations would be drastically different: a) Bands corresponding to either the diphenylhydroxymethyl radical ($\lambda_{max} = 540 \text{ nm}$) or neutral radical derived from the MPTP derivative ($\lambda_{max} = 380$) would *not* be observed because the ring opening would beat out deprotonation, and b) a band corresponding to the benzophenone radical anion ($\lambda_{max} = 650 \text{ nm}$) *would* be observed. Accordingly, these results suggest that the reaction between ^3BP and tertiary aliphatic amines proceed via a HAT reaction. Additionally these models provide evidence that oxidations of N-cyclopropyl derivatives of MPTP catalyzed by MAO-B may not proceed by a pure SET pathway.

Acknowledgments

This work was funded by the National Science Foundation (CHE-0548129). MLG was partially supported by NSF-IGERT (DGE-0333378).

Supplemental Information

Concentration Profiles for MPTP and N-cyclopropyl MPTP

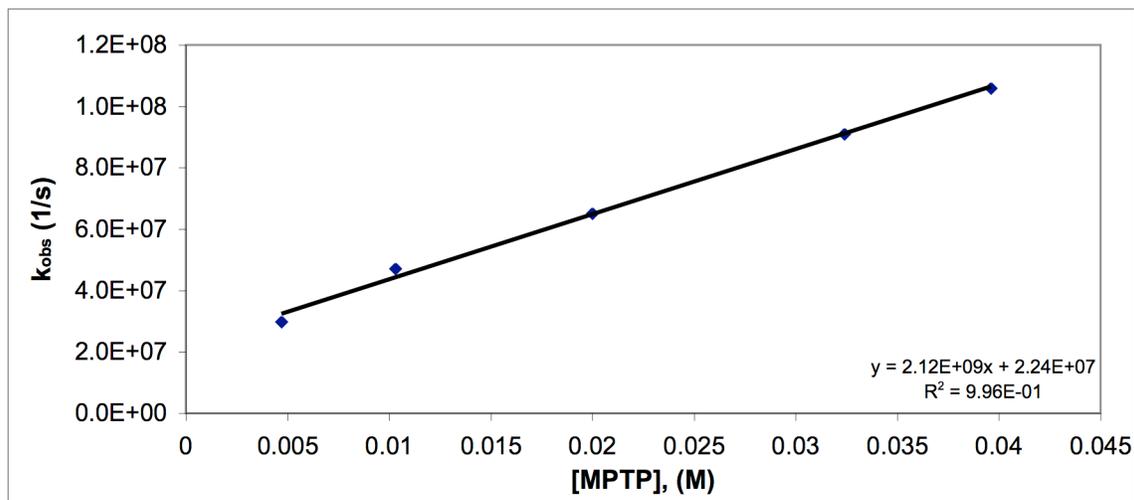


Figure S3.1. Concentration profile for the reaction of benzophenone triplet (^3BP) with MPTP [1] in benzene.

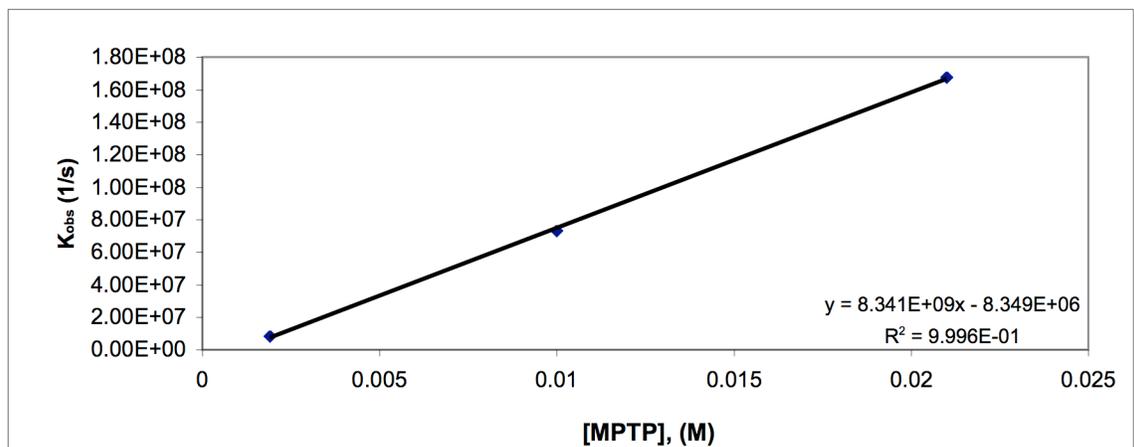


Figure S3.2. Concentration profile for the reaction of benzophenone triplet (^3BP) with MPTP [1] in acetonitrile.

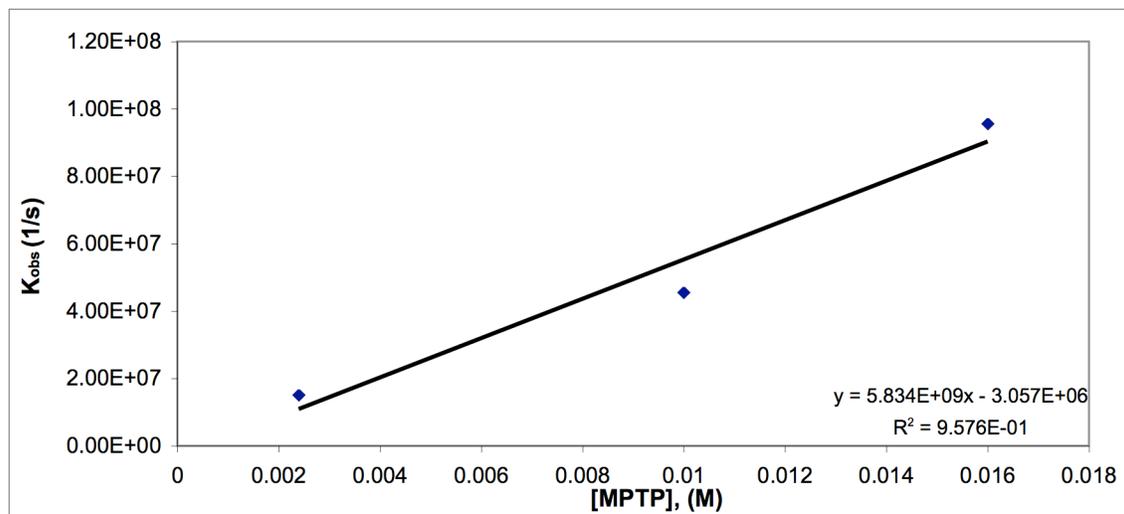


Figure S3.3. Concentration profile for the reaction of benzophenone triplet (³BP) with MPTP [1] in acetonitrile in the presence of 0.5M LiOCl₄.

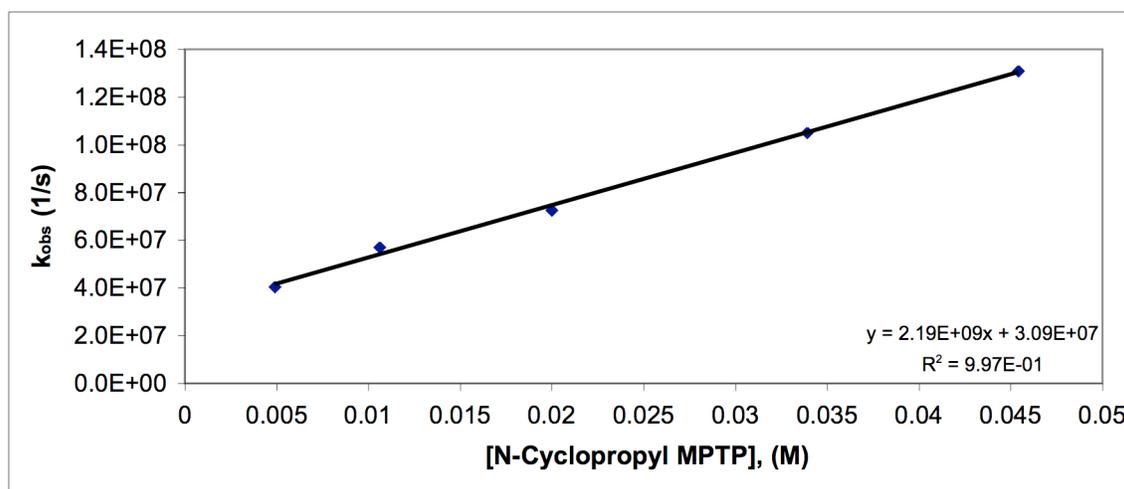


Figure S3.4. Concentration profile for the reaction of benzophenone triplet (³BP) with *N*-cyclopropyl MPTP [2] in benzene.

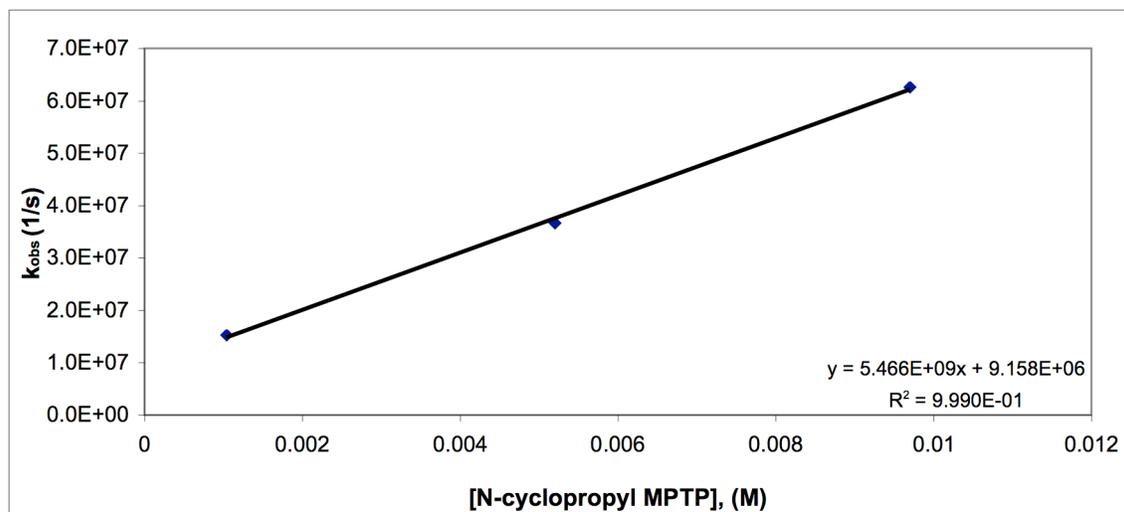


Figure S3.5. Concentration profile for the reaction of benzophenone triplet (^3BP) with *N*-cyclopropyl MPTP [2] in acetonitrile.

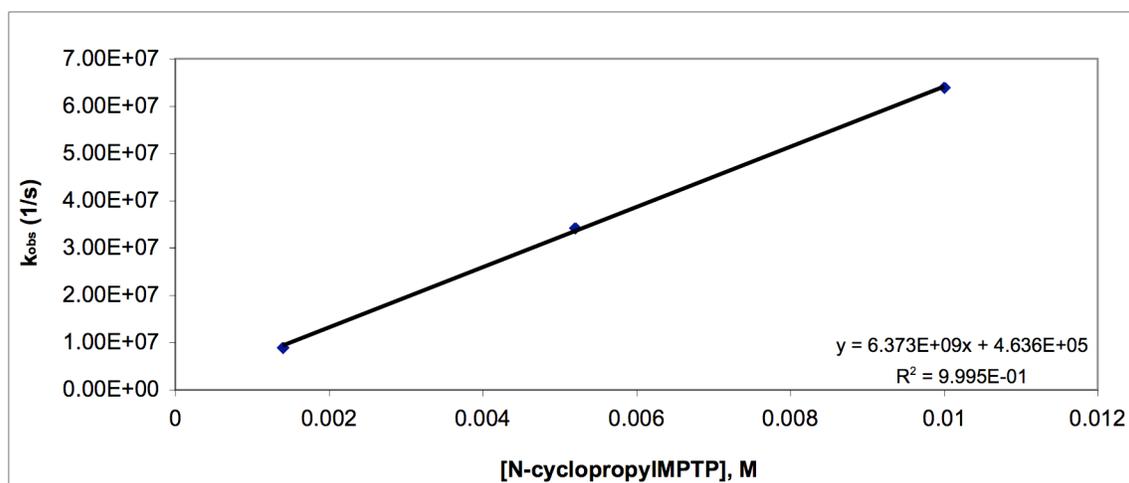


Figure S3.6. Concentration profile for the reaction of benzophenone triplet (^3BP) with *N*-cyclopropyl MPTP [2] in acetonitrile in the presence of 0.5M LiOCl_4 .

References

- (1) Silverman, R. B. Radical Ideas about Monoamine Oxidase. *Acc. Chem. Res.* **1995**, *28*, 335-342.
- (2) Kalgutkar, A. S.; Dalvie, D. K.; Castagnoli, N.; Taylor, T. J. Interactions of Nitrogen-Containing Xenobiotics with Monoamine Oxidase (MAO) Isozymes A and B: SAR Studies on MAO Substrates and Inhibitors. *Chem. Res. Toxicol.* **2001**, *14*, 1139-1162.
- (3) Scrutton, N. S. Chemical Aspects of Amine Oxidation by Flavoprotein Enzymes, *Nat. Prod. Rep.* **2004**, *21*, 722-730.
- (4) Singer, T. P.; Salach, J. I.; Castagnoli, N.; Trevor, A. Interactions of the Neurotoxic Amine 1-Methyl-4-Phenyl-1,2,3,6-Tetrahydropyridine with Monoamine Oxidases. *Biochem. J.* **1986**, *235*, 785-789.
- (5) Kim, J. M.; Bogdan, M. A.; Mariano, P. S. Mechanistic Analysis of the 3-Methylflavin-Promoted Oxidative Deamination of Benzylamine. A Potential Model for Monoamine Oxidase Catalysis. *J. Am. Chem. Soc.* **1993**, *115*, 10591-10595.
- (6) Kim, J. M.; Hoegy, S. E.; Mariano, P. S. Flavin Chemical Models for Monoamine Oxidase Inactivation by Cyclopropylamines, alpha-Silylamines, and Hydrazines. *J. Am. Chem. Soc.* **1995**, *117*, 100-105.
- (7) Miller, J. R.; Edmondson, D. E. Structure-Activity Relationships in the Oxidation of Para-Substituted Benzylamine Analogues by Recombinant Human Liver Monoamine Oxidase A. *Biochemistry* **1999**, *38*, 13670-13683.
- (8) Silverman, R. B.; Hoffman, S. J.; Catus, W. B. A Mechanism for Mitochondrial Monoamine Oxidase Catalyzed Amine Oxidation. *J. Am. Chem. Soc.* **1980**, *102*, 7126-7128.
- (9) Franot, C.; Mabic, S.; Castagnoli, N. Chemical Model Studies on the Monoamine Oxidase-B Catalyzed Oxidation of 4-Substituted 1-Cyclopropyl-1,2,3,6-Tetrahydropyridines. *Bioorg. Med. Chem.* **1998**, *6*, 283-291.
- (10) Markey, S. P.; Johannessen, J. N.; Chiueh, C. C.; Burns, R. S.; Herkenham, M. A. Intraneuronal Generation of a Pyridinium Metabolite May Cause Drug-Induced Parkinsonism. *Nature* **1984**, *311*, 464-467.
- (11) Markey, S. P.; Burns, R. S. MPTP Exposure Victims, *Chem. Eng. News* **1984**, *62*, 2-&.

- (12) Simon, J. D.; Peters, K. S. Picosecond Studies of Organic Photoreactions. *Acc. Chem. Res.* **1984**, *17*, 277-283.
- (13) Peters, K. S. *Advances in Electron Transfer Chemistry*; JAI Press, **1994**; Vol. 4.
- (14) Griller, D.; Howard, J. A.; Marriott, P. R.; Scaiano, J. C. Absolute Rate Constants for the Reactions of tert-Butoxyl, tert-Butylperoxyl, and Benzophenone Triplet with Amines: The Importance of a Stereoelectronic Effect. *J. Am. Chem. Soc.* **1981**, *103*, 619-623.
- (15) Suleman, N. K.; Flores, J.; Tanko, J. M.; Isin, E. M.; N. Castagnoli, J. The tert-Butoxyl Radical Mediated Hydrogen Atom Transfer Reactions of the Parkinsonian Proneurotoxin 1-Methyl-4-Phenyl-1,2,3,6-Tetrahydropyridine and Selected Tertiary Amines. *Bioorg. Med. Chem.* **2008**, *16*, 8557-8562.
- (16) Tanko, J. M.; Friedline, R.; Suleman, N. K.; N. Castagnoli, J. tert-Butoxyl as a Model for Radicals in Biological Systems: Caveat Emptor. *J. Am. Chem. Soc.* **2001**, *123*, 5808-5809.
- (17) Finn, M.; Friedline, R.; Suleman, N. K.; Wohl, C. J.; Tanko, J. M. Chemistry of the t-Butoxyl Radical: Evidence that Most Hydrogen Abstractions from Carbon are Entropy-Controlled. *J. Am. Chem. Soc.* **2004**, *126*, 7578-7584.
- (18) Lalevée, J.; Graff, B.; Allonas, X.; Fouassier, J. P. Aminoalkyl Radicals: Direct Observation and Reactivity toward Oxygen, 2,2,6,6-Tetramethylpiperidine-N-Oxyl, and Methyl Acrylate. *J. Phys. Chem. A* **2007**, *111*, 6991-6998.
- (19) Newcomb, M.; Varick, T. R.; Ha, C.; Manek, M. B.; Yue, X. Picosecond Radical Kinetics. Rate Constants for Reaction of Benzene-selenol with Primary Alkyl Radicals and Calibration of the 6-Cyano-5-Hexenyl Radical Cyclization. *J. Am. Chem. Soc.* **1992**, *114*, 8158-8163.
- (20) Jurva, U.; Bissel, P.; Isin, E. M.; Igarashi, K.; Kuttub, S.; N. Castagnoli, J. Model Electrochemical-Mass Spectrometric Studies of the Cytochrome P450-Catalyzed Oxidations of Cyclic Tertiary Allylamines. *J. Am. Chem. Soc.* **2005**, *127*, 12366-12377.
- (21) Qin, X.-Z.; Williams, F. ESR Studies on the Radical Cation Mechanism of Ring Opening of Cyclopropylamines. *J. Am. Chem. Soc.* **1987**, *109*, 595-597.
- (22) Knolle, W.; Janovský, I.; Naumov, S.; Williams, F. EPR Studies of Amine Radical Cations. Part 2. Thermal and Photo-Induced Rearrangements of Propargylamine and

- Allylamine Radical Cations in Low-Temperature Freon Matrices. *J. Phys. Chem. A* **2006**, *110*, 13816-13826.
- (23) Bouchoux, G.; Alcaraz, C.; Dutuit, O.; Nguyen, M. T. Unimolecular Chemistry of the Gaseous Cyclopropylamine Radical Cation. *J. Am. Chem. Soc.* **1998**, *120*, 152-160.
- (24) Nguyen, M. T.; Creve, S.; Ja, T.-K. On the Formation of the $^*CH_2CH_2CH=NH_2^+$ Distonic Radical Cation upon Ionization of Cyclopropylamine and Allylamine. *Chem. Phys. Lett.* **1998**, *293*, 90-96.
- (25) Jurva, U.; Bissel, P.; Isin, E. M.; Igarashi, K.; Kuttab, S.; Castagnoli, N. Model Electrochemical-Mass Spectrometric Studies of the Cytochrome P450-Catalyzed Oxidations of Cyclic Tertiary Allylamines. *J. Am. Chem. Soc.* **2005**, *127*, 12368-12377.
- (26) Wang, Y.; Luttrull, D. K.; Dinnocenzo, J. P.; Goodman, J. L.; Farid, S.; Gould, I. R. Associative Return Electron Transfer. A Bond-Coupled Electron Transfer in the Photoreactions of Cyclopropylamines. *Photochem. Photobiol. Sci.* **2003**, *2*, 1169-1176.
- (27) Simon, J. D.; Peters, K. S. Sodium(+) Ion and Lithium(-) Ion Effects on the Photoreduction of Benzophenone: A Picosecond Absorption Study. *J. Am. Chem. Soc.* **1983**, *105*, 4875-4882.
- (28) Nimkar, S. K.; Anderson, A. H.; Rimoldi, J. M.; Stanton, M.; Castagnoli, K. P.; Mabic, S.; Wang, Y. X.; Castagnoli, N. Synthesis and Monoamine Oxidase B Catalyzed Oxidation of C-4 Heteroaromatic Substituted 1,2,3,6-Tetrahydropyridine Derivatives. *Chem. Res. Toxicol.* **1996**, *9*, 1013-1022.
- (29) Pitts, S. M.; Markey, S. P.; Murphy, D. L.; Weisz, A.; Lunn, G. Recommended Practices for the Safe Handling of MPTP, in *MPTP: a Neurotoxin Producing a Parkinsonian Syndrome*, Academic: New York, **1986**, p 703-716.
- (30) Wavefunction, Inc.: Irvine, CA 92612, **1998**.
- (31) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; J. A. Montgomery, J.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.;

Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. Gaussian, Inc.: Wallingford CT, **2004**.

Chapter 4

Modeling the Relationship Between the p53 C-Terminal Domain and Its Binding Partners Using Molecular Dynamics

William J. Allen, Daniel G. S. Capelluto, Carla V. Finkielstein, and David R. Bevan

Reproduced with permission from *The Journal of Physical Chemistry B* **2010**, *114* (41), 13201-13213. Copyright 2010, American Chemical Society.

Attribution

Of the work discussed within this chapter, Allen performed all of the simulations, analyses, and wrote and prepared the manuscript. Capelluto, Finkielstein, and Bevan significantly contributed to the editing and review of the manuscript, as well as provided guidance in the research.

Abstract

Fifty percent of all cancer cases result from mutations of the *TP53* gene, which encodes the tumor suppressor p53, and it is hypothesized that the p53-mediated checkpoint pathway is compromised in most of the remaining cases. The p53 C-terminal domain (CTD) is an important site of p53 regulation, but by nature is difficult to study, as it is intrinsically disordered. In this study, we performed molecular dynamics simulations on the p53 CTD and five known regulatory binding partners. We identified distinct trends in fluctuation within and around the p53 CTD binding site on each partner demonstrating a behavior that facilitates association. Further, we present evidence that the size of the hydrophobic pocket in each p53 CTD binding site governs

the secondary structure of the p53 CTD when in the bound state. This information will be useful for predicting new binding partners for the p53 CTD, identifying interacting regions within other known partners, and discovering inhibitors that provide additional points of control over p53 activity.

Introduction

The p53 protein has been extensively studied since it was discovered and identified as an important tumor suppressor protein (1). The p53 protein has the ability to promote cell cycle arrest, apoptosis, and anti-angiogenesis among other functions, which it commonly does in aberrantly growing cells as a means to cease tumorigenesis (2, 3). Non-functioning p53, or a malfunction in the p53-mediated signaling pathway, results in tumor cell formation and abnormal proliferation. Most often, the defect is a single mutation within the *TP53* gene, but it can also result from altered expression or mutation of p53 regulatory proteins (2). The general consensus remains that restoration of p53 function in a malignant cell is of tremendous therapeutic benefit (4).

The p53 C-terminal domain (CTD) represents a widely accepted site for its regulation (5, 6). It is promiscuous in both function and activity, performing many roles in the cell. The p53 CTD binds DNA non-specifically to promote linear diffusion along the DNA (7, 8), but it can also bind DNA in a more sequence-specific manner depending on its lysine acetylation state (9, 10). As a way to regulate this DNA-binding function, the p53 CTD also binds non-specifically to RNA (11). The p53 CTD also binds other proteins, some of which are transcription cofactors to be recruited to specific genes (12-14), some sequester or block p53 activity by directly binding to the C-terminus (15-16), and many others regulate p53 activity through post-translational modification (PTM) (17-20). The p53 CTD contains sites of methylation, acetylation, phosphorylation, and ubiquitination among others. The topic of PTM has been studied extensively and is well reviewed (4, 21, 22), and will not be further discussed here.

A key feature that makes the p53 CTD promiscuous is the fact that the entire domain (residues 363-393) is disordered (23). Disordered proteins have several different functions, a major one being recognition and binding of various biomolecules (24, 25). The nature of being

disordered gives disordered regions the distinct advantage of being able to bind non-specifically to a multitude of different partners, in a sense becoming a “hub” of protein interaction (26). This phenomenon assumes importance when the hub in question is at the center of a cancer-prevention pathway (27).

The association of the p53 CTD to its various partners varies in specificity, type, and mode of binding. For example, when bound to S100 calcium-binding protein B (S100B($\beta\beta$)), the CTD forms an α -helix that spans residues 377-387 (PDB:1DT7) (28), but when it binds to a Sirtuin protein (Sir2), it forms a β -strand from residues 380-385 (PDB:1MA3) (29). In addition to these distinct secondary structure elements, the CTD forms a β -turn when bound to cAMP response element-binding (CREB) binding protein (CBP) (PDB:1JSP), and it lacks any ordered secondary structure when bound to the histone methyltransferase Set9 (PDB:1XQH) (19) or the cyclin A / cyclin dependent protein kinase 2 complex (PDB:1H26) (30). Thus, these five binding partners were chosen for this study for their diverse range of size, function, and nature of interaction with the p53 CTD, as well as for their structure availability.

Our knowledge of the binding process between the p53 CTD and its binding partners is limited. At present, there are two models under consideration: (i) the p53 CTD transiently forms the appropriate secondary structure in solution before binding to its partner (31), or (ii) the p53 CTD adopts the appropriate secondary structure upon coming in close proximity to its binding partner – a process known as induced folding (32, 33). A recent study suggests that the p53 CTD is able to form α -helices transiently in solution, but in the case of binding to S100B($\beta\beta$), it only forms the secondary structure when induced by the binding partner (34). This conformational diversity gives the p53 CTD the advantage of being a viable binding site for its partners at all times, instead of merely when it happens upon the correct configuration. To our knowledge, there is no other evidence, theoretical or physical, that describes the binding event between the p53 CTD and any of its partners other than S100B($\beta\beta$). This limitation makes it very difficult to predict novel protein partners for the p53 CTD, as well as to identify their binding sites and the conformation that the p53 CTD adopts in the complex.

Molecular simulation is a useful technique to study disordered proteins and their interactions. The disordered N-terminus transactivation domain (35-38), the structured DNA-binding domain (39-43), and the tetramerization domain (44), which is immediately N-terminal to the p53 CTD, have all been studied using molecular dynamics (MD) techniques. The p53 CTD has also been

studied by molecular simulation to a limited extent. Namely, there are MD and docking studies on the interaction between p53 CTD and various small molecules in complex with S100B (34, 45, 46), as well as MD and docking studies on the p53 CTD binding region on CREB binding protein (47). Apart from these examples, no one has yet performed a rigorous MD study of the interaction between the p53 CTD and a broad spectrum of binding partners.

This report seeks to answer a few fundamental questions about the nature of the interaction between the p53 CTD and its binding partners. First, what features of the recognition sites within the binding partners promote association to the p53 CTD? Second, what is the nature of the interaction among those two molecules? Third, what effect does the binding event have on the structural dynamics of the p53 CTD and of its interacting protein? Better understanding of these relationships will help the prediction of putative binding sites on other partners, or even in the identification of novel interacting proteins.

Computational Methods

PDB File Preparation

Five files were downloaded from the Research Collaboratory for Structural Bioinformatics Protein Data Bank (PDB, www.pdb.org) (48). Each contains two important elements: some fragment of the p53 CTD, and some binding partner. In each structural file, an N-terminal acetyl cap and C-terminal amide cap were added to the p53 CTD fragment using the LEaP program in AMBER (49). All crystallographic waters were removed. The following subsections provide specific information on further file preparation.

IDT7. This structure file contains forty nuclear magnetic resonance (NMR) spectroscopy models of the interaction between the rat protein S100B($\beta\beta$) homodimer and a human p53 CTD 22-mer fragment (residues 367-388) (28). Of the forty models, the first one was chosen for the simulations (structure validation is discussed in the Analysis section). In the chosen model, the C-terminal carboxyl group was missing from each chain of the S100B($\beta\beta$) homodimer, so it was rebuilt using LEaP. The other two chains in the NMR model (Chain X and Chain Y) are p53 CTD fragments, and each is bound to one of the S100B subunits. Chain X was also missing the

C-terminal carboxyl group, so it was rebuilt in LEaP before the cap was added as described in the previous section. Chain Y was removed for simplicity, and the other three modified chains (S100B($\beta\beta$) homodimer and one p53 CTD chain) were considered for the final model. Because p53 CTD binding to S100B($\beta\beta$) is calcium-dependent, all four calcium ions were also preserved.

1MA3. This crystal structure solved at 2.00 Å contains Sir2 in complex with a p53 CTD 9-mer fragment (residues 379-387) (29). Several residues of the Sir2 protein are missing, including a loop from residues 30-39 and residue 253; in addition there are atoms missing from residues Glu 29, Arg 112, and Glu 252. In order to rebuild the loop and repair the other inaccuracies, we employed Modeller 9v4 (50-53) using the complete protein sequence and the 1MA3 crystal structure as the template. We built ten different conformations for the loop from residues 30-39, and scored the models within the SWISS-MODEL Workspace (54). We chose the best scoring model and relied on the forthcoming MD simulations to further refine the loop.

In the crystallization experiment, an 18-mer fragment of the p53 CTD (residues 372-389) was used, but only a 9-mer fragment (residues 379-387) was solved. The low electron density in the unsolved residues implies a large amount of flexibility and fluctuation. Avalos *et al.* (29) found that the essential interactions are the burying of AcLys 382 in a hydrophobic pocket and the formation of a β -sheet, both of which are conserved in our model. The Arg 379, however, was missing several atoms that were rebuilt in LEaP before adding on the caps as previously described. Further, residue Lys 382 of the p53 CTD is acetylated in the original structure. Finally, two other components had to be considered. First, a zinc ion coordinated by Sir2 was retained. Second, one molecule of 2-(N-morpholino)-ethanesulfonic acid that crystallized with the proteins as an artifact of the crystallization process was deleted.

1H26. This crystal structure solved at 2.24 Å resolution contains several important elements (30). The foundation is formed by a complex of two proteins – cyclin A and phosphorylated cyclin-dependent protein kinase 2 (pCDK2) – of which there are two copies in the crystallographic asymmetric unit. A p53 CTD 9-mer fragment (residues 378-386) is bound to one of the cyclin A chains. The cyclin A / pCDK2 pair that was not bound to the p53 CTD was deleted. The two most C-terminal residues of pCDK2 were missing most of their atoms, and because the location is well removed from the binding site with the p53 CTD, the remaining atoms were deleted and the chain was terminated after Leu 296 with a carboxyl group in LEaP.

IJSP. This structure file contains twenty NMR models of a p53 CTD 20-mer fragment (residues 367-386) in complex with the CREB binding protein (CBP) bromodomain (14). The Lys 382 of the p53 CTD fragment is monoacetylated, a factor that is necessary for binding. The first of the twenty NMR structures was chosen for this study (structure validation is discussed in the Analysis section). No missing residues or atoms needed to be repaired or added other than the caps discussed previously.

IXQH. This crystal structure contains the methyltransferase Set9 in complex with a p53 CTD fragment solved at 1.75 Å resolution (19). In the crystallization experiment, a 10-mer fragment (residues 369-378) of the p53 CTD was used, but only a 6-mer fragment (residues 369-374) was resolved. The crystallographic asymmetric unit contained two copies of the protein complex, and one of the pairs was deleted for simplicity. The N-terminal 14 residues of Set9 were also not distinguished by crystallography, but are located far enough from the p53 CTD binding site that they were not rebuilt. Finally, Lys 372 of the p53 CTD was monomethylated in the original crystal structure.

Molecular Dynamics Simulations

In total, fifteen systems were prepared for simulation. For each of the five PDB files, we simulated (i) the complex of the p53 CTD fragment and binding partner prepared as described in the previous sections, (ii) the p53 CTD fragment alone, and (iii) the binding partner alone. The initial coordinates of simulation groups (ii) and (iii) were taken directly from group (i). All MD simulations were performed with GROMACS 4.0.4 (55) in conjunction with the united atom GROMOS 53a6 force field (56). Methylated and acetylated lysine parameters were reasonably adapted by combining existing parameters in the 53a6 force field, and are included in the Supplemental Information at the end of this chapter (Table S4.1-S4.3, Figure S4.1). The systems were solvated using the explicit simple point charge (SPC) water model (57), and enough sodium and chloride ions were added to neutralize the charge of each system and bring the final concentration to a physiological level of 100 mM. Each system was then energy minimized using a steepest descents integrator (59) either until the maximum force was less than 1000 kJ mol⁻¹ nm⁻¹ on any atom or until additional steps resulted in a potential energy change of less than 1 kJ mol⁻¹.

A 50 ps NVT equilibration was performed at 200 K with position restraints applied to all of the backbone atoms in order to relieve any bad contacts at the side chain-solvent interface. The Berendsen thermostat (59) was used with a temperature coupling time constant (τ_T) of 0.1 ps. All bond lengths were constrained using the linear constraint solver (LINCS) algorithm (60), which allowed for a 2 fs timestep. Long-range electrostatic interactions were approximated using the particle-mesh Ewald (PME) method (61, 62) with a fourth-order spline interpolation and a 0.12 nm Fourier grid spacing.

After the initial equilibration, the position restraints were lifted and the Berendsen thermostat was raised to 310 K for a 100 ps NVT simulation. Once each system was sufficiently equilibrated around the target temperature, the Nosé-Hoover thermostat (63, 64) was applied because it generates a more correct canonical distribution for temperature, and simulated for another 100 ps (NVT) at the same temperature. Finally, a 100 ps NPT simulation was conducted, relaxing the system into an isotropic Parrinello-Rahman barostat (65, 66) set to 1.0 bar of pressure in all directions and a pressure coupling time constant (τ_P) of 1.0 ps. The production MD that followed was performed with the Nosé-Hoover thermostat and Parrinello-Rahman barostat, as well as the LINCS and PME treatments as described.

Replicate systems were generated using the same starting configuration but with different initial velocities applied prior to the first NVT equilibration at 200 K. Each of the fifteen simulations was simulated in triplicate for 100 ns, for a total simulation time of 4.5 μ s across all simulations. Simulations were performed on Virginia Tech's System X Supercomputer, a 12.25 Teraflop computer comprising 1100 Apple PowerMac G5 computers with dual 2.3 GHz PowerPC 970FX processors (67).

Analysis

All analyses were performed using the GROMACS suite of tools and a secondary structure recognition algorithm (DSSP) (68). The hydrophobic solvent-accessible surface area (SASA) of each p53 CTD binding site was determined by measuring the SASA of all hydrophobic residues that lie within the p53 CTD binding site and/or interact closely with the p53 CTD. All of the pertinent residues that were considered are listed in Results and Discussion section. It is important to mention that a bug in the code for the GROMACS analysis tool "*g_rmsf*" was fixed

as outlined on the GROMACS website prior to performing our analyses. The program Grace (69) was used to plot the 2-D data, and the 3-D images were created with either PyMOL (70) or Chimera (71).

In order to verify that our simulation subspace overlaps completely with the NMR structural subspace, thus validating our choice in starting structure for 1DT7 and 1JSP, we employed methods outlined in the literature (72, 73) (Figure 4.1, 4.2, and Table 4.1, 4.2). We found good overlap in the structural ensembles of the receptor proteins, but less overlap in the p53 CTD conformations. In each NMR structure, the solvent-exposed portion of the CTD is disordered and observed to be in many different random solvent-oriented conformations. It is unlikely that traditional MD on this timescale could reproduce the solvent-exposed configurations in the NMR structure, but we are satisfied that the ensemble of structures at the p53 CTD-binding partner interface for each structure was satisfactorily sampled across the three replicates.

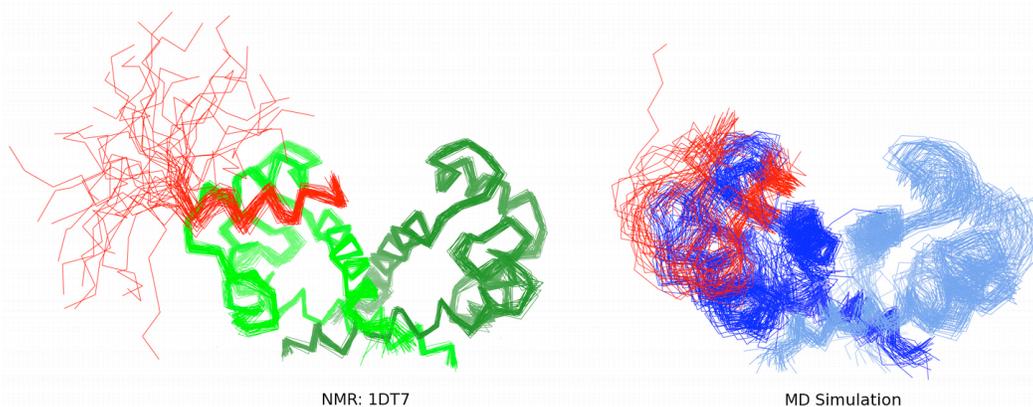


Figure 4.1. NMR and simulation subspace overlap for 1DT7. (Left) Overlay of forty 1DT7 NMR structures. S100B($\beta\beta$) subunits are shown in green and forest green, the p53 CTD is shown in red. (Right) Overlay of forty 1DT7 structures from evenly spaced intervals along the 100 ns MD trajectory. S100B($\beta\beta$) subunits are shown in dark blue and light blue, the p53 CTD is shown in red.

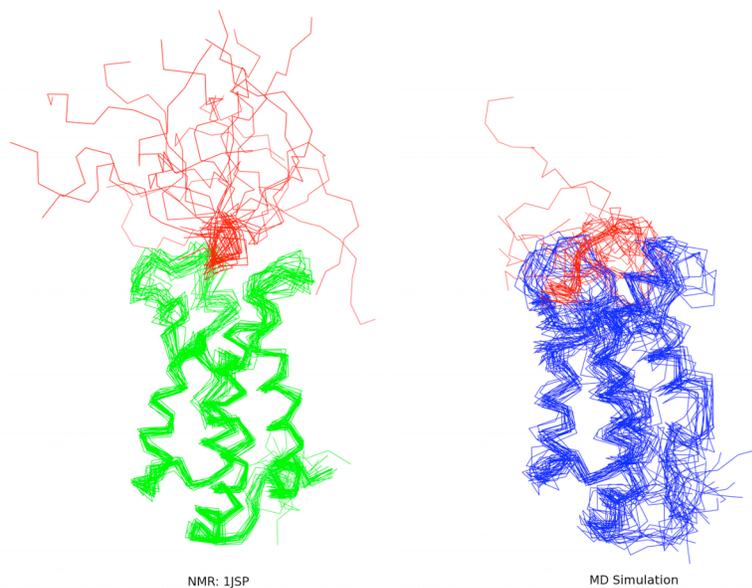


Figure 4.2. NMR and simulation subspace overlap for 1JSP. (Left) Overlay of twenty 1JSP NMR structures. The CBP bromodomain is shown in green and the p53 CTD is shown in red. (Right) Overlay of twenty 1JSP structures from evenly spaced intervals along the 100 ns MD trajectory. The CBP bromodomain is shown in blue and the p53 CTD is shown in red.

Table 4.1. Radius of gyration (R_g). Values are in nm.

		NMR	MD
1DT7	S100B($\beta\beta$)	1.70 \pm 0.01	1.74 \pm 0.02
	p53 CTD	1.14 \pm 0.10	0.91 \pm 0.02
1JSP	CBP bromodomain	1.47 \pm 0.01	1.53 \pm 0.02
	p53 CTD	1.24 \pm 0.10	1.19 \pm 0.12

Table 4.2. Secondary structure content. Values are in number of residues.

			Coil	α -Helix	β -Sheet	Bend	Turn
1DT7	S100B($\beta\beta$)	NMR	21.3 \pm 0.3	112.6 \pm 0.5	11.8 \pm 0.1	20.7 \pm 0.6	16.4 \pm 0.8
		MD	36.8 \pm 3.2	102.1 \pm 2.7	3.0 \pm 2.8	30.7 \pm 3.8	7.6 \pm 3.8
	p53 CTD	NMR	8.1 \pm 1.3	7.3 \pm 0.6	-	1.7 \pm 1.3	3.6 \pm 0.8
		MD	10.5 \pm 1.2	7.0 \pm 1.0	-	4.9 \pm 1.6	0.5 \pm 1.0
1JSP	CBP bromodomain	NMR	21.5 \pm 3.4	51.9 \pm 3.4	-	24.0 \pm 3.6	20.5 \pm 4.3
		MD	35.5 \pm 3.7	53.7 \pm 3.4	-	15.4 \pm 4.0	13.9 \pm 4.4
	p53 CTD	NMR	16.7 \pm 2.2	-	-	2.6 \pm 1.4	0.8 \pm 1.6
		MD	13.6 \pm 2.0	-	-	3.5 \pm 2.0	2.0 \pm 1.5

Results and Discussion

In the following subsections, MD simulations will be discussed as follows. First, we will compare the dynamics of the p53 CTD fragment when bound to a binding partner or unbound in solution. Second, we will compare the dynamics of each binding partner when bound to the p53 CTD fragment and unbound in solution. Finally, we discuss specific interactions of interest for each p53 CTD-binding partner pair. Discussions are organized by the PDB identifier from which the structures originated.

1DT7

We first assessed the stability during MD of the p53 CTD fragment (residues 367-388) when bound to the S100B($\beta\beta$) receptor by analyzing the DSSP profile (Figure 4.3). When bound to the receptor, the p53 CTD underwent little change from the starting structure in all three replicates (Figure 4.3a-c). In the first and third replicates, the α -helix comprising residues 380-387 was well-conserved. In the second replicate, the same helix unwound at the ends slightly and was calculated by the DSSP algorithm to be a turn – one residue short of being considered a true α -helix. When the p53 CTD was simulated unbound to the receptor from the same starting configuration, a major change in secondary structure content occurred that spanned all three replicates (Figure 4.3d-f). The peptide adopted many different secondary structure elements, most notably β -sheets, none of which retained any semblance to the original starting structure of the same peptide in complex with S100B($\beta\beta$).

The root mean square deviation (RMSD) of the p53 CTD also varied between bound and unbound states. Specifically, when the p53 CTD was bound to S100B($\beta\beta$), the average backbone RMSD of the three replicates was markedly lower than when the p53 CTD was unbound (Figure 4.4a). This evidence is in agreement with the DSSP evidence in that the p53 CTD fragment fluctuated less when bound to its receptor.

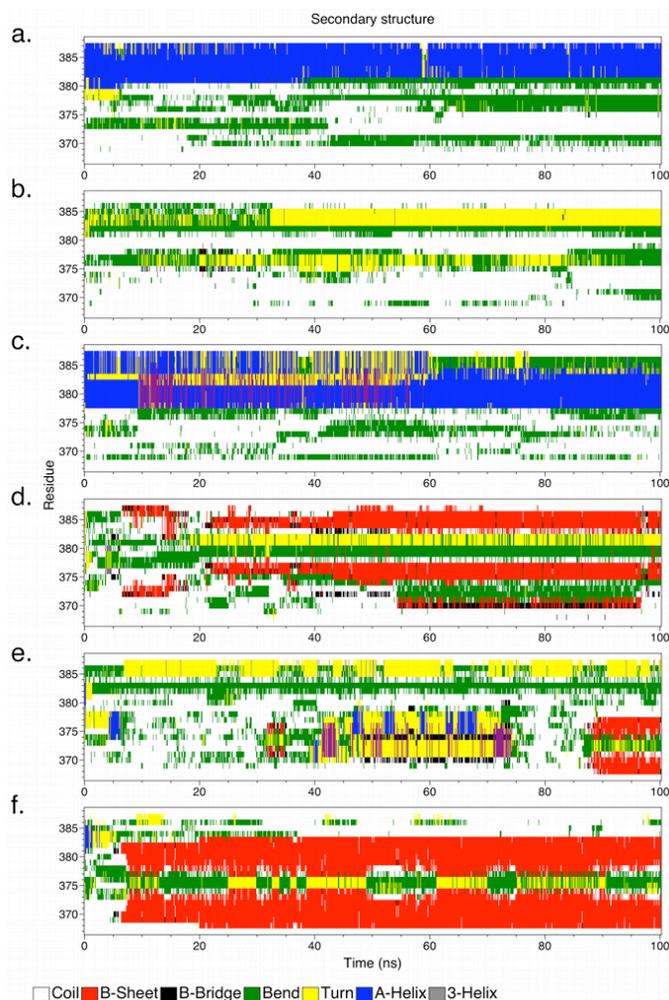


Figure 4.3. DSSP analysis for 1DT7. Panels (a–c) show the secondary structure content of the p53 CTD when bound to S100B($\beta\beta$). Panels (d–f) show the secondary structure content of the unbound p53 CTD.

In the analysis of S100B($\beta\beta$), we find similar results. The average backbone RMSD of the three replicate simulations of S100B($\beta\beta$) bound to the p53 CTD was slightly lower than for the unbound form over the last 50 ns of simulation (Figure 4.4b). These data indicate that when in complex with the p53 CTD fragment, the S100B($\beta\beta$) receptor fluctuates less compared to when it is free in solution. It is also noteworthy that the RMSD of the bound form stabilized in less than 10 ns, whereas the RMSD of the unbound form stabilized in roughly 70 ns. A root mean square fluctuation (RMSF) analysis revealed specific parts of the protein that fluctuated more (or less) in the bound or unbound states (Figure 4.4c). The hinge region (residues 40-49), helix 3 (residues 50-60), and helix 4 (residues 70-88) of the S100B β -subunit comprise the p53 CTD binding site (Figure 4.5), and they all exhibited a higher degree of fluctuation when the p53 CTD

was not bound. In contrast, residues important in calcium binding (residues 20-25) fluctuated more when the p53 CTD was bound.

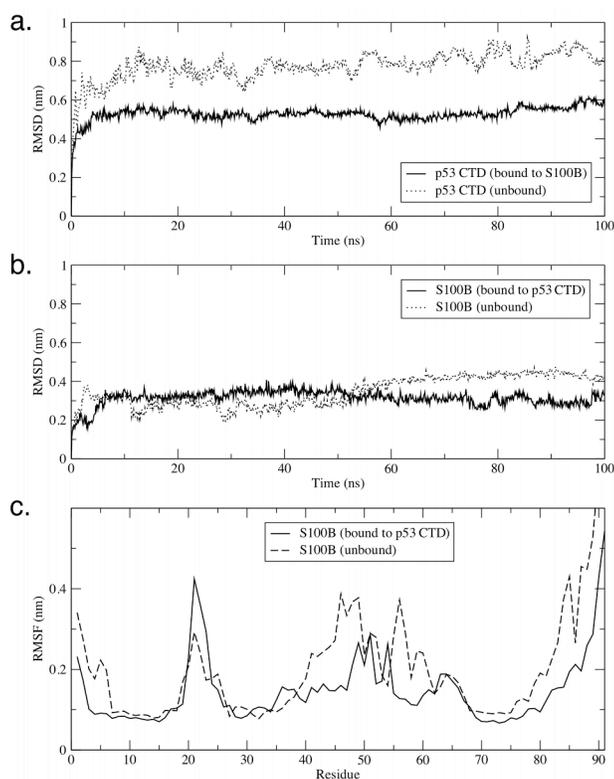


Figure 4.4. RMSD and RMSF analyses for 1DT7. Panel (a) shows the average backbone RMSD for the three replicates of the p53 CTD in complex with S100B($\beta\beta$) (solid line), and in the absence of S100B($\beta\beta$) (dotted line). Panel (b) shows the average backbone RMSD for the three replicates of the S100B($\beta\beta$) receptor in complex with the p53 CTD (solid line) and in the absence of the p53 CTD (dotted line). Panel (c) represents the RMSF of S100B($\beta\beta$) in complex with the p53 CTD (solid line), and in the absence of the p53 CTD (dashed line).

The S100B β -subunit experienced a marked change in configurational principal components (PCs) upon p53 CTD binding. Briefly, PCs are eigenvectors that describe large collective (non-translational and non-rotational) motions of atoms in a given protein that are observed over long time scales. PC vectors are determined in conjunction with a covariance matrix that illustrates correlated and anti-correlated motions of atoms. Figure 4.6a shows a covariance matrix for the two states of the S100B β -subunit, bound to the p53 CTD (bottom right) and in the absence of the p53 CTD (top left). When unbound, the S100B subunit experienced large areas of both correlated and anti-correlated motion, specifically between helix 3 and helix 4 – the helices that

shape the p53 CTD binding site. In Figure 4.6b, the first principal component is illustrated as the beginning (top) and end (bottom) of its eigenvector. In the absence of the p53 CTD, a drastic transition in the shape of helix 3 and of helix 4 and a change in relative position between the two helices is apparent from the two states, a transition which comprised 41.7% of the total motion in the protein. When the p53 CTD was bound, however, both correlated and anti-correlated motion dissipated between helix 3 and helix 4 (Figure 4.6a). The eigenvector that described the largest PC was a relatively small shift in the position of helix 3 (Figure 4.6b). Further, this PC represented only 25.5% of the total motion in the protein.

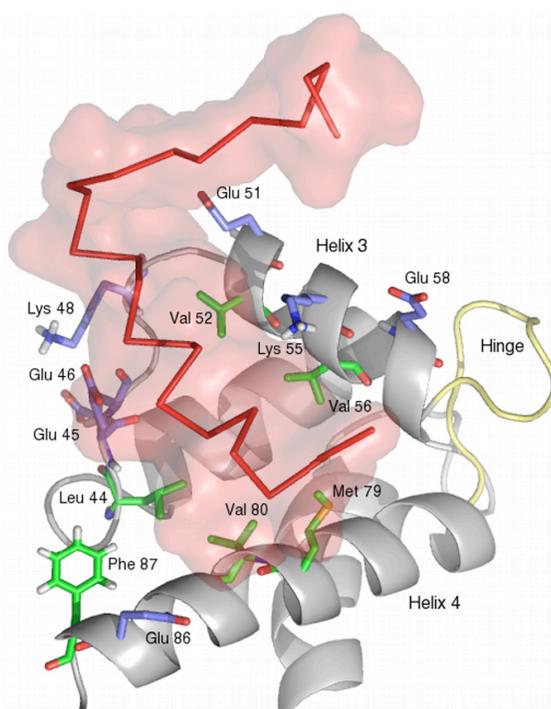


Figure 4.5. Surface rendering and backbone trace of the p53 CTD (colored in red), bound to cartoon representation of one S100B β -subunit (colored in gray). The hinge region is colored in yellow. Polar and acidic residues of the S100B β -subunit are represented with blue carbon atoms. Hydrophobic and aromatic residues are represented with green carbon atoms. Other atoms are colored according to standard schemes (H = white, O = red, N = dark blue, S = yellow).

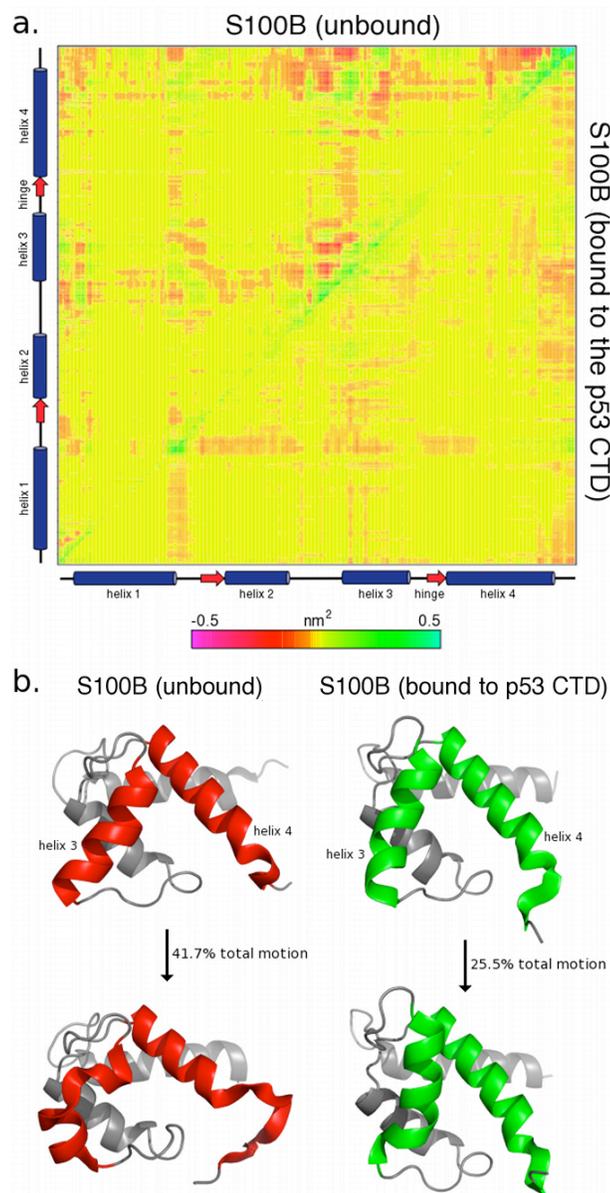


Figure 4.6. PC analysis for 1DT7. Panel (a) shows a covariance matrix illustrating correlated and anti-correlated motions within one S100B β -subunit in the absence of the p53 CTD (top left) and when bound to the p53 CTD (bottom right). The secondary structure of the S100B β -subunit backbone is represented along the axes (from left-to-right, and from bottom-to-top). Panel (b) shows the motion of the largest eigenvector on the S100B($\beta\beta$) structure in the absence of the p53 CTD (left) and when bound to the p53 CTD (right). The sites on S100B($\beta\beta$) where the p53 CTD binds are colored red and green.

These data taken together, we observed a stabilization effect on the p53 CTD fragment upon binding to the receptor S100B($\beta\beta$), a characteristic that is in accordance with previous experimental observations (28). Because these simulation conditions are consistent with

experimental observations, we feel confident in making new observations about the dynamics of the receptor in both the bound and unbound state, as well as the interaction between the receptor and the p53 CTD fragment.

The major contributing factors that stabilize the interaction between the p53 CTD and S100B($\beta\beta$) are summarized in Figure 4.7. Two important residues for this peptide-protein interaction are Arg 379 and Lys 382 of the p53 CTD, which exhibited large negative potential energies of interaction with the receptor. The residue Arg 379 formed persistent hydrogen bonds with His 42 and Glu 45 of S100B($\beta\beta$), whereas Lys 382 formed hydrogen bonds with Glu 86 and the backbone carbonyl oxygen of Phe 88. Residues His 368, Ser 378, and Lys 381 also formed many favorable electrostatic contacts with other S100B($\beta\beta$) residues, especially among those in helix 3 and helix 4 which exhibited a larger fluctuation in the apo-form of the protein, including residues Glu 45, Glu 46, Lys 48, Glu 49, Glu 51, Lys 55, and Glu 89. The Lys 386 residue of the p53 CTD also formed electrostatic interactions with Gln 71 and Glu 86 of the opposite β -subunit.

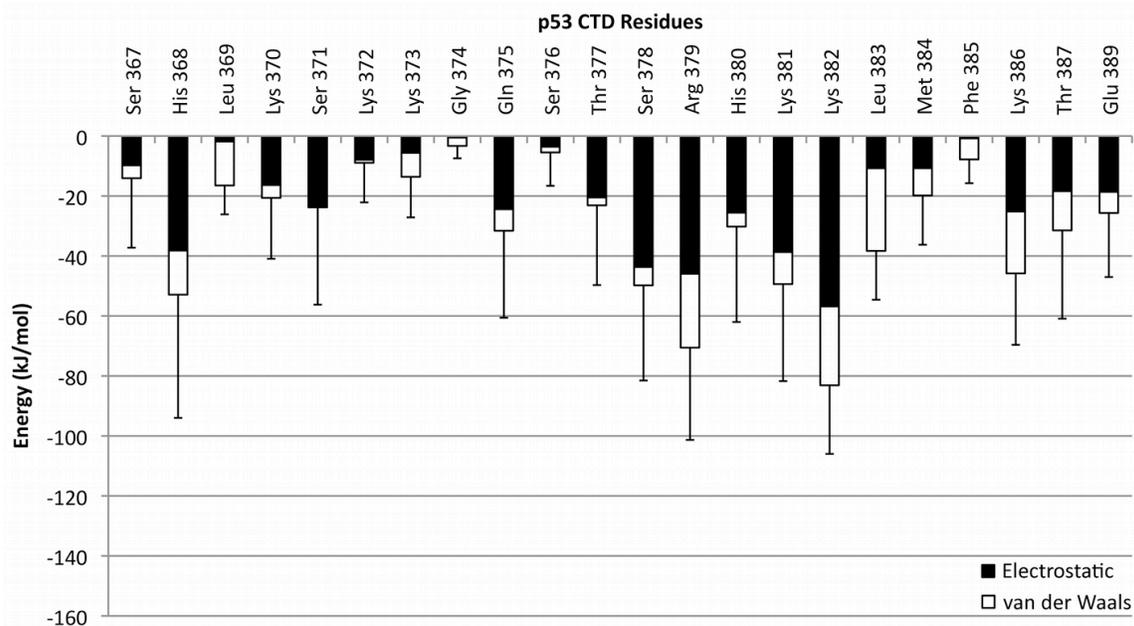


Figure 4.7. Potential energy of interaction between the p53 CTD and S100B($\beta\beta$) by residue. The electrostatic contribution is shown in black, and the van der Waals contribution is shown in white. Error bars represent the standard deviation in the sum of the interactions.

The van der Waals contacts formed by both Arg 379 and Lys 382 further contribute to the binding of the p53 CTD peptide to S100B($\beta\beta$). The hydrophobic binding pocket on S100B($\beta\beta$) is

defined by residues Leu 44, Val 52, Val 56, Met 79, Val 80, and Phe 87. The average hydrophobic solvent-accessible surface area (SASA) of the binding site on S100B($\beta\beta$) was $3.18 \pm 0.22 \text{ nm}^2$ over the last 50 ns of simulation (Figure 4.8). Residue Leu 383 of the p53 CTD faced invariably into this hydrophobic core, and contributed the greatest negative van der Waals potential of interaction to the peptide-protein interaction (Figure 4.7 and Table S4.4). Residues Met 384 and Phe 385 as well as the hydrophobic portion of several other p53 CTD sidechains further contributed to the binding event.

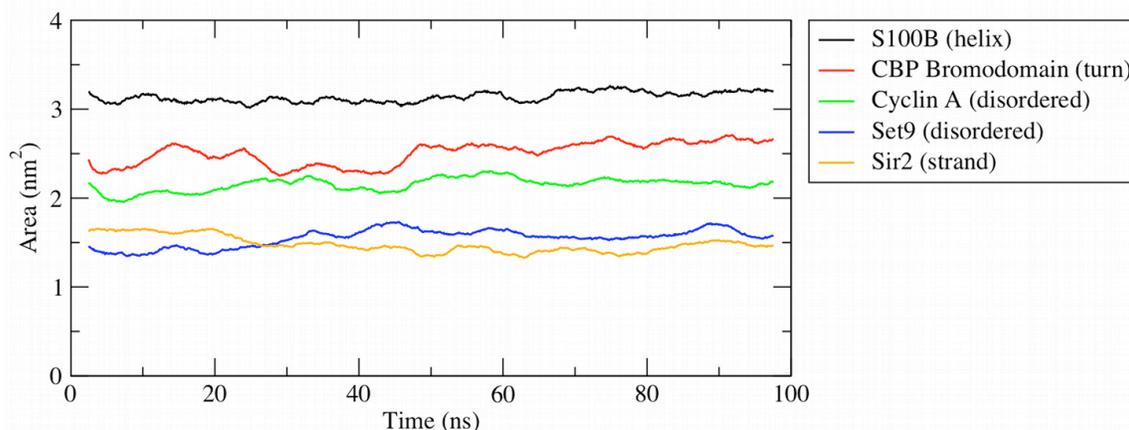


Figure 4.8. Hydrophobic SASA of the p53 CTD binding site on each apo-binding partner. The typical secondary structure content that the p53 CTD adopts when bound to each partner is listed in the legend in parenthesis. The data presented are averaged over the three replicates of the apo-form of the binding partner and smoothed over a 20-point sliding window.

One possible explanation for the specific conformation that the p53 CTD peptide adopts is that an α -helix enables alignment of the hydrophobic residues (Leu 383, Met 384, Phe 385) toward the hydrophobic binding site. This argument would be more convincing if there were additional hydrophobic residues further N- or C-terminus on the peptide that would also be aligned to the binding site only in an α -helical conformation. As it stands, a turn or even random coil could allow for three adjacent residues to be aligned such that they face the same binding site. Another explanation could have to do with the relative shortage of hydrogen bond donors or acceptors within the large hydrophobic core. Most side chain hydrogen bonds are satisfied by the surrounding ring of acidic and polar amino acids, as well as by the solvent, but the large hydrophobic patch does not provide sufficient hydrogen bond donors or acceptors for the p53

CTD backbone (Figure 4.9). Adopting a helical conformation solves this problem, as intra-chain backbone hydrogen bonds are readily formed.

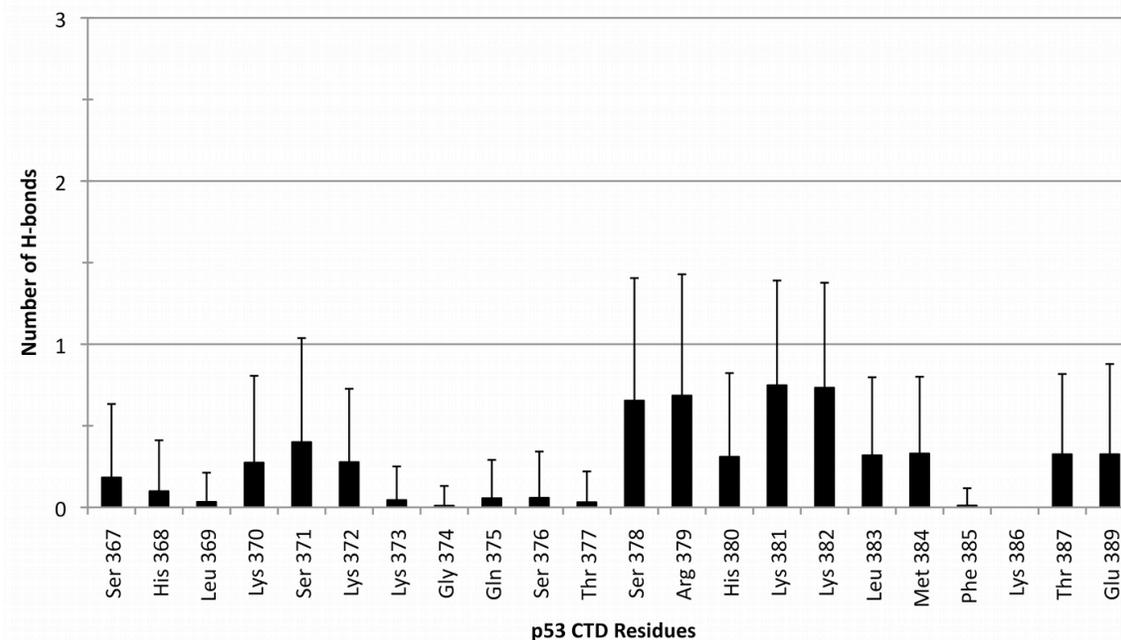


Figure 4.9. Average number of hydrogen bonds formed between the main chain of each p53 CTD residue and the receptor, S100B($\beta\beta$).

The p53 N-terminal transactivation domain (TAD) is a good model for comparison of the intricate peptide-receptor relationship. In a series of recent publications, Dastidar *et al.* (74-76) showed that the MDM2-bound form of the p53 TAD is a helix dominated by van der Waals interactions. The authors suggest that the α -helix is formed so that three hydrophobic residues (Phe 19, Trp 23, and Leu 25) all face into the hydrophobic binding pocket. Further, they found that the initial complex formation is dominated by electrostatic interactions, much like Chen (34) proposes with the “fly-casting” mechanism (77). This fits with our RMSF measurements that describe large fluctuations in acidic and polar residues around the p53 CTD binding site in the apo-form, and increased fluctuations among the hydrophobic residues in the bound form of S100B($\beta\beta$) (Figure 4.4c).

One important observation relates to the role of Ser 376 and Thr 377 of the p53 CTD. Rustandi *et al.* (28) suggested that these two residues are buried against the face of S100B($\beta\beta$), effectively blocking phosphorylation at these sites. Although our simulations agreed that Thr 377

tended to stay buried, forming hydrogen bonds with residues Glu 45, Glu 46, and Glu 49, we found that a small conformational shift that happened fairly quickly in all three replicates caused Ser 376 to become solvent exposed. This is illustrated by the very small potential energy of interaction between Ser 376 and the receptor (Figure 4.7) and by the infrequency of formation of mainchain hydrogen bonds to the receptor (Figure 4.9). Steric hindrance could still be a factor in preventing phosphorylation of that residue given its low angle along the protein surface.

Another interesting feature of S100B($\beta\beta$) is the calcium-binding domain. In their investigation, Rustandi *et al.* (28) determined that calcium binding to each S100B β -subunit induces a major conformational change in helix 3, which forms part of the p53 CTD binding site. This change helps open part of the large hydrophobic core, including S100B($\beta\beta$) residues Met 79, Val 80, Leu 44, and Val 56, which are necessary for p53 CTD binding, corroborating the importance of the hydrophobic binding site suggested by our MD simulations. Interestingly, the RMSF data show that fluctuation in the calcium-binding domain actually increased when the p53 CTD was bound. It has previously been demonstrated that this observed redistribution of fluctuation can be a mechanism of entropic stabilization of the protein complex (78), or as a means for the receptor to bind its substrate (p53 CTD) more tightly.

IMA3

In this crystal structure, the p53 CTD fragment (residues 379-387) is solved in complex with the deacetylase Sir2. From the DSSP profile (Figure 4.10), we observed that the p53 CTD fragment exhibited very little secondary structure content when bound to Sir2. The only notable structure formed was a turn centered at residue Lys 381, which only appeared with any consistency in two of the three replicates (Figure 4.10a-c). In the absence of Sir2, however, the same p53 CTD fragment adopted many different ordered conformations (Figure 4.10d-f), most notably a very short strand-turn-strand structure that persisted for almost 80 ns in one replicate, and appeared for about 10 ns in another replicate. Concurrently, the RMSD analysis revealed an increase in average backbone deviation of the p53 CTD in the absence of Sir2 (Figure 4.11a). These data suggest that our MD simulations successfully reproduced the stabilization effect that Sir2 has on this p53 CTD fragment (29).

In the analysis of Sir2, we found that a trend in backbone deviation was not as apparent as for the p53 CTD (Figure 4.11b). When the p53 CTD was not bound, the average backbone RMSD was only slightly higher than when the p53 CTD was bound. The RMSF analysis, however, clearly indicated two main regions that fluctuated more in the unbound form of the enzyme – residues 45-60 and residues 150-175 (Figure 4.11c). These stretches of amino acids include several regions that form the p53 CTD-Sir2 interface, including Glu 48, Glu 167, and Gln 171, as well as a short extension into the zinc-binding module of Sir2. This region includes the FGE loop (residues 162-169), which, when perturbed, opens up a binding tunnel that acted as a receptacle for AcLys 382 of the p53 CTD (29). In contrast to the more flexible regions that existed in the unbound form of Sir2, there were a few regions that were more flexible in the p53 CTD-bound form of Sir2, most notably residues 69–77, 89–92, 106-112, 190-193, 206-211, and 219-225. Almost all of these regions are loops or helices that comprise the Rossmann fold domain (residues 1-27, 77-118, and 170-253).

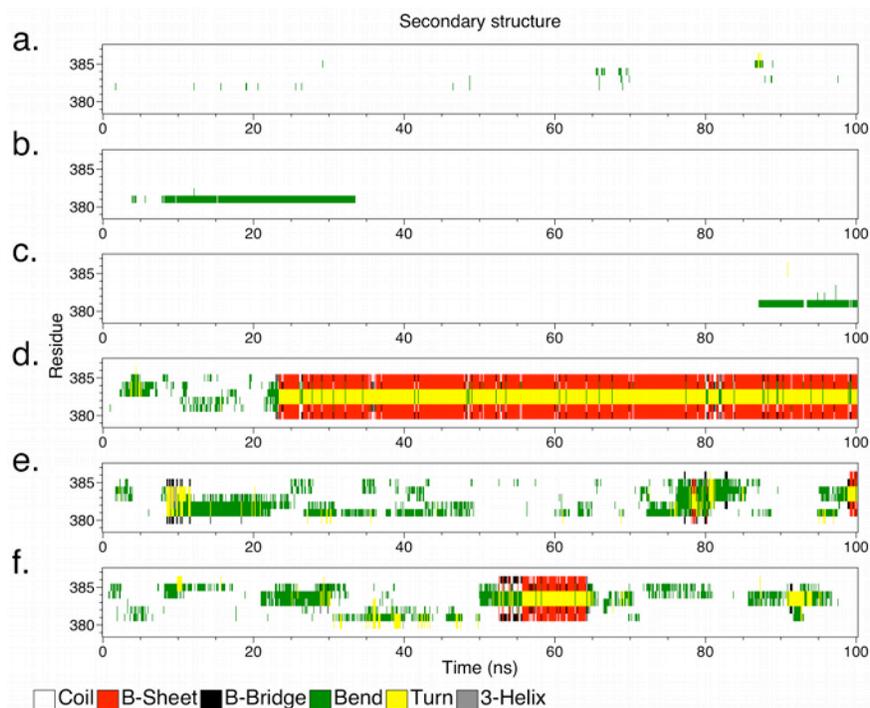


Figure 4.10. DSSP analysis for 1MA3. Panels (a–c) show the secondary structure content of the p53 CTD when bound to Sir2. Panels (d–f) show the secondary structure content of the unbound p53 CTD.

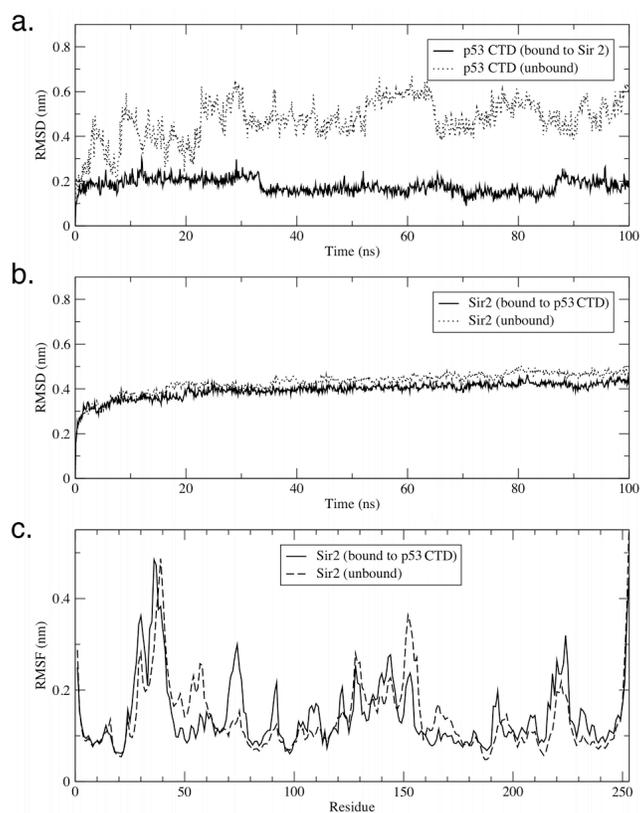


Figure 4.11. RMSD and RMSF analyses for 1MA3. Panel **(a)** shows the average backbone RMSD for the three replicates of the p53 CTD in complex with Sir2 (solid line), and in the absence of Sir2 (dotted line). Panel **(b)** shows the average backbone RMSD for the three replicates of the Sir2 receptor in complex with the p53 CTD (solid line) and in the absence of the p53 CTD (dotted line). Panel **(c)** represents the RMSF of Sir2 in complex with the p53 CTD (solid line), and in the absence of the p53 CTD (dashed line).

The largest PC (23.2%) in Sir2 while bound to the p53 CTD could be described as a pinching and releasing of the Rossmann fold domain (Figure 4.12b). This motion corresponded nicely with an increased fluctuation in this region when bound to the p53 CTD. In contrast to that, the largest PC (29.4%) in the unbound form of Sir2 was a hinging motion at the p53 CTD binding site. This hinging motion is the same motion that was diminished by the formation of a β -sheet upon the p53 CTD binding. Figure 4.12a shows that there was a general increase in both correlated and anti-correlated motion across the entire Sir2 backbone when the p53 CTD was absent.

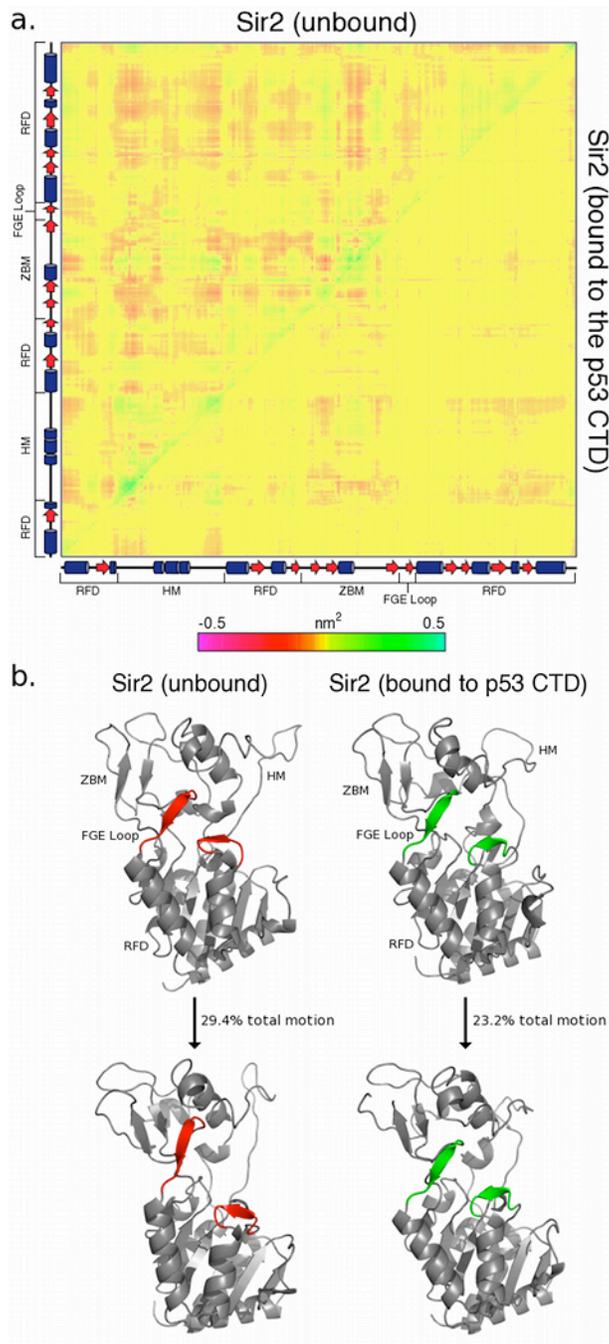


Figure 4.12. PC analysis for 1MA3. Panel (a) shows a covariance matrix illustrating correlated and anti-correlated motions within Sir2 in the absence of the p53 CTD (top left) and when bound to the p53 CTD (bottom right). The secondary structure of the Sir2 backbone is represented along the axes (from left-to-right, and from bottom-to-top). Panel (b) shows the motion of the largest eigenvector on the Sir2 structure in the absence of the p53 CTD (left) and when bound to the p53 CTD (right). The sites on Sir2 that bind the p53 CTD are colored red and green.

Here, we observed another instance in which hydrophilic and acidic residues near the hydrophobic binding site tended to fluctuate more in the apo-form of the receptor, presumably as a method to aid in the formation of initial contacts with the target (76). Also, we detected another case in which the dynamics of a cofactor-binding site changed upon p53 CTD binding. The fluctuation and configurational dynamics (as calculated by the PC analysis) both increased in the NAD-binding domain when the p53 CTD was bound to Sir2. It is well established that increased fluctuations in active sites or cofactor-binding sites can be a means of induced catalysis (79-80). These observations, therefore, are in accordance with the known biochemistry of the p53 CTD-Sir2 interaction, in that the binding event leads to deacetylation of the peptide (29).

One interesting observation regarding the p53 CTD-Sir2 interaction is that the stretch of the p53 CTD from residue 380 to 385 formed a β -strand when in complex with Sir2, while residues from that same stretch formed an α -helix when in complex with S100B($\beta\beta$) (although the DSSP algorithm did not identify the p53 CTD as a β -strand, the molecular geometry and the pattern of hydrogen bond formation was indicative of a β -strand). This β -strand joined two other β -strands of Sir2 (denoted β 7 and β 9, see Figure 4.13), forming a stable β -sheet. Avalos *et al.* (29) termed the role that the p53 CTD plays in this process as a “ β -staple”. One possible explanation for this observation is directly related to the proposed model of the formation of an α -helix in the p53 CTD upon binding S100B($\beta\beta$). In the case of Sir2, the hydrophobic SASA in the binding pocket is relatively small, measuring $1.43 + 0.15 \text{ nm}^2$ averaged over the last 50 ns of simulation (Figure 4.8). An abundance of hydrogen bond donors and acceptors readily interact with the backbone of the p53 CTD (Figure 4.14), eliminating the need to form intra-chain hydrogen bonds, as is the case in an α -helix.

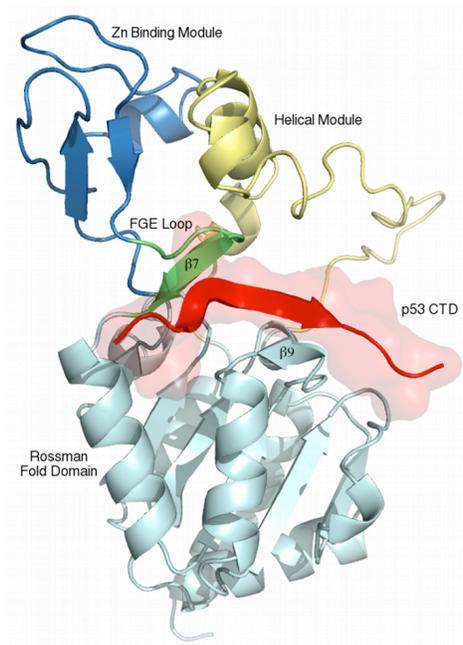


Figure 4.13. Cartoon representation of the p53 CTD (colored in red), bound to Sir2. The Rossman fold domain is colored in cyan, the helical module is colored in yellow, the zinc binding module is colored in dark blue, and the FGE loop is colored in green. The two β -sheets ($\beta 7$ and $\beta 9$) with which the β -strand of the p53 CTD forms a β -staple are also labeled.

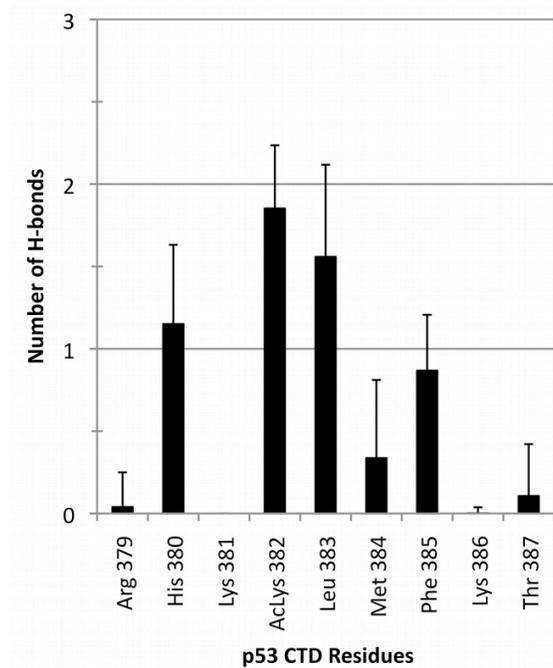


Figure 4.14. Average number of hydrogen bonds formed between the main chain of each p53 CTD residue and Sir2.

The important interactions that maintained this β -sheet structure involved main chain and backbone hydrogen bonds between highly conserved residues on the surface of Sir2 (29). For example, the main chain atoms of His 380 and AcLys 382 (p53 CTD) formed contacts with amino acids in the FGE Loop of Sir2, Gly 166, Glu 167, and Leu 169. Further, the backbone atoms of Leu 383 and Phe 385 of the p53 CTD formed contacts with Val 195 and Tyr 197 of the Rossmann fold domain. We found that these contacts persisted in all three replicate simulations. However, the residue that contributed the largest negative potential energy of interaction was the AcLys 382 of the p53 CTD (Figure 4.15). The modified residue was buried deep within hydrophobic pocket of the enzyme, a location that put it in proximity to the active site. Van der Waals interactions between AcLys 382 and several Sir2 residues (including His 118, Val 163, Phe 165, Leu 169, and Val 196) were the major contribution to the potential energy of interaction. In all three replicate simulations, the positioning of AcLys 382 remained unchanged.

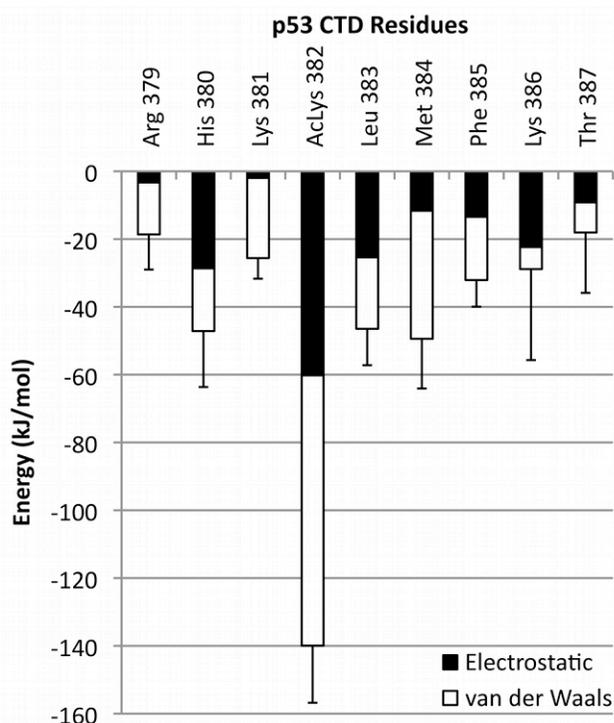


Figure 4.15. Potential energy of interaction between the p53 CTD and Sir2 by residue. The electrostatic contribution is shown in black, and the van der Waals contribution is shown in white. Error bars represent the standard deviation in the sum of the interactions.

Regarding the interaction between the p53 CTD and cyclin A, we began again by assessing the stability during MD of the p53 CTD fragment by measuring the DSSP profile (Figure 4.16). In the three replicates in which the p53 CTD fragment was bound to cyclin A (Figure 4.16a-c), nearly all of the secondary structure manifested as a bend or a turn centered around residue His 380. In some cases, the calculated secondary structure transitioned to coil during the simulation, and only very sparingly did any additional secondary structure other than coil appear within the p53 CTD fragment. In the three replicates in which the p53 CTD was not bound to its receptor (Figure 4.16d-f), the appearance of random secondary structure was much more abundant and sporadic. Bend and turn elements appeared and disappeared between residues 380 to 384 with very little to no observable consistency.

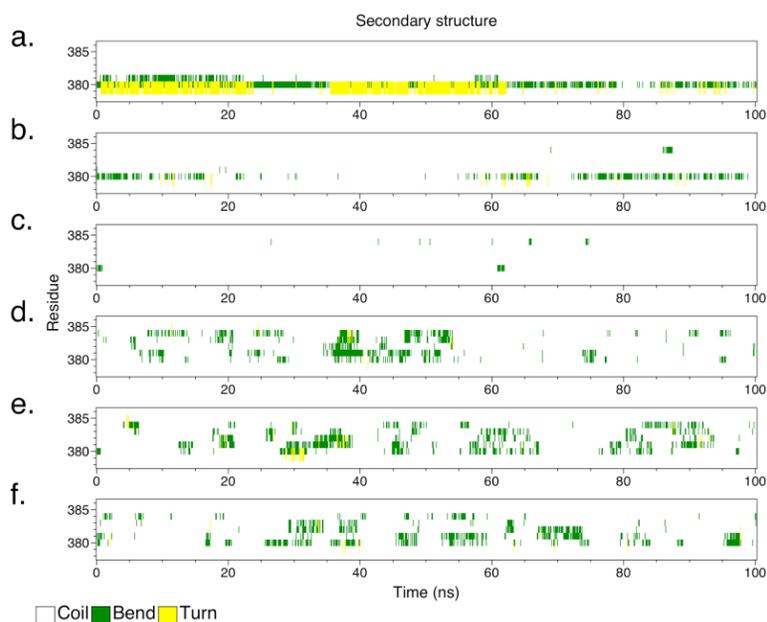


Figure 4.16. DSSP analysis for 1H26. Panels (a–c) show the secondary structure content of the p53 CTD when bound to cyclin A. Panels (d–f) show the secondary structure content of the unbound p53 CTD.

When comparing the average backbone RMSD between the replicates (Figure 4.17a), we found that the deviation in the p53 CTD fragment was slightly lower when it was in complex with the receptor, cyclin A. These data and the DSSP data together show that the bound form of the p53 CTD fragment was calculated to fluctuate less than the unbound form.

The dynamics of cyclin A appeared to change very little between the two states of the protein. The RMSD profile (Figure 4.17b) shows an exceedingly slight increase in deviation between the bound and unbound states. Similarly, a PC analysis did not reveal much change in the large-scale correlated motions in the protein upon p53 CTD binding. There was some increased motion in the loop immediately N-terminus to the $\alpha 1$ helix, as well as in both of the protein termini when the p53 CTD was not bound (Figure 4.18a). Analysis of the greatest eigenvector in each system showed very little change in the actual p53 CTD binding site region (Figure 4.18b). One possible explanation for this observation is that the p53 CTD fragment in question is only nine amino acids in length, very small compared to the size of cyclin A (258 amino acids) and the size of the cyclin A-associated pCDK2 (297 amino acids) (Figure 4.19). On this scale and in this timeframe, it is difficult to measure any broad changes such as total backbone deviation.

Conveniently, we were able to measure local changes in the dynamics of the cyclin A receptor through a RMSF analysis (Figure 4.17c). The RMSF analysis of cyclin A in both the p53 CTD-bound and the unbound states revealed two interesting features. First, a short stretch from residues 241-253 experienced a sharp increase in fluctuation upon p53 CTD binding. This stretch of residues gave shape to the core of the hydrophobic binding pocket, and contains Leu 253, one of the cyclin A residues that interacted with the hydrophobic residues of the p53 CTD. Several other peaks in cyclin A that are also consistently high in RMSF across the replicates in which the p53 CTD fragment was bound are loops that surround the hydrophobic core (residues 194-207, 300-310, and 342-350). When the p53 CTD was not bound, the RMSF in cyclin A increased across residues 280-285, a loop that forms the outer ring of the binding pocket and encompasses Asp 283, a residue that is important in p53 CTD binding.

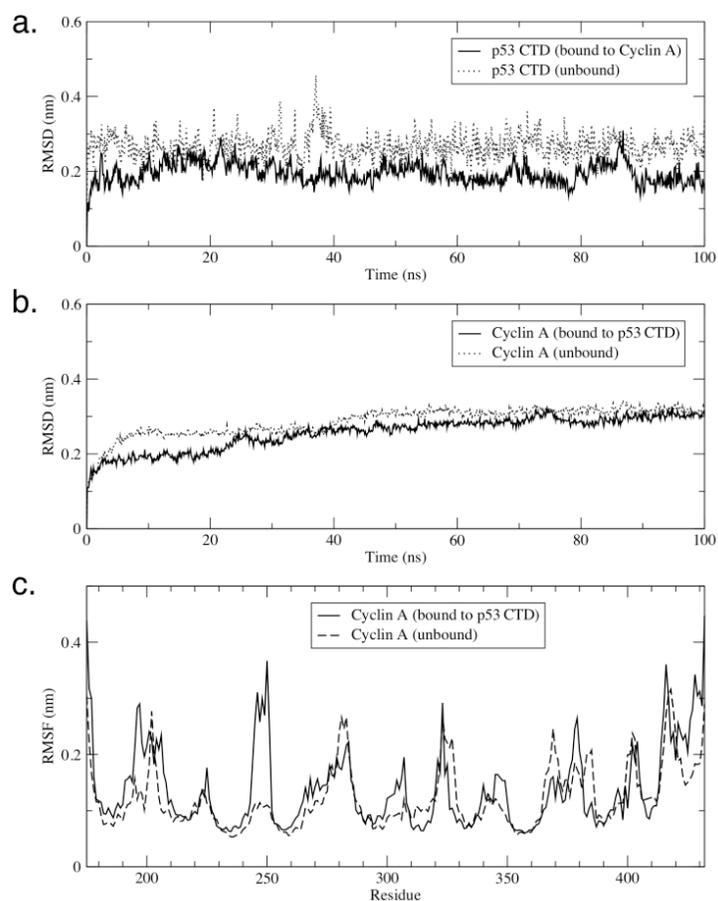


Figure 4.17. RMSD and RMSF analysis for 1H26. Panel (a) shows the average backbone RMSD for the three replicates of the p53 CTD in complex with cyclin A (solid line), and in the absence of cyclin A (dotted line). Panel (b) shows the average backbone RMSD for the three replicates of the cyclin A receptor in complex with the p53 CTD (solid line) and in the absence of the p53 CTD (dotted line). Panel (c) represents the RMSF of cyclin A in complex with the p53 CTD (solid line), and in the absence of the p53 CTD (dashed line).

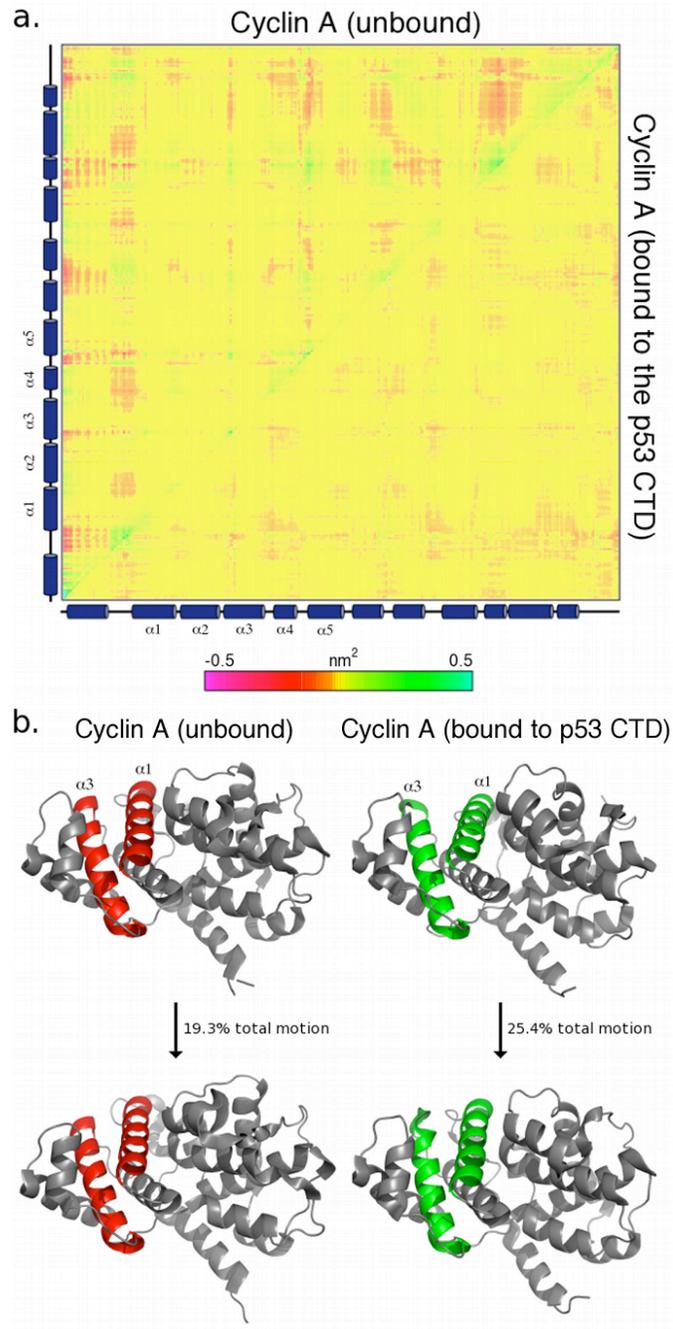


Figure 4.18. PC analysis for 1H26. Panel (a) shows a covariance matrix illustrating correlated and anti-correlated motions within cyclin A in the absence of the p53 CTD (top left) and when bound to the p53 CTD (bottom right). The secondary structure of the cyclin A backbone is represented along the axes (from left-to-right, and from bottom-to-top). Panel (b) shows the motion of the largest eigenvector on the cyclin A structure in the absence of the p53 CTD (left) and when bound to the p53 CTD (right). The sites on cyclin A that bind the p53 CTD are colored red and green.

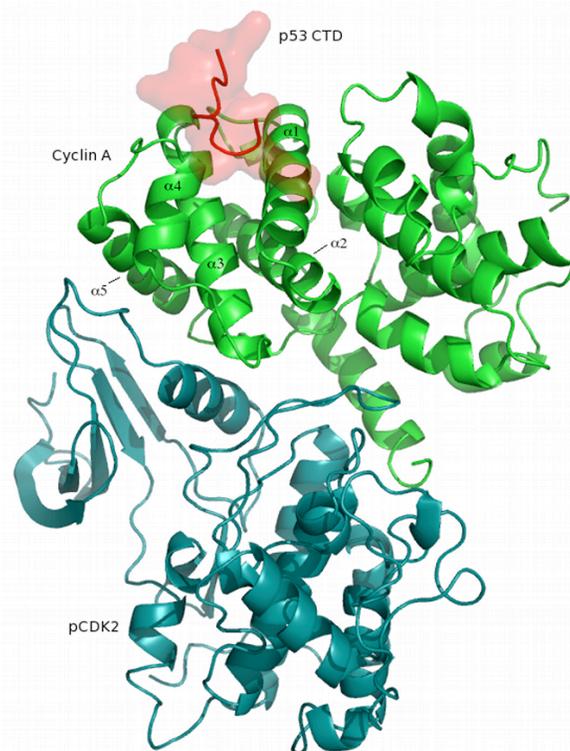


Figure 4.19. Surface rendering and cartoon of the p53 CTD (colored in red). Cyclin A is shown in green cartoon, whereas pCDK2 is shown in blue cartoon. The five helices that form the p53 CTD binding site ($\alpha 1$ - $\alpha 5$) in cyclin A are labeled.

The p53 CTD binding site on cyclin A is characterized by a small hydrophobic binding pocket that averaged $2.20 \pm 0.17 \text{ nm}^2$ of hydrophobic SASA over the last 50 ns of simulation (Figure 4.8). Residues Leu 383 and Phe 385 of the p53 CTD tended to form contacts in this pocket (which includes residues Ile 213, Leu 214, and Leu 253 of Cyclin A), comprising the bulk of the van der Waals contribution to the potential energy of interaction (Figure 4.20). In addition to these hydrophobic contacts, Ser 378, Arg 379, His 380, and Lys 381 of the p53 CTD formed recurring contacts with Glu 220, Glu 224, and Asp 283 of cyclin A, comprising a very large electrostatic potential of interaction (Figure 4.20). Two frequently-modified lysine residues of the p53 CTD (Lys 382 and Lys 386), tended to stay oriented toward the solvent in our simulations. It is noteworthy that the binding site in cyclin A has an abundance of readily available hydrogen bond donors and acceptors. These residues satisfied hydrogen bonds in the p53 CTD backbone, stabilizing the disordered conformation that the peptide maintained. Further, the hydrophobic pocket was much smaller in cyclin A than it was in S100B($\beta\beta$), and thus, there

was less opportunity for the p53 CTD peptide to form intra-chain backbone-backbone hydrogen bonds, and consequently no tendency to form an α -helix structure.

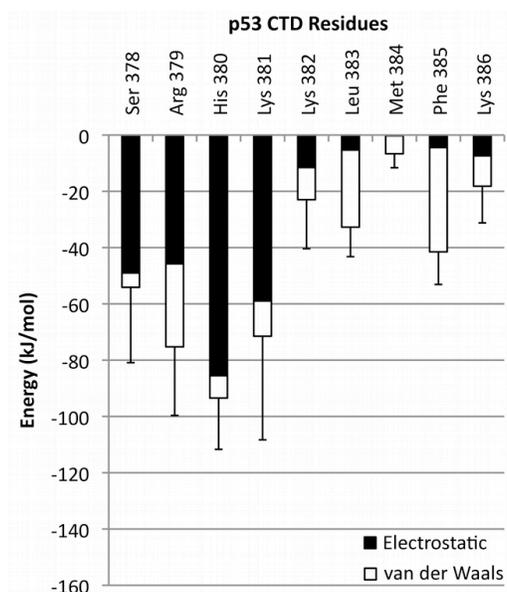


Figure 4.20. Potential energy of interaction between the p53 CTD and cyclin A by residue. The electrostatic contribution is shown in black, and the van der Waals contribution is shown in white. Error bars represent the standard deviation in the sum of the interactions.

The most interesting behavior seen in this binding pocket was the fluctuation pattern as measured by the RMSF analysis. Hydrophobic residues within the binding pocket increased in fluctuation upon p53 CTD binding, and polar and acidic residues near the entrance to the pocket increased in fluctuation when the p53 CTD was not bound. An explanation for this behavior is that when the p53 CTD bound to cyclin A, the hydrophobic residues from the p53 CTD effectively cap the hydrophobic pocket, creating a solvent-free hydrophobic cavity that the cyclin A residues may explore conformationally. When the p53 CTD was not bound, however, residues 280-285 of cyclin A increased in fluctuation. This increase may serve a dual purpose. First, one of these residues, Asp 283, represents another example of an acidic residue that increased in fluctuation in the apo-form of the receptor to favor p53 CTD binding. Secondly, an increased fluctuation in this region partially conceals the hydrophobic binding site from the solvent, helping to preclude the exposure to water molecules.

Overall, our data agree with interactions that have been previously described within the crystal structure of the complex (30). Residues Leu 383 and Phe 385 of the p53 CTD are

important residues for establishing hydrophobic contacts within the hydrophobic binding pocket, whereas Arg 379, His 380, and Lys 381 mediate important electrostatic interactions that occur between the p53 CTD and cyclin A.

1JSP

The DSSP profile of the p53 CTD taken from crystal structure 1JSP was very revealing (Figure 4.21). In the three replicate simulations in which the p53 CTD was bound to the CBP bromodomain, we observed a moderate change in secondary structure content over the course of the MD simulations (Figure 4.21a-c). The peptide fragment was typically characterized by turn or bend conformations along the middle of the chain. In the three replicate simulations of the p53 CTD fragment in the absence of the CBP bromodomain, drastic changes occurred (Figure 4.21d-f). The first 10 to 20 ns of simulation progressed similarly to the receptor-bound replicates, but as the simulations passed the 20 ns mark, the N- and C-terminal ends of the p53 CTD fragment came together as β -strands to form an anti-parallel β -sheet. This conformation was very stable and persisted for the rest of the simulation.

The RMSD analysis of the same p53 CTD fragment (Figure 4.22a) showed that the average backbone deviation for the three replicates in which the peptide was not bound to the receptor was substantially higher than the same RMSD measurement for the three receptor-bound replicates. The secondary structure and RMSD analysis illustrated a considerable change in the structural features of the p53 CTD fragment in the absence of the CBP bromodomain.

The average backbone RMSD of the CBP bromodomain, however, actually increased in the presence of the p53 CTD (Figure 4.22b), which is the only case in all five systems in which this phenomenon was observed. The RMSF analysis showed that the ZA Loop (residues 1115-1138) of the CBP bromodomain fluctuated more when the p53 CTD was not bound (Figure 4.22c). This loop contains several polar and acidic residues (including Asp 1124 and Asp 1127) that were oriented toward the solvent. Two loop regions increased in fluctuation once the p53 CTD was bound. These include the BC Loop (residues 1170-1190) and a short strand-turn-strand structure that is just to the C-terminal side of the ZA Loop (residues 1102-1112). Both of these loops flank the hydrophobic binding pocket (Figure 4.23).

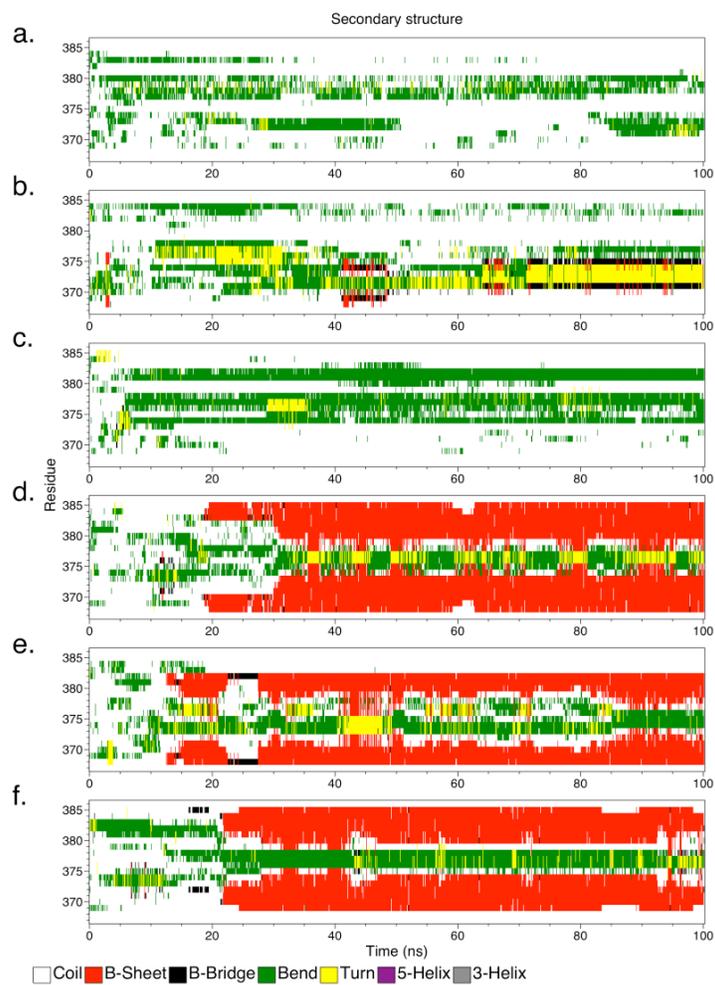


Figure 4.21. DSSP analysis for 1JSP. Panels (a–c) show the secondary structure content of the p53 CTD when bound to the CBP bromodomain. Panels (d–f) show the secondary structure content of the unbound p53 CTD.

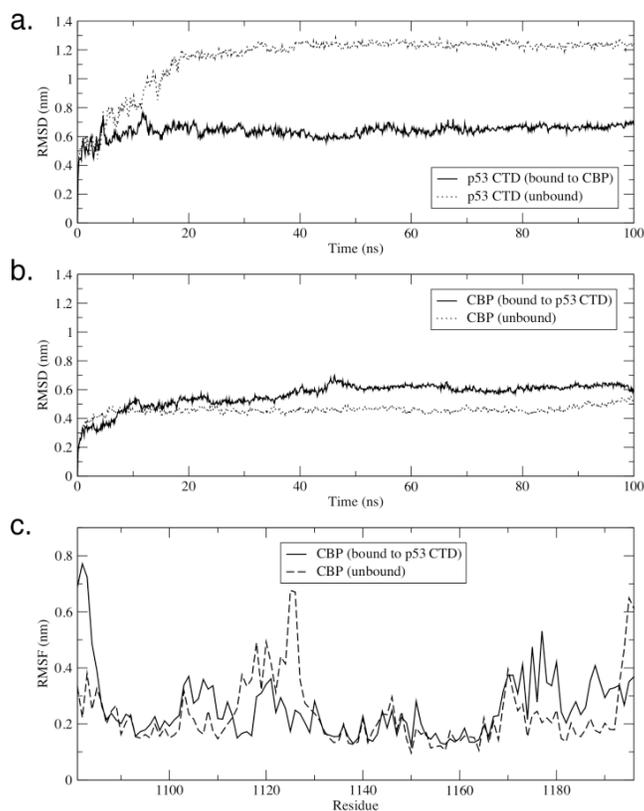


Figure 4.22. RMSD and RMSF analysis for 1JSP. Panel (a) shows the average backbone RMSD for the three replicates of the p53 CTD in complex with the CBP bromodomain (solid line), and in the absence of the CBP bromodomain (dotted line). Panel (b) shows the average backbone RMSD for the three replicates of the CBP bromodomain receptor in complex with the p53 CTD (solid line) and in the absence of the p53 CTD (dotted line). Panel (c) represents of the RMSF of the CBP bromodomain in complex with the p53 CTD (solid line), and in the absence of the p53 CTD (dashed line).

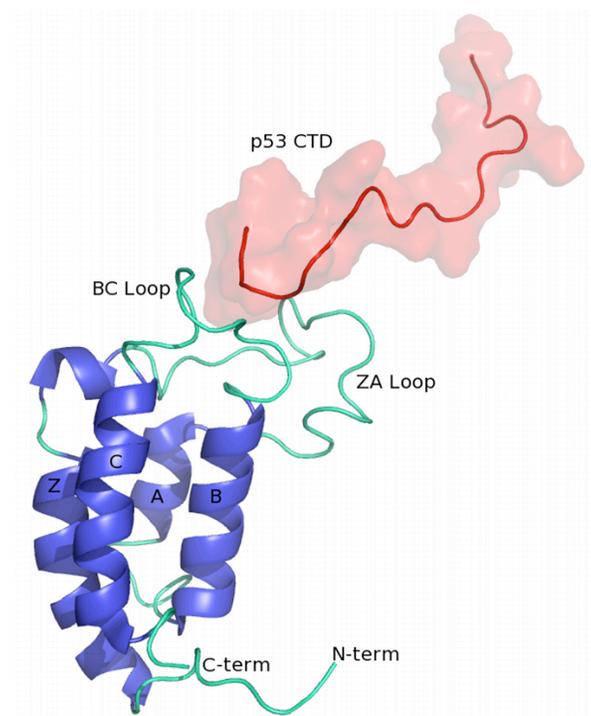


Figure 4.23. Surface rendering and cartoon of the p53 CTD (shown in red). The CBP bromodomain is shown in cartoon. α -Helices are colored dark blue, and loops are colored light blue.

On a larger scale, a PC analysis showed that in the absence of the p53 CTD, the major eigenvector in the CBP bromodomain consisted of an opening and closing of the hydrophobic binding site between the ZA loop and the BC loop (Figure 4.24b). This motion described 58.4% of the total motion in the protein. When the p53 CTD was bound, however, the same motion was dampened and dropped to only 23.8% of the total motion in the protein. From Figure 4.24a it is evident that upon p53 CTD binding, there was a marked increase in both correlated and anti-correlated motion in the CBP bromodomain, especially among the ZA loop, the BC loop, and both of the termini. This observation is in agreement with the marked increase in RMSD that was observed in the CBP bromodomain backbone upon p53 CTD binding.

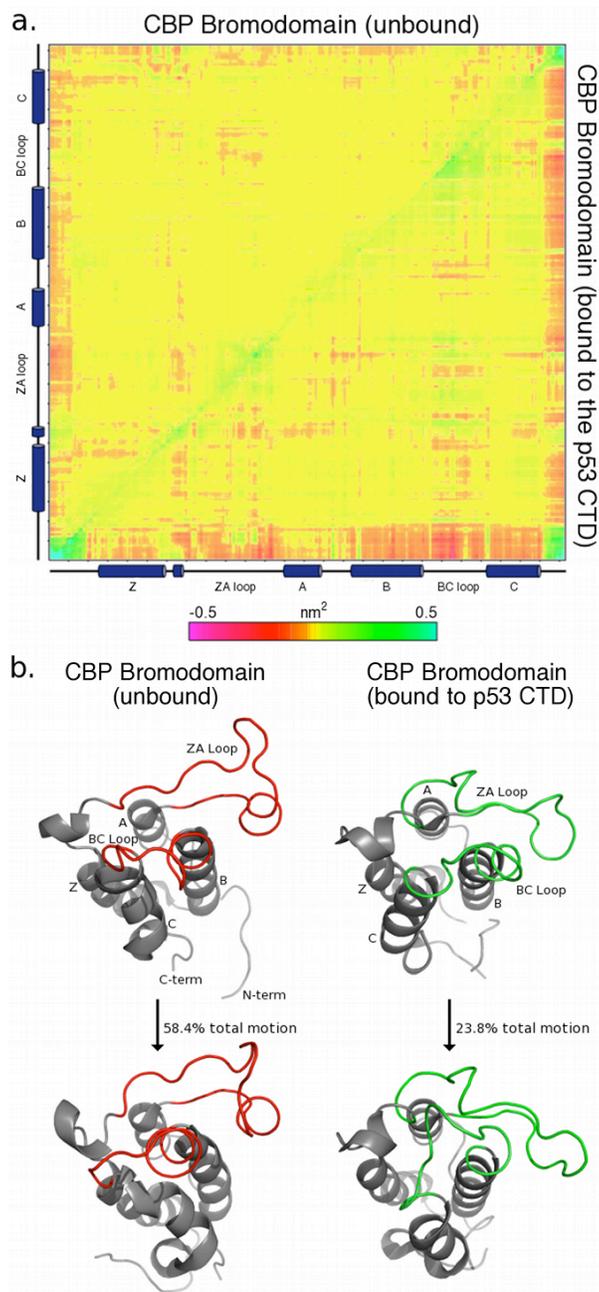


Figure 4.24. PC analysis for 1JSP. Panel (a) shows a covariance matrix illustrating correlated and anti-correlated motions within the CBP bromodomain in the absence of the p53 CTD (top left) and when bound to the p53 CTD (bottom right). The secondary structure of the CBP bromodomain backbone is represented along the axes (from left-to-right, and from bottom-to-top). Panel (b) shows the motion of the largest eigenvector on the CBP bromodomain structure in the absence of the p53 CTD (left) and when bound to the p53 CTD (right). The sites on the CBP bromodomain that bind the p53 CTD are colored red and green.

This case presents another example in which polar or acidic residues near the hydrophobic binding site tends to fluctuate more in the absence of the p53 CTD. This increase in fluctuation in the ZA loop is likely to assist in forming initial electrostatic contacts with the target peptide as was observed in the interaction of MDM2 and the p53 TAD (76). Further, we again observed a case in which hydrophobic residues that shape the binding site increased in fluctuation upon p53 CTD binding. This likely happens because p53 CTD binding effectively caps the binding site such that hydrophobic residues within the site are no longer constrained to avoid water contact and are free to explore a new hydrophobe-friendly conformational space. The PC analysis further confirms this hypothesis because of the observed increase in motion in the hydrophobic binding site upon p53 CTD binding (Figure 4.24).

This molecular system is different from the other four structures considered in this study in that the contact area between the p53 CTD fragment (residues 367-386) and the CBP bromodomain (residues 1081-1196) in the NMR structure is limited to only a few residues. Of the 20-mer p53 fragment, the N-terminal fourteen residues (367-380) are pointing away from CBP and exposed to the solvent, and the C-terminal six residues (381-386) are in close contact with CBP. However, during the simulations, these N-terminal residues of the p53 CTD fragment folded over and formed contacts with the surface of the CBP bromodomain in as quickly as 2 ns, and within 11 ns across all of the replicates. Specifically, polar residues along the p53 CTD backbone (His 368, Ser 371, Lys 373, Gln 375) interacted with Asp 1124 and Asp 1127 of the CBP bromodomain, depending on how the p53 CTD peptide folded during the simulation (Figure 4.25). This shift in the p53 CTD conformation may have induced slight conformational changes along the backbone of the receptor that would not occur in the absence of the peptide, accounting for the increase in RMSD.

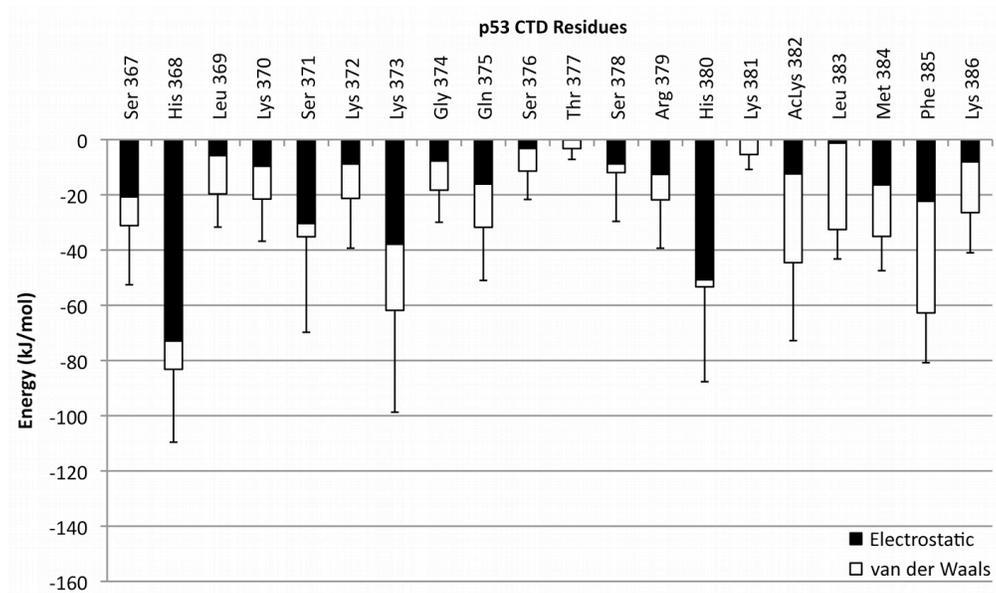


Figure 4.25. Potential energy of interaction between the p53 CTD and the CBP bromodomain by residue. The electrostatic contribution is shown in black, and the van der Waals contribution is shown in white. Error bars represent the standard deviation in the sum of the interactions.

Another key interaction between the p53 CTD and the CBP bromodomain is the location of the AcLys 382 residue. This modified amino acid consistently remained inserted into the hydrophobic cavity in the CBP bromodomain formed by Val 1115, Ile 1122, Tyr 1125, Tyr 1167, Val 1174, and Phe 1177. The two residues to the C-terminal side of the acetylated lysine (Leu 383 and Met 384) also contributed to the negative potential energy of interaction within this pocket, while Phe 285 and Lys 386 formed contacts with the hydrophobic portion of Arg 1112 (Figure 4.25). These hydrophobic contacts remained consistent across all three replicate simulations, indicating that they are likely the major contributing factor in the formation of the p53 CTD-CBP bromodomain complex. We measured the hydrophobic SASA in the binding site to be $2.61 \pm 0.19 \text{ nm}^2$ over the last 50 ns of simulation (Figure 4.8). This area falls closer to that which supports helix formation rather than strand formation, which seems appropriate because the p53 CTD forms a turn structure when bound to the CBP bromodomain, one residue short of an α -helix.

1XQH

According to the DSSP analysis, when the p53 CTD was bound to the methyltransferase Set9, the only secondary structure detected was a bend centered at Ser 371 (Figure 4.26a-c). Ser 371 formed a bend in all three replicates. The secondary structure content of the same fragment consistently changed in the absence of Set9 (Figure 4.26d-f), forming sporadic turn and bend conformations across residues 370-374.

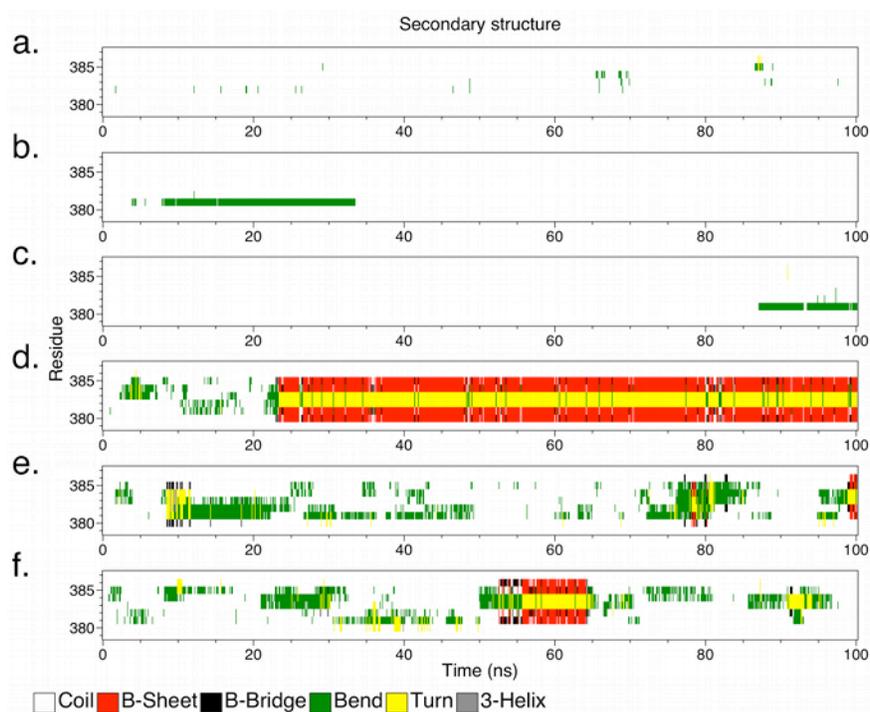


Figure 4.26. DSSP analysis for 1XQH. Panels (a–c) show the secondary structure content of the p53 CTD when bound to Set9. Panels (d–f) show the secondary structure content of the unbound p53 CTD.

The average backbone RMSD of the same p53 CTD fragment was slightly higher when it was unbound compared to when it was bound early in the simulation, although they are comparable for the last 50 ns (Figure 4.27a). These data, along with the DSSP analysis, demonstrate that the fluctuation in the p53 CTD fragment was lower when it was bound to the receptor Set9.

The average RMSD of the Set9 backbone was dependent on whether the p53 CTD fragment was bound (Figure 4.27b). In the uncomplexed form, the average backbone deviation was

slightly higher than in the bound form. A RMSF analysis identified three interesting regions within Set9 (Figure 4.27c). When the p53 CTD was not bound, residues 252-266, a stretch that includes Asp 256 and Thr 266 (residues important in p53 CTD binding), increased in fluctuation. When the p53 CTD was bound, however, two different loops increased in fluctuation. These included residues 217-227 of the Set domain and 278-288 of the Set-I domain, both of which help the formation of the hydrophobic binding pocket in which the methylated Lys 372 of the p53 CTD was inserted (Figure 4.28).

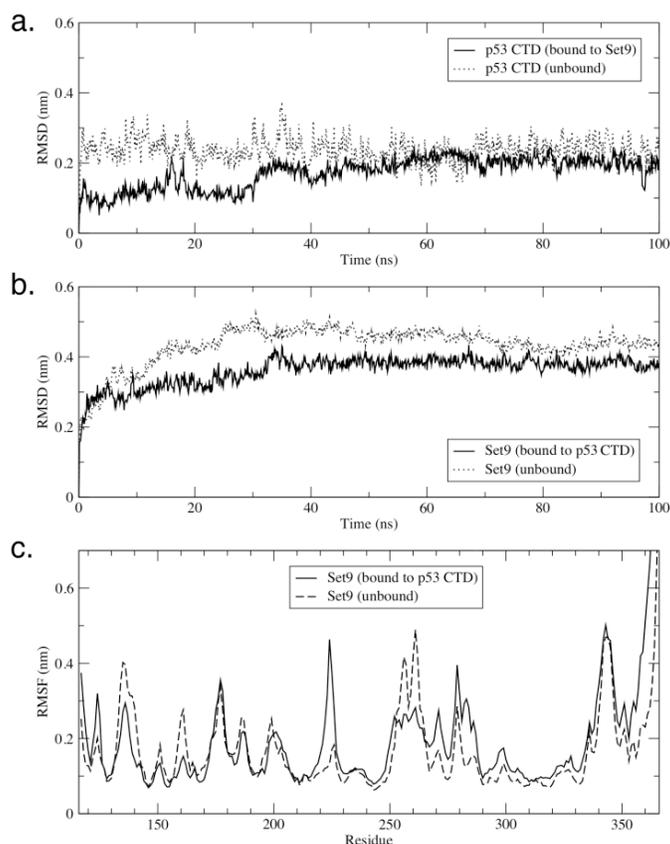


Figure 4.27. RMSD and RMSF analysis for 1XQH. Panel (a) shows the average backbone RMSD for the three replicates of the p53 CTD in complex with Set9 (solid line), and in the absence of Set9 (dotted line). Panel (b) shows the average backbone RMSD for the three replicates of the Set9 receptor in complex with the p53 CTD (solid line) and in the absence of the p53 CTD (dotted line). Panel (c) represents the RMSF of Set9 in complex with the p53 CTD (solid line), and in the absence of the p53 CTD (dashed line).

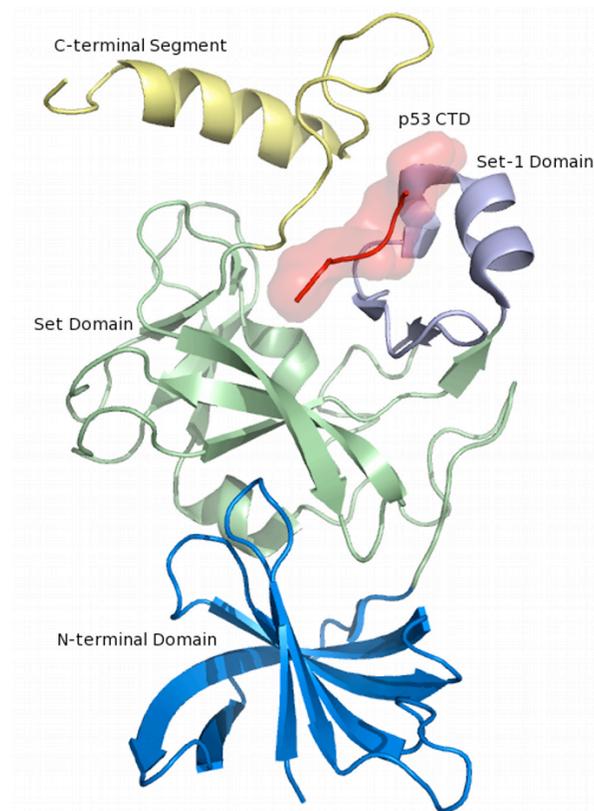


Figure 4.28. Surface rendering and cartoon of the p53 CTD backbone (shown in red). The receptor Set9 is shown in cartoon, highlighting the N-terminal domain (blue), the Set domain (green), the Set-I domain (purple), and the C-terminal segment (yellow).

The PC analysis shows a very drastic difference in the dynamics of Set9 between the bound and unbound forms. When the p53 CTD was unbound, a single vector described 72.0% of the total motion in Set9. That vector corresponded to an opening and closing of the p53 CTD binding site region, or the area between the Set domain and the Set-I domain, through the major shift of the C-terminal segment (Figure 4.29b). When the p53 CTD was bound, the same motion remained the largest PC, but dropped to only a 38.3% contribution to overall motion in the protein. The major eigenvector revealed a much smaller shift of the C-terminal segment between the Set and Set-I domains (Figure 4.29b). Figure 4.29a supports these observations, as it shows a general increase in motion in Set9 when it was not bound to the p53 CTD. Much of this motion could be described by fluctuations in the N- and C-terminus flanking domains, which settled upon p53 CTD binding.

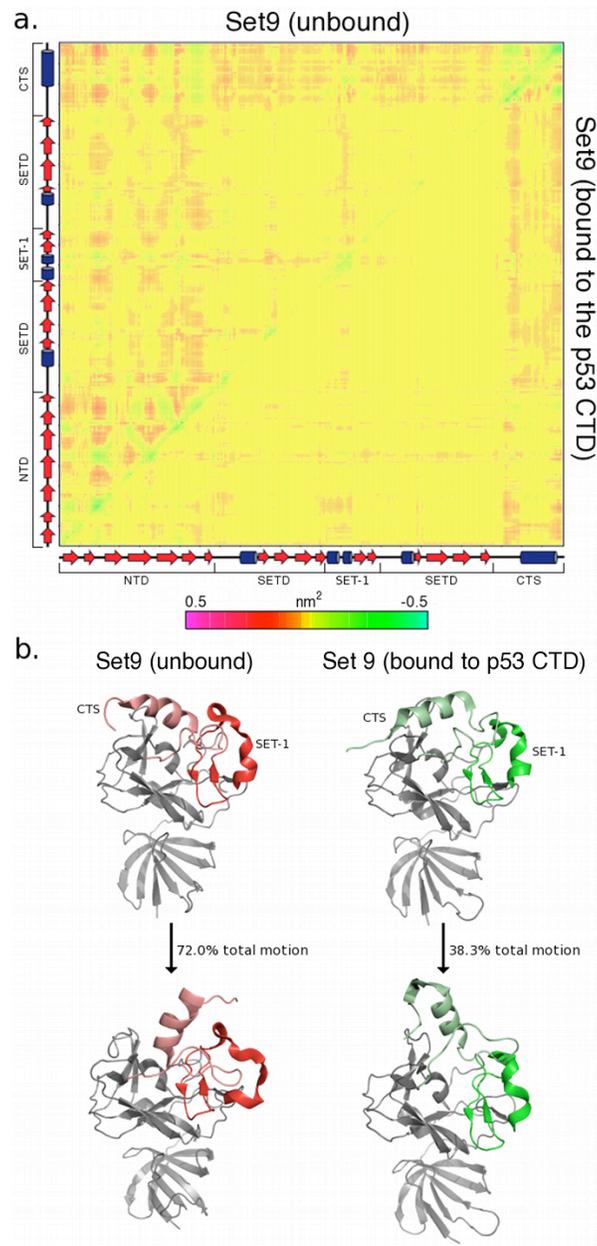


Figure 4.29. PC analysis for 1XQH. Panel (a) shows a covariance matrix illustrating correlated and anti-correlated motions within Set9 in the absence of the p53 CTD (top left) and when bound to the p53 CTD (bottom right). The secondary structure of the Set9 backbone is represented along the axes (from left-to-right, and from bottom-to-top). Panel (b) shows the motion of the largest eigenvector on the Set9 structure in the absence of the p53 CTD (left) and when bound to the p53 CTD (right). The sites on Set9 that bind the p53 CTD are colored red and green.

Again we observed hydrophilic and acidic residues (Asp 256 and Thr 266) with an increased fluctuation near the entrance to the hydrophobic binding site in the absence of the p53 CTD. The PC analysis also described an increased motion in these residues in the absence of the p53 CTD,

indicating a propensity to form initial electrostatic contacts with its target, facilitating binding. Further, upon the binding of the p53 CTD to Set9, we observed increased fluctuation in loops that form both the hydrophobic binding pocket and the cofactor (AdoHcy) binding site. Presumably, residues from the p53 CTD complete the shape of the pocket upon binding in a way that allows hydrophobic residues within the pocket to explore more conformational space (as previously described), and in addition, promote methylation via induced catalysis (79, 80).

Chuikov *et al.* (19) solved the crystal structure of a p53 CTD fragment in complex with methyltransferase Set9. Originally, a 10-mer fragment of the CTD was used in the experiment (residues 369-378), but only the first six residues were well resolved (residues 369-374). Residue Lys 372 of the p53 CTD is methylated, a PTM that regulates p53 by restricting it to the nucleus (19). In our simulations, we found that the MeLys 372 packed against Trp 260 of Set9, as it laid in a channel partly formed by the hydrophobic residues Val 255, Leu 267, and Val 158, with a hydrophobic SASA of roughly $1.60 \pm 0.16 \text{ nm}^2$ (Figure 4.8). It also formed transient hydrogen bonds with Asn 265, Tyr 335, and Tyr 337, adding a substantial electrostatic potential energy contribution to the interaction (Figure 4.30). Chuikov *et al.* proposed that this methylated lysine residue packs hydrophobically against Trp 260, but also they proposed that it forms hydrogen bonds with Arg 258 of Set9, for which we found no evidence in any of our three replicates.

In addition to the interactions of the methylated Lys 372, we found that residues Asp 256, Thr 266, and Ser 268 of Set9 formed stabilizing hydrogen bonds with several backbone atoms of the p53 CTD fragment, as well as with Ser 371 and Lys 373 of the p53 CTD. Chuikov *et al.* (19) suggested Arg 258 of Set9 as an important residue in p53 CTD binding, but we found that this residue tended to orient itself towards the solvent, away from the p53 CTD in all three replicates.

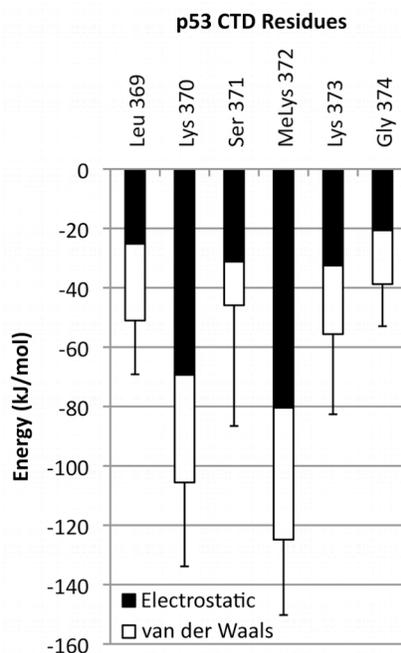


Figure 4.30. Potential energy of interaction between the p53 CTD and Set9 by residue. The electrostatic contribution is shown in black, and the van der Waals contribution is shown in white. Error bars represent the standard deviation in the sum of the interactions.

Conclusions

We studied the dynamics and the interactions of five known p53 CTD-binding partner complexes by MD simulations. We looked for trends in the interactions and conformational dynamics that could be generalized and used to predict new binding partners for the p53 CTD, or locate binding sites on already known partners for which there is no structural information related to protein-protein interactions.

We observed several important trends in the binding partners. Most importantly, the core of each p53 CTD binding site on the partners is a hydrophobic binding pocket that is variable in size. Further, these hydrophobic pockets typically increased in fluctuation upon binding of the p53 CTD. We propose that the binding event completes a cavity from which water is excluded, thereby allowing hydrophobic residues in the receptor to explore more conformational space. In addition, binding sites are typically surrounded by polar and acidic residues, which fluctuate more when the p53 CTD is not bound. This increased fluctuation suggests that the receptors

facilitate p53 CTD binding by forming initial electrostatic contacts with the target peptide, as has been shown previously for the p53 TAD (76). The complex, however, tends to rely heavily on van der Waals type contacts, especially within the hydrophobic cavity, for stabilization. Finally, in the case of receptors that also had a cofactor-binding site, we typically observed an increase in fluctuation at those sites upon p53 CTD binding. This is likely a mechanism to induce catalysis (79, 80) or bind the peptide more tightly by entropic stabilization (78).

We also propose a correlation between the size of the hydrophobic binding site and the secondary structure that the p53 CTD adopts upon binding. The p53 CTD has been observed to be an α -helix, turn, β -strand, or unstructured when in complex with a binding partner. We believe that the larger the hydrophobic area on the receptor, the p53 CTD will have a greater tendency to form a helix. This is due to the lack of hydrogen bond donors in the hydrophobic area. The backbone of the p53 CTD will have unsatisfied hydrogen bonds, thus favoring a helix conformation, which forms intra-chain backbone-backbone hydrogen bonds. If the receptor has a very small hydrophobic area, we anticipate that the p53 CTD will favor a β -strand structure. The availability of hydrogen bond acceptors will be able to satisfy p53 backbone hydrogen bond donors. Receptors with medium-sized hydrophobic patches will induce unstructured or random coil structures on the p53 CTD backbone.

Overall, this information could be useful in designing inhibitors for p53 CTD-binding partner interactions. We would suggest exploiting the mechanism of binding facilitation of the polar and acidic amino acids near the binding site, while keeping in mind with careful consideration the size of the hydrophobic patch and its implication on the structure of the inhibitor. Further, this report will be useful to those trying to determine new binding sites for the p53 CTD among novel partners.

Acknowledgments

The authors thank Justin Lemkul for useful discussions about the methods and results of this project. Also, we thank the administrators of the Terascale Computing Facility at Virginia Tech for technical support and computing hours. This work was supported by the Institute for Critical Technology and Applied Science (ICTAS) at Virginia Tech.

Supplemental Information

Modified Lysine Parameters

Listed below are the parameters for acetylated lysine (Table S4.1), methylated lysine (Table S4.2), and the original lysine parameters from ffG53a6.rtp (Table S4.3). The numbers in the first column correspond to atom numbers as shown in Figure S4.1.

Table S4.1. Acetylated lysine parameters.

Atom Number	Atom Type	Charge Group Number	Charge
1	N	1	-0.31
2	H	1	0.31
3	CH1	2	0
4	CH2	2	0
5	CH2	3	0
6	CH2	3	0
7	CH2	3	0
8	N	4	-0.31
9	H	4	0.31
10	C	5	0.45
11	O	5	-0.45
12	CH3	6	0
13	C	7	0.45
14	O	7	-0.45

Table S4.2. Methylated lysine parameters.

Atom Number	Atom Type	Charge Group Number	Charge
1	N	1	-0.31
2	H	1	0.31
3	CH1	2	0
4	CH2	2	0
5	CH2	3	0
6	CH2	3	0
7	CH2	4	0.127
8	NL	4	0.129
9	H	4	0.3085
10	H	4	0.3085
11	CH3	4	0.127
12	C	5	0.45
13	O	5	-0.45

Table S4.3. Original lysine parameters from ffG53a6.rtp (listed as [LYSH]).

Atom Number	Atom Type	Charge Group Number	Charge
1	N	1	-0.31
2	H	1	0.31
3	CH1	2	0
4	CH2	2	0
5	CH2	3	0
6	CH2	3	0
7	CH2	4	0.127
8	NL	4	0.129
9	H	4	0.248
10	H	4	0.248
11	H	4	0.248
12	C	5	0.45
13	O	5	-0.45

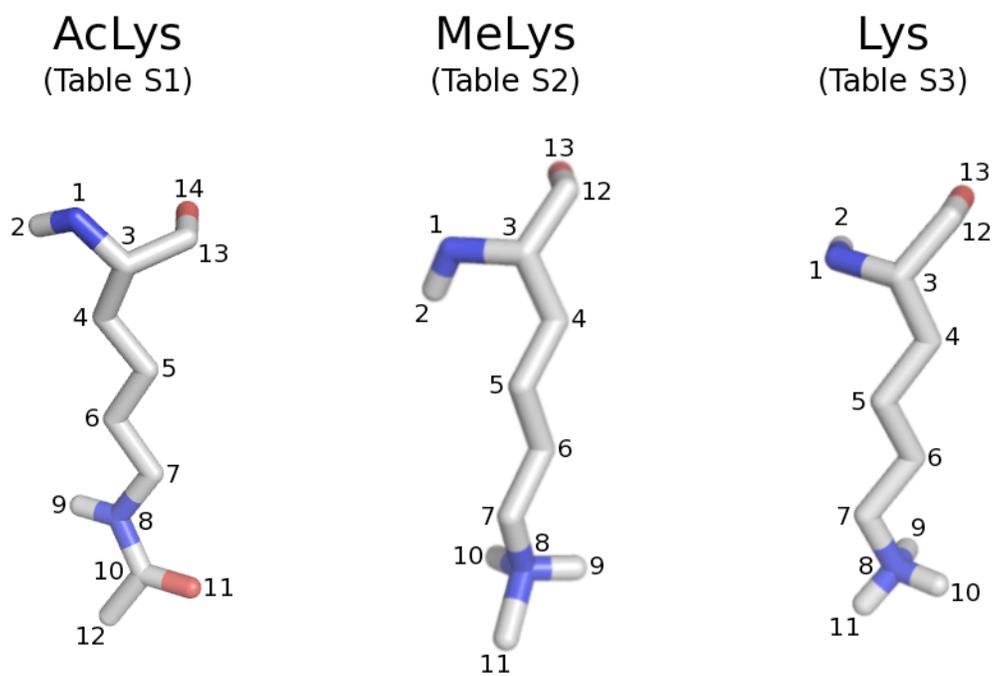


Figure S4.1. Atom numbering scheme for acetylated lysine (AcLys), methylated lysine (MeLys), and unmodified lysine (Lys).

Supplemental Tables for Energy Contributions

Table S4.4. Energy contributions for the interaction between the p53 CTD and S100B($\beta\beta$) (PDB:1DT7). Values are in kJ mol⁻¹.

	Electrostatic	van der Waals	Sum
Ser 367	-9.78±21.96	-4.29±7.38	-14.07±23.17
His 368	-38.21±39.42	-14.69±11.56	-52.90±41.08
Leu 369	-1.83±4.60	-14.66±8.45	-16.49±9.63
Lys 370	-16.35±19.15	-4.29±6.66	-20.64±20.27
Ser 371	-23.16±31.55	-0.59±7.67	-23.75±32.47
Lys 372	-7.95±12.81	-0.98±3.14	-8.93±13.19
Lys 373	-5.62±10.98	-7.99±7.90	-13.61±13.52
Gly 374	-0.53±2.21	-2.77±3.47	-3.30±4.12
Gln 375	-24.42±27.74	-7.17±8.52	-31.58±29.02
Ser 376	-3.59±10.73	-1.89±3.05	-5.47±11.15
Thr 377	-20.51±25.64	-2.60±7.01	-23.11±26.58
Ser 378	-43.67±30.46	-6.12±8.90	-49.79±31.73
Arg 379	-45.94±26.77	-24.62±15.15	-70.56±30.76
His 380	-25.57±30.97	-4.63±7.29	-30.20±31.81
Lys 381	-38.71±30.01	-10.66±12.03	-49.37±32.33
Lys 382	-56.82±19.95	-26.29±11.21	-83.11±22.88
Leu 383	-10.75±15.31	-27.56±5.52	-38.31±16.27
Met 384	-10.74±15.39	-9.10±5.70	-19.84±16.41
Phe 385	-0.80±2.58	-7.01±7.49	-7.81±7.92
Lys 386	-25.20±22.44	-20.60±8.02	-45.80±23.83
Thr 387	-18.38±27.80	-13.09±9.65	-31.47±29.43
Glu 389	-18.64±20.40	-7.01±6.36	-25.65±21.37

Table S4.5. Energy contributions for the interaction between the p53 CTD and Sir2 (PDB:1MA3). Values are in kJ mol⁻¹.

	Electrostatic	van der Waals	Sum
Arg 379	-3.31±6.53	-15.32±8.05	-18.63±10.36
His 380	-28.64±14.60	-18.51±7.65	-47.15±16.49
Lys 381	-1.98±4.97	-23.63±3.52	-25.60±6.09
AcLys 382	-60.22±13.35	-79.73±10.26	-139.95±16.84
Leu 383	-25.40±9.28	-21.07±5.39	-46.47±10.73
Met 384	-11.65±12.56	-37.76±7.54	-49.42±14.65
Phe 385	-13.43±5.97	-18.65±5.10	-32.08±7.85
Lys 386	-22.37±25.19	-6.49±9.22	-28.86±26.82
Thr 387	-9.18±16.17	-8.88±7.47	-18.06±17.81

Table S4.6. Energy contributions for the interaction between the p53 CTD and cyclin A (PDB:1H26). Values are in kJ mol⁻¹.

	Electrostatic	van der Waals	Sum
Ser 378	-49.03±24.72	-5.08±10.30	-54.11±26.78
Arg 379	-45.77±22.86	-29.49±8.49	-75.26±24.38
His 380	-85.52±15.74	-7.93±9.21	-93.45±18.23
Lys 381	-58.98±35.09	-12.51±11.14	-71.49±36.81
Lys 382	-11.55±15.98	-11.44±6.90	-22.99±17.41
Leu 383	-5.34±7.83	-27.39±6.98	-32.73±10.49
Met 384	-0.23±1.52	-6.42±4.72	-6.65±4.96
Phe 385	-4.45±7.13	-37.06±9.13	-41.51±11.58
Lys 386	-7.40±10.35	-10.78±7.95	-18.19±13.05

Table S4.7. Energy contributions for the interaction between the p53 CTD and the CBP bromodomain (PDB:1JSP). Values are in kJ mol⁻¹.

	Electrostatic	van der Waals	Sum
Ser 367	-20.75±20.13	-10.40±7.34	-31.15±21.43
His 368	-72.96±23.12	-10.23±12.71	-83.19±26.38
Leu 369	-5.80±9.80	-13.87±6.97	-19.67±12.02
Lys 370	-9.65±13.95	-11.88±6.23	-21.54±15.27
Ser 371	-30.40±33.74	-4.80±7.61	-35.20±34.58
Lys 372	-8.85±16.22	-12.44±7.89	-21.29±18.04
Lys 373	-37.89±32.07	-23.95±18.23	-61.85±36.89
Gly 374	-7.77±8.50	-10.55±7.91	-18.32±11.61
Gln 375	-16.09±16.83	-15.67±9.29	-31.76±19.23
Ser 376	-3.22±8.73	-8.19±5.43	-11.41±10.28
Thr 377	-0.13±1.85	-3.11±3.47	-3.24±3.93
Ser 378	-8.82±16.53	-3.13±6.27	-11.95±17.67
Arg 379	-12.61±15.67	-9.23±7.80	-21.84±17.50
His 380	-50.79±30.06	-2.52±16.68	-53.32±34.38
Lys 381	-0.04±1.02	-5.35±5.31	-5.39±5.41
AcLys 382	-12.44±15.22	-32.12±23.78	-44.57±28.23
Leu 383	-1.26±1.57	-31.32±10.54	-32.57±10.66
Met 384	-16.39±10.90	-18.67±5.96	-35.06±12.42
Phe 385	-22.34±11.56	-40.41±13.87	-62.76±18.06
Lys 386	-8.06±12.78	-18.39±6.94	-26.46±14.54

Table S4.8. Energy contributions for the interaction between the p53 CTD and Set9 (PDB:1XQH). Values are in kJ mol⁻¹.

	Electrostatic	van der Waals	Sum
Leu 369	-25.20±12.95	-25.83±12.71	-51.03±18.15
Lys 370	-69.35±25.97	-36.23±11.15	-105.58±28.26
Ser 371	-31.21±38.94	-14.70±11.70	-45.91±40.66
MeLys 372	-80.42±21.17	-44.42±14.03	-124.85±25.40
Lys 373	-32.52±24.56	-23.10±11.21	-55.62±27.00
Gly 374	-20.65±12.60	-18.12±6.51	-38.77±14.18

References

- (1) Levine, A. J.; Finlay, C. A.; Hinds, P. W. P53 is a Tumor Suppressor Gene. *Cell* **2004**, *116*, S67-S69.
- (2) Vogelstein, B.; Lane, D.; Levine, A. J. Surfing the p53 Network. *Nature* **2000**, *408*, 307-310.
- (3) El-Deiry, W. S. Regulation of p53 Downstream Genes. *Semin. Canc. Biol.* **1998**, *8*, 345-357.
- (4) Joerger, A. C.; Fersht, A. R. Structural Biology of the Tumor Suppressor p53. *Annu. Rev. Biochem.* **2008**, *77*, 557-582.
- (5) Harms, K. L.; Chen, X. The C Terminus of p53 Family Proteins Is a Cell Fate Determinant. *Mol. Cell. Biol.* **2005**, *25*, 2014-2030.
- (6) Kim, A. L.; Raffo, A. J.; Brandt-Rauf, P. W.; Pincus, M. R.; Monaco, R.; Abarzua, P.; Fine, R. L. Conformation and Molecular Basis for Induction of Apoptosis by a p53 C-terminal Peptide in Human Cancer Cells. *J. Biol. Chem.* **1999**, *274*, 34924-34931.
- (7) McKinney, K.; Mattia, M.; Gottifredi, V.; Prives, C. p53 Linear Diffusion along DNA Requires Its C Terminus. *Mol. Cell* **2004**, *16*, 413-424.
- (8) Tafvizi, A.; Huang, F.; Leith, J. S.; Fersht, A. R.; Mirny, L. A.; van Oijen, A. M. Tumor Suppressor p53 Slides on DNA with Low Friction and High Stability. *Biophys. J.* **2008**, *95*, L01-L03.
- (9) Luo, J.; Li, M.; Tang, Y.; Laskowska, M.; Roeder, R. G.; Gu, W. Acetylation of p53 Augments its Site-Specific DNA Binding Both *in Vitro* and *in Vivo*. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 2259-2264.
- (10) Friedler, A.; Veprintsev, D. B.; Freund, S. M. V.; von Glos, K. I.; Fersht, A. R. Modulation of Binding of DNA to the C-Terminal Domain of p53 by Acetylation. *Structure* **2005**, *13*, 629-636.
- (11) Riley, K. J.-L.; Ramirez-Alvarado, M.; Maher III, L. J. RNA-p53 Interactions in Vitro. *Biochemistry* **2007**, *46*, 2480-2487.
- (12) Gu, W.; Shi, X.-L.; Roeder, R. G. Synergistic Activation of Transcription by CBP and p53. *Nature* **1997**, *387*, 819-823.

- (13) Espinosa, J. M.; Emerson, B. M. Transcriptional Regulation by p53 through Intrinsic DNA/Chromatin Binding and Site-Directed Cofactor Recruitment. *Mol. Cell* **2001**, *8*, 57-69.
- (14) Mujtaba, S.; He, Y.; Zeng, L.; Yan, S.; Plotnikova, O.; Sachchidanand; Sanchez, R.; Zeleznik-Le, N. J.; Ronai, Z. e.; Zhou, M.-M. Structural Mechanism of the Bromodomain of the Coactivator CBP in p53 Transcriptional Activation. *Mol. Cell* **2004**, *13*, 251-263.
- (15) Fernandez-Fernandez, M. R.; Veprintsev, D. B.; Fersht, A. R. Proteins of the S100 Family Regulate the Oligomerization of p53 Tumor Suppressor. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 4735-4740.
- (16) Fernandez-Fernandez, M. R.; Rutherford, T. J.; Fersht, A. R. Members of the S100 Family Bind p53 in Two Distinct Ways. *Prot. Sci.* **2008**, *17*, 1663-1670.
- (17) Gu, W.; Roeder, R. G. Activation of p53 Sequence-Specific DNA Binding by Acetylation of the p53 C-Terminal Domain. *Cell* **1997**, *90*, 595-606.
- (18) Chernov, M. V.; Bean, L. J. H.; Lerner, N.; Stark, G. R. Regulation of Ubiquitination and Degradation of p53 in Unstressed Cells through C-terminal Phosphorylation. *J. Biol. Chem.* **2001**, *276*, 31819-31824.
- (19) Chuikov, S.; Kurash, J. K.; Wilson, J. R.; Xiao, B.; Justin, N.; Ivanov, G. S.; McKinney, K.; Tempst, P.; Prives, C.; Gamblin, S. J.; Bariev, N. A.; Reinberg, D. Regulation of p53 Activity through Lysine Methylation. *Nature* **2004**, *432*, 353-360.
- (20) Feng, L.; Lin, T.; Uranishi, H.; Gu, W.; Xu, Y. Functional Analysis of the Roles of Posttranslational Modifications at the p53 C Terminus in Regulating p53 Stability and Activity. *Mol. Cell. Biol.* **2005**, *25*, 5389-5395.
- (21) Bode, A. M.; Dong, Z. Post-Translational Modification of p53 in Tumorigenesis. *Nat. Rev. Canc.* **2004**, *4*, 793-805.
- (22) Lavin, M. F.; Gueven, N. The Complexity of p53 Stabilization and Activation. *Cell Death Differ.* **2006**, *13*, 941-950.
- (23) Bell, S.; Klein, C.; Müller, L.; Hansen, S.; Buchner, J. p53 Contains Large Unstructured Regions in its Native State. *J. Mol. Biol.* **2002**, *322*, 917-927.
- (24) Tompa, P. The Interplay between Structure and Function in Intrinsically Unstructured Proteins. *FEBS Lett.* **2005**, *579*, 3346-3354.

- (25) Dyson, H. J.; Wright, P. E. Intrinsically Unstructured Proteins and Their Functions. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 197-208.
- (26) Dunker, A. K.; Cortese, M. S.; Romero, P.; Iakoucheva, L. M.; Uversky, V. N. Flexible Nets. The roles of Intrinsic Disorder in Protein Interaction Networks. *FEBS J.* **2005**, *272*, 5129-5148.
- (27) Iakoucheva, L. M.; Brown, C. J.; Lawson, J. D.; Obradovic, Z.; Dunker, A. K. Intrinsic Disorder in Cell-signaling and Cancer-associated Proteins. *J. Mol. Biol.* **2002**, *323*, 573-584.
- (28) Rustandi, R. R.; Baldisseri, D. M.; Weber, D. J. Structure of the Negative Regulatory Domain of p53 Bound to S100B($\beta\beta$). *Nat. Struct. Biol.* **2000**, *7*, 570-575.
- (29) Avalos, J. L.; Celic, I.; Muhammad, S.; Cosgrove, M. S.; Boeke, J. D.; Wolberger, C. Structure of a Sir2 Enzyme Bound to an Acetylated p53 Peptide. *Mol. Cell* **2002**, *10*, 523-535.
- (30) Lowe, E. D.; Tews, I.; Cheng, K. Y.; Brown, N. R.; Gul, S.; Noble, M. E. M.; Gamblin, S. J.; Johnson, L. N. Specificity Determinants of Recruitment Peptides Bound to Phospho-CDK2/Cyclin A. *Biochemistry* **2002**, *41*, 15625-15634.
- (31) Fuxreiter, M.; Simon, I.; Friedrich, P.; Tompa, P. Preformed Structural Elements Feature in Partner Recognition by Intrinsically Unstructured Proteins. *J. Mol. Biol.* **2004**, *338*, 1015-1026.
- (32) Receveur-Bréchet, V.; Bourhis, J.-M.; Uversky, V. N.; Canard, B.; Longhi, S. Assessing Protein Disorder and Induced Folding. *Proteins* **2006**, *62*, 24-45.
- (33) Dyson, H. J.; Wright, P. E. Coupling of Folding and Binding for Unstructured Proteins. *Curr. Opin. Struct. Biol.* **2002**, *12*, 54-60.
- (34) Chen, J. Intrinsically Disordered p53 Extreme C-Terminus Binds to S100B($\beta\beta$) through "Fly-Casting". *J. Am. Chem. Soc.* **2009**, *131*, 2088-2089.
- (35) Lowry, D. F.; Stancik, A.; Shrestha, R. M.; Daughdrill, G. W. Modeling the Accessible Conformations of the Intrinsically Unstructured Transactivation Domain of p53. *Proteins* **2008**, *71*, 587-598.
- (36) Espinoza-Fonseca, L. M. Leucine-rich Hydrophobic Clusters Promote Folding of the N-terminus of the Intrinsically Disordered Transactivation Domain of p53. *FEBS Lett.* **2009**, *583*, 556-560.

- (37) Taranta, M.; Bizzarri, A. R.; Cannistraro, S. Modeling the Interaction Between the N-terminal Domain of the Tumor Suppressor p53 and Azurin. *J. Mol. Recognit.* **2009**, *22*, 215-222.
- (38) Michel, J.; Harker, E. A.; Tirado-Rives, J.; Jorgensen, W. L.; Schepartz, A. In Silico Improvement of β 3-Peptide Inhibitors of p53•hDM2 and p53•hDMX. *J. Am. Chem. Soc.* **2009**, *131*, 6356-6357.
- (39) Lu, Q.; Tan, Y.-H.; Luo, R. Molecular Dynamics Simulations of p53 DNA-Binding Domain. *J. Phys. Chem. B* **2007**, *111*, 11538-11545.
- (40) De Grandis, V.; Bizzarri, A. R.; Cannistraro, S. Docking Study and Free Energy Simulation of the Complex Between p53 DNA-binding Domain and Azurin. *J. of Mol. Recognit.* **2007**, *20*, 215-226.
- (41) Tomoda, K.; Takahashi, N.; Hibi, Y.; Asamitsu, K.; Ishida, H.; Kondo, T.; Yoshitaka, F.; Okamoto, T. Molecular Docking Analysis of the Protein-protein Interaction between RelA-associated Inhibitor and the Tumor Suppressor Protein p53 and its Inhibitory Effect on p53 Action. *Canc. Sci.* **2008**, *99*, 615-622.
- (42) Madhumalar, A.; Jun, L. H.; Lane, D. P.; Verma, C. S. Dimerization of the Core Domain of the p53 Family. *Cell Cycle* **2009**, *8*, 137-148.
- (43) Khalili, M.; Wales, D. J. Computer Simulations of Peptides from the p53 DNA Binding Domain. *J. Chem. Theory Comput.* **2009**, *5*, 1380-1392.
- (44) Lwin, T. Z.; Durant, J. J.; Bashford, D. A Fluid Salt-bridging Cluster and the Stabilization of p53. *J. Mol. Biol.* **2007**, *373*, 1334-1347.
- (45) Geldon, A.; Mori, M.; Del Conte, R. Theoretical Study on Binding of S100B Protein. *J. Mol. Model.* **2007**, *13*, 1123-1131.
- (46) Withlow, J. L.; Varughese, J. F.; Zhou, Z.; Bartolotti, L. J.; Li, Y. Computational Screening and Design of S100B Ligand to Block S100B-p53 Interaction. *J. Mol. Graph. Model.* **2009**, *27*, 969-977.
- (47) Sachchidanand; Resnick-Silverman, L.; Yan, S.; Mutjaba, S.; Liu, W.-j.; Zeng, L.; Manfredi, J. J.; Zhou, M.-M. Target Structure-Based Discovery of Small Molecules that Block Human p53 and CREB Binding Protein Association. *Chem. Biol.* **2006**, *13*, 81-90.

- (48) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235-242.
- (49) Case, D. A.; Darden, T. A.; Cheatham III, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Crowley, M.; Walker, R. C.; Zhang, W.; Merz, K. M.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Kollman, P. A. *AMBER 10*, University of California, San Francisco, **2008**.
- (50) Eswar, N.; Marti-Renom, M.-A.; Webb, B.; Madhusudhan, M. S.; Eramian, D.; Shen, M.; Pieper, U.; Sali, A. Comparative Protein Structure Modeling With MODELLER. In *Current Protocols in Bioinformatics*, John Wiley & Sons, Inc.: **2006**; Vol. Supp 15, pp 5.6.1-5.6.30.
- (51) Sali, A.; Blundell, T. L. Comparative Protein Modelling by Satisfaction of Spatial Restraints. *J. Mol. Biol.* **1993**, *234*, 779-815.
- (52) Marti-Renom, M. A.; Stuart, A.; Fiser, A.; Sánchez, R.; Melo, F.; Sali, A. Comparative Protein Structure Modeling of Genes and Genomes. *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 291-325.
- (53) Fiser, A.; Do, R. K.; Sali, A. Modeling of Loops in Protein Structures. *Prot. Sci.* **2000**, *9*, 1753-1773.
- (54) Arnold, K.; Bordoli, L.; Kopp, J.; Schwede, T. The SWISS-MODEL Workspace: a Web-based Environment for Protein Structure Homology Modelling. *Bioinformatics* **2006**, *22*, 195-201.
- (55) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435-447.
- (56) Oostenbrink, C.; Villa, A.; Mark, A. E.; van Gunsteren, W. F. A Biomolecular Force Field Based on the Free Enthalpy of Hydration and Solvation: The GROMOS Force-Field Parameter Sets 53A5 and 53A6. *J. Comput. Chem.* **2004**, *25*, 1656-1657.

- (57) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. *Intermolecular Forces, Interaction Models for Water in Relation to Protein Hydration*. D. Reidel Publishing Co.: Dordrecht, The Netherlands, **1981**.
- (58) van der Spoel, D.; Lindahl, E.; Hess, B.; van Buuren, A. R.; Apol, E.; Meulenhobb, P. J.; Tieleman, D. P.; Sijbers, A. L. T. M.; Feenstra, K. A.; van Drunen, R.; Berendsen, H. J. C. *Gromacs User Manual version 4.0*, www.gromacs.org, **2005**.
- (59) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81*, 3684-3690.
- (60) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997**, *18*, 1463-1472.
- (61) Darden, T.; York, D.; Pedersen, L. G. Particle Mesh Ewald: An N-log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089-10092.
- (62) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577-8593.
- (63) Nosé, S. A Molecular Dynamics Method for Simulations in the Canonical Ensemble. *Mol. Phys.* **1984**, *52*, 255-268.
- (64) Hoover, W. G. Canonical Dynamics: Equilibrium Phase-Space Distributions. *Phys. Rev. A* **1985**, *31*, 1695-1697.
- (65) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52*, 7182-7190.
- (66) Nosé, S.; Klein, M. L. Constant Pressure Molecular Dynamics for Molecular Systems. *Mol. Phys.* **1983**, *50*, 1055-1076.
- (67) Virginia Tech Advanced Research Computing: System X, **2007-2010**. <<http://www.arc.vt.edu/arc/SystemX/>>
- (68) Kabsch, W.; Sander, C. Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-Bonded and Geometrical Features. *Biopolymers* **1983**, *22*, 2577-2637.
- (69) Turner, P. J. *Grace*, Center for Coastal and Land-Margin Research Oregon Graduate Institute of Science and Technology, Beaverton, Oregon, **2008**.
- (70) DeLano, W. L. *The PyMOL User's Manual*. DeLano Scientific, Palo Alto, CA, USA, **2002**.

- (71) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera—A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004**, *25*, 1605-1612.
- (72) Abseher, R.; Horstink, L.; Hilbers, C. W.; Nilges, M. Essential Spaced Defined by NMR Structure Ensembles and Molecular Dynamics Simulation Show Significant Overlap. *Proteins* **1998**, *31*, 370-382.
- (73) Philippopoulos, M.; Lim, C. Exploring the Dynamic Information Content of a Protein NMR Structure: Comparison of a Molecular Dynamics Simulation With the NMR and X-Ray Structures of *Escherichia coli* Ribonuclease HI. *Proteins* **1999**, *36*, 87-110.
- (74) Dastidar, S. G.; Lane, D. P.; Verma, C. S. Multiple Peptide Conformations Five Rise to Similar Binding Affinities: Molecular Simulations of p53-MDM2. *J. Am. Chem. Soc.* **2008**, *130*, 13514-13515.
- (75) Dastidar, S. G.; Lane, D. P.; Verma, C. S. Modulation of p53 Binding to MDM2: Computational Studies Reveal Important Roles of Tyr100. *BMC Bioinf.* **2009**, *10* (Suppl 15), S6.
- (76) Dastidar, S. G.; Madhumalar, A.; Fuentes, G.; Lane, D. P.; Verma, C. S. Forces Mediating Protein-Protein Interactions: A Computational Study of p53 “Approaching” MDM2. *Theor. Chem. Acc.* **2010**, *125*, 621-635.
- (77) Shoemaker, B. A.; Portman, J. J.; Wolynes, P. G. Speeding Molecular Recognition by Using the Folding Funnel: The Fly-Casting Mechanism. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 8868-8873.
- (78) Kuroki, R.; Kawakita, S.; Nakamura, H.; Yutani, K. Entropic Stabilization of a Mutant Human Lysozyme Induced by Calcium Binding. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 6803-6807.
- (79) Young, L.; Post, C. B. Catalysis by Entropic Guidance from Enzymes. *Biochemistry* **1996**, *35*, 15129-15133.
- (80) Yon, J. M.; Perahia, D.; Ghélis, C. Conformational Dynamics and Enzyme Activity. *Biochimie* **1998**, *80*, 33-42.

Chapter 5

Using Modeling to Predict Attenuated Mutants of the Viral Protein B2, a Potent Suppressor of RNA Silencing

Michael R. Wiley, William J. Allen, Kevin M. Myles, David R. Bevan, and Zachary N. Adelman

To be submitted.

Attribution

Of the work discussed within this chapter, Wiley performed all of the wet-lab experiments, Allen performed all of the simulations, and Wiley and Allen contributed roughly equally to the writing and preparation of the manuscript. Myles provided guidance in the research. Bevan and Adelman significantly contributed to the editing and review of the manuscript, as well as provided guidance in the research.

Abstract

The protein B2 from the Nodamura virus is a potent suppressor of RNA silencing, a function that allows the virus to proliferate in host cells. In order to study how the protein functions, it is convenient to design mutant proteins that have attenuated or knocked-out ability to bind its target, double-stranded (ds)RNA. Here, we describe computational techniques, including conventional and steered molecular dynamics simulations, that help us to understand how the B2 protein binds dsRNA, and that provide a highly specific predictive insight, resulting in an approach that can decrease time and monetary costs for designing mutants that have desired

effects, and a better understanding of a potent anti-viral defense pathway that has important applications in drug design.

Introduction

RNA interference (RNAi) refers to an RNA-guided gene regulatory process found in most eukaryotic organisms ranging from plants to mammals. In general, intracellular double-stranded RNA (dsRNA) is cleaved into small interfering RNAs (siRNAs) by the RNase III endonuclease, Dicer (1). These siRNAs are loaded into the RNA silencing complex (RISC), made primarily of Argonaute-2 (AGO2), which facilitates sequence-specific silencing of homologous transcripts (2, 3). In plants and invertebrates, this process is an important part of the antiviral response (4).

In response to RNAi, some viruses express suppressors of RNA silencing (SRSs) upon infecting plant and invertebrate hosts (5). SRSs are proteins that block the RNAi pathway by binding to the inherent protein machinery, to long dsRNA, or to siRNAs (5, 6). The most well-characterized SRS identified in an animal virus thus far is the B2 protein of flock house virus (FHV) (6). Another B2 protein of similar function, yet has less than 30% identity, is also present in the distantly related Nodamura virus (NoV) (5). FHV and NoV are members of *Nodaviridae*, a family of positive-sense RNA viruses with bipartite genomes. During replication of RNA1, a subgenomic RNA3 is transcribed containing the open reading frame (ORF) for B2, which overlaps with the viral replicase, protein A. The B2 protein is required for FHV and NoV replication in cells that depend on RNAi as an antiviral response and when components of the RNAi pathway are depleted or knocked down viral RNA accumulation increases (7-10). B2 suppresses the RNAi pathway by binding to both siRNA duplexes and long dsRNAs (11-15). Thus, B2 acts by preventing the protein components of the RNAi pathway access to siRNAs and dsRNAs associated with FHV and NoV infection (6, 12-15).

To better understand how B2 functions, the crystal structure of FHV B2 was solved both bound and unbound to dsRNA (12, 13). The crystal structure of FHV B2 complexed with dsRNA revealed key binding residues that were involved in interacting with dsRNA. Mutational analysis has been conducted on FHV B2 binding residues C44 and R54. FHV B2 C44S and C44A mutants revealed a 100-fold reduce binding affinity to dsRNA (12). Mutational analysis of

FHV B2 R54Q revealed a 20-fold reduction in FHV viral RNA accumulation compared to wild-type virus, as well as the inability of the FHV B2 R54Q to bind long dsRNA and block dicer cleavage (15). Unlike previous experiments that used untranslatable B2 mutants, a direct connection was made between RNA accumulation and the ability of B2 to bind dsRNA. The crystal structure of NoV B2 has also recently been solved and its structure has been superimposed onto FHV B2 revealing potential equivalent binding residues (11). NoV B2 may prove to be of greater importance when compared to FHV B2, as NoV naturally infects mammalian and insect hosts (9, 16, 17).

In order to confirm or refute the importance of the NoV B2 predicted binding residues, as well as identify additional residues that may be important to dsRNA binding, we employed a series of molecular dynamics (MD) techniques (18). Using the wealth of available structural data, we constructed a model of WT NoV B2 in complex with dsRNA, as well as models of several NoV B2 mutants in complex with dsRNA. Previously, Xia *et al.* (19) used steered MD techniques to simulate the unbinding of the SRS p19 from dsRNA, and in doing so they were able to distinguish weakly-binding mutants from the more tightly-binding wild type protein. We used a variant of their technique to perform a similar analysis, and we were ultimately able to create a rank-order of binding affinity for the NoV B2 WT and mutants that we later confirmed experimentally. In addition, we expanded on the project by measuring specific electrostatic and van der Waals interactions between individual NoV B2 residues and the dsRNA, thus identifying the specific residues that contributed most to the binding interaction and creating unique energy of binding profiles for the WT and mutant proteins. These data were applied to the formulation of NoV RNA1 mutants that were tested for self-directed replication in culture cells. Herein, we demonstrate a molecular model for predicting the effects of point mutations in proteins that correlates with *in vivo* experimental results.

Materials and Methods

Computational

Structure and Topology. Crystal structures were obtained from the Protein Data Bank (20) including FHV B2 in complex with dsRNA (PDB ID: 2AZ0) (12), and unbound NoV B2 (PDB ID: 3G80) (11). The backbone atoms of the conserved residues were used to align the NoV B2 protein to FHV B2. The resulting NoV B2 coordinates were combined with the original coordinates of the dsRNA (18 nucleic acids in each chain) from 2AZ0 to make a complete starting structure for our simulations. In addition to the wild type (WT) NoV B2 simulation, parallel systems containing NoV B2 mutants were also simulated; each point mutation was prepared manually. The GROMACS 4.0.7 (21) suite of programs was used for all simulations in conjunction with the ffamber03 all-atom force field (22). Each protein-dsRNA complex was solvated with the TIP3P water model (23) and sufficient ions were added to neutralize the charge of the system and simulate *in vivo* conditions. Each system was minimized using a steepest descent integrator until the maximum force on any atom was less than 1000 kJ mol⁻¹.

Traditional MD Simulations. During all simulations, explicit electrostatic interactions were restricted to 0.8 nm and VDW interactions to 1.4 nm, and the particle mesh Ewald (PME) method (24, 25) was used to approximate long-range interactions. In addition, the linear constraint solver (26) was used to constrain all bond lengths, allowing for a 2 fs timestep. Following minimization, 200 ps of constant volume-temperature (NVT) MD simulations were performed at 301 K using the Berendsen thermostat (27) and with weak position restraints on the protein and dsRNA backbone atoms. The systems were further equilibrated under constant pressure-temperature (NPT) conditions for 200 ps at 1 bar of pressure and the same temperature. During NPT simulations, the Nosé-Hoover thermostat (28, 29) and Parrinello-Rahman barostat (30, 31) were employed. After lifting the position restraints on the protein backbone atoms, 60 ns of production MD was performed on all NoV B2 mutant systems, and 120 ns of production MD was performed on the NoV B2 WT system.

Steered MD Simulations. Constant-velocity steered MD runs were performed on the WT and each mutant system in quintuplicate. Full velocity and coordinate snapshots taken at 200 ps intervals at the end of the production MD served as the starting points for steered MD

simulations. Position restraints were maintained on the dsRNA backbone atoms. An imaginary point was tethered to the center-of-mass of all protein atoms via a spring with a spring constant of $1000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$; that point then traveled directly away from the dsRNA at a constant rate of 0.01 nm ps^{-1} . All simulations were performed on Virginia Tech's SystemX supercomputer (www.arc.vt.edu). Using 24 processors, we were able to collect one nanosecond of data in approximately 5.5 hours of simulation.

Experimental

Cells and Viruses. Baby hamster kidney (BHK- 21) and fruit fly (S2) cells were obtained from ATCC. BHK-21 cells were maintained in DMEM supplemented with penicillin, streptomycin, L-glutamine, and 10% fetal bovine serum at 37°C ; S2 cells were maintained in *Drosophila* media supplemented with penicillin, streptomycin, L-glutamine, and 10% fetal bovine serum at 28°C . The pNoV RNA1 plasmid and mutants were synthesized with a T7 promoter placed at the start of the 5'UTR.

Transfections. NoV RNA1 was *in vitro* transcribed and cleaned using the MEGAclear™ kit (Ambion). One μg of RNA was electroporated in $400 \mu\text{L}$ of BHK-21 (460 V , 725Ω , $75 \mu\text{F}$) or S2 (715 V , 1575Ω , $50 \mu\text{F}$) cells at a concentration of $1 \times 10^7 \text{ cells mL}^{-1}$ using a BTX ECM 630. Electroporated cells were divided between 3 wells of a 6 well plate. Cells were harvested and RNA was extracted using TRI Reagent® at 1, 2, and 3 days post-transfection.

Northern Blots. RNA was analyzed by Northern blot using standard procedures. Probes were generated with the Megaprime™ DNA Labeling System (Amersham) from a fragment spanning the *NotI* and *PflmI* sites of pNoV RNA1. Detection was achieved using a Storm 840 phosphorimager (GE Healthcare). Images were analyzed with ImageQuant software (GE Healthcare).

Results

Energy Profiles for NoV B2–dsRNA Interaction

We constructed a model of NoV B2 in complex with dsRNA by superimposing the crystal structure of the NoV B2 protein from 3G80 over the coordinates of the FHV B2 protein from 2AZ0. The FHV B2 atoms were deleted, leaving the NoV B2 atoms in complex with dsRNA. We simulated WT NoV B2 in complex with dsRNA for 120 ns. The starting configuration of the complex is given in Figure 5.1A. Excluding the first 20 ns of production MD, we measured the average energy of interaction between each NoV B2 residue and the dsRNA, broken down into individual electrostatic and VDW components. Because NoV B2 is a homodimer, the calculated energies were averaged between equivalent residues on opposite chains. Residues that are N-terminal to D36 or C-terminal to A71 were distant enough from the dsRNA during the simulations that only negligible interactions occurred. The results are plotted in Figure 5.2.

The energy profile of WT NoV B2 binding to dsRNA revealed important structural information for studying the interaction between the two biomolecules and for designing attenuated or knockout mutants. Peaks in the energy profile appeared every 3-4 residues, a characteristic that is reflective of the helical conformation of NoV B2 (Figure 5.2). Two residues near the middle of the binding interface, R56 and R59 (Figure 5.1B), exhibited the greatest average potential energy of interaction, -62 ± 20 and -99 ± 26 kJ mol⁻¹ respectively, indicating that the greatest contribution to dsRNA binding is occurring in this part of the protein. However, in their study, Körber *et al.* only identified R59 as an important binding residue, not R56 (11). In addition to R56, we also identified E48, P68, and R69 as residues important to binding dsRNA, although Körber *et al.* did not. Further, Körber *et al.* identified A45 as a residue important to binding dsRNA, for which we found no evidence. Many other residues predicted by Körber *et al.* and predicted by our model were in agreement. These data attest to the limitations of static crystal structures when predicting biophysical properties of a dynamic interaction. In designing an attenuated mutant rather than a knockout mutant it is logical then to begin by modifying those residues that impart a smaller contribution to the total potential of interaction, yet still predicted to be an important binding residue, such as F49.

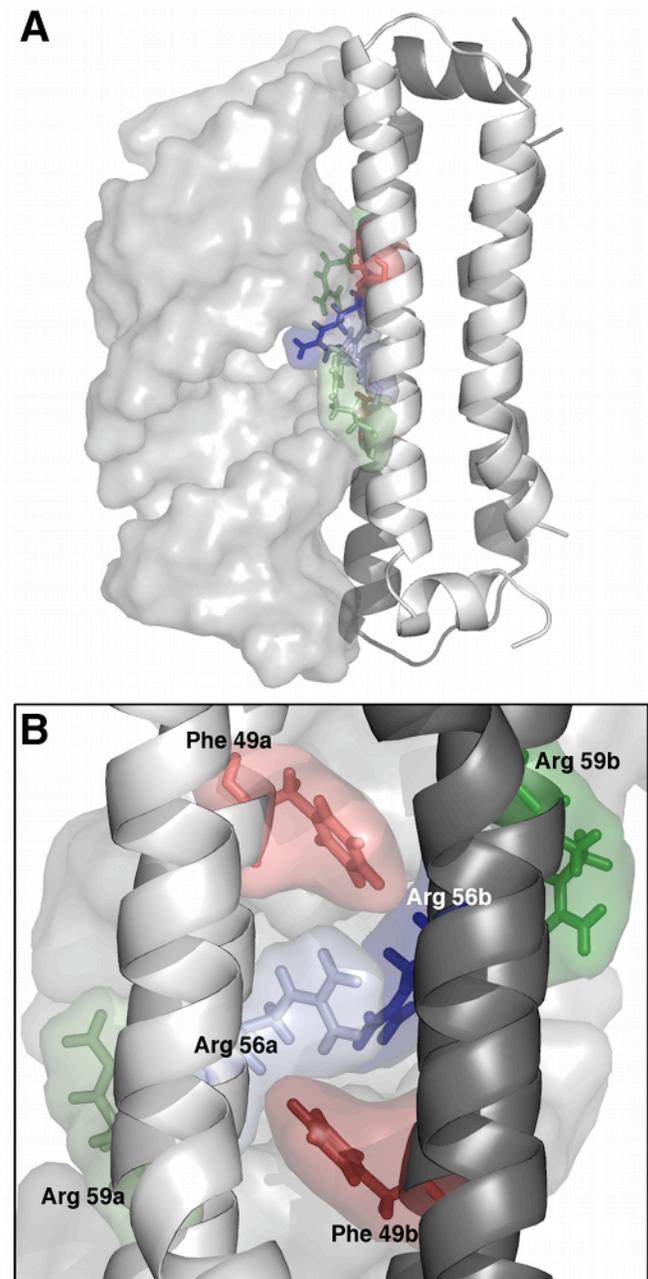


Figure 5.1. Model of NoV B2 WT protein in complex with dsRNA. (A) Protein homodimer is shown as a cartoon, chain A is colored light gray and chain B is colored dark gray. dsRNA is shown as a transparent surface rendering. (B) Alternate view of important residues. R56 and R59 on each chain lay in dsRNA major groove. F49 residues are adjacent to dsRNA backbone.

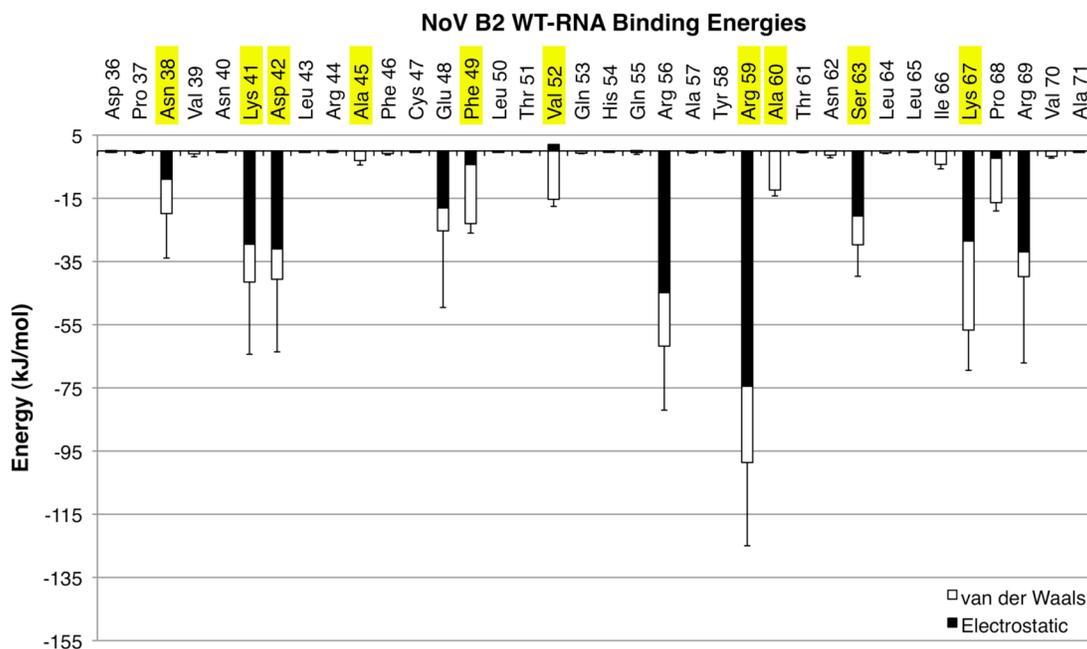


Figure 5.2. Energy profile for NoV B2 WT broken down by individual residue. Electrostatic contributions to dsRNA binding are shown as black bars, VDW contributions are shown as white bars. Highlighted residues are residues that were predicted to be important by Körber *et al.* (11).

We constructed four models of mutant NoV B2 systems: F49C, F49S, F49Y, and an R56L / R59L double mutant. The number of possible NoV B2 mutations was limited by the presence of an overlapping reading frame containing protein A and B1. The mutants that we created only altered the amino acid sequence of the B2 protein – not that of protein A or B1 (Figure 5.3). Each mutant system was simulated for 60 ns, and the last 40 ns of production MD were used to calculate potential energies of interaction as described above.

The simulations of the mutant proteins demonstrated that the addition of single point mutations can cause global changes in how the protein and dsRNA interact. In Figure 5.4, we plot the difference in potential energy of interaction between the NoV B2 mutants and the WT; positive values indicate a *loss* in potential energy of interaction, and negative values indicate a *gain* of potential energy of interaction. For instance, the mutated residue F49C exhibited a small gain in electrostatic interactions and a small loss of VDW interactions with the dsRNA substrate, which is to be expected when mutating from a non-polar to a polar residue. More interestingly, we observed much larger gains and losses among other residues of NoV B2. In the F49C mutant, R69 lost nearly 20 kJ mol⁻¹ potential of interaction, R56 bound more tightly in some regards, and

D42 also lost about 8 kJ mol^{-1} potential of interaction. These results are reflected in the two other F49 mutants. There are changes in the way many residues interacted with the substrate, not just the mutated residue itself. The residues D42, R56, R59 and R69 exhibited the most notable changes. Whether the resulting mutant protein is attenuated or not is yet uncertain, but it is clear that the F49 point mutations changed how the two biomolecules interacted by reducing the magnitude of interaction in some areas while increasing the magnitude of interactions in other areas.

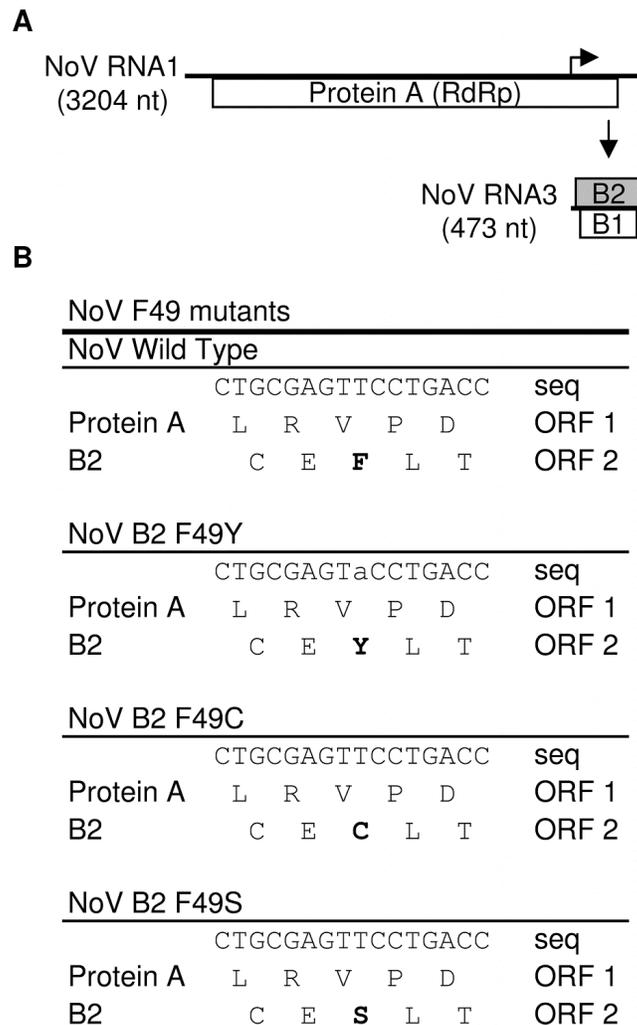


Figure 5.3. NoV F49 mutants change B2 sequence without affecting the protein A/B1 amino acid sequence. (A) NoV RNA1 is shown with the subgenomic promoter depicted by a black arrow. During replication, RNA3 is made and codes for B1 (a protein of unknown function) and B2. Genes in reading frame 1 are white boxes and the gene in reading frame 2 is in a gray box. (B) F49 point mutations made in the NoV RNA1 sequence (nt 2881-2896 are shown). Mutated nucleotides are in lower case and the residues that were changed are in bold.

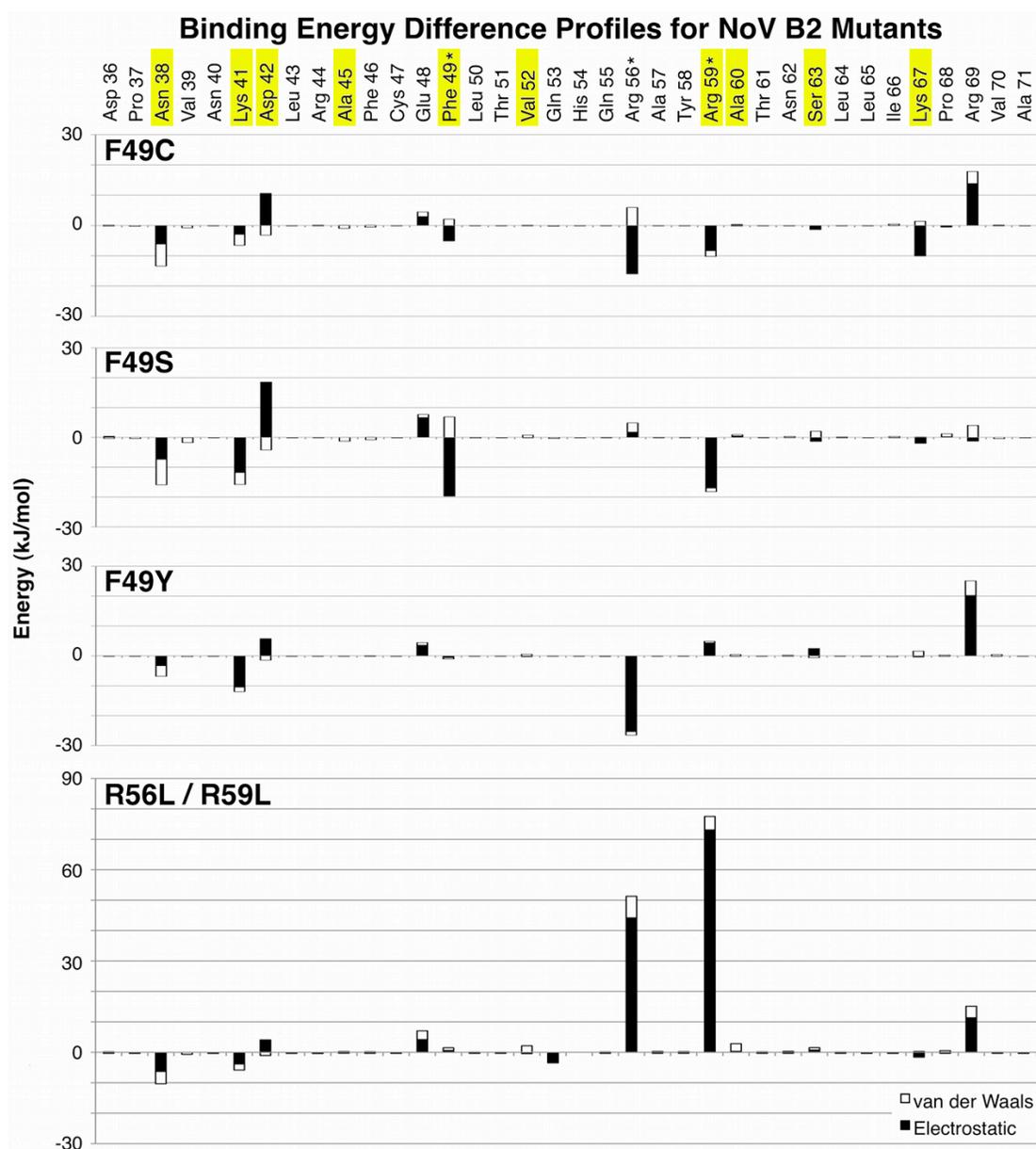


Figure 5.4. Energy *difference* profiles for NoV B2 mutants. Positive values indicate a loss in energy of binding, or a weaker interaction; negative values indicate a gain in energy of binding, or a stronger interaction. Highlighted residues are residues that were predicted to be important by Körber *et al.* (11). Residues marked with asterisks were mutated in this study.

The NoV B2 R56L / R59L double mutant behaved much differently from the F49 mutants. There was a substantial loss of potential of interaction between the two mutated residues, of a combined value of 130 kJ mol^{-1} . Residue N38 experienced a comparatively small gain in potential of interaction, 11 kJ mol^{-1} . These results taken together suggest that, in practice, the

ability of these particular F49 mutants to bind dsRNA may be attenuated, and the R56L / R59L double mutant may be knocked out. Whether this holds true is yet to be seen, but we may say with confidence that the nature of the interaction has changed. The high specificity of these types of interaction leads us to believe that any sort of change in binding energy profiles as we see here could lead to a change in *in vivo* binding affinity.

Steered MD Simulations of NoV B2–dsRNA Complex

At the conclusion of the traditional MD simulations, we took five full-coordinate, full-velocity snapshots of each system in 200 ps intervals (as described in the Materials and Methods) to be used as the starting configurations for the steered MD simulations. In each steered MD simulation, the force on the spring during pulling was plotted by time, and the resulting profiles are plotted in Figure 5.5.

In the case of the WT NoV B2 protein, we observed very little disparity in the pulling resistance from the dsRNA among the replicates. The mutant proteins behaved very similarly to the WT protein during the first 100 – 150 ps of the pulled simulations, as the tension on the spring grew. During the last 100 – 150 ps of the pulled simulations the force curves typically re-converged with the WT protein curves as it required the same amount of force to pull each protein through bulk solvent. During the middle of the simulations, however, especially between 150 – 300 ps, we observed a change in the slope of the force curves when compared to the WT. In most cases, the force curve of mutant NoV B2 proteins decreased in slope more quickly and peaked at a lower overall force value. This indicates that movement of the center of mass of protein atoms in the direction away from the dsRNA occurred sooner for the mutant proteins than it did for the WT protein, thus relieving the tension on the spring more quickly. Physiologically, this is comparable to a weaker affinity of the mutant proteins to binding dsRNA. Among the different mutant simulations, we found that the weaker binding affinity was in part caused by a weaker interaction between the mutant protein and dsRNA, and in some cases, by a destabilization of the 4-helix bundle of the protein structure.

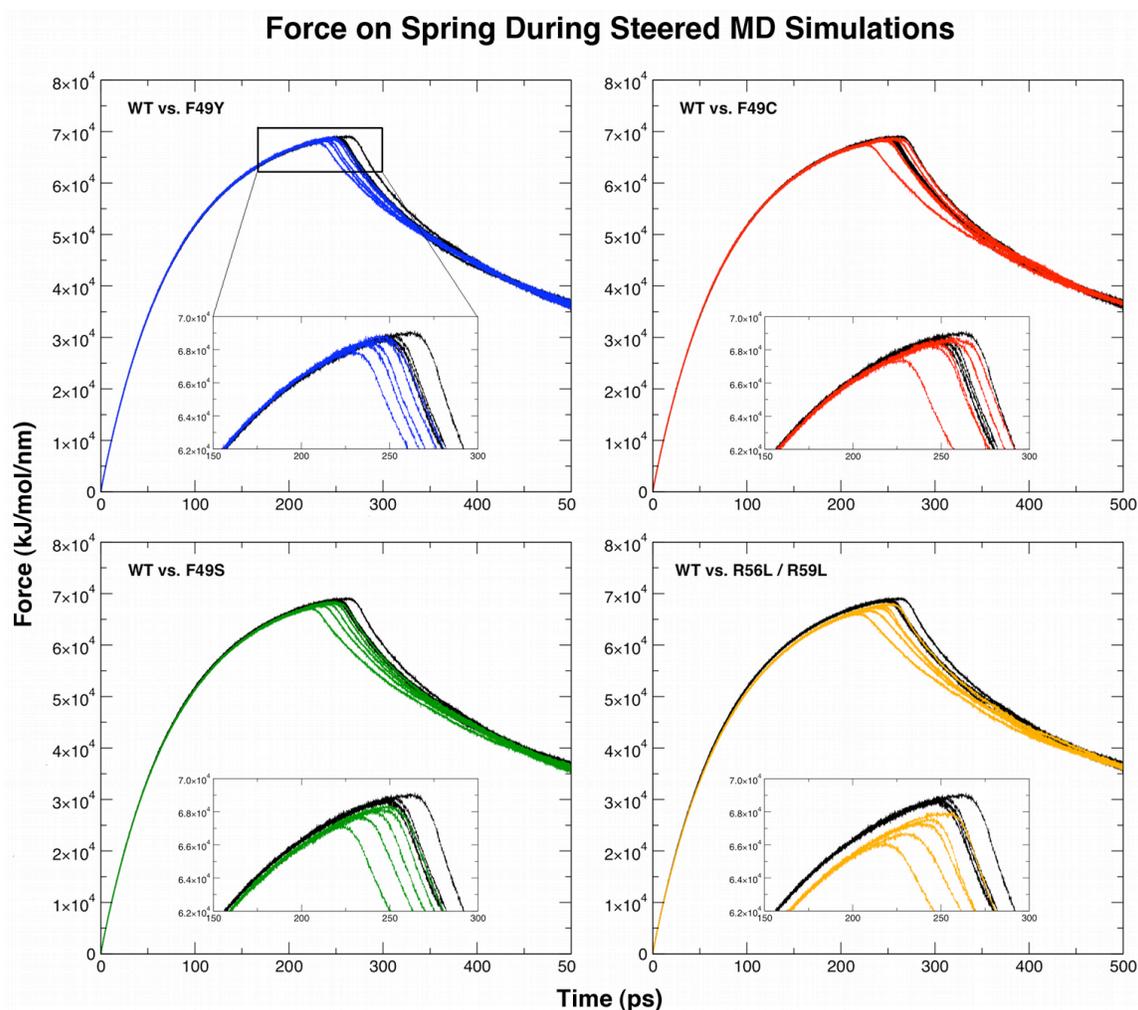


Figure 5.5. Force profiles for steered MD simulations. Each set of black lines are the force profiles for pulling NoV B2 WT protein away from dsRNA. The colored lines are the force profiles for the same pulling procedure of the mutant proteins noted on the plots. The inserts are magnified views of the peak force region.

The peak amount of force required to cause the dissociation of NoV B2 (WT or mutants) from dsRNA is reflective of the degree of attenuation. The average peak force required to pull the WT protein from dsRNA is $68881 \pm 177 \text{ kJ mol}^{-1} \text{ nm}^{-1}$ (or nano-Newtons) (Figure 5.6A). The F49Y and F49C mutants required slightly less peak force (respectively) to pull from the dsRNA, although not significantly less. The F49S mutant required a significantly larger decrease in peak force to pull from the dsRNA when compared to the WT protein according to a Student's t-test ($p < 0.05$). Finally, the R56L / R59L double mutant experienced the greatest decrease in peak force, also a significant decrease by the same test ($p < 0.05$). In analyzing these results, we are able to predict a rank order of binding strength as follows: WT > F49Y > F49C > F49S > R56L /

R59L. We also analyzed the time at which the peak occurred for the WT and each mutant (Figure 5.6B). By that convention we created a rank order of binding strength as: WT > F49C > F49Y > F49S > R56L / R59L; a juxtaposition of the F49Y and F49C residues when compared to the previous rank order. It is important to note that using that measure, no mutants were found to be significantly different.

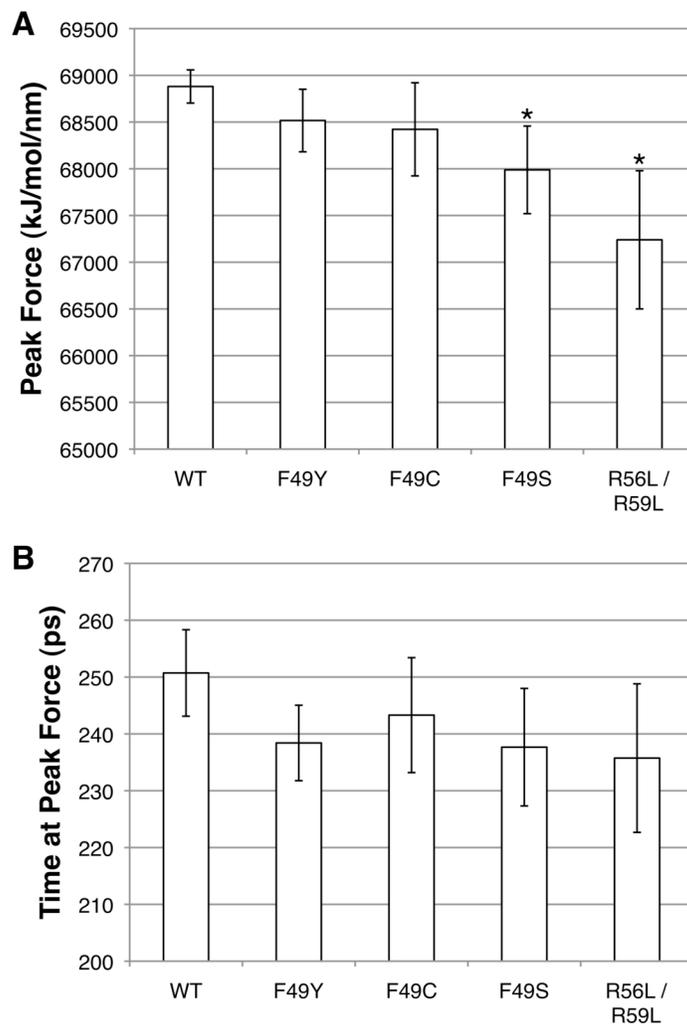


Figure 5.6. (A) Average peak force on spring from pulling simulations. The mutants F49S and R56L / R59L experience a significant decrease in binding affinity. (B) Average time required to reach peak force during pulling simulations.

Self-Directed NoV RNA1 F49 Mutants in Cells

Previous work with a FHV B2 mutant demonstrated that a reduction in viral RNA accumulation coincided with a loss of the mutant's ability to bind long dsRNA (15). In order to validate our predictions, we assessed the ability of NoV RNA1 F49 mutants to self-replicate in cells where functional B2 protein is required for replication (7, 9). To ensure that these mutations were not affecting replication directly, we also transfected NoV RNA1 F49 mutants into cells that did not require B2 to replicate (17). In BHK-21 cells, all of the NoV RNA1 F49 mutants were able to replicate to levels equivalent to wild-type NoV RNA1 (Figure 5.7A). The Δ B2 mutant was able to replicate at every time point but to lower levels when compared to wild type NoV RNA1, which has also been shown previously in mammalian cells.

Drosophila S2 cells were electroporated with WT, Δ B2 (non-translatable mutant), or F49 mutant *in vitro*-transcribed NoV RNA1. Total RNA was analyzed for RNA1 and RNA3 replication at 1, 2, and 3 days post-transfection. The residue to which F49 was mutated altered the ability of RNA1 self-direct replication in S2 cells (Figure 5.7B). The F49Y mutant showed no significant difference between wild type RNA1 and RNA3 replication. The F49S mutant was not able to replicate in S2 cells and exhibited identical replication as the NoV RNA1 mutant that did not translate B2 (Δ B2). The F49C mutant was an intermediate between F49Y and F49S. It was able to replicate, but by three days post-transfection, a significant difference was observed when compared to WT RNA1 accumulation (Student's t-test, $p < 0.05$). Thus we saw a similar ranked order in viral RNA accumulation by three days post infection (dpi): WT = F49Y > F49C > F49S.

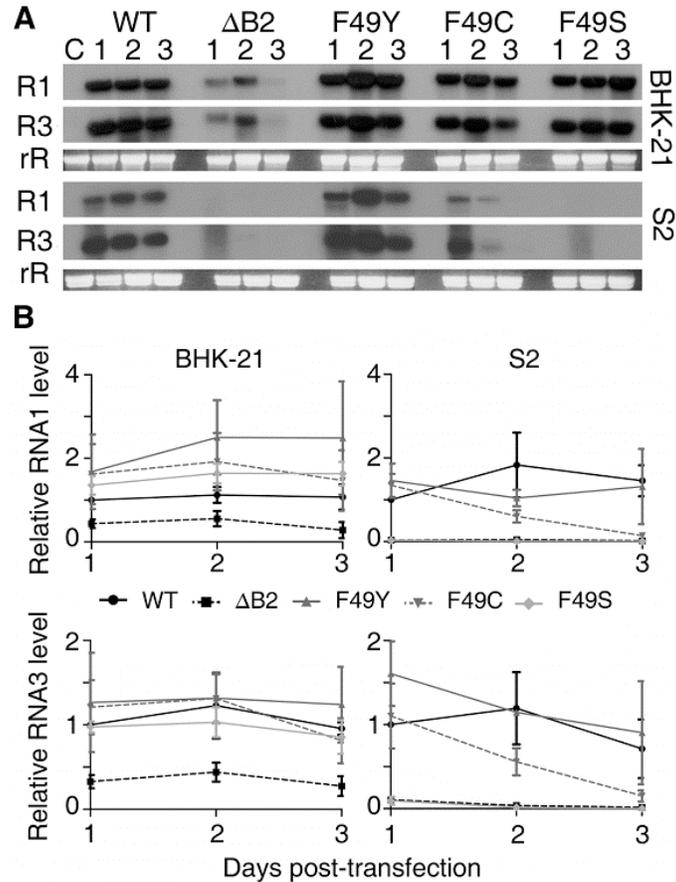


Figure 5.7. NoV F49 B2 mutants affect replication in S2 cells. (A) Detection of RNA1 (R1) and RNA3 (R3) in BHK-21 (top) and S2 (bottom) cells. RNA was extracted at days 1, 2, and 3 post transfection as indicated above the lane. rRNA (rR) indicates equivalent loading of total RNA in each lane. (B) Average relative RNA levels of three biological replicates determined using a phosphorimager. All samples were normalized to the RNA1 band of the wild-type day1 sample.

Discussion

In the previous MD study another SRS, p19, the authors predicted that the mutant protein would lose some of its capacity to bind siRNA and that the tertiary structure would be affected, observations that were confirmed by experiment (19). In our study, we used similar non-equilibrium MD techniques to predict how certain mutations would affect the affinity of NoV B2 to binding dsRNA. However, we expanded the model one step further through careful observations of individual amino acid contributions to the potential of interaction, thus creating an energy profile of protein-dsRNA interaction that has broad applicability in designing

attenuated proteins *in silico*. We found that our computational model of NoV B2 mostly supported the predictions of Körber *et al.*, who first identified putatively important binding residues through aligning the crystal structure of NoV B2 with that of FHV B2, which, in turn, had been crystallized in complex with dsRNA. Further, our model identified additional residues that contributed significantly to dsRNA binding, especially residue R56; residues that were not identified through examination of the crystal structure alone (Figure 5.2). Part of the reason for this deficit of information is likely due to the fact that NoV B2 has not been crystallized in complex with dsRNA, and it is difficult to make conjectures of dynamic interactions through analysis of static structures alone.

We used the energy profile data to try to design a mutant NoV B2 protein that would not have the same capacity for binding dsRNA as does the WT. Knockout mutants have been demonstrated previously with FHV B2, in which the R54 residue was mutated, causing long dsRNA binding to be eliminated (15). From this data we presumed that, in our attempt to design an attenuated mutant, it would be fruitless to mutate residues that had large binding contributions in the energy profile, such as R56 (the predicted equivalent of FHV B2 R54). We chose instead to mutate the NoV B2 F49 residue for several reasons: (i) when mutated, the equivalent residue in FHV B2 resulted in reduced binding affinity to dsRNA, (ii) our model showed that the binding contribution was small, yet significant, and (iii) we were able to mutate the residue to three distinct amino acids without affecting the overlapping reading frame (Figure 5.3). We used constant velocity steered MD to analyze the effects of three F49 mutations, as well as a R56 / R59 double mutation, on dsRNA binding. Using the peak force during the steered MD simulations, we identified a trend in binding strength for the different mutants and created a rank order as follows: WT > F49Y > F49C > F49S > R56L/R59L. We also note that the peak force required is a better rubric than the time required to reach the peak force.

The trend obtained from the steered MD simulations gave us a testable model for studies in a physiologically relevant situation. The RNA accumulation in cells with a function siRNA pathway reflected the rank order that we predicted from our computational data.

This supports our assumption that the F49C and F49S mutants had lower viral RNA accumulation due to the reduced affinity of mutant protein binding to dsRNA, allowing the RNAi pathway to target the dsRNA replicative intermediates produced during viral replication. This was further supported by our BHK-21 experiments. Previous studies showed that at 48 hrs

post-transfection there was no major difference between NoV RNA accumulation for WT B2 and Δ B2 in BHK-21 cells (16). Our results suggested that when RNA accumulation is examined on a time period of 3 days, B2 is involved in replication in BHK-21 cells. The disparity between these findings could be the difference in method of transfection: we directly transfected RNA into cells whereas previous study used DNA constructs and a T7 expressing cell line. Regardless of this discrepancy, our results suggest that replication in BHK-21 cells is not affected by the F49 mutants, revealing that these mutants are not affecting replication of the virus itself, but instead the reduced RNA accumulation in S2 cells must be caused by the targeting of the virus by the RNAi pathway.

The results from the simulations demonstrate that the affinity for NoV B2 to binding dsRNA is integral to the suppression of the RNAi pathway. The small differences observed from the steered MD simulations were paralleled by differences in accumulations of RNA. For example, the F49C mutant did not show significant difference to WT in the steered MD, but did have a significant difference in RNA accumulation. Conversely, we were able to see a significant difference between the F49S mutant and WT in the steered MD simulations as well as in RNA accumulation at every time point in S2 cells. Taken together, these data illustrate the utility of modeling in predicting the effects of specific mutations on attenuated binding capacity.

As researchers try to utilize the mammalian RNAi pathway as a way to naturally defend against viruses, we need to fully understand the different components that are involved and how they function. Using suppressors from pathogens that naturally infect mammalian hosts, like NoV B2, can help in our elucidation and eventually our exploitation of these mechanisms to help fight disease and understand other cellular pathways.

Mutagenesis techniques are commonly used to aid in the determination of protein function. For example, the appropriate mutations at predicted binding or active sites of a protein can reveal important information about protein-substrate interactions. Traditionally, many different mutant forms based on crystallographic or phylogenetic data are synthesized and characterized to determine how each specific mutation affects the protein function (whether the function has been knocked out, attenuated, or unaffected). This approach, however, often results in creating mutations that have no effect on protein function. The ability to predict how a specific mutation or combination of mutations will affect a protein-substrate interaction would save valuable time and resources, notably in the design of attenuated RNA-binding viral proteins.

References

- (1) Bernstein, E.; Caudy, A. A.; Hammond, S. M.; Hannon, G. J. Role for a Bidentate Ribonuclease in the Initiation Step of RNA Interference. *Nature* **2001**, *409*, 363-366.
- (2) Zamore, P. D.; Tuschl, T.; Sharp, P. A.; Bartel, D. P. RNAi: Double-Stranded RNA Directs the ATP-Dependent Cleavage of mRNA at 21 to 23 Nucleotide Intervals. *Cell* **2000**, *101*, 25-33.
- (3) Hammond, S. M.; Bernstein, E.; Beach, D.; Hannon, G. J. An RNA-Directed Nuclease Mediates Post-Transcriptional Gene Silencing in *Drosophila* Cells. *Nature* **2000**, *404*, 293-296.
- (4) Ding, S.-W.; Voinnet, O. Antiviral Immunity Directed by Small RNAs. *Cell* **2007**, *130*, 413-426.
- (5) Li, F.; Ding, S.-W. Virus Counterdefense: Diverse Strategies for Evading the RNA-Silencing Immunity. *Annu. Rev. Microbiol.* **2006**, *60*, 503-531.
- (6) Marques, J. T.; Carthew, R. W. A Call to Arms: Coevolution of Animal Viruses and Host Innate Immune Responses. *Trends Genet.* **2007**, *23*, 359-364.
- (7) Wang, X.-H.; Aliyari, R.; Li, W.-X.; Li, H.-W.; Kim, K.; Carthew, R.; Atkinson, P.; Ding, W. RNA Interference Directs Innate Immunity Against Viruses in Adult *Drosophila*. *Science* **2006**, *312*, 452-454.
- (8) Galiana-Arnoux, D.; Dostert, C.; Schneemann, A.; Hoffmann, J. A.; Imler, J.-L. Essential Function in vivo for Dicer-2 in Host Defense Against RNA Viruses in *Drosophila*. *Nat. Immunol.* **2006**, *7*, 590-597.
- (9) Li, W.-X.; Li, H.; Lu, R.; Li, F.; Dus, M.; Atkinson, P.; Brydon, E. W. A.; Johnson, K. L.; Garcia-Sastre, A.; Ball, L. A.; Palese, P.; Ding, S.-W. Interferon Antagonist Proteins of Influenza and Vaccinia Viruses are Suppressors of RNA Silencing. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 1350-1355.
- (10) Li, H.; Li, W. X.; Ding, S.-W. Induction and Suppression of RNA Silencing by an Animal Virus. *Science* **2002**, *296*, 1319-1321.
- (11) Körber, S.; Shaik Syed Ali, P.; Chen, J. C. H. Structure of the RNA-Binding Domain of Nodamura Virus Protein B2, a Suppressor of RNA Interference. *Biochemistry* **2009**, *48*, 2307-2309.

- (12) Chao, J. A.; Lee, J. H.; Chapados, B. R.; Debler, E. W.; Schneemann, A.; Williamson, J. R. Dual Modes of RNA-Silencing Suppression by Flock House Virus Protein B2. *Nat. Struct. Mol. Biol.* **2005**, *12*, 952-957.
- (13) Lingel, A.; Simon, B.; Izaurralde, E.; Sattler, M. The Structure of the Flock House Virus B2 Protein, a Viral Suppressor of RNA Interference, Shows a Novel Mode of Double-Stranded RNA Recognition. *Euro. Molec. Biol. Org. Rep.* **2005**, *6*, 1149-1154.
- (14) Sullivan, C. S.; Ganem, D. A Virus-Encoded Inhibitor That Blocks RNA Interference in Mammalian Cells. *J. Virol.* **2005**, *79*, 7371-7379.
- (15) Lu, R.; Maduro, M.; Li, F.; Li, H. W.; Broitman-Maduro, G.; Li, W. X.; Ding, S.-W. Animal Virus Replication and RNAi-Mediated Antiviral Silencing in *Caenorhabditis elegans*. *Nature* **2005**, *436*, 1040-1043.
- (16) Johnson, K. L.; Price, B. D.; Ball, L. A. Recovery of Infectivity from cDNA Clones of Nodamura Virus and Identification of Small Nonstructural Proteins. *Virology* **2003**, *305*, 436-451.
- (17) Johnson, K. L.; Price, B. D.; Eckerle, L. D.; Ball, L. A. Nodamura Virus Nonstructural Protein B2 Can Enhance Viral RNA Accumulation in both Mammalian and Insect Cells. *J. Virol.* **2004**, *78*, 6698-6704.
- (18) Izrailev, S.; Stepaniants, S.; Balsera, M.; Oono, Y.; Schulten, K. Molecular Dynamics Study of Unbinding of the Avidin-Biotin Complex. *Biophys. J.* **1997**, *72*, 1568-1581.
- (19) Xia, Z.; Zhu, Z.; Zhu, J.; Zhou, R. Recognition Mechanism of siRNA by Viral p19 Suppressor of RNA Silencing: A Molecular Dynamics Study. *Biophys. J.* **2009**, *96*, 1761-1769.
- (20) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucl. Acids Res.* **2000**, *28*, 235-242.
- (21) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435-447.
- (22) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. A Point-Charge Force Field for

- Molecular Mechanics Simulations of Proteins Based on Condensed-Phase Quantum Mechanical Calculations. *J. Comp. Chem.* **2003**, *24*, 1999-2012.
- (23) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Phys. Chem.* **1993**, *79*, 926-935.
- (24) Darden, T.; York, D.; Pedersen, L. G. Particle mesh Ewald: An N-log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089-10092.
- (25) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577-8593.
- (26) Hess, B. P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 116-122.
- (27) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81*, 3684-3690.
- (28) Nosé, S. A Molecular Dynamics Method for Simulations in the Canonical Ensemble. *Mol. Phys.* **1984**, *52*, 255-268.
- (29) Hoover, W. G. Canonical Dynamics: Equilibrium Phase-Space Distributions. *Phys. Rev. A* **1985**, *31*, 1695-1697.
- (30) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52*, 7182-7190.
- (31) Nosé, S.; Klein, M. L. Constant Pressure Molecular Dynamics for Molecular Systems. *Mol. Phys.* **1983**, *50*, 1055-1076.

Chapter 6

Modeling Loop Dynamics of Antigenic Peanut Protein Orthologs from the Genus *Arachis*

William J. Allen, Sheena A. Friend, Khidir W. Hilu, and David R. Bevan

To be submitted.

Attribution

Of the work discussed within this chapter, Allen performed all of the simulations, analyses, and wrote and prepared the document. Friend generated the homology models, and her dissertation provided the foundation of information that became the introduction of this chapter. Hilu and Bevan significantly contributed to the guidance of this research.

Abstract

The protein Ara h 2 is an antigenic peanut protein that causes a response in people with peanut allergies. Different orthologs of Ara h 2 among the species of peanut that fall under the *Arachis* genus cause different levels of allergic response. The disordered loop that falls between helices $\alpha 2$ and $\alpha 3$ likely contains the epitopes that cause this varying level of response. Here, we use molecular dynamics simulations to study Ara h 2 proteins from eight different species under the genus *Arachis*. Our aim is to correlate some behavior or dynamic property of that specific loop with antigenicity, thus enabling crop producers to grow safer peanuts and improving the overall understanding of *Arachis* phylogeny.

Introduction

The genus *Arachis* describes a group of 69 recognized species of flowering plants that includes the common peanut (1). *Arachis* is native to South America where it is an economically important crop. Peanut allergies, however, make the consumption of most of the species dangerous or even fatal due to a specific peanut allergen known as Ara h 2 (2). Although the 69 species have previously been well documented, not much is understood about the phylogenetic relationship between Ara h 2 proteins in different species of *Arachis*. It seems that Ara h 2 can cause varying levels of allergic response based on the species of origin. The major difference between Ara h 2 proteins of different species falls within one disordered loop between helices $\alpha 2$ and $\alpha 3$. A better understanding of this protein allergen, and an improved overall phylogenetic knowledge of the genus *Arachis*, will aid crop breeders in their attempt to grow a safer, more economically viable peanut.

In this study we used molecular dynamics (MD) simulations to study the behaviors of nine Ara h 2 proteins from eight different species (two different starting models of the protein from the same species were studied). The proteins we studied were from the species *A. batizocoi*, *A. dardani*, *A. duranensis*, *A. glabrata*, *A. hypogaea*, *A. ipaensis*, *A. macedoi*, and *A. triseminata*.

Methods

Initially, nine protein structures were obtained for molecular dynamics (MD) simulation. Herein, different Ara h 2 orthologs will be referred to by the first three or four letters of the species from which they originate. When a protein name is paired with a number, *i.e.* 'bat2', the '2' is in reference to a specific model number for that protein. The models are *bat2*, *dar*, *dur* (more extended loop), *dur2* (more collapsed loop), *glab*, *hypo*, *ipa*, *mace*, and *tris2* (original names of files are listed in Table 6.1). Briefly, each protein consists of five α -helices (denoted $\alpha 1$, $\alpha 2$...) joined by loops of variable size. The N- and C-termini, as well as a large loop connecting $\alpha 2$ and $\alpha 3$, are all disordered regions of the protein. A multiple sequence alignment (ClustalW 2.0.12) was performed, and the results are shown in Figure 6.1. Helix locations (as

calculated by PyMOL) are highlighted in red. It is apparent that the major variation in sequence falls in the large loop connecting $\alpha 2$ and $\alpha 3$.



Figure 6.1. Sequence alignment of different Ara h 2 proteins used in this study. Red residues are part of an α -helix.

All simulations were performed using GROMACS version 4.0.5 (3) in conjunction with the GROMOS 43a1 force field (4). First, each structure was solvated using the simple point charge (SPC) water model in an octahedral box with at least 1.5 nm of space between all protein atoms and the edge of the box. Enough chloride and sodium ions were added to neutralize the total charge of the system, as well as bring the total ionic concentration of the system to 0.1 M. Table 6.1 contains information on specific systems. Each of the nine systems (*bat2*, *dar*, *dur*, *dur2*, *glab*, *hypo*, *ipa*, *mace*, and *tris2*) was energy minimized using a steepest descents integrator until the maximum force on any atom was less than $1.0 \text{ kJ mol}^{-1} \text{ nm}^{-1}$, or until subsequent steps produced less than 1.0 kJ mol^{-1} change in energy.

A 25 ps simulation with a 1 fs timestep was performed on each system with position restraints on all of the protein backbone atoms in order to best equilibrate the water atoms around the sidechains. The linear constraint solver (LINCS) algorithm (5) was used to constrain all bond

lengths. The particle-mesh Ewald (PME) method (6, 7) was used to approximate long-range electrostatics for the system, with a fourth-order spline interpolation and a 0.12 nm Fourier grid spacing. The Berendsen thermostat (8) was applied to the system with a temperature coupling time constant (τ_T) of 0.1 ps at a temperature of 200 K.

After the initial equilibration, each single simulation diverged into three replicate simulations. (In the discussion, the different replicate simulations will be referred to as *bat2-1*, *bat2-2*, *bat2-3*, etc.). The Berendsen thermostat was adjusted to generate new random velocities for each system that approximated a temperature of 310 K. This simulation was carried out for 50 ps with a 2 fs timestep. Another 50 ps simulation was performed after switching to the Nose-Hoover thermostat (9, 10), which generates a more correct distribution of velocities.

Once each system was sufficiently equilibrated around the target temperature, the Parrinello-Rahman barostat (11, 12) was implemented with an isotropic reference pressure of 1 bar and a pressure coupling time constant (τ_p) of 1.0 ps. This simulation was carried out for another 50 ps with a 2 fs timestep.

Finally, the position restraints on the protein backbone atoms were lifted and each system was simulated for 60 ns, for a total simulation time of 1.44 μ s. All analyses were performed using GROMACS tools. Clusters were formed by combining the trajectories of replicate simulations into a single trajectory, followed by clustering with the GROMOS algorithm with a 0.12 nm cut-off. Data was visualized using Grace, and structure images were generated using PyMOL (13).

Table 6.1: System-specific simulation conditions.

	<i>bat2</i>	<i>glab</i>	<i>dar</i>
Original Filename	bat2_104-108.pdb	glab_127-129.pdb	dar_final.pdb
# Residues	165	165	155
# Protein Atoms	1,750	1,750	1,630
# Water Residues	24,630	24,642	24,674
# Na+ Atoms	48	48	48
# Cl- Atoms	46	46	46
# Total Atoms	75,734	75,770	75,815
	<i>dur</i>	<i>dur2</i>	<i>mace</i>
Original Filename	dur_122-124.pdb	A_duranensis.B99990005.pdb	mace_final.pdb
# Residues	160	160	156
# Protein Atoms	1,688	1,688	1,685
# Water Residues	33,070	24,659	24,674
# Na+ Atoms	64	48	47
# Cl- Atoms	62	46	46
# Total Atoms	101,024	75,759	75,800
	<i>ipa</i>	<i>tris2</i>	<i>hypo</i>
Original Filename	ipa134-136.pdb	tris2_54-67_67-80.pdb	1W2Q.pdb
# Residues	172	158	127
# Protein Atoms	1,812	1,702	1,324
# Water Residues	24,610	24,676	24,861
# Na+ Atoms	48	46	49
# Cl- Atoms	46	47	46
# Total Atoms	75,736	75,823	76,002

Results and Discussion

Part I: RMSD Analysis

In order to verify sufficient sampling of each simulation, we measured the backbone root mean square deviation (RMSD) of each protein over the course of 60 ns. The results are shown in Figure 6.2. At first glance, all three replicates of *bat2*, *glab*, *dar*, *tris2*, and *hypo* equilibrate fairly quickly (within the first 5-10 ns), to about 1 nm of backbone deviation, a reasonable value for proteins this size. These systems remained this way for the duration of the simulations.

Four sets of proteins, however, required further investigation: *mace*, *ipa*, *dur*, and *dur2*. The replicate *mace-1* seems to equilibrate fairly quickly to a value that is similar to the proteins mentioned in the previous paragraph. *Mace-2* and *mace-3*, however each spike between 10 and 20 ns to a slightly higher value of about 1.25 nm. This amount of RMSD is still within a reasonable range, especially considering that there are large unstructured regions in the protein. After these replicates spike, they do level off, which is encouraging. The replicates *ipa-1*, *ipa-2*, and *ipa-3* behave very similar to *mace*. They equilibrate to somewhere between 1 and 1.25 nm, which are still reasonable values for proteins this size.

The three replicates of *dur* and *dur2* seem to be experiencing the most backbone deviation. At first, *dur-1* seems to level off between 1 and 1.25 nm, but then undergoes some conformational shift that causes a spike in RMSD at 25 ns. It appears that the shift in structure has moved the protein into a fairly stable conformation. *Dur-2* has a lower RMSD overall, but the value fluctuates greatly, indicating that the protein cannot find a stable conformation (quite the opposite of what happens in *dur-1*). The final replicate, *dur-3*, does equilibrate at approximately 1.8 nm backbone deviation, a value that is approaching unreasonably high, indicating that the structure has changed significantly from the starting structure. The *dur2* replicates, despite the starting conformation being slightly more collapsed than *dur*, also exhibit high RMSD values. The replicate *dur2-1* takes nearly a significantly long time (30 ns) to reach a very high 1.7 nm equilibrium, *dur2-2* actually looks okay, despite the RMSD value being on the high-end of acceptable, and *dur2-3* also happened to equilibrate a little high, at 1.5 nm. If we only consider the three *dur* replicates, the RMSD data would indicate that the *dur* starting structure was not very stable, or not a biologically relevant conformation to begin with. However, because two different *dur* starting structures (*dur* and *dur2*) both display unusual RMSD values, we might conclude that the increased fluctuation we observe is indeed biologically relevant. Further analyses will need to be performed to confirm this.

In summation, all 27 replicates are likely equilibrated enough for analysis, with the possible exception of the *dur* or *dur2* replicates. However, there is no indication that further simulation of the *dur* or *dur2* replicates will lead to better equilibration.

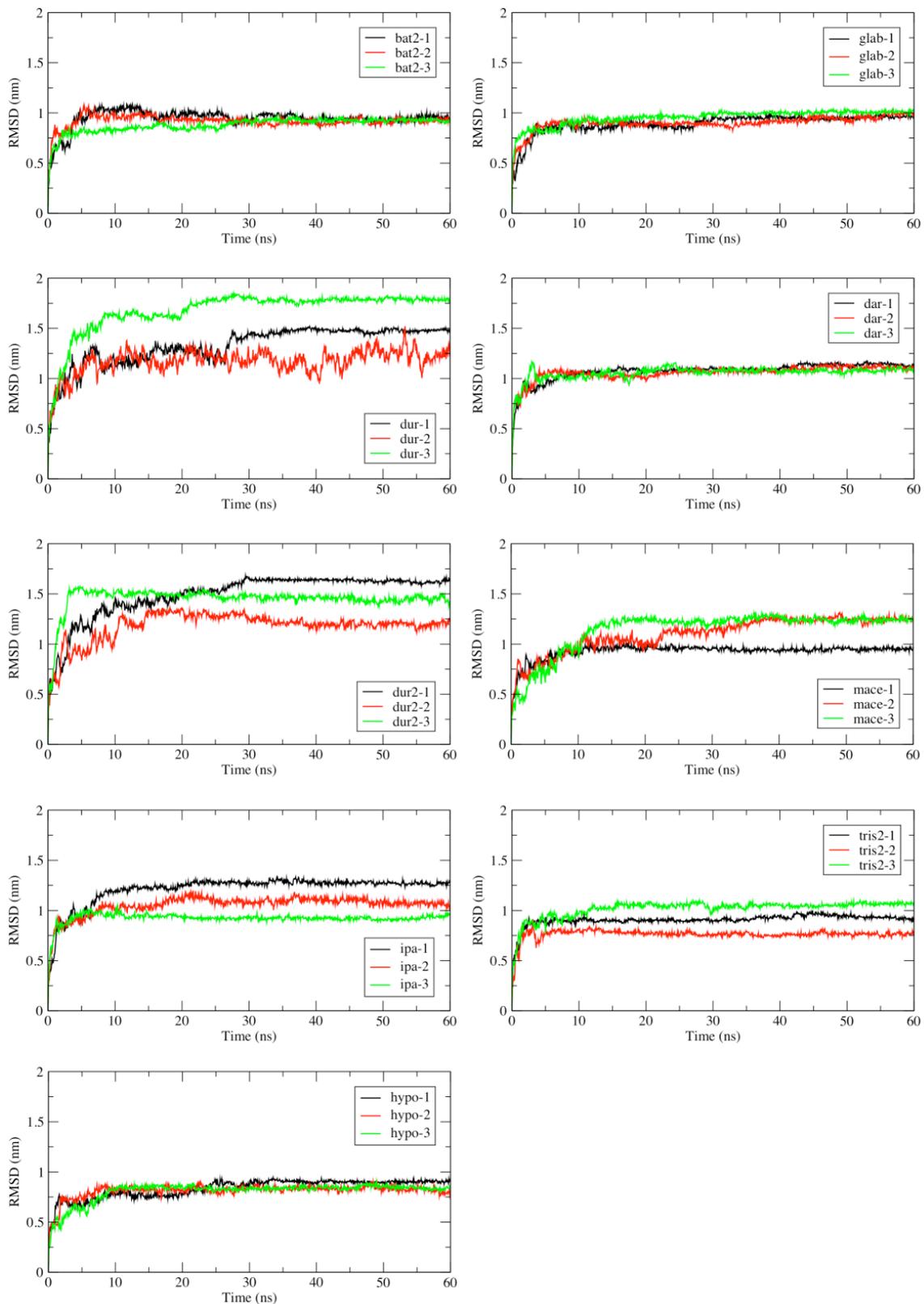


Figure 6.2. RMSD values measuring backbone deviation. All three replicates of each protein simulation are overlaid as different colored lines.

Part II: Radius of Gyration (R_g)

The radius of gyration (R_g) is measured as:

$$R_g = \left(\frac{\sum_i \|r_i\|^2 m_i}{\sum_i m_i} \right)^{\frac{1}{2}}$$

Where r_i is the position of atom i with respect to the center of mass of the protein, and m_i is the mass of atom i . In short, it is a measure of how folded or unfolded a protein is. A protein with a very low R_g is more folded, as in a spherical, globular protein, and a protein with a very high R_g is more unfolded, as in an unfolded, disordered protein. In the case of our eight proteins, their starting conformations reflect some ‘unfolded’ characteristics, especially in the disordered regions (the N-terminus, the C-terminus, and the loop between $\alpha 2$ and $\alpha 3$). We expect that as we perform MD simulation, the protein should become slightly more globular in the disordered regions in order to preclude water from hydrophobic residues. What we would see then, is a trend toward decrease in R_g that should level off over the course of the simulation. This, conveniently, can also be used as a measure of how well each replicate has equilibrated, much like the RMSD values from the *Part I* of the results.

In Figure 6.3, plots of R_g over the course of 60 ns are shown for each of the 27 replicates. It appears that the proteins *dar*, *tris2*, and *hypo* each reach a more globular, folded state quickly in the simulations that have a fairly low R_g value. *Dur2* also reaches a folded state with low R_g values, although in two of the replicates (*dur2-1* and *dur2-2*), it takes about 20 ns to reach that point. *Bat2*, *glab*, and *ipa* also reach a globular state quickly (with the exception of *bat2-1*), at just a slightly higher R_g . This could be due to the fact that the size of the disordered loop between $\alpha 2$ and $\alpha 3$ is much larger in these three proteins.

Two of the same proteins that display equilibration issues in *Part I* (*dur* and *mace*) appear to have some folding problems. *Mace-2* and *mace-3* reach a globular state with reasonable R_g values within the first 20 ns, about 10 to 15 ns longer than it took for *bat2*, *dar*, *glab*, and *tris2* to reach the same state. The other replicate, *mace-1*, seems to trend toward more folded, but then reverses and becomes more unfolded around the 30 ns mark. Referring back to Figure 6.2, *mace-*

1 seemed to have the best RMSD of all of the replicates, so this is quite counter-intuitive and difficult to explain. A closer look at the trajectory reveals that the N-terminal disordered region in the *mace-1* replicate undergoes a small transition at around the 30 ns mark, which could affect the R_g .

The *dur* replicates *dur-1* and *dur-3* each reach a low R_g value by about 30 ns. The R_g of *dur-2*, however, remains high and fluctuates considerably throughout the simulation. This indicates (in concert from the RMSD results of Figure 6.2), that this particular replicate is not folded well, and likely the *dur* starting structure was not very good. However, the fact that two of the *dur* replicates (*dur-1* and *dur-3*) reach the same R_g value as all three of the *dur2* replicates (between 1.5 and 1.6 nm), suggests that the starting structure may actually not entirely influence the ensuing dynamics simulation. Previously, we had believed that we may be observing stronger fluctuations in the *dur* protein because the loop between $\alpha 2$ and $\alpha 3$ started in an extended position. Now, it seems as though the extended structure (*dur*) and the collapsed structure (*dur2*) can reach the same conformation during production dynamics anyway, implying that the large fluctuations are not an artifact of the starting structure, but rather they could be of biological importance.

Whatever the case, it seems that, as was the conclusion from the RMSD data, all 27 replicates are probably ready for analysis. Further simulation could lead to the convergence of some of the different replicates, including *dur-2*, but it is impossible to predict how long that transition would take, or whether it would happen at all.

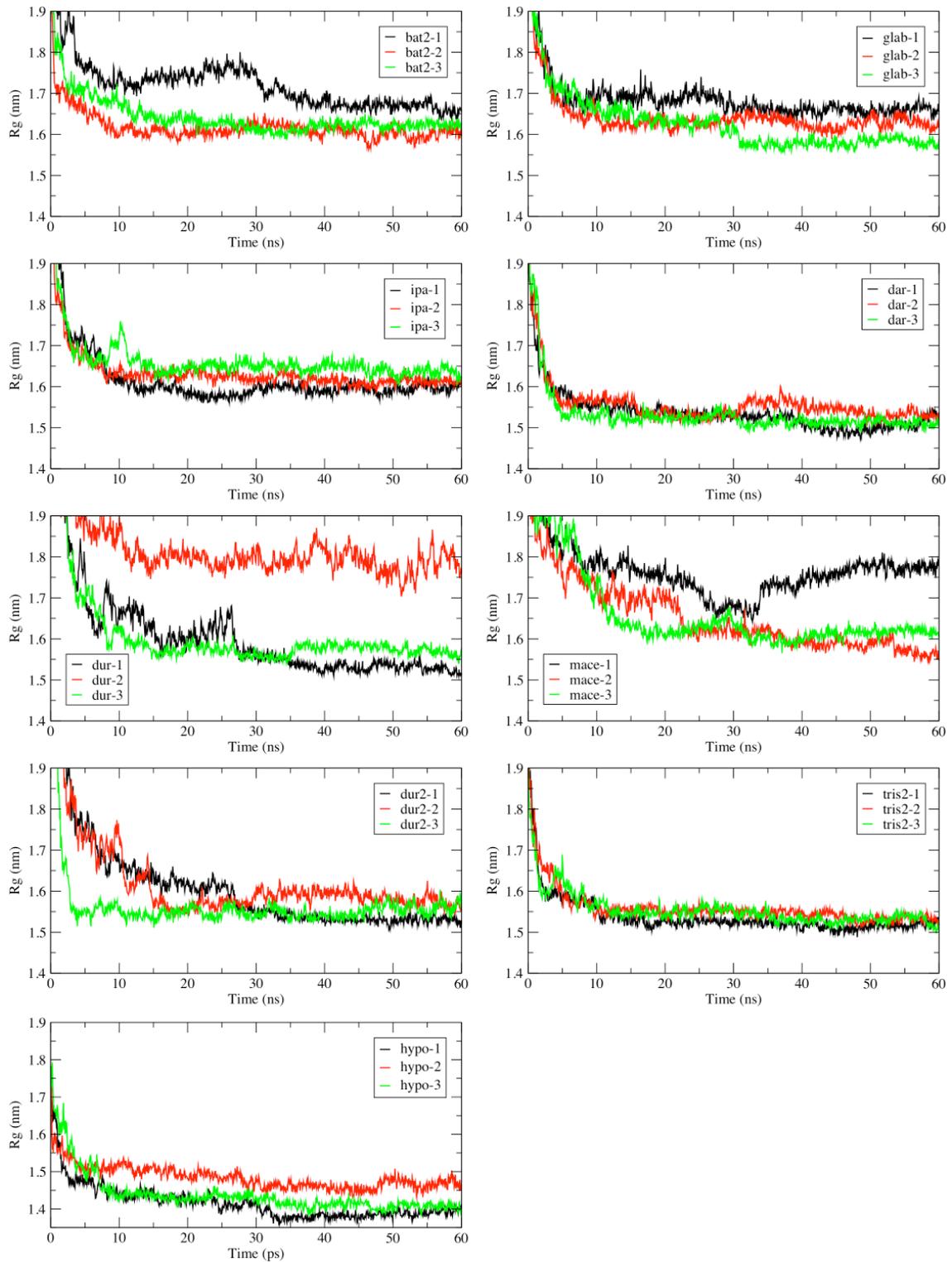


Figure 6.3. Radius of gyration, or ‘foldedness’ of protein backbone. All three replicates of each protein simulation are overlaid as different colored lines.

Part III: RMSF Analysis

Next, once we determined that we have sufficient sampling for at least most of the replicates, we examined root mean square fluctuation (RMSF) in the protein side chains. Briefly, RMSF is a measure of how much a protein side chain oscillates or fluctuates over the course of a simulation. This is similar to the ‘ β -factor’ or ‘temperature factor’ reported in PDB files as a result of x-ray crystallography. Regions with very high fluctuation often are regions of importance to the protein, for example it may be a site that a substrate binds, or a site of interaction between two proteins or protein and DNA.

The initial folding process of the large loop structures in each of the proteins influenced the total RMSF in an artificial way. It is more reasonable to measure RMSF in the protein only after it is equilibrated, i.e., the last 20 ns of simulation. Figure 6.4 illustrates this concept. It shows the RMSF of *bat2-1* during the entire 60 ns simulation, during the last 40 ns of simulation, and during the last 20 ns of simulation. It is plain to see that the total RMSF was, on average, much higher across virtually every residue if the first 40 ns of simulation are considered. When the RMSF was only measured over the last 20 ns, the values tended to be lower in most areas, but stay peaked in some, presumably important areas (residues 66-75, for example).

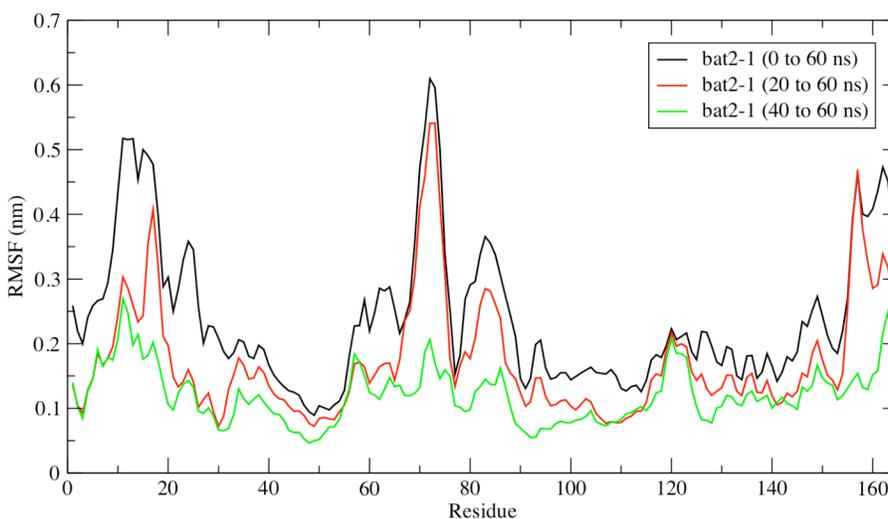


Figure 6.4. RMSF analysis of *bat2-1* across all 60 ns of simulation (black line), during the last 40 ns of simulation (red line), and during the last 20 ns of simulation (green line).

The RMSF measurements tended to change slightly among the different replicate simulations of a single protein. Figure 6.5 contains RMSF data for all 27 replicate simulations. At first glance, some of the RMSF peaks do not seem to line up in the different replicate simulations. However, it is easy to identify stretches that exhibit higher fluctuation across all replicate simulations, or adjacent sections of fluctuation in different replicate simulations.

A simple analysis at Figure 6.5 is actually very informative. In general, there are four areas across all eight proteins that tend to be high in fluctuation. Three of these areas are the N-terminus, the C-terminus, and the large loop between $\alpha 2$ and $\alpha 3$, which is expected because these are all disordered regions. Another peak consistently shows up among all six proteins in the range of residues 115-130 (depending on the length of the protein). This peak is centered about the short loop connecting $\alpha 4$ and $\alpha 5$, which indicates that this may be another functional area of the protein that is worth further investigation. Figure 6.6 contains a structure of *bat2* with the regions of fluctuation highlighted from Figure 6.5.

Referring to the *dur* panel in Figure 6.5, it is plain to see that the second replicate, *dur-2* exhibits a very high degree of fluctuation (as high as 0.8 nm). This is in direct correspondence with the RMSD and R_g data shown earlier for that replicate. What is interesting, though, is that the largest peaks are still in functionally relevant areas – the N-terminus, the loop between $\alpha 2$ and $\alpha 3$, and the loop between $\alpha 4$ and $\alpha 5$. This could suggest that the increased fluctuation may not be an artifact of the simulation or of the starting structure, but in fact a functionally relevant fluctuation.

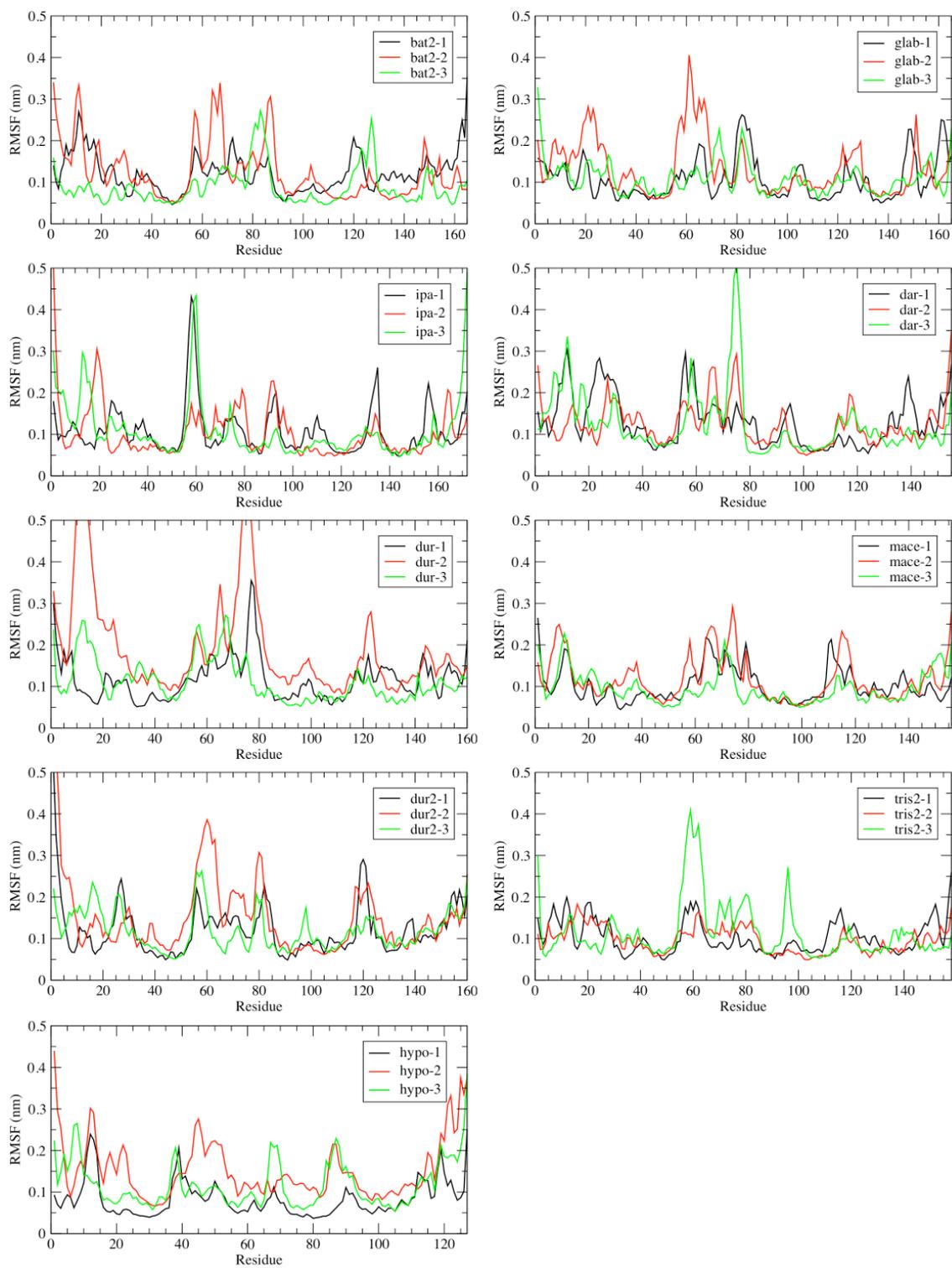


Figure 6.5. RMSF values measuring sidechain fluctuation for the last 20 ns of simulation. All three replicates of each protein simulation are overlaid as different colored lines.

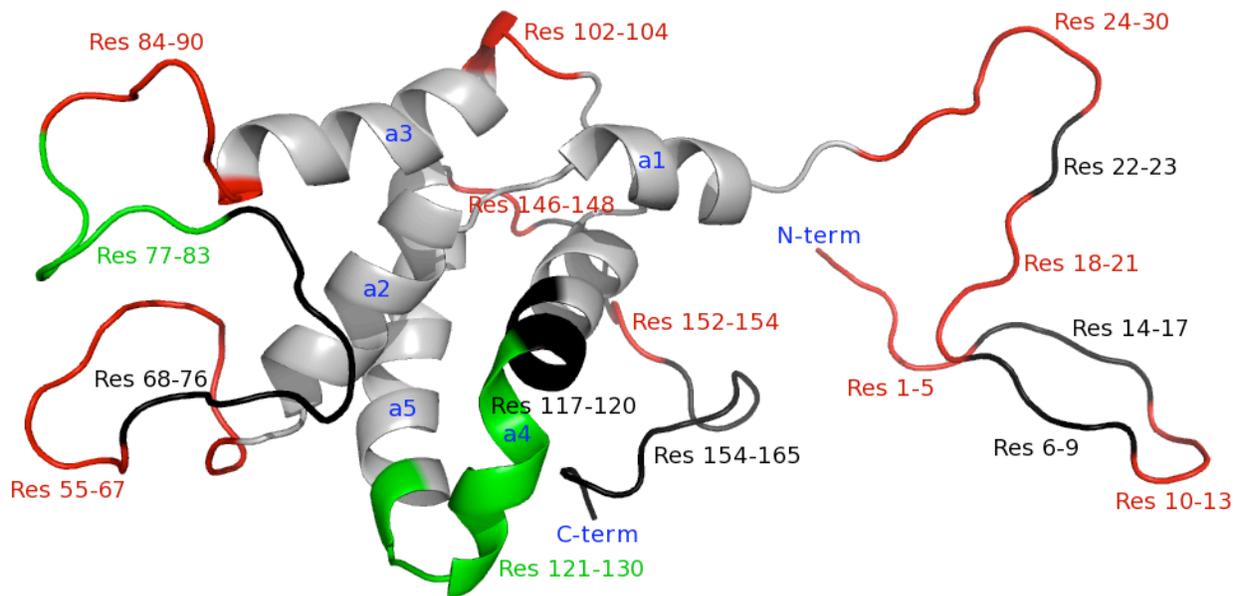


Figure 6.6. Initial conformation of *bat2* protein. Peaks from Figure 6.5 are highlighted.

Part IV: Clustering Analysis

Clustering can give us an idea of what structural conformations appear more frequently than others. For this analysis, the replicate trajectories were concatenated and clustered as described in the Methods section. In all, 3,001 frames were considered for each protein. Some pertinent data on the resulting clusters is listed in Table 6.2. Figure 6.7 shows the ‘most average’ structures from the top 4 clusters of *bat2*. In other words, the structures shown in Figure 6.7 are what we hypothesize to be the most likely conformation *bat2* adopts *in vivo*.

The clustering results from *bat2* and the eight other proteins clearly show that a collapse of the N-terminal unstructured region is favored for all proteins; in addition, the N-terminal region seems to sporadically form some secondary structural elements. Another notable observation when looking through the clustering results is that the 5 α -helices in each of the original starting structures do not necessarily stay intact, especially $\alpha 3$. This raises an interesting question about the secondary structure behavior of the protein during simulation. Regions that tend to freely change secondary structure content can often infer protein function. It is difficult to observe

changes in protein structure by looking at a finite number of cluster members, so the secondary structural elements will be explored further in the next section.

Table 6.2. Cluster results.

	<i>bat2</i>	<i>glab</i>	<i>dar</i>
# Clusters found	544	518	438
# Clusters with ≥ 20 members	31	28	42
# Members in largest cluster	238	232	138
Time step (ps)	176,460	160,860	26,700
	<i>dur</i>	<i>dur2</i>	<i>mace</i>
# Clusters found	1,353	797	645
# Clusters with ≥ 20 members	17	23	27
# Members in largest cluster	127	268	211
Time step (ps)	163,380	43,620	173,580
	<i>ipa</i>	<i>tris2</i>	<i>hypo</i>
# Clusters found	555	404	455
# Clusters with ≥ 20 members	32	33	30
# Members in largest cluster	305	232	322
Time step (ps)	144,120	97,800	40620

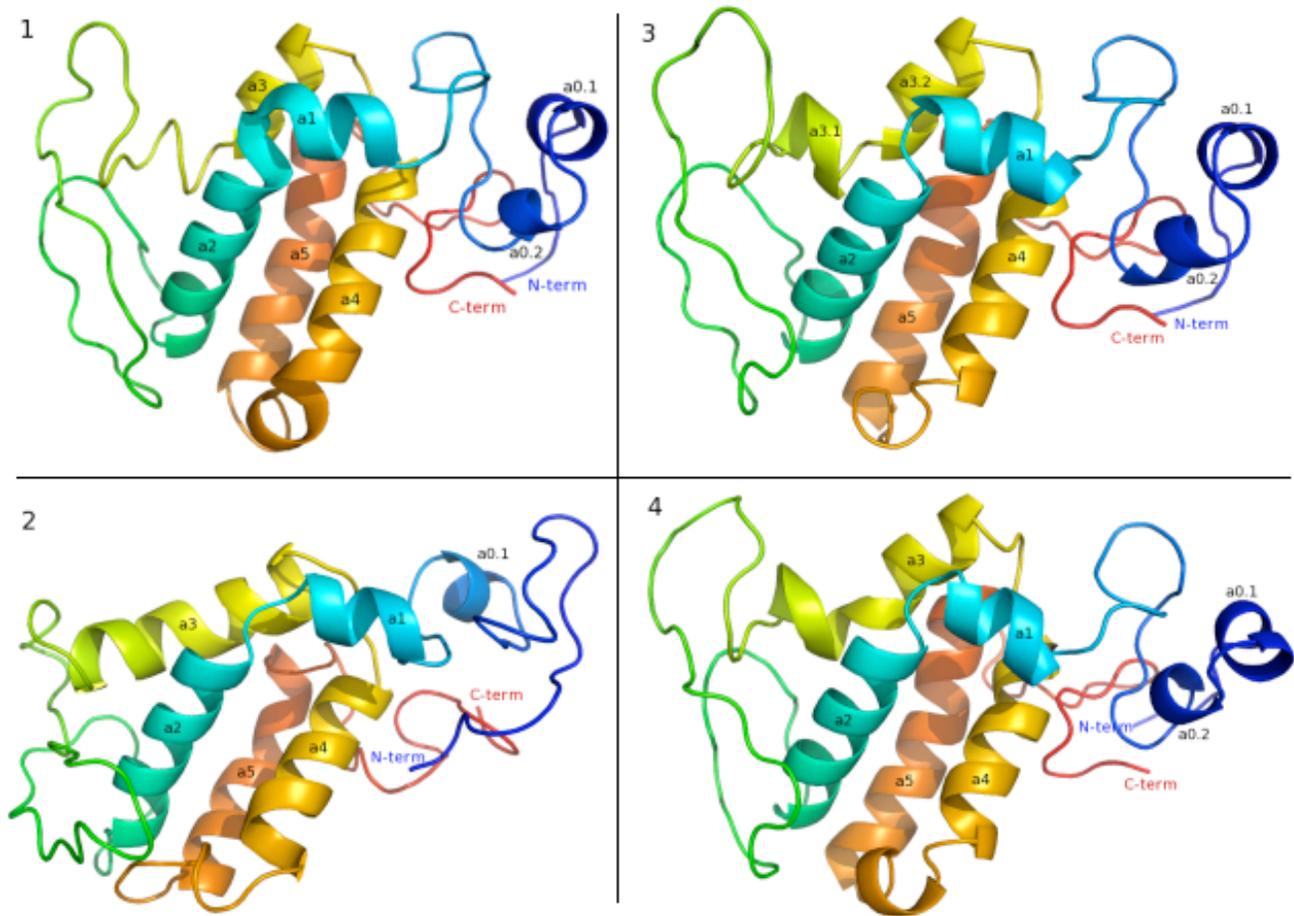


Figure 6.7. Top members of each of the 4 largest *bat2* clusters. Proteins are colored as rainbows, starting with blue at the N-terminus, and ending with red at the C-terminus. Notice the variability in length of $\alpha 3$, and the appearance of extra helices in the N-terminal disordered region.

Part V: DSSP Analysis

The “define secondary structure of proteins” (DSSP) algorithm was implemented to monitor the secondary structure content of each replicate system over the 60 ns of simulation. This algorithm plots the secondary structure content (i.e. α -helix, β -strand, random coil) of each individual residue as a function of time. The results are compiled in Figure 6.8 – Figure 6.16. The locations of the original 5 helices are summarized in Table 6.3.

To summarize the results from the DSSP analysis, we see two main recurring themes. First, $\alpha 3$ seems to be the least stable helix, undergoing complete destabilization in many of the replicates. The second major theme is the formation of secondary structure elements in the N-terminal region, especially short β -strands and short helices. This behavior is not observed as frequently in the disordered C-terminus or in the disordered loop between $\alpha 2$ and $\alpha 3$.

Table 6.3. Helix locations in starting structures (as determined by PyMOL).

	<i>bat2</i>	<i>glab</i>	<i>dar</i>
$\alpha 1$	33-40	33-40	33-40
$\alpha 2$	43-53	43-53	43-53
$\alpha 3$	90-103	89-103	79-94
$\alpha 4$	110-127	110-127	98-117
$\alpha 5$	129-145	131-145	119-135
	<i>dur</i>	<i>dur2</i>	<i>mace</i>
$\alpha 1$	33-40	33-40	33-40
$\alpha 2$	43-53	43-54	43-54
$\alpha 3$	84-98	84-99	81-95
$\alpha 4$	105-122	103-122	99-118
$\alpha 5$	124-140	124-140	120-136
	<i>ipa</i>	<i>tris2</i>	<i>hypo</i>
$\alpha 1$	33-39	33-40	15-21
$\alpha 2$	43-53	43-53	25-38
$\alpha 3$	96-110	82-97	52-67
$\alpha 4$	117-134	101-120	73-85
$\alpha 5$	138-152	122-138	94-110

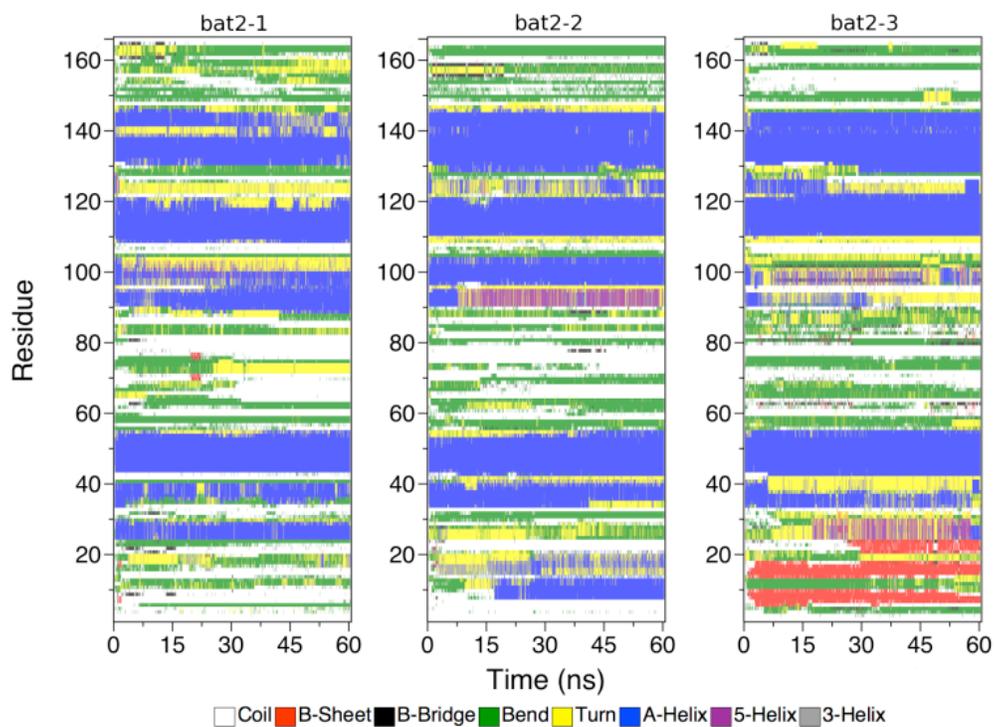


Figure 6.8. Secondary structure content of *bat2* as a function of time for each replicate.

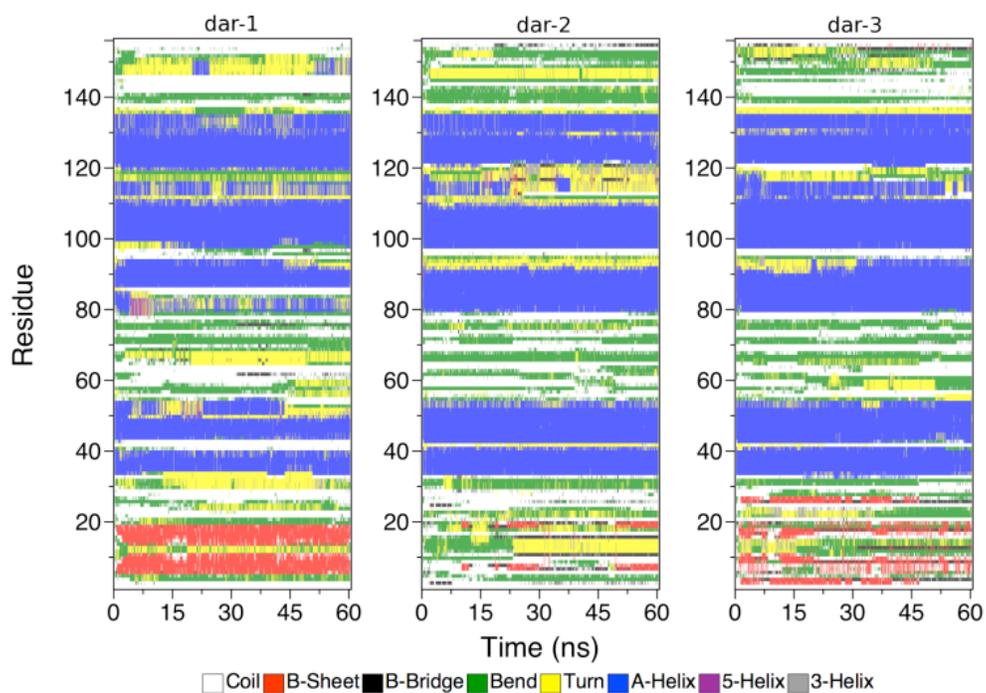


Figure 6.9. Secondary structure content of *dar* as a function of time for each replicate.

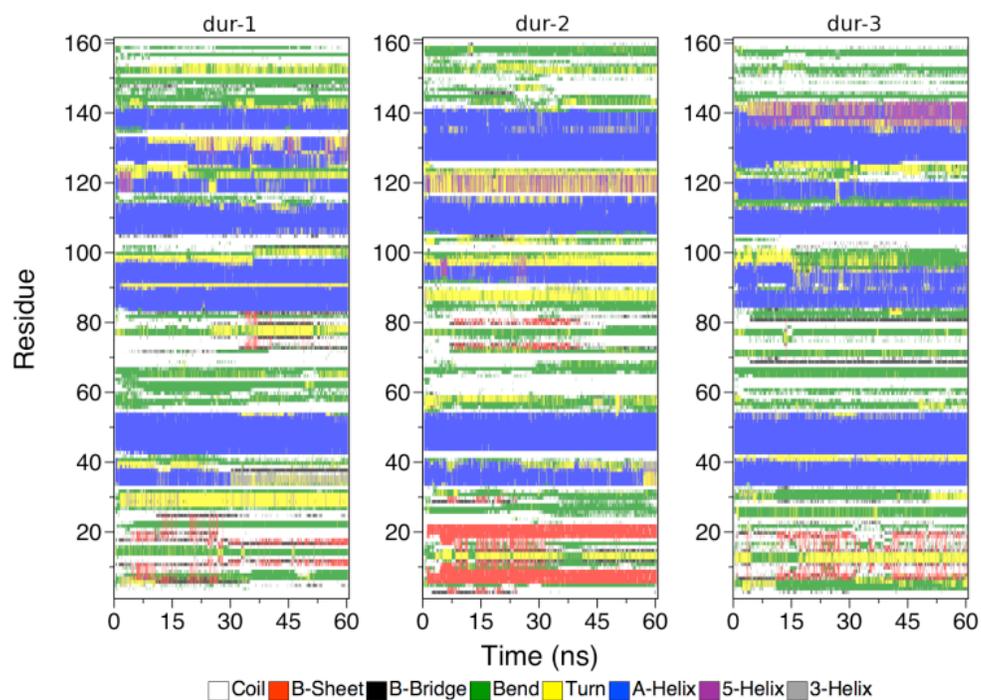


Figure 6.10. Secondary structure content of *dur* as a function of time for each replicate.

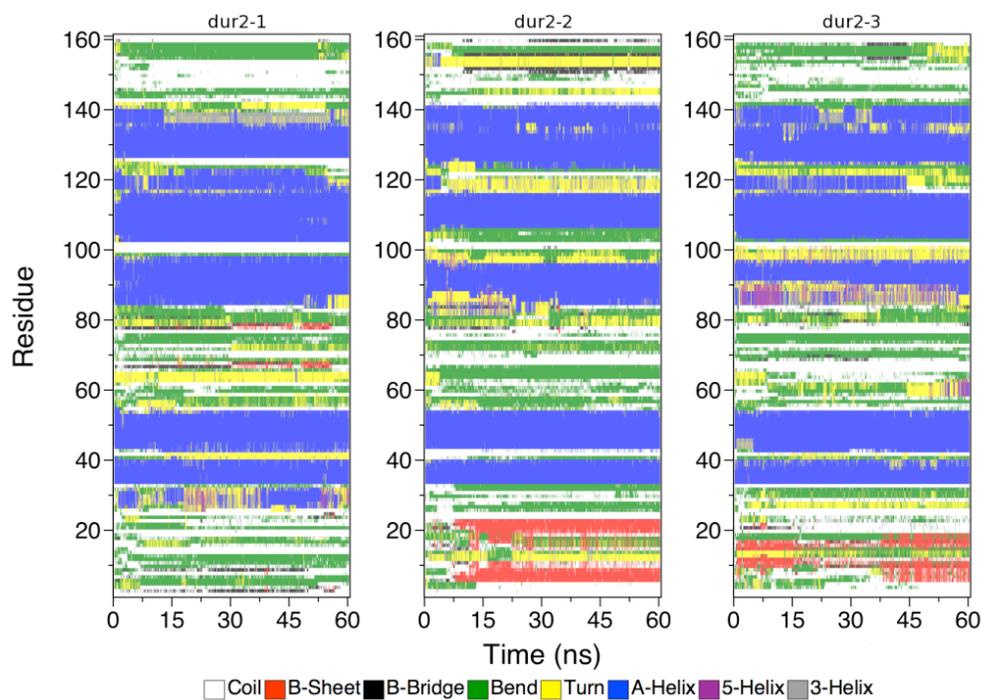


Figure 6.11. Secondary structure content of *dur2* as a function of time for each replicate.

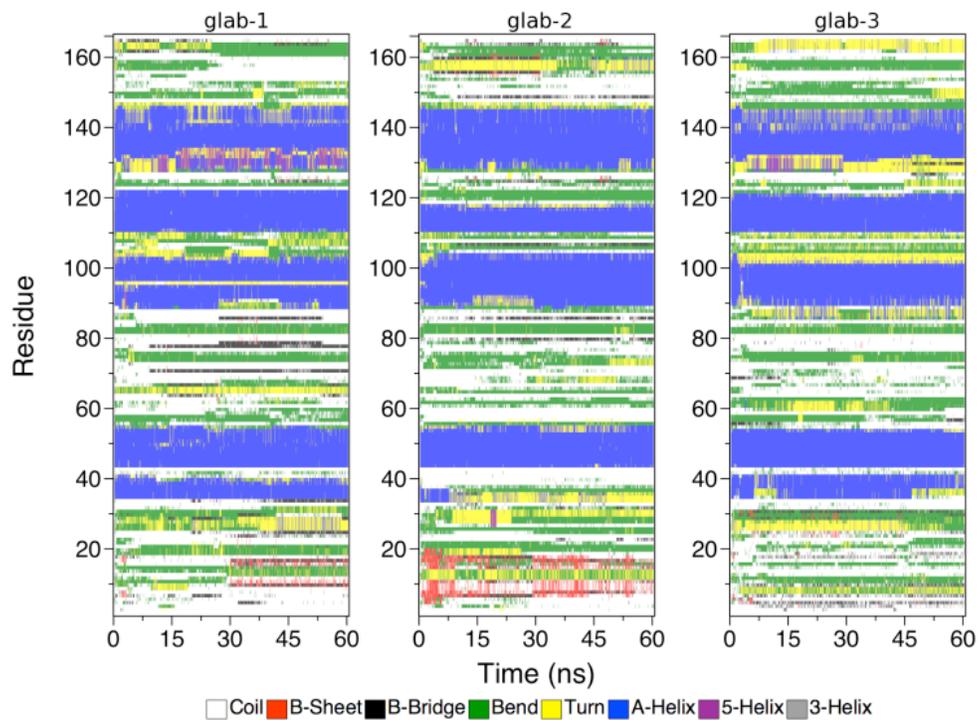


Figure 6.12. Secondary structure content of *glab* as a function of time for each replicate.

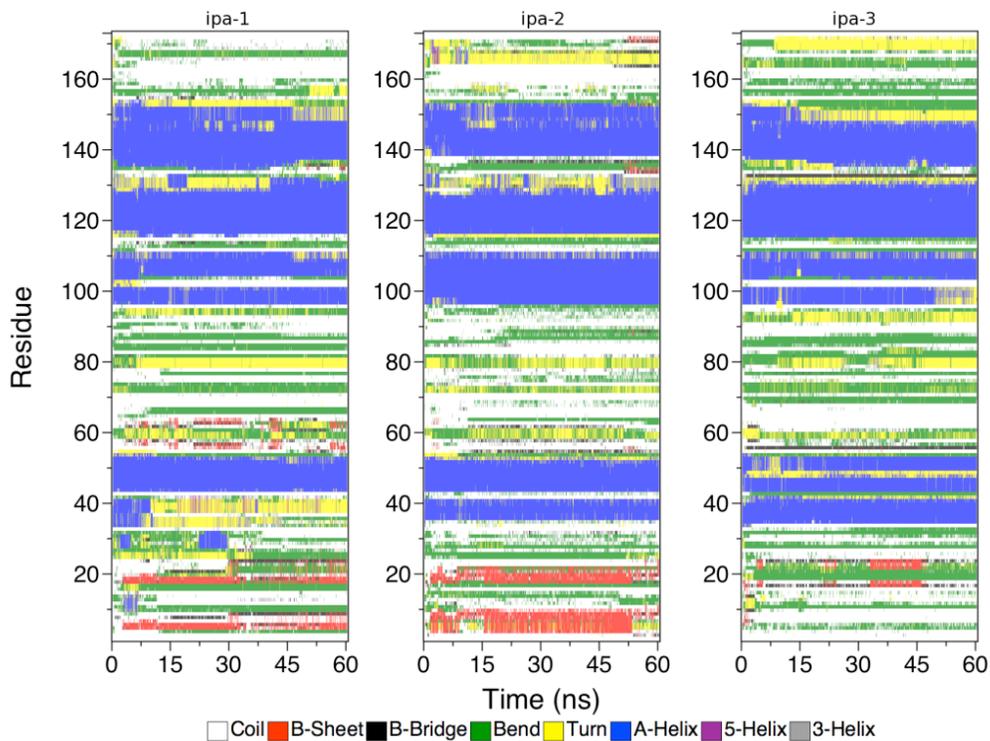


Figure 6.13. Secondary structure content of *ipa* as a function of time for each replicate.

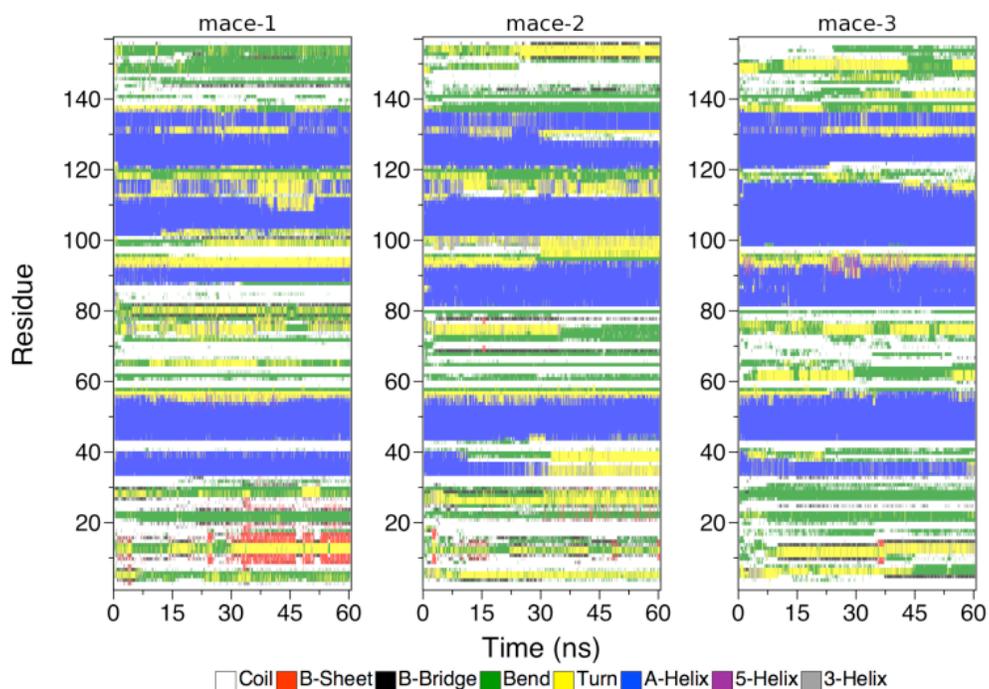


Figure 6.14. Secondary structure content of *mace* as a function of time for each replicate.

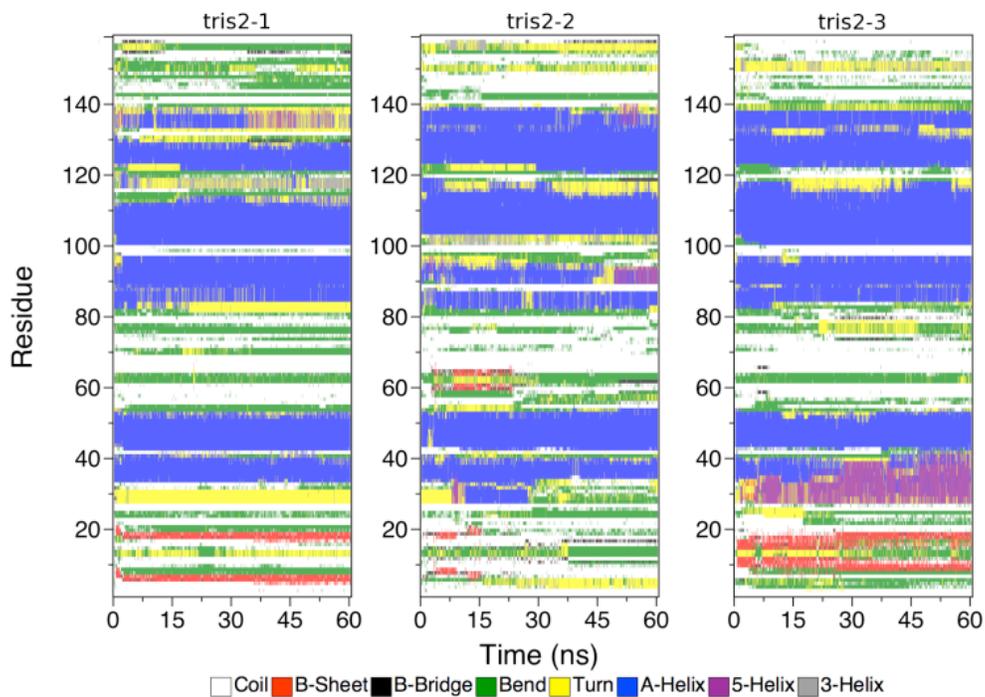


Figure 6.15. Secondary structure content of *tris2* as a function of time for each replicate.

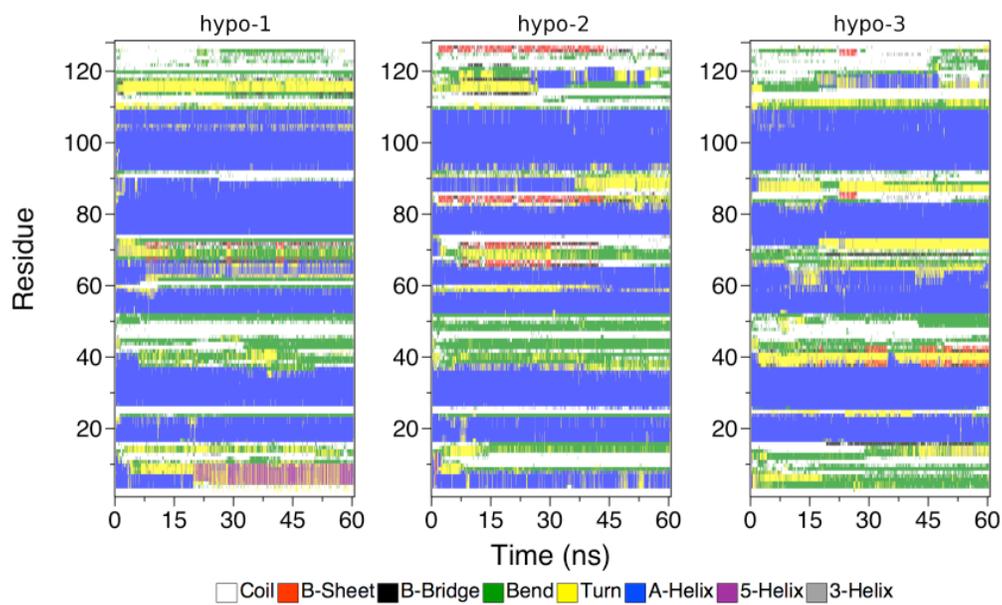


Figure 6.16. Secondary structure content of *hypo* as a function of time for each replicate.

Part VI: Solvent-Accessible Surface Area

As another measure to see how unfolded or how folded the loop is between $\alpha 2$ and $\alpha 3$, we can measure the solvent accessible surface area (SASA). Simply put, a ball 0.14 nm in diameter (about the size of one water molecule) is rolled over the entire protein, generating what is known as a Connolly surface. This surface represents what parts of the protein are exposed to the solvent. We then choose a specific group, like the loop between $\alpha 2$ and $\alpha 3$ for example, and measure the surface area of the Connolly surface of those residues. The SASA measured over the last 20 ns of simulation are shown in Figure 6.17.

It is easy to see that $\alpha 2$ - $\alpha 3$ loop in the *ipa* protein simulations is exposed to the most amount of solvent, and that the same loop in the *dar* protein simulations is exposed to the least amount of solvent. However, this information can be deceiving because not all of the loops are the same length. In order to normalize this data for a more direct comparison, we divided the absolute SASA value by the number of residues in the loop for each specific protein, resulting in a normalized (n)SASA. The location of the loop and the number of residues in each loop are recorded in Table 6.4, and the (n)SASA is reported in Figure 6.18.

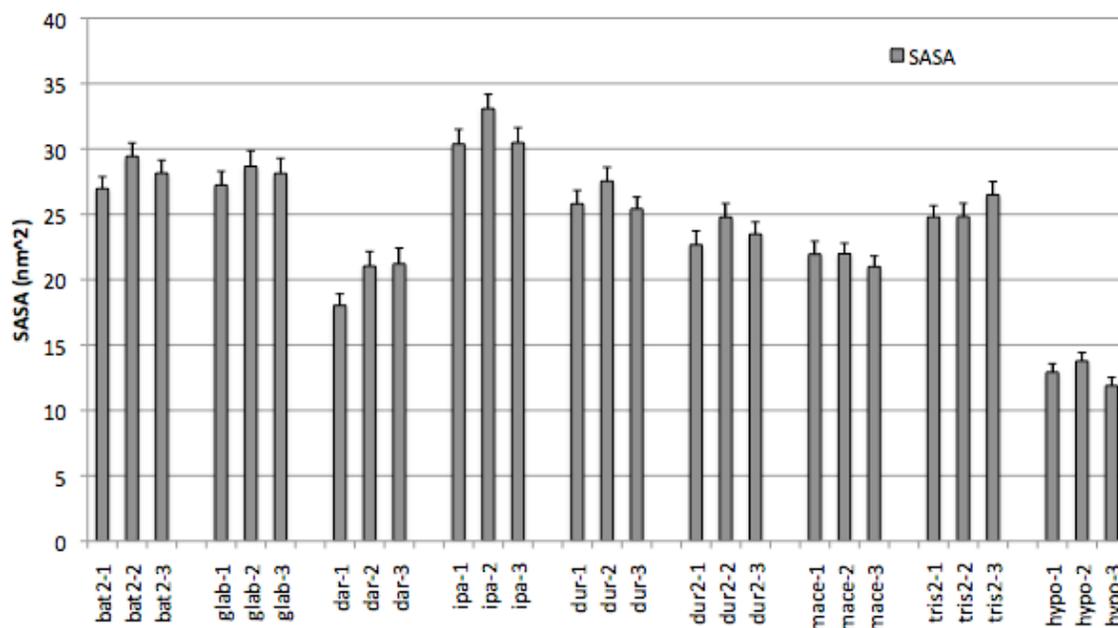


Figure 6.17. The SASA of the disordered loop between $\alpha 2$ and $\alpha 3$ for the last 20 ns of each replicate simulation.

Table 6.4. Loop location and length for each protein.

	Loop Residues	# of Residues
<i>bat2</i>	54-89	36
<i>glab</i>	54-88	35
<i>dar</i>	54-78	25
<i>ipa</i>	54-95	42
<i>dur</i>	54-83	30
<i>dur2</i>	54-83	30
<i>mace</i>	55-80	26
<i>tris2</i>	54-81	28
<i>hypo</i>	39-51	13

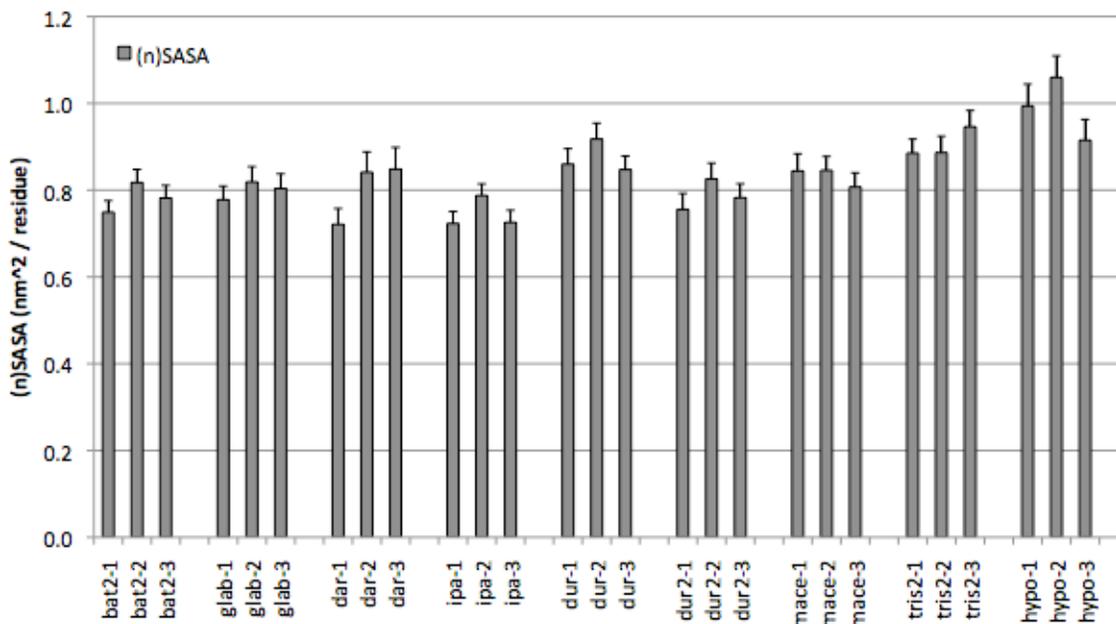


Figure 6.18. The (n)SASA of the disordered loop between $\alpha 2$ and $\alpha 3$ for the last 20 ns of each replicate simulation.

From Figure 6.18, it appears as if the $\alpha 2$ - $\alpha 3$ loop in *hypo* has emerged as the most solvent exposed (per residue), or the most unfolded, followed closely by *tris2*, *dur*, *mace*, and *dar*. From the other end, *ipa* looks to be the most folded, or have the least amount of contact with water per residue. At this point, it is not readily apparent what the significance of this observation is. It is important to note that this data might be biased in the following way: If the $\alpha 2$ - $\alpha 3$ loop in a given protein contains smaller residues, the data may be skewed low. And if the loop is rich in bulky residues, the data may be skewed too high. However, this type of information could prove to be important in characterizing this loop.

Conclusions

In this project we simulated nine Ara h 2 protein orthologs from eight different species within the *Arachis* genus. From the RMSD and radius of gyration data, we deduced that the 60 ns of simulation that was performed in triplicate for each protein was a sufficient amount. One of

the *dur* replicates (*dur-2*), however, is the only replicate that did not seem to converge. Without knowing how long it would take for that protein to converge, or whether it would converge at all, could make it quite time-consuming to attempt any further simulation. Consequently, we proceeded with the analysis of the data that we collected in 60 ns.

In all of the systems, the N-terminal and C-terminal domains, and the loop between $\alpha 2$ and $\alpha 3$ all fluctuated significantly. This is normal for disordered regions, and may indicate (especially for the loop between $\alpha 2$ and $\alpha 3$) that there may be an important function at those locations. The short loop between $\alpha 4$ and $\alpha 5$ also tended to fluctuate at a higher frequency, indicating this may be an important functional structure in the protein.

Clustering results show tendency for collapse in all disordered regions, as well as destabilization of some helices (especially $\alpha 3$). Further, there seems to be spontaneous secondary structure formation in the N-terminal region. DSSP results are in agreement with clustering and RMSD data. Some helices tend to destabilize (especially $\alpha 3$), and there seems to be a high propensity for β -strand formation in the N-terminal disordered region. Finally, the solvent-accessible surface area, especially when normalized to the number of residues in the loop between $\alpha 2$ and $\alpha 3$, could be a good indicator of the antigenicity of the protein.

The most obvious limitation of this project is the lack of supporting experimental data. Before the dynamics of the loop make any sense, we need to be able to measure the antigenicity of each protein ortholog experimentally. Unfortunately, at the time of this writing, the assays have yet to be completed. However, when they are completed, this Chapter will provide an excellent theoretical supplement to the wet lab observables.

References

- (1) Krapovickas, A.; Gregory, W. C. Taxonomía del Género *Arachis* (Leguminosae). *Bonplandia* **1994**, *8*, 1-186.
- (2) Koppelman, S. J.; Knulst, A. C.; Helbling, A.; Hefle, S. L., *et al.* Ara h 8, a Bet v 1-Homologous Allergen From Peanut, Is a Major Allergen in Patients With Combined Birch Pollen and Peanut Allergy. *J. Allergy Clin. Immunol.* **2004**, *114*, 1410-1417.
- (3) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435-447.
- (4) van Gunsteren, W. F.; Billeter, S. R.; Eising, A. A.; Hunenberger, P. H.; Kruger, P.; Mark, A. E.; Scott, W. R. P.; Tironi, I. G. Biomolecular Simulation: The GROMOS96 Manual and User Guide. Zurich, Switzerland: Hochschulverlag AG an der ETH Zurich. **1996**.
- (5) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997**, *18*, 1463-1472.
- (6) Darden, T.; York, D.; Pedersen, L. G. Particle Mesh Ewald: An N-log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089-10092.
- (7) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577-8593.
- (8) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81*, 3684-3690.
- (9) Nosé, S. A Molecular Dynamics Method for Simulations in the Canonical Ensemble. *Mol. Phys.* **1984**, *52*, 255-268.
- (10) Hoover, W. G. Canonical Dynamics: Equilibrium Phase-Space Distributions. *Phys. Rev. A* **1985**, *31*, 1695-1697.
- (11) Parrinello, M.; Rahman, A. (1981) Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52*, 7182-7190.
- (12) Nosé, S.; Klein, M. L. Constant Pressure Molecular Dynamics for Molecular Systems. *Mol. Phys.* **1983**, *50*, 1055-1076.

- (13) DeLano, W. L. *The PyMOL User's Manual*, DeLano Scientific, Palo Alto, CA, USA, **2002.**

Chapter 7

Practical Considerations for Building GROMOS-Compatible Small-Molecule Topologies

Justin A. Lemkul, William J. Allen, and David R. Bevan

Reproduced with permission from *Journal of Chemical Information and Modeling* **2010**, 50 (12), 2221-2235. Copyright 2010, American Chemical Society.

Attribution

Of the work discussed within this chapter, Lemkul derived parameters for all of the molecules and performed the simulations of the interfacial amino acid systems; Allen performed the simulations of UGM and the condensed phase / gas phase small molecules, and performed the calculations in Spartan; Lemkul and Allen contributed roughly equally to the writing and preparation of the manuscript. Bevan significantly contributed to the editing and review of the manuscript, as well as provided guidance in the research.

Abstract

Molecular dynamics simulations are being applied to increasingly complex systems, including those involving small endogenous compounds and drug molecules. In order to obtain meaningful and accurate data from these simulations, high-quality topologies for small molecules must be generated in a manner that is consistent with the derivation of the force field applied to the system. Often, force fields are designed for use with macromolecules such as proteins, making their transferability to other species challenging. Investigators are increasingly

attracted to automated topology generation programs, although the quality of the resulting topologies remains unknown. Here we assess the applicability of the popular PRODRG server that generates small molecule topologies for use with the GROMOS family of force fields. We find that PRODRG does not reproduce topologies for even the most well-characterized species in the force field due to inconsistent charges and charge groups. We assessed the effects of PRODRG-derived charges on several systems: pure liquids, amino acids at a hydrophobic-hydrophilic interface, and an enzyme-cofactor complex. We found that partial atomic charges generated by PRODRG are largely incompatible with GROMOS force fields, and the behavior of these systems deviates substantially from that of simulations using GROMOS parameters. We conclude by proposing several points as “best practices” for parameterization of small molecules under the GROMOS force fields.

Introduction

Preparation of small molecule topologies for use in molecular dynamics (MD) simulations is of critical importance in studies of drug-enzyme and drug-receptor interactions, partitioning of drugs into and across membranes, among others. The principal challenge in including small molecules in these simulations is the effort necessary to produce a topology that is consistent with the underlying theory of the force field, which is often based on common macromolecules such as proteins. Thus, less common functional groups present in drug molecules and synthetic compounds are difficult to parameterize. Different force fields frequently used for biomolecular simulation, such as AMBER (1-4), CHARMM (3-7), OPLS (8, 9), and GROMOS (10-14), have different functional forms and underlying assumptions, requiring different parameterization procedures and validation protocols. While parameterization of small molecules for AMBER and CHARMM force fields has been aided by the development of the general AMBER force field (GAFF) (15) and the CHARMM general force field (16), topology generation for OPLS and GROMOS remains more challenging. Procedures involving quantum mechanical calculations and geometry optimization are thoroughly described for the OPLS-AA parameter set (8, 9), but such detailed information is not publicly available for the GROMOS96 parameter sets. Thus, investigators seeking to use the GROMOS96 force fields must rely on empirical parameter

assignment and subsequent validation through the time-consuming process of thermodynamic integration (14), a challenge that has long been recognized within the simulation field (17).

The GROMOS96 force fields offer an advantage over the AMBER, CHARMM, and OPLS parameter sets in that they are based on the united-atom approach to parameterization. That is, nonpolar hydrogens are not explicitly represented, decreasing the number of atoms in the system appreciably and speeding up simulations. These force fields are attractive for long-time simulations and large sets of simulations that may be utilized *in silico* to analyze the dynamics of small molecules and potential drugs.

Automated topology generation is possible through a number of tools, including the very popular PRODRG server (18, 19). Topologies produced by this program have been used widely in studies of protein-peptide (20), protein-ligand (21, 22), protein-lipid (23), and small molecule-lipid interactions (24, 25), drug partitioning into lipid membranes (26), and simulations of small molecules that inhibit amyloid aggregation (27). Careful inspection of the topologies produced by PRODRG reveals that the charges and charge groups assigned to the functional groups of small molecules are often inconsistent with the same groups present in the GROMOS96 43A1 parameter library. The effects of these inconsistencies are unknown, providing the motivation for the present study. Several methods for generating GROMOS-compatible charges for refining PRODRG topologies have been proposed in the literature (28-31), but thus far, to the best of our knowledge, no standard has been set for the accuracy and applicability of any of these methods.

Here, we evaluate the quality of small-molecule topologies generated by the automated server PRODRG (18, 19), using small molecules whose functional groups are described by the GROMOS96 43A1 parameter set (11). We explore a variety of systems, including pure liquids, amino acids at a hydrophobic-hydrophilic interface, and a cofactor bound to an enzyme. By demonstrating the implications of the inherent deficiencies in automated parameter generation and by making recommendations for proper parameter development, we hope to elaborate on the best practices in small molecule topology generation under the GROMOS force fields.

Methods

To assess the quality of small-molecule topologies in a range of scenarios, several systems were constructed: (i) condensed-phase and gas-phase systems of hexane, ethanol, and *p*-cresol, (ii) several amino acids (alanine, asparagine, aspartate, isoleucine, lysine, and serine) at the interface of water and cyclohexane, and (iii) the protein UDP-galactopyranose mutase (UGM) with bound flavin adenine dinucleotide (FAD). All systems were prepared and simulations conducted using facilities present in the GROMACS package, version 4.0.7 (32). One set of small molecule topologies was generated using the PRODRG 2.5 server (18, 19), which is designed to produce topologies consistent with the GROMOS96 43A1 parameter set. The other topologies were generated using known GROMOS96 43A1 (11) functional group charges and charge groups, as implemented in the GROMACS distribution. Bonded parameters and atom types for each molecule were kept the same between the different topologies, such that the comparisons involved only charges and charge groups assigned by PRODRG relative to those implemented in the GROMOS96 43A1 parameter set. Unless otherwise noted, all simulations employed a twin-range cutoff scheme for short-range nonbonded interactions, with the real-space contribution to Coulombic terms truncated at 0.8 nm and short-range van der Waals interactions truncated at 1.4 nm. All bond lengths were constrained using the LINCS method (33), allowing a 2-fs time step. The neighbor list was updated every 5 simulation steps (10 fs). Long-range electrostatic interactions were calculated with the smooth particle mesh Ewald (PME) method (34, 35). Equilibration simulations utilized the Berendsen weak coupling method (36) to control temperature and/or pressure, while production simulations utilized the Nosé-Hoover thermostat (37, 38) and Parrinello-Rahman barostat (39, 40) to generate a rigorous NPT ensemble.

Condensed-Phase and Gas-Phase Systems

To measure various thermodynamic and physical properties of pure liquid systems as a function of the molecular topology, systems containing hexane, ethanol, and *p*-cresol were constructed by replicating a single molecule of each compound in three dimensions to create a cubic grid of 512 molecules. For gas-phase systems, molecules were separated by 50 nm and kept in place with weak position restraints, in accordance with GROMOS parameterization

methodology, and plain cutoffs were used to calculate electrostatic interactions (14). Steepest descent minimization was performed on gas-phase and condensed-phase systems alike, followed by 100 ps of isochoric-isothermal (NVT) equilibration at 298 K. Further equilibration under an isothermal-isobaric (NPT) ensemble was performed for 300 ps at the same temperature and 1 bar of pressure. Finally, each system was simulated for another 20 ns production run under the same conditions. Charges and charge groups for these molecules are illustrated in Figure 7.1.

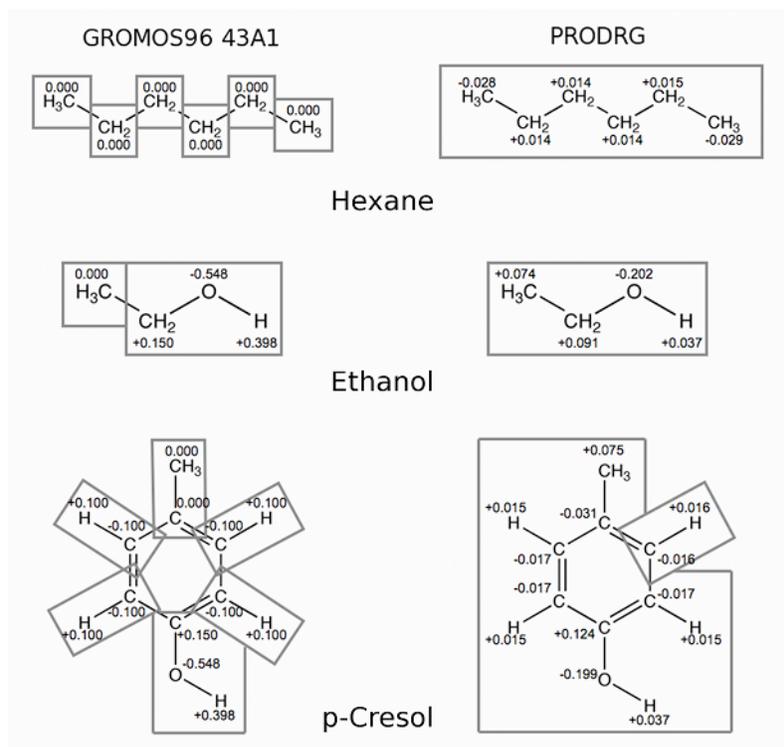


Figure 7.1. Comparison of GROMOS96 43A1 and PRODRG charges and charge groups for hexane (top), ethanol (middle), and *p*-cresol (bottom). Charges are given in units of e and charge groups are indicated by gray boxes.

Interfacial Amino Acid Systems

Amino acids represent some of the most common components of biomolecular force fields and thus should serve as reliable test systems. By using amino acids, we can assess the quality of PRODRG topologies against groups that have been rigorously defined. Six amino acids (alanine, asparagine, aspartate, isoleucine, lysine, and serine) were chosen to represent the twenty common amino acids. The chosen molecules encompass hydrophobic (Ala and Ile), polar

uncharged (Asn and Ser), and charged (Asp and Lys) species. Each amino acid was modeled as a dipeptide, with N- and C-terminal groups capped with acetyl and N-methyl groups, respectively, to give uncharged termini. Since the goal of the GROMOS96 parameter sets is to accurately describe partitioning of amino acids between polar (water) and hydrophobic (cyclohexane) media (14), we sought to produce a biphasic system that would model partitioning behavior of the different amino acids.

A 5-nm cubic box of cyclohexane was generated, containing a total of 466 molecules. Standard bond, angle, and dihedral parameters from the GROMOS96 43A1 library were assigned in the cyclohexane topology. All atoms were represented as united-atom CH₂ with zero charge. Following steepest descent minimization, the cyclohexane box was equilibrated first under an isochoric-isothermal (NVT) ensemble for 100 ps (298 K), and then under an isothermal-isobaric (NPT) ensemble for 500 ps (298 K, 1 bar), at which point the box vectors and density had stabilized. The system was simulated for an additional 10 ns to generate the cyclohexane layer used in building the interfacial systems.

Each amino acid was placed at the center of a rectangular box such that both the plane of the amino acid backbone and the C α -C β bond were coincident with the x-y plane of the unit cell (Figure 7.2). The equilibrated cyclohexane box was placed in the lower half of the simulation cell, while the remainder of the volume was filled with SPC water (41). A single Cl⁻ counterion was added to the aqueous phase of the lysine system to compensate for the net charge of the amino acid. In the case of aspartate, a single Na⁺ ion was added. No ions were added to systems containing uncharged amino acids. Each system was energy-minimized and equilibrated in the same manner as the cyclohexane box discussed above, with position restraints applied to amino acid heavy atoms during equilibration. Three replicates of each system were generated from different random initial velocities. The data collection period lasted for 20 ns. Charges and charge groups used for the simulations of amino acids are given in Figures 7.3 and 7.4.

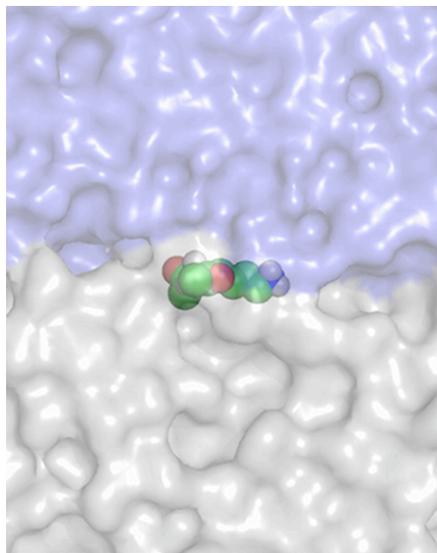


Figure 7.2. Initial position of lysine in the cyclohexane-water interfacial system. The amino acid backbone is perpendicular to the page and the sidechain is extending to the right in the image. Lysine is shown with its atoms colored by element and rendered as van der Waals spheres. The co-solvents are rendered as transparent surfaces, with water in blue and cyclohexane in gray. The image was rendered with PyMOL (56).

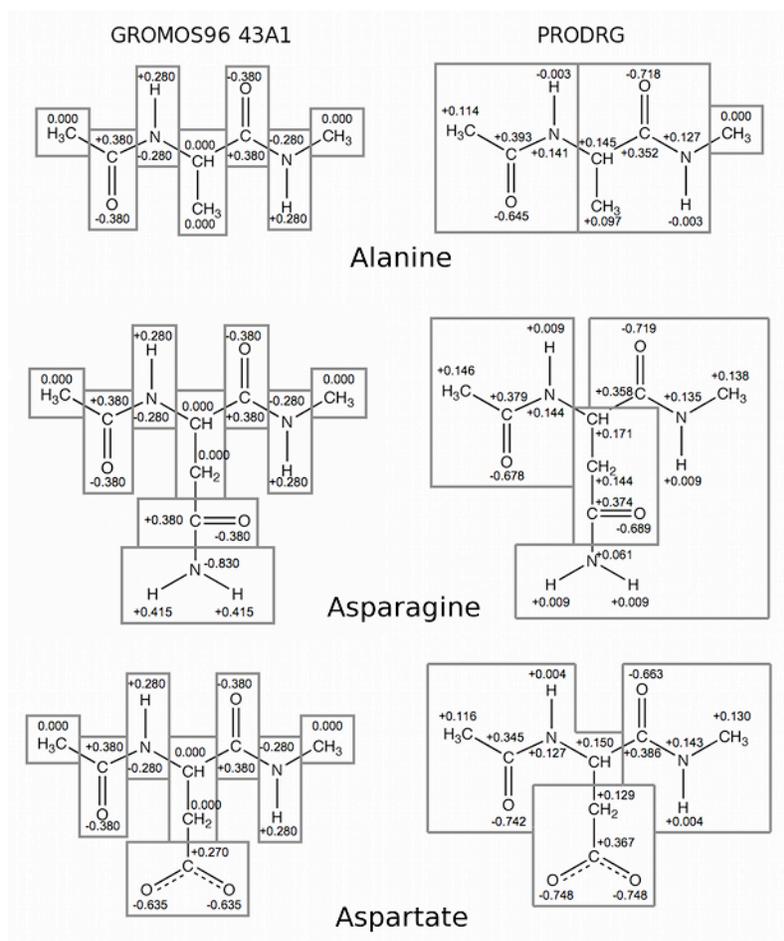


Figure 7.3. Comparison of GROMOS96 43A1 and PRODRG charges and charge groups for alanine (top), asparagine (middle), and aspartate (bottom). Charges are given in units of e and charge groups are indicated by gray boxes.

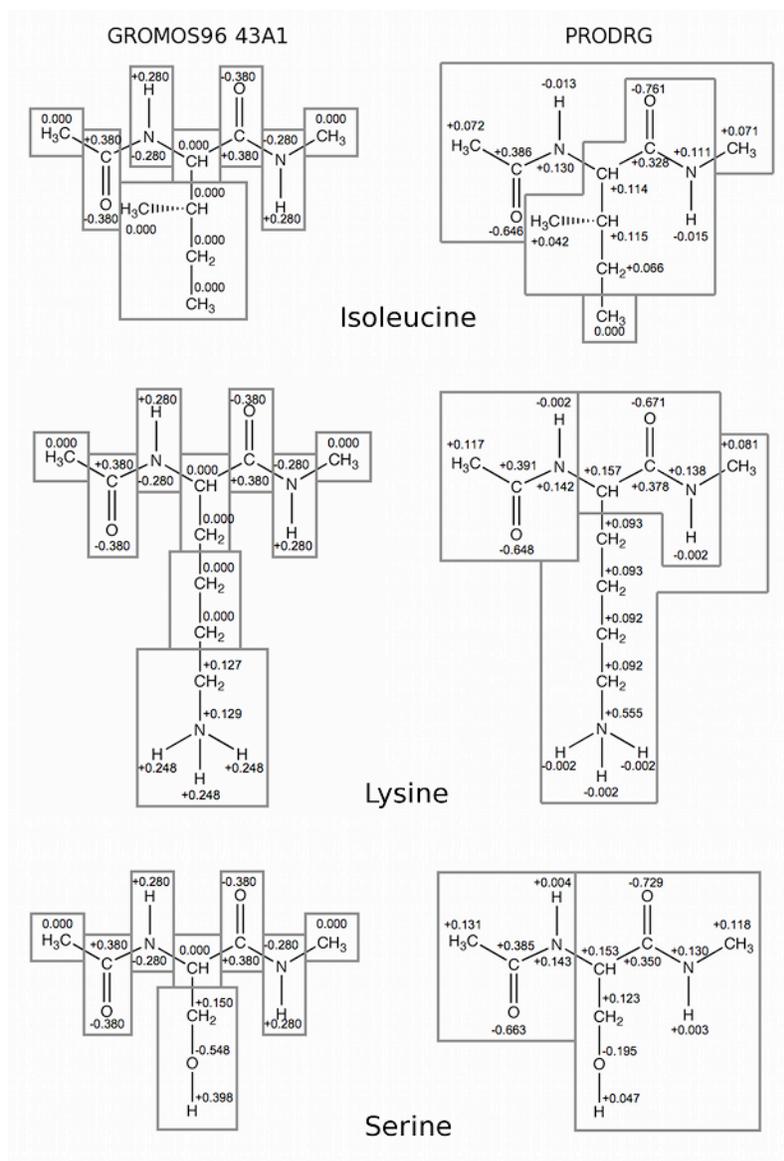


Figure 7.4. Comparison of GROMOS96 43A1 and PRODRG charges and charge groups for isoleucine (top), lysine (middle), and serine (bottom). Charges are given in units of e and charge groups are indicated by gray boxes.

UDP-Galactopyranose Mutase

To examine the stability of a cofactor bound to a protein, the structure of UDP-galactopyranose mutase (UGM) bound non-covalently with flavin adenine dinucleotide (FAD) was utilized. The starting structure for these simulations was taken from the crystal structure of the *K. pneumoniae* enzyme-cofactor complex determined by Beis *et al* (PDB: 2BI7) (42), which contains an oxidized FAD cofactor. Parameters for the protein were assigned from the

GROMOS96 43A1 parameter set, and FAD parameters were either taken from this same force field or generated by PRODRG (Figure 7.5). The protein-cofactor complex was centered in a rhombic dodecahedral unit cell that was subsequently filled with SPC water (41) and sufficient counter ions to balance the charge and bring the final ionic concentration to a near-physiological 100 mM. Equilibration procedures similar to the previous systems were implemented. Steepest descent minimization was performed, followed by 100 ps of isochoric-isothermal (NVT) equilibration. Next, 500 ps of equilibration under an isothermal-isobaric (NPT) ensemble were performed, followed by 40-ns production simulations. Position restraints were applied to the backbone of the protein during equilibration, and released for the production run. Different random initial velocities were used to generate independent trajectories.

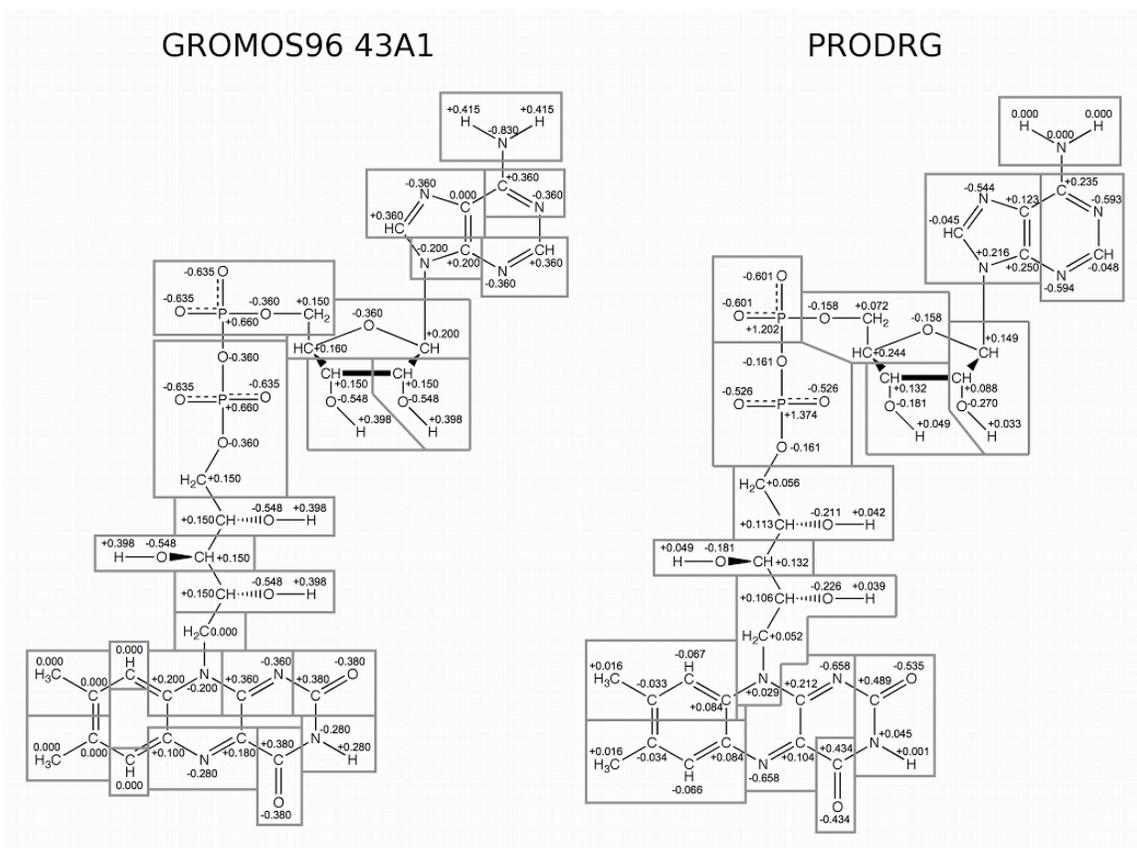


Figure 7.5. Comparison of GROMOS96 43A1 and PRODRG charges and charge groups for FAD. Charges are given in units of e and charge groups are indicated by gray boxes.

Results and Discussion

The goals of our approach include the assessment of several physical and thermodynamic parameters in the liquid- and gas-phase systems, behavior of amino acids in the presence of hydrophobic and polar media, and the interactions of an enzyme-bound molecule with its partner protein. These scenarios represent some of the most common uses of small molecule topologies in the literature, and provide us with versatile systems in which to test the quality of these topologies. We provide an analysis of the various intermolecular interactions that give rise to the observed behavior, focusing primarily on electrostatic and van der Waals interactions, hydrogen bonding, and relative orientation of the different molecules. While these quantities are largely interdependent, a comprehensive analysis of these terms is appropriate as they relate directly to quantities that are often evaluated in MD simulations.

For all comparisons between the results of using PRODRG and GROMOS96 43A1 topologies, the behavior observed in the GROMOS96 43A1 simulations is considered the benchmark for all comparisons. Parameters for novel species must be derived in a manner consistent with the existing, validated force field, which is taken as the standard for accuracy, and thus we evaluate the results of using PRODRG parameters in this context.

Condensed-Phase and Gas-Phase Systems

The parameterization of the GROMOS96 force field parameter sets relies on accurate reproduction of condensed-phase behavior and thermodynamic data. In the most recent edition of the force field (53A6) (14), the criteria used to assess the accuracy of the derived parameters included density and the heat of vaporization, ΔH_{vap} . These quantities can be calculated from MD simulations of hexane, ethanol, and *p*-cresol.

Table 7.1 summarizes the results of density and ΔH_{vap} calculations. The equation used to calculate ΔH_{vap} was:

$$\Delta H_{\text{vap}} = \langle E_{\text{gas}} \rangle - \langle E_{\text{liquid}} \rangle + RT$$

In this equation, the quantities in angle brackets denote the time average of the total energy of the gaseous and liquid systems. R is the gas constant, and T is the absolute temperature in Kelvin.

Table 7.1. Density and ΔH_{vap} for small molecules (54, 55).

System	Density [g mL ⁻¹]			ΔH_{vap} [kJ mol ⁻¹]		
	43A1	PRODRG	Expt.	43A1	PRODRG	Expt.
Ethanol	0.766 ± 0.008	0.972 ± 0.007	0.789 ^a	20.3 ± 0.2	9.4 ± 0.4	42.3 ^d
Hexane	0.681 ± 0.005	0.679 ± 0.005	0.661 ^b	17.2 ± 0.3	17.2 ± 0.3	31.6 ^e
<i>p</i> -Cresol	1.097 ± 0.004	1.045 ± 0.005	1.019 ^c	48.7 ± 0.5	35 ± 2	73.1 ^f

^aRef 54, pp 3-206; ^bpp 3-308; ^cpp 3-134; ^dpp 6-111; ^epp 6-117; ^fRef 55.

We found relatively good agreement between the GROMOS96 43A1 topologies and experimental data with respect to density (Table 7.1). Densities were slightly overestimated for hexane and *p*-cresol (by 3.0% and 7.7%, respectively), and slightly underestimated for ethanol (by 3.0%). For PRODRG-produced topologies, the densities of hexane and *p*-cresol were in close agreement with both the GROMOS96 43A1 results and experimental data (overestimated by 2.7% for hexane and 2.6% for *p*-cresol), but that of ethanol was significantly overestimated (by 23.2%).

With respect to ΔH_{vap} values, neither GROMOS96 43A1 nor PRODRG parameters closely reproduced experimental data, with all simulation results producing values that were well below the experimentally-determined ΔH_{vap} . The values of ΔH_{vap} generated by GROMOS96 43A1 topologies, however, generally were in somewhat closer agreement with experimental values than were values generated using PRODRG parameters (Table 7.1).

To explain these results, we examined the structures of the liquids themselves in terms of both hydrogen bonding and radial distribution functions (RDF), which describe the density of chosen particles as a function of distance, a calculation that can be applied to individual atoms or to whole molecules. In an RDF plot, closely interacting groups are separated by shorter distances, such that the resulting density peak is higher, indicating a greater probability of interaction at a given distance.

The decreased value of ΔH_{vap} for ethanol and *p*-cresol resulting from PRODRG parameters can be attributed mainly to a significant decrease in hydrogen bonding capacity. For ethanol, the

use of GROMOS96 43A1 parameters resulted in 0.93 ± 0.01 hydrogen bonds per molecule (averaged over time). For *p*-cresol, this value was 0.94 ± 0.01 . Use of PRODRG parameters gave substantially lower hydrogen bonding for these two species, 0.12 ± 0.02 hydrogen bonds per ethanol molecule and only 0.05 ± 0.01 per *p*-cresol molecule. The RDF of these two species (Figure 7.6) provides insight into the reason for this observed behavior.

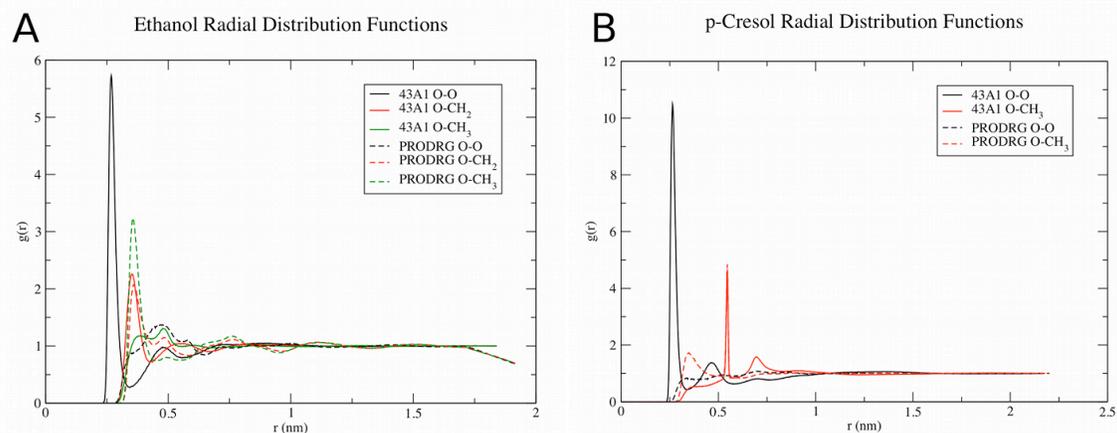


Figure 7.6. RDF for different atom pairs in (A) ethanol and (B) *p*-cresol. In (B), the intramolecular O-CH₃ peak referred to in the text appears as a single, solid line since it is present in both curves and is thus overlapping.

PRODRG parameters caused ethanol molecules to align head-to-tail, that is, with the hydroxyl group associating most prominently with the methyl group of a nearby molecule, at a distance of approximately 0.3 nm (Figure 7.6A). Interactions between hydroxyl groups occurred at a much greater distance, nearly 0.5 nm, exceeding the optimal distance for hydrogen bonding. The combination of a relatively large charge ($+0.071 e$) on the terminal methyl group and the comparatively small charge on the hydroxyl hydrogen ($+0.037 e$) leads to this behavior. The association of the hydroxyl oxygen atoms with the terminal methyl groups disrupted hydrogen bonding, decreased ΔH_{vap} , and increased the density relative to the results obtained using GROMOS96 43A1 parameters. Proper hydrogen bonding was observed using GROMOS96 43A1 parameters for ethanol, with hydroxyl-hydroxyl interactions occurring at approximately 0.3 nm, indicating strong hydrogen bonding and comparing closely with the RDF of previous simulations using different parameters (43).

In the case of *p*-cresol, we also observed a reduction in hydrogen bonding when PRODRG parameters were applied, as well as an increase in hydroxyl-hydroxyl distance relative to the

results obtained using GROMOS96 43A1 parameters (Figure 7.6B). Under PRODRG parameters, there was no distinct hydroxyl-hydroxyl RDF peak, but there was a probable interaction between the hydroxyl group and the *p*-methyl group occurring at approximately 0.35 nm, indicating head-to-tail orientation similar to the case of ethanol. Use of GROMOS96 43A1 parameters showed this behavior as well, but at a much larger distance, approximately 0.7 nm, while hydrogen bonding was clearly occurring due to hydroxyl-hydroxyl interactions at approximately 0.25 nm. A peak at approximately 0.55 nm occurred in the RDF produced by both PRODRG and GROMOS96 43A1 parameters; this peak is the intramolecular distance between the hydroxyl group and the *p*-methyl group.

RDF plots for hexane (data not shown) indicate that there was no distinguishable difference in the liquid structure due to the application of GROMOS96 43A1 or PRODRG parameters. The average end-to-end distance of hexane was 0.607 ± 0.001 nm in both cases, with a radius of gyration of 0.216 ± 0.003 nm. Thus, the presence of small partial charges on hexane assigned by PRODRG did not significantly influence the configurations of hexane or the structure of the bulk liquid.

Interfacial Amino Acid Systems

To assess the quality of PRODRG-generated amino acid topologies relative to those present in the GROMOS96 43A1 force field library, several parameters were analyzed: the position and orientation of the amino acids in the biphasic cyclohexane-water system, the nonbonded energies (Coulombic and Lennard-Jones) between the amino acids and the two solvents, and the ability of relevant functional groups in the amino acids to hydrogen bond with water. The balance of these terms will dictate the behavior of the amino acids in these systems and thus is an important measure of the quality of the small molecule topologies.

Configurations and Trajectory Observations. The position of amino acids parameterized under the GROMOS96 43A1 force field was largely consistent over time, with all amino acids remaining at the cyclohexane-water interface. The only exception was one simulation of isoleucine, in which the molecule diffused into the cyclohexane layer to localize at the cyclohexane-water interface that is formed at the periodic boundary of the system in the *z*-dimension (Figure 7.7). The orientation of the sidechains varied depending on the chemical

nature of the amino acid. Hydrophobic amino acids (alanine and isoleucine) were oriented such that their side chains were either coincident with the interface or buried in the cyclohexane layer. Polar uncharged amino acids (asparagine and serine) were positioned with their sidechains coincident with the interface or directed towards the aqueous solvent. Charged amino acids (aspartate and lysine) consistently oriented their sidechains towards the aqueous solvent.

The behavior of the amino acids whose parameters were assigned by the PRODRG server was less consistent (Figure 7.8). In all three simulations of alanine, the amino acid diffused freely into the water layer. Simulations of isoleucine parameterized by PRODRG performed comparably to the simulations under the GROMOS96 43A1 parameters, with isoleucine remaining at the cyclohexane-water interface and its sidechain buried in the cyclohexane layer. The position of asparagine across three simulations was inconsistent; in two cases, asparagine remained largely interfacial, occasionally diffusing as much as 3 nm into the aqueous phase before returning to the interface. In the third simulation of asparagine, the amino acid diffused to the top of the unit cell to associate with the cyclohexane-water interface at the periodic boundary. In the simulations of serine, two systems remained interfacial with their polar sidechains buried in the cyclohexane layer. In the third serine simulation, the amino acid diffused to the top of the unit cell to ultimately bury its sidechain into the periodic cyclohexane layer. In simulations of both lysine and aspartate, the amino acids diffused freely into the aqueous phase; their positions were variable over time.

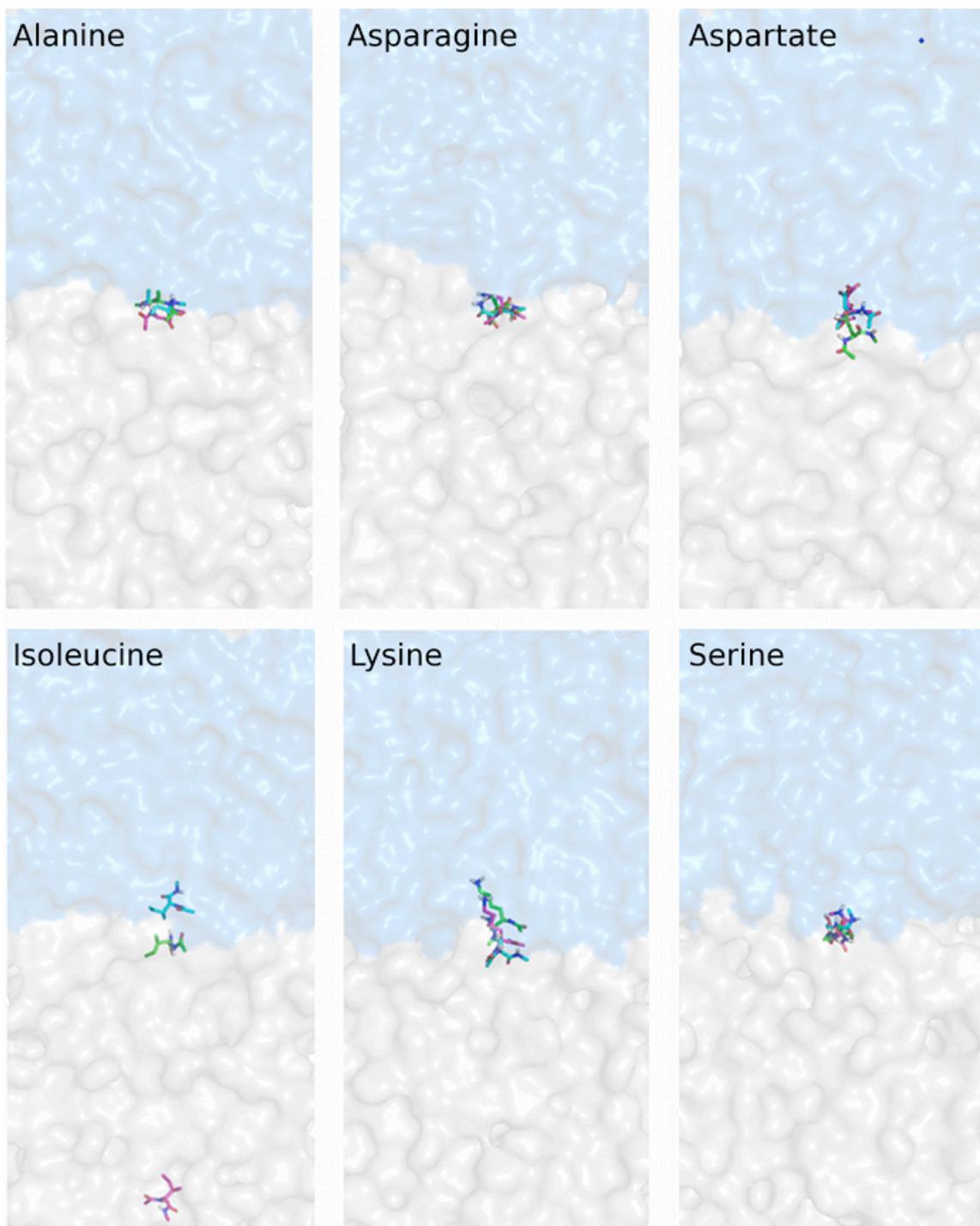


Figure 7.7. Overlays of amino acid positions generated by GROMOS96 43A1 parameters, taken from the final snapshot (20 ns) of each simulation after fitting for global x-y translation. Amino acids are shown as sticks and colored separately by replicate. Water and cyclohexane are rendered as in Figure 7.2. The images were generated with PyMOL (56).

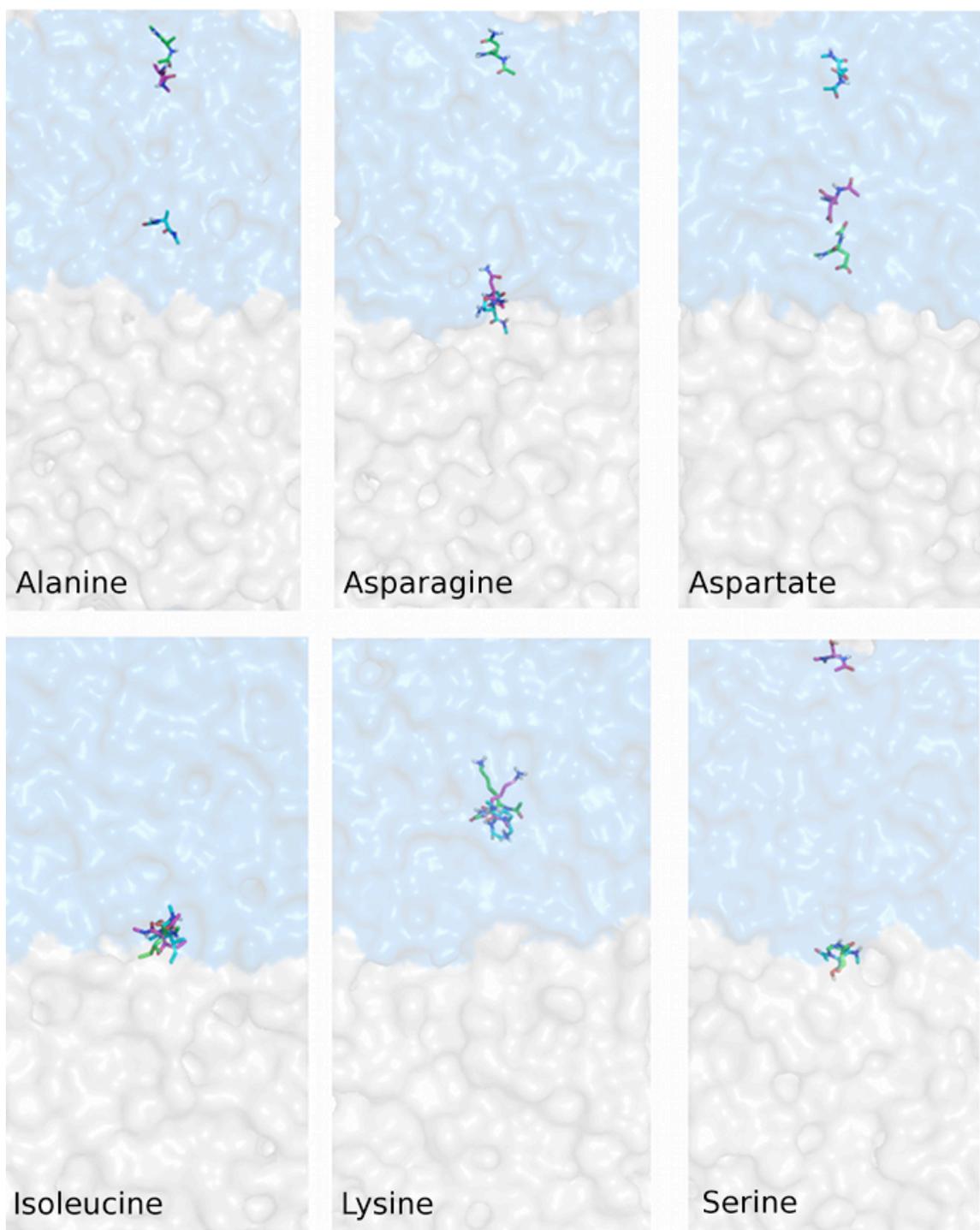


Figure 7.8. Overlays of amino acid positions generated by PRODRG parameters, taken from the final snapshot (20 ns) of each simulation after fitting for global x-y translation. Amino acids are shown as sticks and colored separately by replicate. Water and cyclohexane are rendered as in Figure 7.2. The images were generated with PyMOL (56).

Energetics of Interactions with Co-Solvents. The interactions of the amino acids with the two co-solvents of the interfacial systems were strongly influenced by the nature of the electrostatic and van der Waals interactions between the small molecules and the aqueous and hydrophobic environments. The balance between these terms dictated the partitioning behavior of the amino acids. Since the cyclohexane topology included no charges on the constituent atoms, the only electrostatic interactions in these systems (aside from water-water interactions) occurred between the amino acid and water. In all cases, the charges assigned in PRODRG-generated topologies led to significantly lower-energy (more favorable) electrostatic interactions with water than the results obtained using GROMOS96 43A1 parameters (Figure 7.9A).

As a consequence of the greater interaction with the aqueous phase in the PRODRG-parameterized systems, atoms were drawn closer together by stronger electrostatic interactions. As a result, van der Waals interactions with the aqueous phase became more repulsive for all amino acids except for lysine, which experienced a greater net Lennard-Jones attraction (Figure 7.9B). RDF analysis (Figure 7.10) indicates that more water molecules interacted with the sidechain methylene groups when PRODRG charges were applied. These groups are normally uncharged in the GROMOS96 43A1 parameter set, leading to weaker solvent interactions. The oxygen atom in water interacted with each of the sidechain methylene groups when small partial charges (assigned by PRODRG) were present such that favorable van der Waals interactions were established, a behavior that was promoted by favorable dipole interactions from the partial charges present.

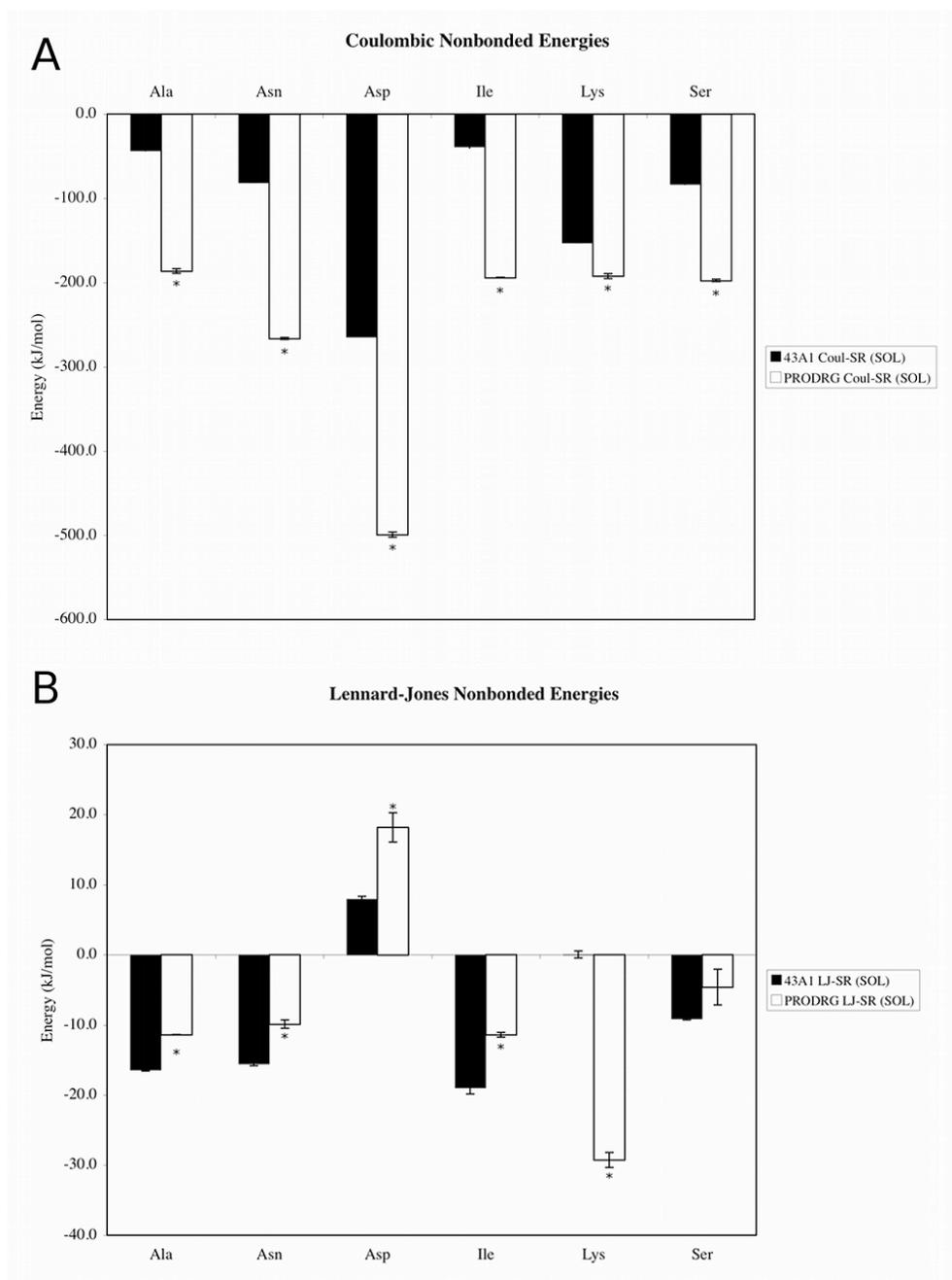


Figure 7.9. (A) Electrostatic and (B) Lennard-Jones interaction energies between the amino acids (listed on the x-axis) and water. Asterisks (*) indicate a significant difference based on a two-tailed Student's t-test ($p < 0.05$). Error bars represent standard deviations of the three simulations of each system.

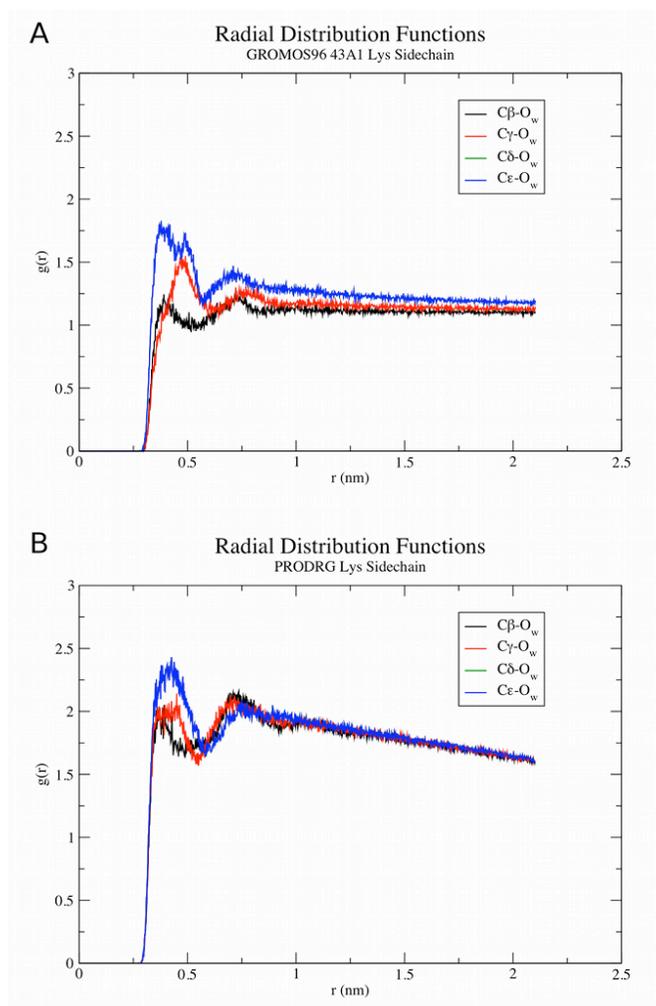


Figure 7.10. RDF for SPC oxygen (O_w) atoms with respect to lysine sidechain atoms using (A) GROMOS96 43A1 and (B) PRODRG parameters. In both panels, the curves for Cd and Ce atoms are overlapping.

For all the other amino acids considered here, the use of PRODRG parameters caused water molecules to be distributed around the solute molecules at either a closer distance or at the same distance but with higher probability than with GROMOS96 43A1 parameters, based on RDF analysis (data not shown). The attraction of water molecules to these species derived largely from the over-polarization of the backbone and/or sidechains of each amino acid in the PRODRG topologies. The result was a net repulsion due to an increased density of water around these amino acids, particularly at distances between atomic pairs above the energy minimum on the Lennard-Jones curve.

Hydrogen Bonding. All amino acids parameterized under GROMOS96 43A1 formed essentially the same number of backbone hydrogen bonds with water (approximately 2), averaged over all simulations and all snapshots in each simulation (Figure 7.11A). The consistency across simulations is likely due to the fact that the backbone amide groups had identical parameters, in accordance with the transferability of functional groups within the GROMOS parameter sets. That is, in the GROMOS force fields, the same functional groups have the same charges, even if they occur in different species, since charges are derived from model species that represent a small portion of an existing compound, typically an amino acid sidechain (14). For all the amino acids examined here, the PRODRG-produced topologies led to a significant increase in the number of backbone-water hydrogen bonds relative to the results obtained with GROMOS96 43A1 parameters, nearly 6 in the case of aspartate. This effect can be attributed to the greater separation of charges in the backbone amide groups (Figures 7.3 and 7.4). These charges vary somewhat in their magnitude depending on the amino acid, but in all cases the resulting dipoles are greater than those present in the GROMOS96 43A1 topologies.

Sidechain hydrogen bonds are another important consideration when examining these systems. Illustrated in Figure 7.11B are hydrogen bonds between sidechain groups (where applicable) and water. In the cases of asparagine and aspartate, significantly more hydrogen bonds formed between sidechain groups (amide and carboxylate, respectively) in the case of PRODRG-generated topologies. The principal reason for this increase is the greater magnitude of charge separation on the amide carbonyl in asparagine and carboxylate in aspartate. The stronger dipoles in these groups led to more hydrogen bonding with nearby water. The opposite effect is observed in the case of lysine and serine systems, wherein hydrogen bonds were significantly decreased relative to the system simulated with GROMOS96 43A1 parameters. In the case of lysine, the charges on the ϵ -amino group hydrogens assigned by PRODRG are $-0.002 e$, opposite in sign and significantly reduced in magnitude compared to the charges that are assigned under GROMOS96 43A1 ($+0.248 e$). The partial negative charges assigned by PRODRG likely inhibit the ability of this ϵ -amino group to effectively serve as a hydrogen bond donor. As described above, in the simulations of serine, the PRODRG-derived charges caused the amino acid to bury its sidechain in the cyclohexane layer, rather than in water. The dipole on the sidechain hydroxyl group is reduced compared to the GROMOS96 43A1 topology, and thus hydrogen bonding to water was less of a driving force in the interactions in these systems.

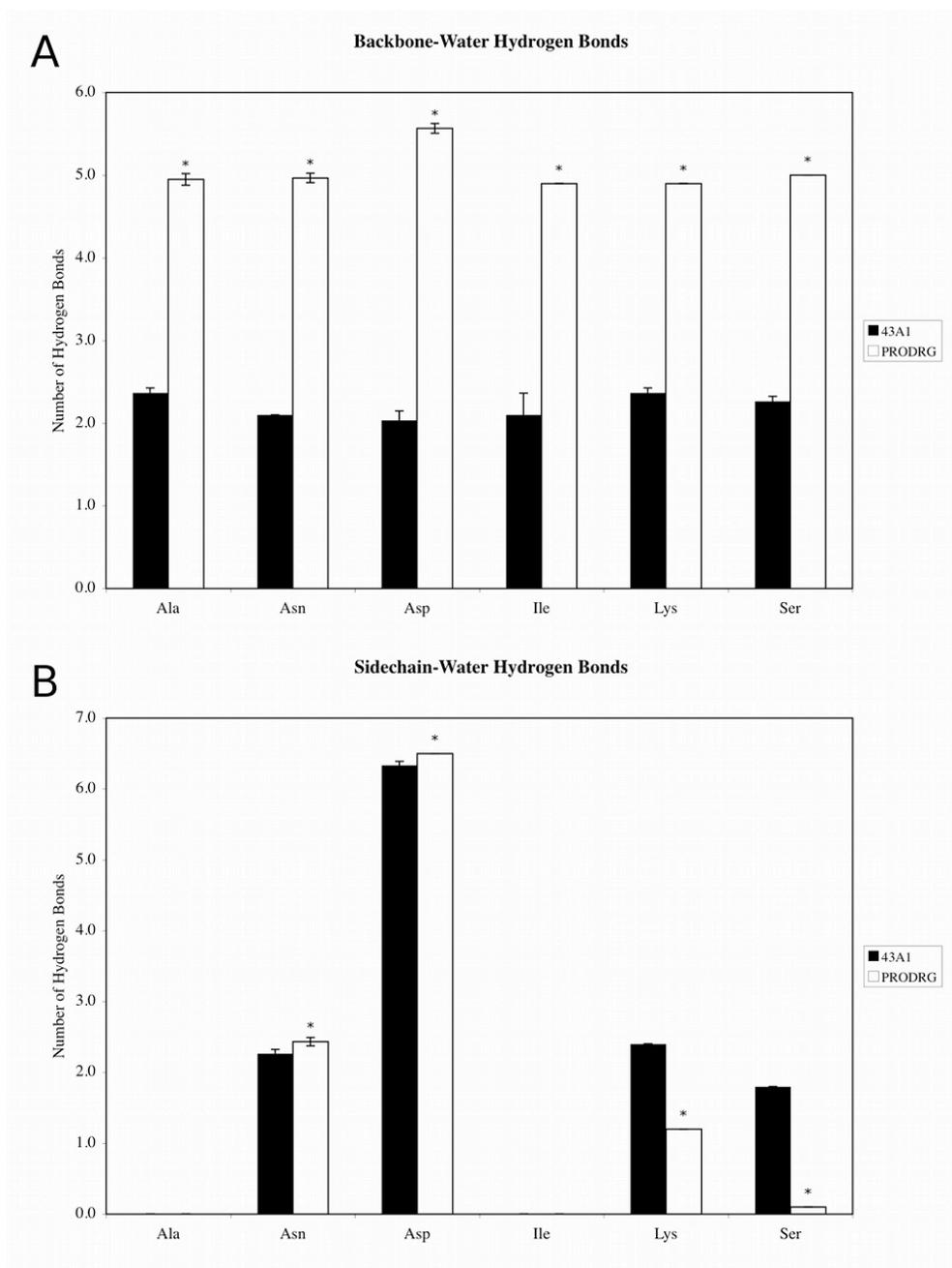


Figure 7.11. (A) Backbone and (B) sidechain hydrogen bonds to water. Asterisks (*) indicate a significant difference based on a two-tailed Student's t-test ($p < 0.05$). Error bars represent standard deviations of the three simulations of each system.

The results from the interfacial amino acid simulations call into question the use of unrefined PRODRG topologies for small molecules. PRODRG was unable to replicate the partial charges

of the GROMOS96 43A1 parameter set for the atoms of any of the amino acids, which are the most well-characterized species in most biomolecular force fields. Very little information is publicly available regarding the method by which PRODRG assigns charges to the input molecules, aside from a brief statement in the original paper stating that charges are referenced from a database of common functional groups (18). In fact, it seems that, in the case of uncharged molecules, PRODRG assigns charges that are a hybrid of polar and nonpolar groups. That is, polar groups (Asn and Ser sidechains, for example) are insufficiently polar, while normally uncharged groups (Ala and Ile sidechains, methylene groups in Lys) are assigned small partial charges, rendering them more hydrophilic. Some other polar groups, such as backbone amides in all the amino acids, have much larger dipoles assigned by PRODRG than would be expected under the GROMOS96 43A1 parameter set. In the case of charged groups, such as the carboxylate group of the Asp sidechain and the amino group of the Lys sidechain, the dipole on the carboxylate group is too large, while that of the amino group is too small.

UDP-Galactopyranose Mutase

The PRODRG server was designed to produce ligand topologies for use in modeling protein-ligand structures (18, 19). Thus, it is important to assess its applicability to this purpose. To design a test system for which ligand parameters are already present in the GROMOS96 43A1 force field library, we chose to model the UGM enzyme in complex with its FAD cofactor. Modeling other small molecules in protein binding sites would be an interesting exercise, but the lack of a clear method for generating their topologies has limited the focus of this study.

Over the course of triplicate 40-ns trajectories under both GROMOS96 43A1 parameters and those generated by PRODRG, the FAD molecule remained bound to the protein (which, in all cases, was assigned parameters under the GROMOS96 43A1 force field), with its position largely consistent over time. That is, there was no global translation of the cofactor out of its binding site, or rotation therein. However, we did observe a number of important differences with respect to the configuration of FAD and its interactions with the UGM enzyme. Structural overlays, taken from the final snapshot of each trajectory, are shown in Figure 7.12. While the position of the isoalloxazine ring was largely consistent across the simulations under both

GROMOS96 43A1 and PRODRG, the location of the adenine ring was more variable in the simulations utilizing PRODRG parameters.

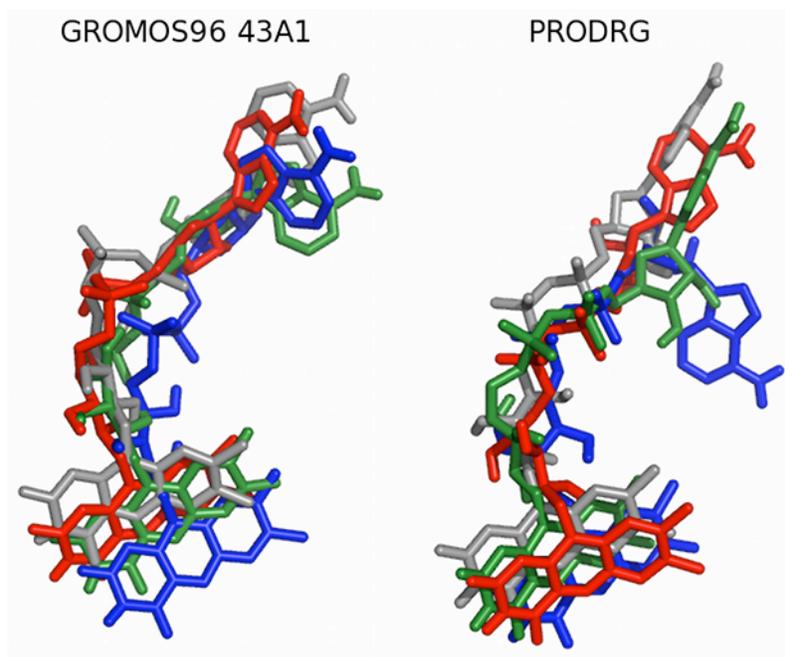


Figure 7.12. Alignment of FAD structures, with the reference position (the configuration of FAD following equilibration) for each simulation shown in gray. The final snapshot of each 40-ns simulation is shown and is colored by simulation. Overlays were produced by fitting the structure of the protein to remove global translation and rotation and extracting the resulting FAD coordinates. The images were generated with PyMOL (56).

The configurations generated in these simulations would seem to indicate that the PRODRG parameters assigned to FAD are sufficiently accurate for simulation. However, a conclusion regarding accuracy cannot be made based on structures alone. The nature of the interactions between FAD and UGM is an important criterion for assessing the accuracy of these parameters, and is also critically important in simulations of small molecules docked to receptors. MD simulations can be used to assess the stability of docking results and can be used to calculate the energetics of the resulting interactions over time.

We found that the parameters assigned to FAD by PRODRG led the cofactor to participate in significantly fewer hydrogen bonds to UGM. In the original crystal structure, eight hydrogen bonds between FAD and UGM were detected, using a cutoff angle of 30° and a cutoff radius of 3.5 \AA . Under GROMOS96 43A1, 10.3 ± 0.4 hydrogen bonds were formed (averaged over time

and over three replicate simulations). Using PRODRG parameters resulted in the formation of just 5.0 ± 0.9 hydrogen bonds. Several of these missing hydrogen bonds can likely be attributed to the assignment of zero charge to the atoms of the C⁶-amino group of the adenine ring (Figure 7.5). The absence of hydrogen bonding between this moiety and UGM also explains the variability of the position of the adenine ring across the three simulations.

MD simulations are also commonly used to predict binding energies between small molecules and their target proteins (44). By measuring non-bonded interaction energies between FAD and UGM, the nature of the interactions between the protein and its cofactor can be assessed. When GROMOS96 43A1 charges were assigned to FAD, the average total interaction energy between UGM and FAD was -866 ± 52 kJ mol⁻¹. When PRODRG-derived charges were assigned to FAD, the average interaction energy was -588 ± 30 kJ mol⁻¹. While the Lennard-Jones energies were similar in both simulation sets, the Coulombic energies differed drastically. With GROMOS96 43A1 charges, the magnitude of the intermolecular Coulombic interactions was, on average, -427 ± 36 kJ mol⁻¹, but with PRODRG charges, the same term was -158 ± 22 kJ mol⁻¹. Thus, despite the apparent similarity in the positions of FAD within the UGM enzyme, the energetics of this interaction were remarkably different, particularly with respect to the Coulombic interactions.

It is apparent that the inconsistent charge and charge group assignments generated by PRODRG had an adverse effect on the dynamics and behavior of the FAD cofactor in its binding site. However, it is even more of a concern that the PRODRG-assigned FAD charges might influence the dynamics of the protein itself. A measure of the root mean square fluctuation (RMSF) in UGM indicated that helix $\alpha 9$ and the loop following it (residues 215-224) significantly increased in fluctuation ($p < 0.05$) in one simulation in which the FAD molecule was assigned PRODRG charges (Figure 7.13). The average RMSF over the last 20 ns of simulation using GROMOS96 43A1 parameters for FAD was 0.18 ± 0.04 nm, while that of the simulation using PRODRG parameters was 0.26 ± 0.08 . This loop is important in binding the adenine moiety of the FAD, and we attribute the increased fluctuation to a lack of hydrogen bonding between UGM and FAD in this region. The two other trajectories using PRODRG charges for FAD resulted in reasonable configurations for the cofactor, similar to those obtained by assigning charges from GROMOS96 43A1. Nonetheless, we have demonstrated that the energetics of these interactions deviate substantially from the expected results using

GROMOS96 43A1. The implication that PRODRG topology charges extend beyond the small molecule for which they were originally intended raises concerns about the validity of any simulations that have used unrefined PRODRG topologies.

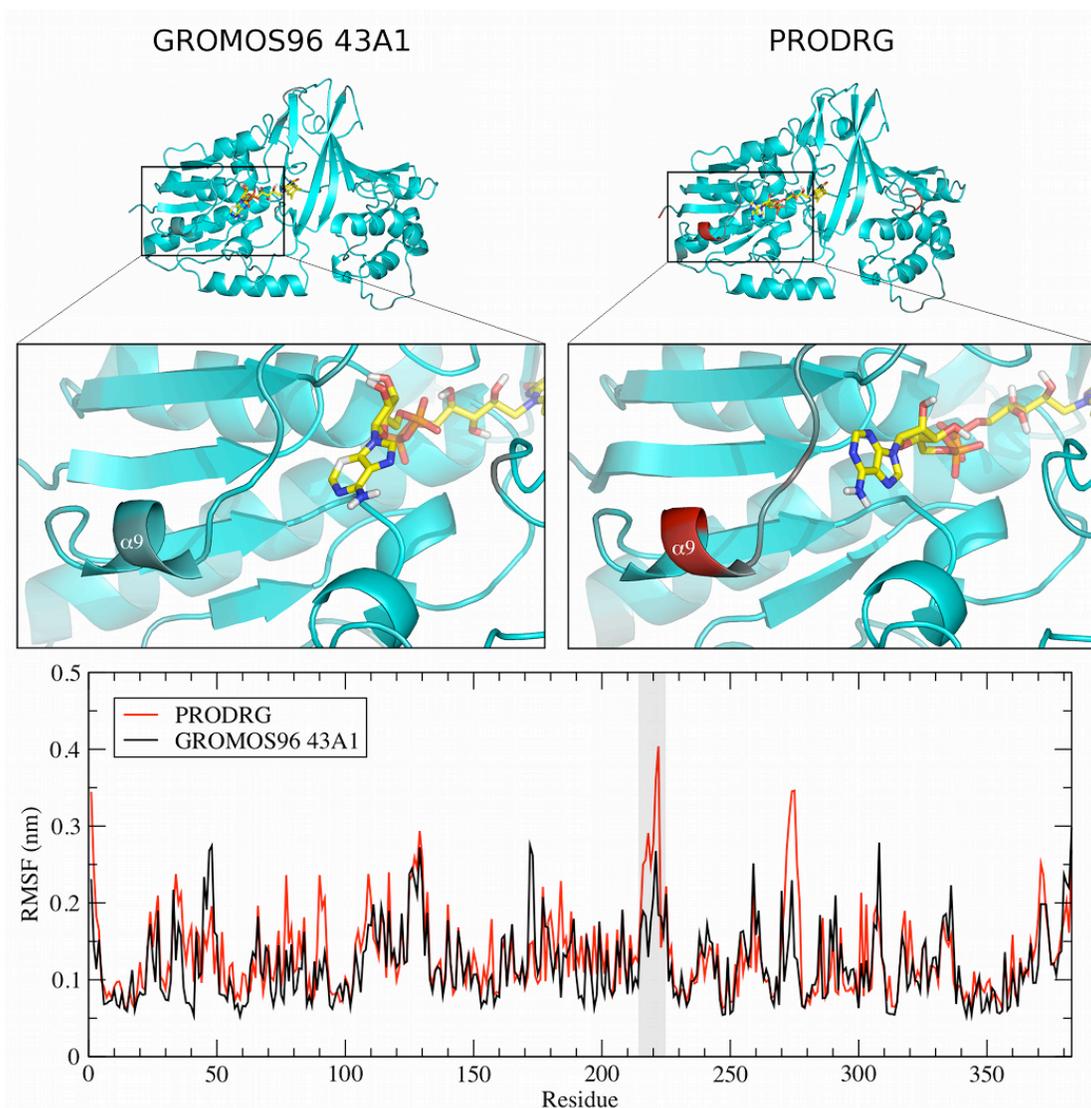


Figure 7.13. RMSF of the UGM backbone. Only the charges on the FAD molecule differ between the two models. The protein (cartoon) is colored according to the backbone RMSF, using a color gradient from blue to red, with blue regions indicating less fluctuation and red regions indicating more fluctuation. The FAD cofactor is shown as sticks and colored by element. Structures are taken from representative snapshots of each simulation set (GROMOS96 43A1 and PRODRG FAD parameters). The images were generated with PyMOL (56). The RMSF of residues referred to in the main text is indicated in the shaded box on the RMSF plot.

Strategies for Charge Assignment in GROMOS Topologies

The empirical charge assignment applied in the GROMOS96 force field parameter sets presents one of the principal challenges in deriving small molecule topologies. There is no complete description in the literature regarding the proper derivation of charges for molecules within the GROMOS96 parameter sets, aside from empirical fitting. It is for this reason that automated methods of charge calculation and assignment are so attractive. Having concluded that the charges assigned by PRODRG did not reproduce the expected behavior for even the most well-defined molecules and common functional groups, we offer strategies for initial charge assignment that are relatively fast and easily applied to GROMOS topologies. For application to small molecules in MD simulations, these initial charges should be evaluated and refined as necessary to reproduce proper condensed-phase behavior using the methodology described by the authors of the GROMOS96 parameter sets and perhaps the strategies employed here. The charge-derivation strategies we discuss here should be considered a starting point, not an ending point, for this process.

Quantum mechanical calculations are not explicitly stated as a source of GROMOS96 charges, although they present one method for deriving these parameters. We utilized the Antechamber program within AmberTools (version 1.0) (45) to calculate charges for all-atom versions of several of the small molecules studied here by applying several different semi-empirical charge calculation methods. The results of these calculations are summarized in Table 7.2. Overall, the charge calculation method that best reproduced the GROMOS96 43A1 partial charges is AM1-BCC (46, 47). Charges on O and H atoms of hydroxyl groups were in good agreement with expected force field values for ethanol, *p*-cresol, and serine. For primary alcohol functional groups in these molecules, if the AM1-BCC partial charges of the C and H atoms in the CH₂ group are combined, the resulting charge was similar to the force field value of +0.150 *e*, +0.1677 *e* in the case of ethanol and +0.1919 *e* in the case of serine.

Aromatic C and H atoms had charges of comparable magnitude as well, a result that was also produced by Mulliken charges (48). The Mulliken charges for amide groups were in closest agreement with the expected values in the GROMOS96 43A1 parameter set. The AM1-BCC method tended to over-polarize amide functional groups, but in all cases the direction of the dipole was correct. That is, none of the N atoms were assigned partial positive charges, as was

the case in all of the PRODRG amino acid topologies. All of the methods utilized here also had a tendency to place small partial charges on nonpolar groups, an inconsistency with the GROMOS96 43A1 force field shared by PRODRG. The Gasteiger method (49) produced charges that were in poorest agreement with expected GROMOS96 43A1 parameters.

Antechamber is freely available within the AmberTools distribution, but it is limited to semi-empirical calculations. Spartan '04 (Wavefunction, Inc., Irvine, CA) is a commercially-available quantum mechanical program that can perform some higher level of theory calculations, and we sought to determine what level of theory would be necessary to best reproduce GROMOS96 43A1 charges. We employed three methods: the semi-empirical AM1 method (50), the *ab initio* Hartree-Fock (HF) method (51) with the 6-31G** basis set, and a common *ab initio* density functional theory (DFT) hybrid method (B3LYP) (52), also with the 6-31G** basis set. Charges were derived using the CHELP algorithm (53) within Spartan, which is a type of electrostatic charge (ESP) method suitable for describing intermolecular interactions.

The charges calculated by the Spartan methods (Table 7.3) were, in general, all in good agreement with those charges found in the GROMOS96 43A1 force field. In the case of ethanol, the AM1 calculations yielded a near-neutral charge on the CH₃ group (-0.003 *e*), whereas the HF and DFT derived charges were somewhat more negative (-0.075 *e* and -0.057 *e*, respectively). The net charge in the CH₂ group calculated by the AM1 method (+0.240 *e*) was again closest to the GROMOS96 43A1 parameters (+0.150 *e*), and the HF- and DFT-derived charges were slightly greater (+0.341 *e* and +0.263 *e*, respectively). The magnitudes of the charges on the oxygen and hydrogen of the hydroxyl groups of all Spartan methods were in close agreement with the GROMOS96 43A1 charges, with the closest approximation coming from the DFT calculations. The important result is that the dipoles derived by these methods were greater in the hydroxyl group than in the methyl group, likely resulting in a correct hydrogen bonding behavior, which was not observed in the case of PRODRG-derived charges.

In the case of *p*-cresol, the *p*-methyl group was assigned a net negative charge by all methods, which was offset by the net positive charge on the carbon in the 4-position. The four equivalent carbons (at the 2-, 3-, 5-, and 6- positions) were all slightly over-polarized with the closest approximation to the GROMOS96 43A1 parameters coming from the AM1 calculations. The dipole on the hydroxyl group, however, was similar to the magnitude of this same group in the GROMOS96 43A1 parameter set in all cases, with the closest approximation derived by the

DFT method. It appears that the magnitude of the dipole on the hydroxyl group compared to that of the *p*-methyl group was large enough such that a reasonable pattern of hydrogen bonding behavior would result.

In the case of serine, many of the same patterns are observed. The net charges on the N- and C-terminal methyl groups were in close agreement with the accepted GROMOS96 43A1 charges, whereas both the N- and C-terminal amide atoms were calculated to have the correct sign, but with a slight over-polarization by all methods. The largest disparity among these methods was the charge calculated for the α -carbon. The AM1 method predicted a small negative charge ($-0.074 e$), while the HF and DFT methods both calculated large net positive charges ($+0.570 e$ and $+0.534 e$, respectively). The dipole on the hydroxyl group, however, was in close agreement with GROMOS96 43A1 parameters when calculated by all methods, with the closest being the AM1 method.

It is imperative to note that simple calculation of charges by any of the methods here would be insufficient for parameterizing new functional groups that are compatible with the GROMOS96 force fields. Topologies still must be validated using thermodynamic integration and an assessment of other condensed-phase criteria (14). We do believe, however, that the application of any of the aforementioned methods, especially the semi-empirical AM1 methods discussed here, provides a reasonable starting point for deriving charges for novel functional groups, taking much of the guesswork out of the empirical derivation strategy. Assignment of known functional group charges to small molecules is appropriate in the case of GROMOS96 parameter sets, as they have been designed to contain transferable functional groups that have the same parameters (with respect to partial charges), regardless of the remainder of the molecule.

Table 7.2. Charges calculated by Antechamber semi-empirical methods.

Molecule	Group	Atom	Partial Charges (<i>e</i>)				
			AMI-BCC	Mulliken	Gasteiger	GROMOS96	
Ethanol	CH ₃	C	-0.0971	-0.2150	-0.041839		
		H1	+0.0380	+0.0773	+0.025375		
		H2	+0.0481	+0.0874	+0.025375		
		H3	+0.0482	+0.0875	+0.025375		
		Sum	+0.0372	+0.0372	+0.034286	0.000	
	CH ₂	C	+0.1312	-0.0192	+0.040211		
		H1	+0.0183	+0.0576	+0.056075		
		H2	+0.0182	+0.0575	+0.056075		
		Sum	+0.1677	+0.0959	+0.152361	+0.150	
	OH	O	-0.6024	-0.3296	-0.396675	-0.548	
H		+0.3976	+0.1996	+0.210027	+0.398		
<i>p</i> -Cresol	CH ₃	C	-0.0487	-0.1739	-0.039771		
		H1	+0.0399	+0.0792	+0.027965		
		H2	+0.0437	+0.0830	+0.027965		
		H3	+0.0437	+0.0830	+0.027965		
		Sum	+0.0786	+0.0713	+0.044214	0.000	
	C	C	-0.1123	-0.1050	-0.050945	0.000	
	CH*	C	-0.1381	-0.1381	-0.044274	-0.100	
		H	+0.1371	+0.1371	+0.063842	+0.100	
	COH	C	+0.1187	+0.0736	+0.071255	+0.150	
		O	-0.4987	-0.2526	-0.360943	-0.548	
		H	+0.4179	+0.2169	+0.218239	+0.398	
	Serine	Acetyl CH ₃	C	-0.1766	-0.2445	+0.012705	
			H1	+0.0829	+0.1222	+0.032806	
H2			+0.0783	+0.1176	+0.032806		
H3			+0.0674	+0.1067	+0.032806		
Sum			+0.0520	+0.1029	+0.111123	0.000	
N-terminal amide		C	+0.6657	+0.3096	+0.209208	+0.380	
		O	-0.6022	-0.3631	-0.278059	-0.380	
		N	-0.5854	-0.3945	-0.301933	-0.280	
		H	+0.3153	+0.2288	+0.150258	+0.280	
CaH		C	+0.0447	+0.0180	+0.122426		
		H	+0.1033	+0.1426	+0.062991		
		Sum	+0.1480	+0.1606	+0.185417	0.000	
CbH ₂		C	+0.1238	-0.0266	+0.070315		
		H1	+0.0283	+0.0676	+0.058897		
		H2	+0.0398	+0.0791	+0.058897		
		Sum	+0.1919	+0.1201	+0.188109	+0.150	
OH		O	-0.5855	-0.3127	-0.393908	-0.548	
		H	+0.4101	+0.2091	+0.210159	+0.398	
C-terminal amide		C	+0.6454	+0.2893	+0.234251	+0.380	
		O	-0.6069	-0.3678	-0.275489	-0.380	
		N	-0.5679	-0.3770	-0.316830	-0.280	
		H	+0.3207	+0.2342	+0.149048	+0.280	
C-terminal N-CH ₃		C	+0.0873	-0.0680	-0.000395		
	H1	+0.0491	+0.0884	+0.043014			
	H2	+0.0269	+0.0662	+0.043014			
	H3	+0.0357	+0.0750	+0.043014			
	Sum	+0.1990	+0.1616	+0.128647	0.000		

*Averaged over all equivalent groups in the ring.

Table 7.3. Charges calculated by methods available in Spartan '04.

Molecule	Group	Atom	Partial Charges (<i>e</i>)			GROMOS96
			Semi-empirical AM1	Hartree-Fock 6-31G**	DFT (B3LYP) 6-31G**	
Ethanol	CH ₃	C	-0.437	-0.502	-0.504	
		H1	+0.154	+0.158	+0.162	
		H2	+0.154	+0.156	+0.161	
		H3	+0.126	+0.113	+0.124	
		Sum	-0.003	-0.075	-0.057	0.000
	CH ₂	C	+0.135	+0.344	+0.219	
		H1	+0.052	-0.001	+0.022	
		H2	+0.053	-0.002	+0.022	
		Sum	+0.240	+0.341	+0.263	+0.150
OH	O	-0.518	-0.668	-0.572	-0.548	
	H	+0.318	+0.402	+0.367	+0.398	
<i>p</i> -Cresol	CH ₃	C	-0.547	-0.758	-0.778	
		H1	+0.163	+0.205	+0.208	
		H2	+0.166	+0.211	+0.217	
		H3	+0.167	+0.213	+0.217	
		Sum	-0.051	-0.129	-0.136	0.000
	C	C	+0.075	+0.307	+0.304	0.000
	CH*	C	-0.211	-0.324	-0.277	-0.100
		H	+0.144	+0.211	+0.178	+0.100
	COH	C	+0.417	+0.479	+0.386	+0.150
		O	-0.524	-0.648	-0.566	-0.548
		H	+0.354	+0.442	+0.412	+0.398
	Serine	Acetyl CH ₃	C	-0.547	-0.623	-0.600
H1			+0.181	+0.205	+0.173	
H2			+0.162	+0.165	+0.199	
H3			+0.163	+0.178	+0.150	
		Sum	-0.041	-0.075	-0.078	0.000
N-terminal amide		C	+0.716	+0.875	+0.718	+0.380
		O	-0.556	-0.660	-0.562	-0.380
		N	-0.554	-0.852	-0.723	-0.280
		H	+0.291	+0.364	+0.331	+0.280
CaH		C	-0.065	+0.542	+0.500	
		H	+0.139	+0.028	+0.034	
		Sum	+0.074	+0.570	+0.534	0.000
CbH ₂		C	+0.042	-0.092	-0.161	
		H1	+0.078	+0.094	+0.102	
		H2	+0.088	+0.097	+0.100	
		Sum	+0.208	+0.099	+0.041	+0.150
OH		O	-0.537	-0.689	-0.607	-0.548
		H	+0.349	+0.453	+0.422	+0.398
C-terminal amide		C	+0.637	+0.598	+0.438	+0.380
		O	-0.553	-0.646	-0.544	-0.380
		N	-0.443	-0.490	-0.360	-0.280
		H	+0.299	+0.353	+0.320	+0.280
C-terminal N- CH ₃		C	-0.323	-0.410	-0.447	
		H1	+0.157	+0.171	+0.194	
	H2	+0.130	+0.194	+0.151		
	H3	+0.146	+0.146	+0.173		
	Sum	+0.110	+0.101	+0.071	0.000	

* Averaged over all equivalent groups in the ring.

Conclusions

We have demonstrated the effects of incorrect small molecule topologies in a variety of simulated systems. We conclude that the automated PRODRG 2.5 server is a valuable tool for the preparation of small molecule topologies for molecular simulation. It is fast, versatile, and easy to use. Bonded parameters and atom types assigned by PRODRG are correctly assigned in all cases. Unfortunately, the resulting topologies often suffer from deficiencies in the charges and charge groups that are assigned. These parameters often deviate substantially from the functional groups defined in the GROMOS96 43A1 parameter set, with the same functional groups in different molecules having different charges, leading to incorrect behavior in a variety of well-characterized systems. In general, we find that hydrophobic groups parameterized by PRODRG are inappropriately assigned partial charges, making them overly hydrophilic. As a consequence, the amino acid alanine inappropriately diffused into the aqueous solvent of our biphasic water-cyclohexane systems and the aliphatic portion of the lysine sidechain experienced overly-attractive van der Waals interactions with water. Dipoles in polar uncharged and charged species were somewhat unpredictable, as in the case of asparagine and aspartate, the sidechain dipole was too large, resulting in increased hydrogen bonding, while for lysine and serine, the dipole of the relevant polar group was too small, decreasing the hydrogen bonding capacity of these molecules. Application of small partial charges to hydrophobic moieties of *p*-cresol and ethanol led to substantially reduced hydrogen bonding capacity and altered liquid structural properties.

Perhaps the most troubling result of all comes from the simulations of the UGM-FAD complex. The PRODRG topology for FAD appeared to give rise to reasonable configurations for FAD, despite some increased flexibility in the adenine moiety. However, the nature and strength of the interactions between FAD and UGM were significantly different between the GROMOS96 43A1-generated FAD topology and the topology from the PRODRG server. Hydrogen bonds were significantly reduced, due in large part to a completely uncharged amine group, and the value of nonbonded interaction energies was drastically different. Further, the negative effects of the PRODRG charges and charge groups assignments extended to the dynamics of UGM itself, conferring greater flexibility to nearby residues relative to what should be expected with GROMOS96 43A1 parameters applied to FAD. Thus, the applicability of unrefined PRODRG topologies to high-throughput screening methods is questionable, at best.

The validity of any program that purports to produce small molecule topologies deserves some scrutiny, and the resulting topologies should always be rigorously validated before using these parameters in any simulation study. We caution all investigators to properly calculate partial charges for the molecules in their simulations, first by analogy to similar chemical entities already in the force field library, then by semi-empirical quantum mechanical calculations for any group not present in the force field library, and to always validate their small molecule topologies in accordance with the prescribed force field methodology. Topologies generated by PRODRG are increasingly used in diverse applications in the literature, but the results obtained may not be consistent with the parent force field, GROMOS96 43A1. That is, the results obtained may suffer from many of the limitations and inaccuracies we have reported here. This fact raises significant concerns about the widespread use of this automated tool in the absence of any further topology modification or validation. The original intent for the PRODRG server was to be used in conjunction with crystal structure refinement and energy minimization, and extrapolation to more complicated systems may not be appropriate in the absence of substantial topology refinement.

The group concept employed in the GROMOS force fields allows for versatile transferability of chemical moieties between molecules; that is, parameters of known functional groups can be applied to different molecules, independent of other structural differences between these species. Thus, it is reasonable to construct a topology for an arbitrary small molecule under the GROMOS force fields in a way that is consistent with the original force field derivation, with minimal effort, by piecing together existing building blocks. For functional groups that are not yet defined, fast charge calculation methods like AM1-BCC, for which extensive geometry optimization is not required, should suffice as a starting point for further topology refinement, but we emphasize that partial charges for any new chemical functional groups should be iteratively refined as necessary, and rigorously validated before use in production simulations.

Acknowledgments

The authors thank Michelle Oppenheimer for useful discussions regarding UGM, Prof. J. M. Tanko for allowing us the use of Spartan '04, and the Virginia Tech Terascale Computing

Facility for computing time on the SystemX supercomputer. The material is based upon work supported by the Macromolecular Interfaces with Life Sciences (MILES) Integrative Graduate Education and Research Traineeship (IGERT) of the National Science Foundation under Agreement No. DGE-0333378, and by the Institute for Critical Technology and Applied Science (ICTAS) at Virginia Tech.

Supplemental Information

Results of the Antechamber and Spartan '04 charge calculations for the five remaining amino acids are included below.

Table S7.1. Amino acid charges calculated by Antechamber semi-empirical methods (full).

Molecule	Group	Atom	Partial Charges (<i>e</i>)			
			AM1-BCC	Mulliken	Gasteiger	GROMOS96
Alanine	Acetyl CH ₃	C	-0.1760	-0.2439	+0.012703	
		H1	+0.0782	+0.1175	+0.032805	
		H2	+0.0657	+0.1050	+0.032805	0.000
		H3	+0.0797	+0.1190	+0.032805	
	N-terminal amide	C	+0.6647	+0.3086	+0.209132	+0.380
		O	-0.6045	-0.3654	-0.278060	-0.380
		N	-0.5889	-0.3980	-0.304237	-0.280
		H	+0.3170	+0.2305	+0.150160	+0.280
	C α H	C	+0.0409	+0.0142	+0.096349	
		H	+0.0891	+0.1284	+0.060201	0.000
	C β H ₃	C	-0.1083	-0.2262	-0.037773	
		H1	+0.0631	+0.1024	+0.025456	
		H2	+0.0547	+0.0940	+0.025456	0.000
		H3	+0.0406	+0.0799	+0.025456	
	C-terminal amide	C	+0.6266	+0.2705	+0.231571	+0.380
		O	-0.6072	-0.3681	-0.275604	-0.380
		N	-0.5638	-0.3729	-0.316918	-0.280
		H	+0.3124	+0.2259	+0.149047	+0.280
	C-terminal N-CH ₃	C	+0.0823	-0.0730	-0.000397	
		H1	+0.0407	+0.0800	+0.043014	
H2		+0.0363	+0.0756	+0.043014	0.000	
H3		+0.0568	+0.0961	+0.043014		
Asparagine	Acetyl CH ₃	C	-0.1735	-0.2414	+0.012704	
		H1	+0.0752	+0.1145	+0.032805	
		H2	+0.0623	+0.1016	+0.032805	0.000
		H3	+0.0614	+0.1007	+0.032805	
	N-terminal amide	C	+0.6557	+0.2996	+0.209155	+0.380
		O	-0.6149	-0.3758	-0.278060	-0.380
		N	-0.5534	-0.3625	-0.303433	-0.280
		H	+0.3234	+0.2369	+0.150190	+0.280
	C α H	C	-0.0286	-0.0553	+0.107680	
		H	+0.0998	+0.1391	+0.061037	0.000
	C β H ₂	C	+0.0614	+0.0328	+0.051398	
		H1	+0.0590	+0.0983	+0.038395	0.000
		H2	+0.0754	+0.1147	+0.038395	
	Sidechain amide	C	+0.5935	+0.2374	+0.211443	+0.380
		O	-0.6521	-0.4130	-0.277893	-0.380
		N	-0.6593	-0.4193	-0.329083	-0.830
		H1	+0.3313	+0.2268	+0.145484	+0.415
		H2	+0.3264	+0.2399	+0.145484	+0.415
	C-terminal amide	C	+0.6302	+0.2741	+0.232385	+0.380
		O	-0.6330	-0.3939	-0.275574	-0.380
N		-0.5693	-0.3784	-0.316895	-0.280	
H		+0.3390	+0.2525	+0.149047	+0.280	
C-terminal N-CH ₃	C	+0.0868	-0.0685	-0.000396		
	H1	+0.0381	+0.0774	+0.043014		
	H2	+0.0229	+0.0622	+0.043014	0.000	
	H3	+0.0603	+0.0996	+0.043014		

Aspartate	Acetyl CH ₃	C	-0.1932	-0.2619	+0.012704	
		H1	+0.0502	+0.0898	+0.032805	0.000
		H2	+0.1394	+0.1789	+0.032805	
		H3	+0.0499	+0.0888	+0.032805	
	N-terminal amide	C	+0.6678	+0.3120	+0.209153	
		O	-0.6685	-0.4291	-0.278060	-0.380
		N	-0.5656	-0.3748	-0.303570	-0.280
		H	+0.3217	+0.2351	+0.150187	+0.280
	C α H	C	+0.0653	+0.0390	+0.104340	0.000
		H	+0.0906	+0.1297	+0.060901	
	C β H ₂	C	-0.2249	-0.2537	+0.014249	0.000
		H1	+0.0513	+0.0905	+0.035313	
		H2	+0.0475	+0.0866	+0.035313	
	Sidechain carboxylate	C	+0.9029	+0.3223	+0.043753	+0.270
		O1	-0.8321	-0.5667	-0.550083	-0.635
		O2	-0.8471	-0.5818	-0.550083	-0.635
	C-terminal amide	C	+0.6527	+0.2963	+0.232248	+0.380
		O	-0.6593	-0.4202	-0.275577	-0.380
		N	-0.5642	-0.3734	-0.316897	-0.280
		H	+0.2999	+0.2136	+0.149047	+0.280
	C-terminal N-CH ₃	C	+0.0547	-0.0997	-0.000396	0.000
H1		+0.0240	+0.0635	+0.043014		
H2		+0.0197	+0.0592	+0.043014		
H3		+0.1173	+0.1559	+0.043014		
Isoleucine	Acetyl CH ₃	C	-0.1730	-0.2409	+0.012704	0.000
		H1	+0.0778	+0.1171	+0.032805	
		H2	+0.0619	+0.1012	+0.032805	
		H3	+0.0647	+0.1040	+0.032805	
	N-terminal amide	C	+0.6622	+0.3061	+0.209151	+0.380
		O	-0.6148	-0.3757	-0.278060	-0.380
		N	-0.5600	-0.3691	-0.303683	-0.280
		H	+0.3265	+0.2400	+0.150185	+0.280
	C α H	C	+0.0411	+0.0144	+0.101681	0.000
		H	+0.0979	+0.1372	+0.060787	
	C β H	C	-0.0669	-0.1062	-0.017161	0.000
		H	+0.0494	+0.0887	+0.032411	
	C γ H ₃	C	-0.0973	-0.2116	-0.060239	0.000
		H1	+0.0434	+0.0827	+0.023402	
		H2	+0.0353	+0.0746	+0.023402	
		H3	+0.0449	+0.0842	+0.023402	
	C γ H ₂	C	-0.0743	-0.1529	-0.051301	0.000
		H1	+0.0452	+0.0854	+0.026729	
		H2	+0.0416	+0.0809	+0.026729	
	C δ H ₃	C	-0.1046	-0.2225	-0.065045	0.000
		H1	+0.0527	+0.0920	+0.023047	
		H2	+0.0360	+0.0753	+0.023047	
		H3	+0.0291	+0.0684	+0.023047	
	C-terminal amide	C	+0.6375	+0.2814	+0.232134	+0.380
		O	-0.6166	-0.3775	-0.275579	-0.380
		N	-0.5736	-0.3827	-0.316889	-0.280
		H	+0.3141	+0.2276	+0.149047	+0.280
	C-terminal N-CH ₃	C	+0.0873	-0.0680	-0.000396	0.000
		H1	+0.0454	+0.0847	+0.043014	
		H2	+0.0394	+0.0787	+0.043014	
		H3	+0.0439	+0.0832	+0.043014	

Lysine	Acetyl CH ₃	C	-0.1762	-0.2399	+0.012705	
		H1	+0.0948	+0.1341	+0.032808	0.000
		H2	+0.0647	+0.1040	+0.032808	
		H3	+0.0597	+0.0990	+0.032808	
	N-terminal amide	C	+0.6577	+0.3016	+0.209136	
		O	-0.5695	-0.3304	-0.278068	-0.380
		N	-0.6005	-0.4096	-0.303951	-0.280
		H	+0.3257	+0.2392	+0.150176	+0.280
	C α H	C	+0.0523	+0.0256	+0.099156	0.000
		H	+0.0843	+0.1236	+0.060503	
	C β H ₂	C	-0.0858	-0.1644	-0.025604	0.000
		H1	+0.0693	+0.1086	+0.029088	
		H2	+0.0622	+0.1015	+0.029088	
	C γ H ₂	C	-0.0682	-0.1468	-0.046073	0.000
		H1	+0.0710	+0.1103	+0.026951	
		H2	+0.0542	+0.0935	+0.026951	
	C δ H ₂	C	-0.1081	-0.1867	-0.013759	0.000
		H1	+0.0646	+0.1039	+0.032059	
		H2	+0.0877	+0.1270	+0.032059	
	C ϵ H ₂	C	+0.1022	-0.1346	-0.038079	+0.127
		H1	+0.1066	+0.1459	+0.081283	
		H2	+0.1205	+0.1598	+0.081283	
	Sidechain amine	N	-0.8393	-0.0667	+0.231559	+0.129
		H1	+0.4575	+0.2527	+0.196024	+0.248
		H2	+0.4604	+0.2556	+0.196024	+0.248
		H3	+0.4646	+0.2598	+0.196024	+0.248
	C-terminal amide	C	+0.6356	+0.2795	+0.231852	+0.380
O		-0.6717	-0.4326	-0.275598	-0.380	
N		-0.5425	-0.3516	-0.316909	-0.280	
H		+0.3300	+0.2435	+0.149049	+0.280	
C-terminal N-CH ₃	C	+0.0735	-0.0818	-0.000399	0.000	
	H1	+0.0634	+0.1027	+0.043016		
	H2	+0.0554	+0.0947	+0.043016		
	H3	+0.0395	+0.0788	+0.043016		

Table S7.2. Charges calculated by methods available in Spartan '04 (full).

Molecule	Group	Atom	Partial Charges (<i>e</i>)			GROMOS96
			Semi-empirical AM1	Hartree-Fock 6-31G**	DFT (B3LYP) 6-31G**	
Alanine	Acetyl CH ₃	C	-0.538	-0.786	-0.757	0.000
		H1	+0.156	0.208	0.200	
		H2	+0.159	0.222	0.212	
		H3	+0.170	0.230	0.221	
	N-terminal amide	C	+0.721	+0.973	+0.816	+0.380
		O	-0.559	-0.672	-0.575	-0.380
		N	-0.569	-0.897	-0.775	-0.280
		H	+0.301	+0.399	+0.364	+0.280
	CaH	C	-0.062	+0.360	+0.334	0.000
		H	+0.139	+0.053	+0.052	
	CbH ₃	C	-0.373	-0.578	-0.570	0.000
		H1	+0.121	+0.158	+0.178	
		H2	+0.133	+0.149	+0.158	
		H3	+0.138	+0.184	+0.146	
	C-terminal amide	C	+0.605	+0.597	+0.437	+0.380
		O	-0.529	-0.607	-0.507	-0.380
		N	-0.451	-0.478	-0.345	-0.280
		H	+0.290	+0.332	+0.296	+0.280
	C-terminal N- CH ₃	C	-0.269	-0.365	-0.415	0.000
		H1	+0.142	+0.181	+0.186	
		H2	+0.142	+0.176	+0.181	
H3		+0.133	+0.160	+0.165		
Asparagine	Acetyl CH ₃	C	-0.579	-0.841	-0.804	0.000
		H1	+0.157	+0.212	+0.200	
		H2	+0.162	+0.220	+0.207	
		H3	+0.178	+0.240	+0.228	
	N-terminal amide	C	+0.708	+0.882	+0.732	+0.380
		O	-0.567	-0.635	-0.543	-0.380
		N	-0.477	-0.572	-0.449	-0.280
		H	+0.310	+0.367	+0.333	+0.280
	CaH	C	-0.252	-0.259	-0.301	0.000
		H	+0.210	+0.211	+0.200	
	CbH ₂	C	-0.433	-0.493	-0.458	0.000
		H1	+0.169	+0.227	+0.209	
		H2	+0.209	+0.171	+0.159	
	Sidechain amide	C	+0.670	+0.889	+0.726	+0.380
		O	-0.563	-0.641	-0.541	-0.380
		N	-0.662	-1.013	-0.872	-0.830
		H1	+0.327	+0.428	+0.411	+0.415
		H2	+0.298	+0.450	+0.392	+0.415
	C-terminal amide	C	+0.696	+0.736	+0.600	+0.380
		O	-0.558	-0.590	-0.500	-0.380
		N	-0.449	-0.483	-0.354	-0.280
H		+0.304	+0.352	+0.318	+0.280	
C-terminal N- CH ₃	C	-0.243	-0.323	-0.383	0.000	
	H1	+0.121	+0.161	+0.174		
	H2	+0.131	+0.139	+0.149		
	H3	+0.134	+0.165	+0.167		

Aspartate	Acetyl CH ₃	C	-0.521	-0.817	-0.779	
		H1	+0.170	+0.200	+0.187	0.000
		H2	+0.138	+0.225	+0.211	
		H3	+0.147	+0.213	+0.199	
	N-terminal amide	C	+0.695	+0.976	+0.810	+0.380
		O	-0.602	-0.705	-0.614	-0.380
		N	-0.557	-0.974	-0.817	-0.280
		H	+0.318	+0.419	+0.374	+0.280
	CaH	C	+0.164	+0.736	+0.629	0.000
		H	+0.126	+0.039	+0.047	
	CbH ₂	C	-0.585	-0.607	-0.571	
		H1	+0.168	+0.142	+0.132	0.000
		H2	+0.157	+0.146	+0.135	
	Sidechain carboxylate	C	+0.723	+0.847	+0.667	+0.270
		O1	-0.744	-0.787	-0.686	-0.635
		O2	-0.775	-0.840	-0.731	-0.635
	C-terminal amide	C	+0.577	+0.473	+0.348	+0.380
		O	-0.597	-0.659	-0.567	-0.380
		N	-0.423	-0.483	-0.357	-0.280
		H	+0.291	+0.348	+0.310	+0.280
	C-terminal N-CH ₃	C	-0.319	-0.347	-0.379	
H1		+0.136	+0.149	+0.146	0.000	
H2		+0.102	+0.099	+0.105		
H3		+0.212	+0.207	+0.200		
Isoleucine	Acetyl CH ₃	C	-0.581	-0.886	-0.844	
		H1	+0.163	+0.226	+0.214	0.000
		H2	+0.183	+0.250	+0.206	
		H3	+0.155	+0.218	+0.238	
	N-terminal amide	C	+0.697	+0.969	+0.812	+0.380
		O	-0.561	-0.657	-0.561	-0.380
		N	-0.505	-0.709	-0.578	-0.280
		H	+0.307	+0.400	+0.366	+0.280
	CaH	C	-0.082	-0.112	-0.159	0.000
		H	+0.130	+0.122	+0.114	
	CbH	C	-0.002	+0.355	+0.335	0.000
		H	+0.088	+0.001	+0.005	
	C _γ H ₃	C	-0.499	-0.851	-0.798	
		H1	+0.145	+0.206	+0.196	0.000
		H2	+0.146	+0.220	+0.209	
		H3	+0.164	+0.208	+0.195	
	C _γ H ₂	C	-0.148	+0.017	0.000	
		H1	+0.094	+0.029	+0.062	0.000
		H2	+0.104	+0.059	+0.032	
	C _δ H ₃	C	-0.483	-0.636	-0.613	
		H1	+0.171	+0.154	+0.180	0.000
		H2	+0.136	+0.190	+0.152	
		H3	+0.141	+0.154	+0.157	
	C-terminal amide	C	+0.627	+0.740	+0.587	+0.380
		O	-0.545	-0.610	-0.512	-0.380
		N	-0.476	-0.588	-0.447	-0.280
		H	+0.288	+0.354	+0.324	+0.280
	C-terminal N-CH ₃	C	-0.257	-0.306	-0.372	
		H1	+0.134	+0.162	+0.174	0.000
H2		+0.129	+0.150	+0.160		
H3		+0.139	+0.164	+0.170		

Lysine	Acetyl CH ₃	C	-0.599	-0.681	-0.667	
		H1	+0.154	+0.221	+0.218	
		H2	+0.183	+0.211	+0.207	
		H3	+0.191	+0.165	+0.153	
	N-terminal amide	C	+0.703	+0.840	+0.691	+0.380
		O	-0.522	-0.626	-0.527	-0.380
		N	-0.563	-0.803	-0.684	-0.280
		H	+0.316	+0.404	+0.371	+0.280
	CaH	C	-0.056	+0.294	+0.255	0.000
		H	+0.112	+0.041	+0.042	
	CbH ₂	C	-0.198	-0.252	-0.236	
		H1	+0.128	+0.116	+0.128	0.000
		H2	+0.129	+0.133	+0.110	
	CγH ₂	C	-0.303	-0.322	-0.308	
		H1	+0.162	+0.124	+0.125	0.000
		H2	+0.135	+0.123	+0.122	
	CδH ₂	C	-0.179	-0.118	-0.121	
		H1	+0.124	+0.088	+0.092	0.000
		H2	+0.113	+0.104	+0.102	
	CεH ₂	C	-0.186	+0.173	+0.097	
		H1	+0.156	+0.083	+0.096	+0.127
		H2	+0.152	+0.098	+0.108	
	Sidechain amine	N	-0.018	-0.701	-0.588	+0.129
		H1	+0.264	+0.422	+0.395	+0.248
		H2	+0.270	+0.413	+0.387	+0.248
		H3	+0.264	+0.414	+0.387	+0.248
	C-terminal amide	C	+0.631	+0.623	+0.470	+0.380
		O	-0.551	-0.566	-0.475	-0.380
N		-0.484	-0.588	-0.452	-0.280	
H		+0.305	+0.367	+0.338	+0.280	
C-terminal N-CH ₃	C	-0.241	-0.281	-0.335		
	H1	+0.144	+0.167	+0.172	0.000	
	H2	+0.139	+0.169	+0.177		
	H3	+0.127	+0.144	+0.152		

References

- (1) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz Jr., K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A., A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179-5197.
- (2) Wang, J.; Cieplak, P.; Kollman, P. A., How Well Does a Restrained Electrostatic Potential (RESP) Model Perform in Calculating Conformational Energies of Organic and Biological Molecules? *J. Comput. Chem.* **2000**, *21*, 1049-1074.
- (3) Sorin, E. J.; Pande, V. S., Exploring the Helix-Coil Transition via All-Atom Equilibrium Ensemble Simulations. *Biophys. J.* **2005**, *88*, 2472-2493.
- (4) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P., A Point-Charge Force Field for Molecular Mechanics Simulations of Proteins Based on Condensed-Phase Quantum Mechanical Calculations. *J. Comput. Chem.* **2003**, *24*, 1999-2012.
- (5) Neria, E.; Fischer, S.; Karplus, M., Simulation of activation free energies in molecular systems. *J. Chem. Phys.* **1996**, *105*, 1902.
- (6) MacKerell Jr., A. D.; Bashford, D.; Bellott, M.; Dunbrack Jr., R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher III, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M., All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586-3616.
- (7) MacKerell Jr., A. D.; Feig, M.; Brooks III, C. L., Extending the Treatment of Backbone Energetics in Protein Force Fields: Limitations of Gas-Phase Quantum Mechanics in Reproducing Protein Conformational Distributions in Molecular Dynamics Simulations. *J. Comput. Chem.* **2004**, *25*, 1400-1415.
- (8) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J., Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225-11236.

- (9) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L., Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. *J. Phys. Chem. B* **2001**, *105*, 6474-6487.
- (10) Scott, W. R. P.; Hünenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennen, J.; Torda, A. E.; Huber, T.; Krüger, P.; van Gunsteren, W. F., The GROMOS Biomolecular Simulation Program Package. *J. Phys. Chem. A* **1999**, *103*, 3596-3607.
- (11) van Gunsteren, W. F.; Billeter, S. R.; Eising, A. A.; Hünenberger, P. H.; Mark, A. E.; Scott, W. R. P.; Tironi, I. G., *Biomolecular Simulation: The GROMOS96 Manual and User Guide*. Verlag de Fachvereine, Zürich, Switzerland, 1996.
- (12) Daura, X.; Mark, A. E.; van Gunsteren, W. F., Parametrization of Aliphatic CH_n United Atoms of GROMOS96 Force Field. *J. Comput. Chem.* **1997**, *19*, 535-547.
- (13) Schuler, L. D.; Daura, X.; van Gunsteren, W. F., An Improved GROMOS96 Force Field for Aliphatic Hydrocarbons in the Condensed Phase. *J. Comput. Chem.* **2001**, *22*, 1205-1218.
- (14) Oostenbrink, C.; Villa, A.; Mark, A. E.; van Gunsteren, W. F., A Biomolecular Force Field Based on the Free Enthalpy of Hydration and Solvation: The GROMOS Force-Field Parameter Sets 53A5 and 53A6. *J. Comput. Chem.* **2004**, *25*, 1656-1676.
- (15) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Development and testing of a general Amber force field. *J. Comput. Chem.* **2004**, *25*, 1157-1174.
- (16) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; MacKerell Jr., A. D., CHARMM General Force Field: A Force Field for Drug-Like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. *J. Comput. Chem.* **2010**, *31*, 671-690.
- (17) Kandt, C.; Ash, W. L.; Tieleman, D. P., Setting up and running molecular dynamics simulations of membrane proteins. *Methods* **2007**, *41*, 475-488.
- (18) van Aalten, D. M. F.; Bywater, R.; Findlay, J. B. C.; Hendlich, M.; Hooft, R. W. W., PRODRG, a program for generating molecular topologies and unique molecular descriptors from coordinates of small molecules. *J. Comput.-Aided Mol. Des.* **1996**, *10*, 255-265.

- (19) Schüttelkopf, A. W.; van Aalten, D. M. F., PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr., Sect D: Biol. Crystallogr.* **2004**, *D60*, 1355-1363.
- (20) Yang, C.; Zhu, X.; Li, J.; Shi, R., Exploration of the mechanism for LPFFD inhibiting the formation of β -sheet conformation of A β (1-42) in water. *J. Mol. Model.* **2010**, *16*, 813-821.
- (21) Li, J.; Zhu, X.; Yang, C.; Shi, R., Characterization of the binding of angiotensin II receptor blockers to human serum albumin using docking and molecular dynamics simulation. *J. Mol. Model.* **2010**, *16*, 789-798.
- (22) Sharma, M.; Khanna, S.; Bulusu, G.; Mitra, A., Comparative modeling of thioredoxin glutathione reductase from *Schistosoma mansoni*: A multifunctional target for antischistosomal therapy. *J. Mol. Graphics Model.* **2009**, *27*, 665-675.
- (23) Levin, L. B.-A.; Ganoth, A.; Amram, S.; Nachliel, E.; Gutman, M.; Tsfadia, Y., Insight into the interaction sites between fatty acid binding proteins and their ligands. *J. Mol. Model.* **2010**, *16*, 929-938.
- (24) Boggara, M. B.; Faraone, A.; Krishnamoorti, R., Effect of pH and Ibuprofen on the Phospholipid Bilayer Bending Modulus. *J. Phys. Chem. B* **2010**, *114*, 8061-8066.
- (25) Cordoní, A.; Prades, J.; Frau, J.; Vögler, O.; Funari, S. S.; Perez, J. J.; Esribá, P. V.; Barceló, F., Interactions of fatty acids with phosphatidylethanolamine membranes: X-ray diffraction and molecular dynamics studies. *J. Lipid Res.* **2010**, *51*, 1113-1124.
- (26) Boggara, M. B.; Krishnamoorti, R., Partitioning of Nonsteroidal Antiinflammatory Drugs in Lipid Membranes: A Molecular Dynamics Simulation Study. *Biophys. J.* **2010**, *98*, 586-595.
- (27) Liu, F.-F.; Ji, L.; Dong, X.-Y.; Sun, Y., Molecular Insight into the Inhibition Effect of Trehalose on the Nucleation and Elongation of Amyloid β -Peptide Oligomers. *J. Phys. Chem. B* **2009**, *113*, 11320-11329.
- (28) Turner, D. C.; Yin, F.; Kindt, J. T.; Zhang, H., Molecular Dynamics Simulations of Glycocholate-Oleic Acid Mixed Micelle Assembly. *Langmuir* **2010**, *26*, 4687-4692.
- (29) Cilpa, G.; Hyvönen, M. T.; Koivuniemi, A.; Riekkola, M.-L., Atomistic Insight into Chondroitin-6-Sulfate Glycosaminoglycan Chain Through Quantum Mechanics

- Calculations and Molecular Dynamics Simulations. *J. Comput. Chem.* **2010**, *31*, 1670-1680.
- (30) Gandhi, N. S.; Mancera, R. L., Can current force fields reproduce ring puckering in 2-O-sulfo- α -L-iduronic acid? A molecular dynamics simulation study. *Carbohydr. Res.* **2010**, *345*, 689-695.
- (31) Sanjeeva, R.; Weerasinghe, S., Development of a molecular mechanics force field for caffeine to investigate the interactions of caffeine in different solvent media. *J. Mol. Struct. THEOCHEM* **2010**, *944*, 116-123.
- (32) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E., GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435-447.
- (33) Hess, B., P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 116-122.
- (34) Darden, T.; York, D.; Pedersen, L., Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089-10092.
- (35) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G., A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577-8593.
- (36) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R., Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684-3690.
- (37) Nosé, S., A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.* **1984**, *81*, 511-519.
- (38) Hoover, W. G., Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A: At. Mol. Opt. Phys.* **1985**, *31*, 1695-1697.
- (39) Parrinello, M.; Rahman, A., Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182-7190.
- (40) Nosé, S.; Klein, M. L., Constant pressure molecular dynamics for molecular systems. *Mol. Phys.* **1983**, *50*, 1055-1076.
- (41) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J., Interaction models for water in relation to protein hydration. In *Intermolecular Forces*, Pullman, B., Ed. Reidel, Dordrecht, **1981**.

- (42) Beis, K.; Srikannathasan, V.; Liu, H.; Fullerton, S. W. B.; Bamford, V. A.; Sanders, D. A. R.; Whitfield, C.; McNeil, M. R.; Naismith, J. H., Crystal Structures of *Mycobacteria tuberculosis* and *Klebsiella pneumoniae* UDP-Galactopyranose Mutase in the Oxidised State and *Klebsiella pneumoniae* UDP-Galactopyranose Mutase in the (Active) Reduced State. *J. Mol. Biol.* **2005**, *348*, 971-982.
- (43) Zhang, C.; Yang, X., Molecular dynamics simulation of ethanol/water mixtures for structure and diffusion properties. *Fluid Phase Equilib.* **2005**, *231*, 1-10.
- (44) Åqvist, J.; Luzhkov, V. B.; Brandsal, B. O., Ligand Binding Affinities from MD Simulations. *Acc. Chem. Res.* **2002**, *35*, 358-365.
- (45) Case, D. A.; Cheatham III, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Jr., K. M. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J., The Amber Biomolecular Simulation Programs. *J. Comput. Chem.* **2005**, *26*, 1668-1688.
- (46) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I., Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: I. Method. *J. Comput. Chem.* **2000**, *21*, 132-146.
- (47) Jakalian, A.; Jack, D. B.; Bayly, C. I., Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: II. Parameterization and Validation. *J. Comput. Chem.* **2002**, *23*, 1623-1641.
- (48) Mulliken, R. S., A New Electroaffinity Scale; Together with Data on Valence States and on Valence Ionization Potentials and Electron Affinities. *J. Chem. Phys.* **1934**, *2*, 782-792.
- (49) Gasteiger, J.; Marsili, M., Iterative partial equalization of orbital electronegativity - a rapid access to atomic charges. *Tetrahedron* **1980**, *36*, 3219-3228.
- (50) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P., AM1: A New General Purpose Quantum Mechanical Molecular Model. *J. Am. Chem. Soc.* **1985**, *107*, 3902-3909.
- (51) Szabo, A.; Ostlund, N. S., *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*. Dover Publications: New York, **1996**.
- (52) Becke, A. D., Density-Functional Thermochemistry. III. The Role of Exact Change. *J. Chem. Phys.* **1993**, *98*, 5648-5652.

- (53) Chirlian, L. E.; Francl, M. M., Atomic charges derived from electrostatic potentials: a detailed study. *J. Comput. Chem.* **1987**, *8*, 894-905.
- (54) Lide, D. R., *CRC Handbook of Chemistry and Physics*. 84th ed.; CRC Press LLC: Boca Raton, **2004**.
- (55) Richard, L. S.; Bernardes, C. E. S.; Diogo, H. P.; Leal, J. P.; Minas da Piedade, M. E., Energetics of Cresols and of Methylphenoxy Radicals. *J. Phys. Chem. A* **2007**, *111*, 8741-8748.
- (56) DeLano, W. L., *The PyMOL Molecular Graphics System*. DeLano Scientific LLC: Palo Alto, CA, **2009**.

Chapter 8

GridMAT-MD: A Grid-Based Membrane Analysis Tool for Use with Molecular Dynamics

William J. Allen, Justin A. Lemkul, and David R. Bevan

Reproduced with permission from *Journal of Computational Chemistry* **2009**, 30 (12), 1952-1958. Copyright 2009, Wiley.

Attribution

Of the work discussed within this chapter, Allen wrote the program in its entirety, Lemkul performed the validation simulations with the KALP peptide, and Allen and Lemkul contributed roughly equally to the writing and preparation of the manuscript. Bevan significantly contributed to the editing and review of the manuscript, as well as provided guidance in the research.

Abstract

GridMAT-MD is a new program developed to aid in the analysis of lipid bilayers from molecular dynamics simulations. It reads a GROMACS coordinate file and generates two types of data: a two-dimensional contour plot depicting membrane thickness, and a polygon-based tessellation of the individual lipid headgroups. GridMAT-MD can also account for proteins or small molecules within the headgroups of the lipids, closely approximating their occupied lateral area. The program requires no installation, is fast, and is freely available.

Introduction

Molecular dynamics (MD) simulations of membranes and membrane proteins have proven useful to investigate structure-function relationships, and they provide researchers a unique perspective at the atomic level, one that cannot be easily seen experimentally (1-5). Converting three-dimensional coordinate files, however, to useful figures for analysis and publication can be challenging, so concise methods of visualization are needed. GridMAT-MD is a free, open source program that has been developed to conduct the analysis of two important lipid bilayer parameters: thickness and area per lipid headgroup.

Three-dimensional coordinate files of membranes and membrane proteins can be rendered such that they produce detailed figures, but comparison among multiple images is time-consuming and difficult. A two-dimensional rendering of a three-dimensional structure also has its pitfalls. For example, trends in the thickness of the bilayer can be hard to distinguish, especially near the core of the bilayer where the view is obstructed by other lipid atoms. The area per lipid head group is simple enough to calculate for a pure-lipid model system. It is the lateral surface area of the lipid bilayer divided by the number of lipids in a given leaflet. However, this value can be difficult to obtain when there is a protein present. The lateral area of ideal proteins such as helices or barrels can be estimated, but irregularly shaped proteins are a challenge.

GridMAT-MD was designed to facilitate analysis of lipid bilayer thickness and area per lipid headgroup. On characterizing bilayer thickness, the program generates the data for a top-down perspective of the lipid bilayer, which can be colored according to the thickness of the bilayer using an external viewer program, much like a contour map. From this view, it is easier to examine fluctuations in bilayer dimensions across the unit cell. GridMAT-MD also has the ability to not only calculate the area per lipid headgroup when proteins are present, but it will also report the specific lateral area for each lipid from a polygon-based tessellation, and the lateral area for the protein.

GridMAT-MD is designed to readily interface with coordinate files generated as part of the output of an MD simulation conducted with GROMACS (6). Current GROMACS analysis tools for lipid bilayers include those that analyze deuterium order parameters, center of mass distances, and densities, all of which are commonly used in the literature when analyzing the results of membrane simulations. The limitation in reporting center of mass distances, averaged

over time, is that the calculation is often applied to a large group of atoms, such as phosphorus, limiting the level of detail that can be obtained through this analysis.

A method for analyzing bilayer thickness exists and has been successfully applied to large, coarse-grain membrane protein systems (7). The method relies on a fitting algorithm and coordinate extraction and processing from external software, working well for the large system in that study. However, its applicability to smaller systems is limited (X. Periole and T. Huber, personal communication). Thus, we sought to design a generally-applicable tool to analyze bilayer thickness.

Some investigators have rightly noted that it is difficult to calculate the explicit lateral area of irregularly shaped proteins in a membrane (8, 9). An essential tool that is still missing in bilayer analysis is one that can efficiently calculate the area per lipid headgroup, even in the presence of an embedded protein. To our knowledge, no such tool is freely and generally available. Existing methods for conducting such calculations have been reported on various online forums (e.g., http://www.ks.uiuc.edu/Research/vmd/mailling_list/), but often require extensive scripting knowledge and commercial software.

GridMAT-MD is a superior tool for membrane analysis because (i) it is free to use and modify under the terms of the GNU public license (10), (ii) no installation is required, provided the Perl interpreter (11) is installed, (iii) it calculates an explicit area for proteins, making no assumptions of “ideal proteins,” and (iv) it is very fast to run and easy to operate.

Description

Only two files are needed as input for this program; a GROMACS-format coordinate file that contains a lipid bilayer that is oriented normal to the z -axis, and a parameter file that contains user-specified information about the coordinate file, desired output, and other parameters, an example of which is distributed with the program. A User’s Guide is also distributed with the program that provides clear definitions of all of the options in the parameter file, as well as detailed descriptions on how to run specific analyses. In this section, the logic of the program and the type of data it generates will be discussed.

Processing the Coordinate File

A flowchart of the GridMAT-MD algorithm is shown in Figure 8.1. Initially, GridMAT-MD parses through the GROMACS coordinate file in order to separate all of its components. The lipid and solvent atoms are stored in separate arrays for later processing. The ions are deleted, and anything remaining is treated as a “protein” atom, which can more broadly be interpreted as a protein, small molecule, or other macromolecule. The first two lines in the GROMACS coordinate file (the title and the number of atoms) are discarded, but the last line (box vectors) is stored for later use.

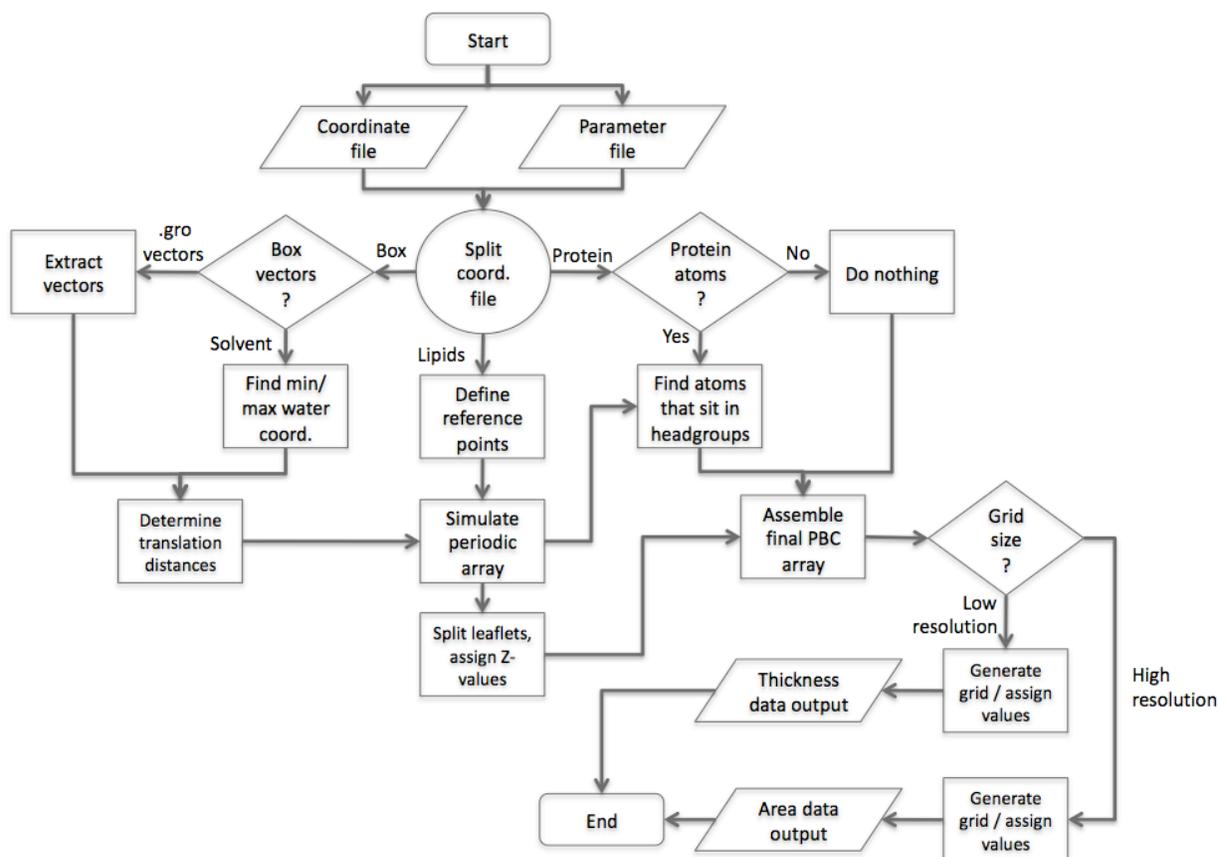


Figure 8.1. Flowchart of the GridMAT-MD logic structure.

Not all of the lipid atoms are needed to determine the thickness of the membrane, so most can be deleted. Those that the user selects are preserved, and can range from one atom per lipid

(such as phosphorus), to several (such as the glycerol backbone). In the case of a user choosing multiple atoms, the geometric center of mass is determined and the original coordinates are discarded. Ultimately, the program stores one point in space per lipid (a “reference point”), representative of the lipid to which it belongs. After that, the reference points are further divided into an upper and a lower leaflet.

The next step generates a 3x3 periodic array in the x - and y -directions of the reference points. The goal of this step is to have a central image for analysis that is immediately surrounded by eight identical images, since periodic boundary conditions are routinely applied in almost all MD simulations. The unit-cell translation distances for generating such an array can come from one of two sources, the upper and lower x - and y -limits of the box of solvent, or the last line of the GROMACS coordinate file – the box vectors. Since post-processing of coordinates may be necessary to account for effects such as periodicity, centering, and structural fitting, among others, sometimes the box vectors printed to the coordinate file may not reflect the actual dimensions of the system. Thus, we implemented a feature that allows the user to choose how the system size should be determined – from the printed box vectors, or from the dimensions of the solvent, which are often representative of a rectangular unit cell.

Once a periodic array of nine images is built, the next step is to determine which protein atoms are contained within the lateral area of the lipid headgroups (Figure 8.2). For each protein atom, the program first finds all lipid reference points within a user-defined radius of that protein atom on the x - y plane. If at least one of those reference points has a greater z -coordinate value and one has a lower z -coordinate value than the protein atom, that protein atom is said to fall within the lipid headgroups. This process is repeated for each protein atom, and the total array of those protein atoms that fall within the lipid headgroups is also translated into a 3x3 periodic array in the x - and y -directions.

Calculating Thickness

The most important concept in calculating the thickness of the bilayer is the Z -value. Each reference point in the top and the bottom leaflet is assigned a Z -value that represents the thickness (in nanometers) of the bilayer at that point. In order to determine the Z -value, the program considers one reference point at a time, starting in the top leaflet. It will search for the

nearest neighbor of that reference point in the opposite leaflet, in just the x - and y -directions. Then it will take the difference of the z -coordinates between the two reference points, and that value becomes the Z -value. A Z -value is calculated for every reference point in the top leaflet, then this process is repeated for every point in the bottom leaflet. The user defines an arbitrary Z -value for the protein atoms. The goal here is not to try to represent the thickness of the bilayer with the protein atoms, but just to set them apart. So if most of the bilayer will have thickness values ranging from 3 to 4 nm, the Z -value of the protein might be set to 10. This way, protein atoms stand out in contrast against the lipid reference points.

Once the Z -values have been calculated, a grid is aligned to each leaflet in the center image of the periodic array. The resolution and shape of the grid are also user-defined. Each grid point in the top grid then searches for the nearest reference atom in the top leaflet, in just the x - and y -directions. Once found, that grid point is assigned the Z -value from that reference atom. This process is repeated for the bottom grid in the bottom leaflet. The periodicity of the bilayer allows reference atoms on one side of the system to be “seen” by grid points on the opposite side of the system, as would be the case in a molecular dynamics simulation using true periodic boundary conditions. The results of this process are two similar matrices of data: the thickness of the bilayer from the top leaflet down, and the thickness of the bilayer from the bottom leaflet up. Each matrix can be visualized using an external viewer by assigning a color to each grid point depending on the magnitude of the Z -value, easily distinguishing thin and thick parts of the lipid bilayer (Figure 8.3). Visually, the top and bottom matrices typically look very similar with some minor differences, so the program averages the data in order to give the best representation. For typical thickness calculations, we have generated ideal images using grids with points spaced every 3-4 Angstroms. This corresponds to a grid of about 20x20 points for a bilayer system with 50-60 lipids per leaflet.

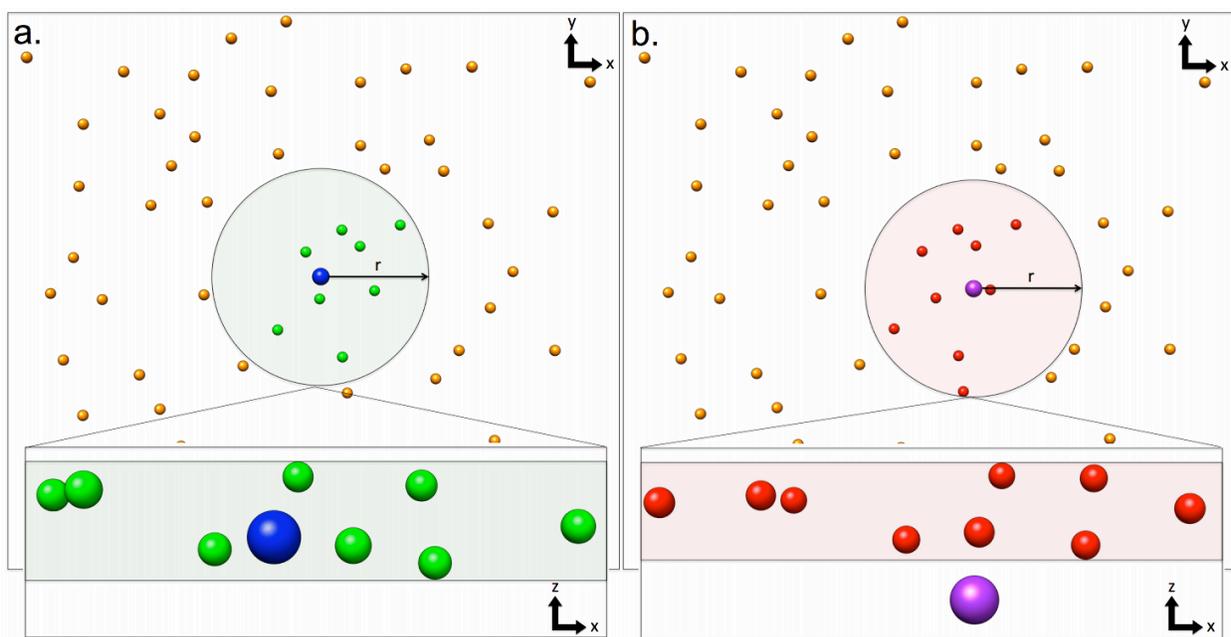


Figure 8.2. Diagram illustrating GridMAT-MD decision structure in choosing protein atoms. Atoms represented using USCF Chimera (23). (a.) A top-down view of the top leaflet. The phosphorus atoms are the reference points and are shown as orange spheres. A protein atom is depicted as a blue sphere. Within a given radius, 8 reference points are considered. (insert) A side view of the top leaflet reference atoms. The protein atom falls within the upper and lower z-coordinates of the reference atoms, so its coordinates are saved. (b.) A top-down view of the top leaflet, with a different atom represented as a purple sphere. In this case, 9 reference points fall within the search radius. (insert) From the side view, one can see that the protein atom does not fall within the lipid headgroups. Its coordinates are discarded.

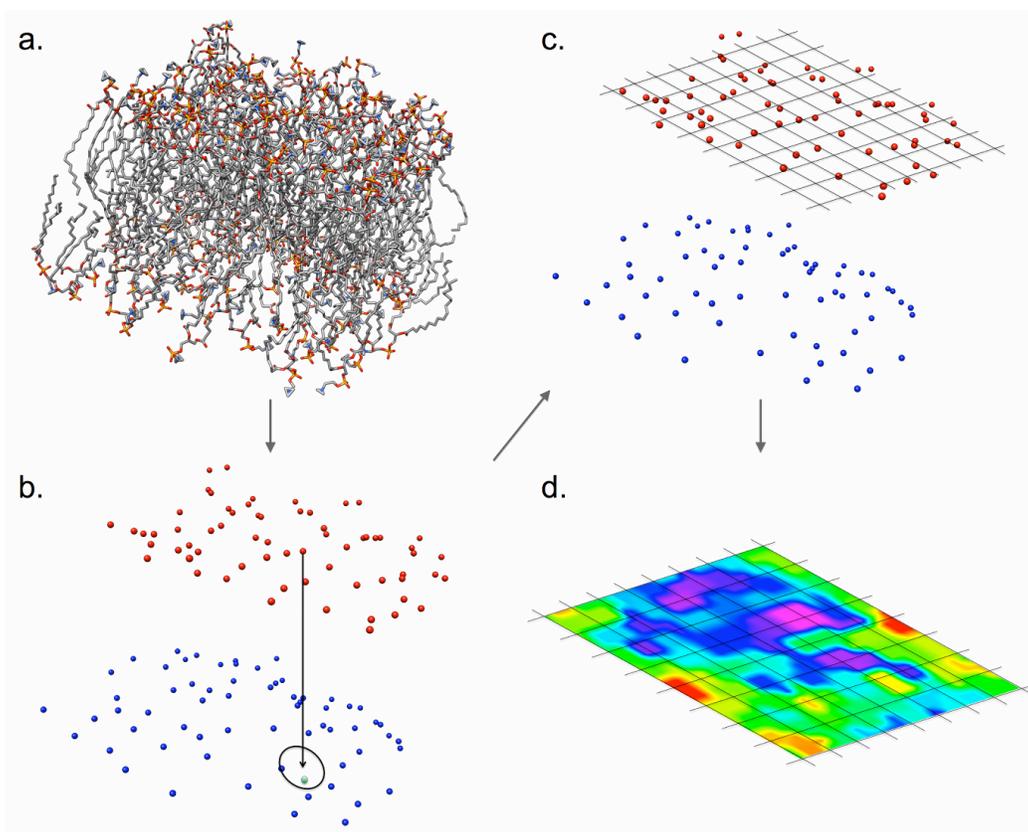


Figure 8.3. Schematic diagram of the extraction of thickness from a lipid bilayer. (a.) A rendering of a 128-lipid POPC (24) bilayer created with USCF Chimera (23). (b.) The phosphorus atoms are chosen as reference points and each searches for its nearest neighbor in the opposite leaflet. The distance of separation becomes the Z -value of that reference point. (c.) Once all Z -values are assigned, a grid is overlaid to the top leaflet. For clarity, the grid assigned to the bottom leaflet is not shown. (d.) Every vertex in the grid is assigned the Z -value of the nearest reference point, and is colored based on the magnitude of the value.

As described, GridMAT-MD will provide an image of bilayer thickness for a single coordinate file, however one snapshot from a long MD does not provide an accurate representation of the fluctuations of the lipid bilayer over time. A better approach is to take a snapshot every few timesteps, and average the results. The raw data format printed by GridMAT-MD makes further manipulation with scripts or spreadsheet software very straightforward.

Calculating Area Per Lipid Headgroup

In developing GridMAT-MD, we observed that as the resolution of the grid is increased to about 200x200 grid points, the membrane system could be divided into polygons. Further, the

number of polygons and lipid molecules always occurred in a one-to-one ratio. Upon generating the polygons, calculating not only the area per lipid headgroup, but also the lateral area of a specific lipid, becomes a simple task. All of the grid points within a certain polygon are assigned the same Z -value, and correspond to the same reference point. Subsequently, the number of grid points within a given polygon, divided by the total number of grid points, is equivalent to the area of a specific lipid headgroup as a fraction of the total lateral area of the bilayer system. Further, this reasoning extends to the protein atoms that are assigned an arbitrary Z -value by the user. The protein atoms compete for grid points with the lipid reference points, so a simple count of the points with the user-defined arbitrary Z -value will yield the lateral surface area that the protein occupies in the lipid bilayer. We believe that this is a much more accurate way to estimate the lateral area of a membrane protein, rather than making assumptions or estimations for “ideal proteins,” based on ideal geometry.

In calculating bilayer thickness, it is useful to average the matrix data from the top and bottom leaflets as previously mentioned. However, in calculating area per lipid headgroup, it may only be meaningful to look at one leaflet at a time, especially in the case of an asymmetrically-oriented or surface-adsorbed peptide, protein, or other molecule. GridMAT-MD was designed in such a way that the data for the top and the bottom leaflets can be printed separately. The data output from area per lipid calculations is a generic text file containing the area of each individual lipid, printed with the lipid residue number to which it corresponds. Then it is quite easy for the user to look at trends in area for a specific lipid (or set of lipids, for example, that are within a certain distance from a protein) over the span of a trajectory.

Validation Simulations

Methods

To test the ability of GridMAT-MD to reproduce previous reports of bilayer thickness, we replicated the work of Kandasamy and Larson by reconstructing one of their systems, the KALP₁₅ peptide in a di-palmitoyl-phosphatidyl-choline (DPPC) membrane. This system

(identified as L1615-16 in their work) served the original authors as part of an examination of hydrophobic mismatch (12).

The KALP₁₅ peptide was built within the xLeap module of AmberTools, version 1.0 (13), imposing the backbone geometry of an ideal α -helix ($\phi = -60^\circ$, $\psi = -40^\circ$) on the constituent residues. The N- and C-termini were capped with acetyl and amide groups, respectively, to give uncharged termini.

The peptide-bilayer system was prepared based on the procedure originally described by Kandasamy and Larson, with some small differences. The previous work employed the methodology proposed by Faraldo-Gómez et al. (14) for inserting the peptide into the membrane. We chose to use a more recent “InflateGRO” methodology proposed by Kandt et al. (15), as it does not rely on multiple programs or modifications to the GROMACS source code. By using the InflateGRO method, 4 lipids were deleted from the bilayer (2 from each leaflet), leaving 124 DPPC molecules packed around the KALP₁₅ peptide.

Parameters from the Gromos96 53a6 parameter set were applied to the peptide, water, and ions in the system. Lipids were described by parameters derived by Berger et al. (16). The equilibration procedures used were the same as those described in the original work. Briefly, an equilibration phase of 5 ns duration was conducted under an NPT ensemble, using the Berendsen weak coupling method for both temperature and pressure (17). The temperature of the peptide, membrane, and solvent and ions were maintained at 323 K separately using a coupling constant of 0.1 ps. The pressure was regulated semi-isotropically at 1 bar in all directions, using a coupling constant of 1.0 ps. Position restraints were applied to the peptide backbone during the equilibration phase. After equilibration, restraints were removed from the peptide and the production phase began. The treatment of nonbonded calculations was identical to that of the original work, cutting off short-range Coulombic and van der Waals interactions at 1.2 nm, and calculating long-range electrostatics with the Particle Mesh Ewald (PME) method (18, 19). Production simulations were allowed to proceed for 50 ns. All simulations were conducted with the GROMACS package, version 3.3.3, using 16 CPUs of Virginia Tech’s SystemX supercomputer (20).

Results

With respect to the bilayer thickness in the vicinity of the KALP₁₅ peptide, our results are in good agreement with those previously reported (12). The hydrophobic region of the peptide is shorter than the transmembrane dimension of the bilayer, a situation called negative hydrophobic mismatch. As a result, the authors of the original work reported a total depression of bilayer dimensions of ~ 15 Å from data collected over the last 25 ns of 50-ns simulation. Over this same period, we note a depression of ~ 13 Å in the vicinity of the KALP₁₅ peptide. Figure 8.4 illustrates the capabilities of GridMAT-MD to display these data concisely. Snapshots are shown from configurations at the start and end of the trajectory, as well as the average over the last 25 ns of the trajectory. The peptide is shown in each panel for reference regarding its location and conformation. The average peptide structure over the last 25 ns, based on RMSD clustering analysis, is shown in the final panel. For comparison, a traditional rendering of the system (in its final configuration, at 50 ns) is shown in Figure 8.5.

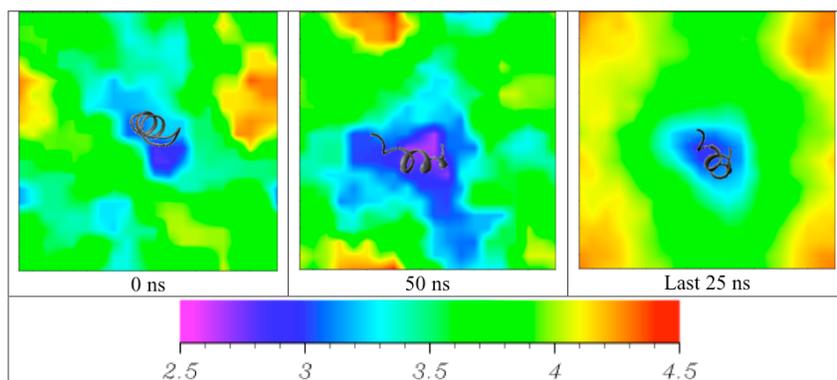


Figure 8.4. Snapshots from the KALP₁₅-DPPC trajectory. The legend shows the bilayer thickness values mapped to a rainbow color gradient, in units of nm. Data were visualized and images rendered using the Xmatrix viewer (version 2.8.0), distributed as part of the Matpack C++ Numerics and Graphics Library (25). The peptide was rendered using UCSF Chimera (23).

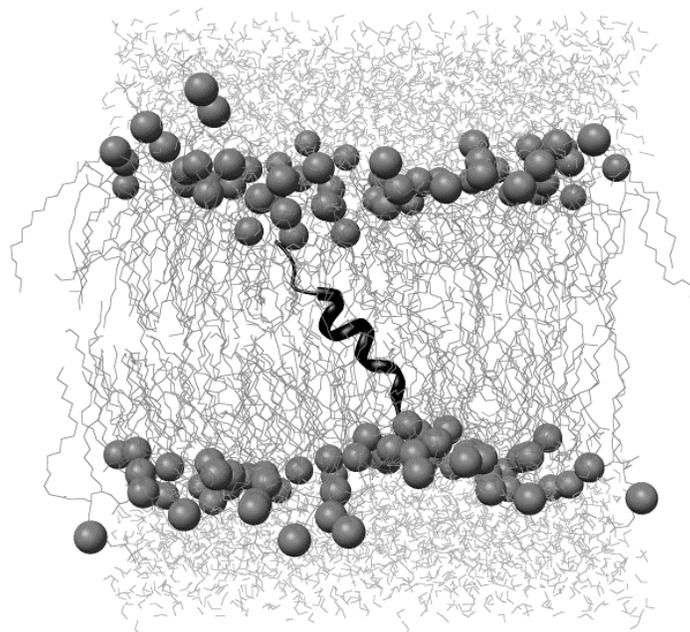


Figure 8.5. The KALP₁₅-DPPC system, at 50 ns, in a traditional rendering. The peptide is shown as a ribbon, phosphorus atoms as spheres, and lipid chains and water shown as lines. The image was rendered using UCSF Chimera (23).

GridMAT-MD produces bilayer thickness data very quickly. Each frame analyzed to produce the data in Figure 8.4 was processed in less than 5 seconds on a MacBook laptop (2.4 GHz Intel Core 2 Duo, 2 GB RAM).

GridMAT-MD also provides a concise representation of the lipid molecules in the system. Each lipid is assigned to a polygon, as discussed above. Figure 8.6 shows a representation of this output, also from the final configuration of the KALP₁₅-DPPC system. Over the last 25 ns of the 50-ns trajectory, the average area per lipid headgroup for the top and bottom leaflets of the DPPC bilayer was $62.2 \pm 1.3 \text{ \AA}^2$ and $61.8 \pm 0.6 \text{ \AA}^2$, respectively. Our own simulations of pure DPPC bilayers under the same conditions give an average area per lipid headgroup of $63.0 \pm 0.9 \text{ \AA}^2$ (data not shown), in agreement with previous simulation (16, 21) and experimental results (22). Area per lipid calculations for the KALP₁₅-DPPC system were also completed quickly by GridMAT-MD, with each frame processed in, on average, 2.5 minutes using the same hardware described above.

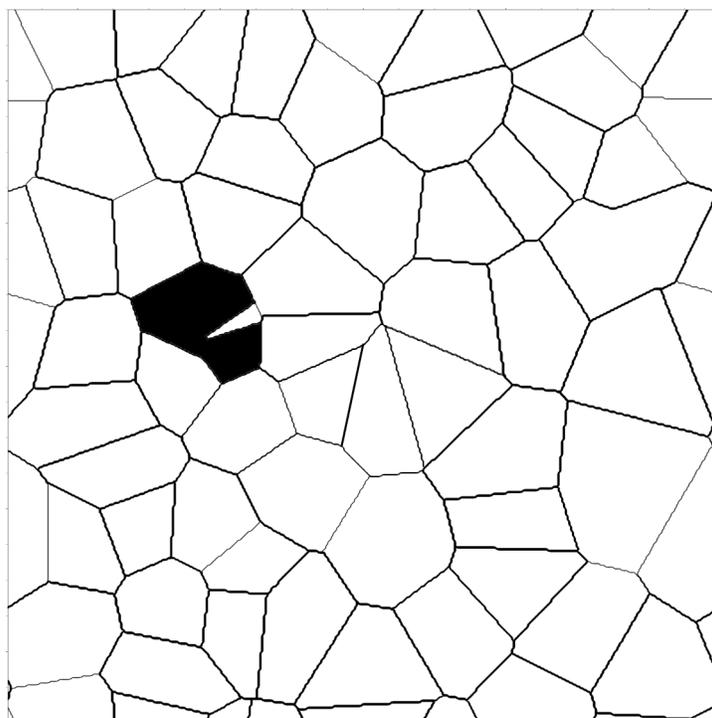


Figure 8.6. Representation of area per lipid headgroup in the top leaflet, from the final configuration of the system. Each polygon represents an individual lipid, determined using the headgroup phosphorus as a reference point. The dark shaded region corresponds to the lateral area of the KALP₁₅ peptide that extends into the lipid headgroup region. Image rendered using the Xmatrix viewer (version 2.8.0), distributed as part of the Matpack C++ Numerics and Graphics Library (25).

Summary

In summary, GridMAT-MD is capable of fast calculations of membrane thickness and area per lipid headgroup with no installation necessary, making it accessible to any of the major operating systems that provide a Perl interpreter. The resulting contour plots are a convenient way to represent the thickness of a lipid bilayer, and the area per lipid calculations are the most convenient and most rigorous that we believe to be publicly available. GridMAT-MD is available for download at <http://www.bevanlab.biochem.vt.edu/GridMAT-MD>.

Acknowledgments

The authors thank Xavier Periole and Thomas Huber for insights into their method of membrane thickness calculations, Tim Sirk and Stephanie Lewis for helpful discussions, and the administrators of the Terascale Computing Facility at Virginia Tech for technical support and computing hours. The material is based upon work supported by the Macromolecular Interfaces with Life Sciences (MILES) Integrative Graduate Education and Research Traineeship (IGERT) of the National Science Foundation under Agreement No. DGE-0333378, and by the Institute for Critical Technology and Applied Science (ICTAS) at Virginia Tech.

References

- (1) Klauda J. B.; Kucerka N.; Brooks B. R.; Pastor R. W.; Nagle J. F. Simulation-Based Methods for Interpreting X-ray Data from Lipid Bilayers. *Biophys. J.* **2006**, *90*, 2796-2807.
- (2) Murtola T.; Kupiainen M.; Falck E.; Vattulainen I. Conformational Analysis of Lipid Molecules By Self-Organizing Maps. *J. Chem. Phys.* **2007**, *126*, 054707.
- (3) Shinoda, W.; Okazaki, S. A Voronoi Analysis of Lipid Area Fluctuation in a Bilayer. *J. Chem. Phys.* **1998**, *109*, 1517-1521.
- (4) Sung B. J.; Yethiraj A. Lateral Diffusion of Proteins in the Plasma Membrane: Spatial Tessellation and Percolation Theory. *J. Phys. Chem. B* **2008**, *112*, 143-149.
- (5) Esteban-Martín S.; Salgado J. The Dynamic Orientation of Membrane-Bound Peptides: Bridging Simulations and Experiments. *Biophys. J.* **2007**, *93*, 4278-4288.
- (6) van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, Flexible, Free. *J. Comp. Chem.* **2005**, *26*, 1701-1718.
- (7) Periole, X.; Huber, T.; Marrink, S-J.; Sakmar, T. P. G Protein-Coupled Receptors Self-Assemble in Dynamics Simulations of Model Bilayers. *J. Am. Chem. Soc.* **2007**, *129*, 10126-10132.
- (8) Tieleman, D. P.; MacCallum, J. L.; Ash, W. L.; Kandt, C.; Xu, Z.; Monticelli, L. Membrane Protein Simulations with a United-Atom Lipid and All-Atom Protein Model: Lipid-Protein Interactions, Side Chain Transfer Free Energies and Model Proteins. *J. Phys.: Condens. Matter* **2006**, *18*, S1221-S1234.
- (9) Shepherd, C. M.; Vogel, H. J.; Tieleman, D. P. Interactions of the Designed Antimicrobial Peptide MB21 and Truncated Dermaseptin S3 with Lipid Bilayers: Molecular-Dynamics Simulations. *Biochem. J.* **2003**, *370*, 233-243.
- (10) The GNU General Public License V3.0 – GNU Project – Free Software Foundation, **2007**. <<http://www.gnu.org/copyleft/gpl.html>>
- (11) The Perl Programming Language, **2002-2011**. <<http://www.perl.com/>>
- (12) Kandasamy, S. K.; Larson, R. G. Molecular Dynamics Simulations of Model Trans-Membrane Peptides in Lipid Bilayers: A Systematic Investigation of Hydrophobic Mismatch. *Biophys. J.* **2006**, *90*, 2326-2343.

- (13) Case, D. A.; Cheatham, T.; Darden, T.; Gohlke, H.; Luo, R.; Merz, Jr., K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. The Amber Biomolecular Simulation Programs. *J. Comp. Chem.* **2005**, *26*, 1668-1688.
- (14) Faraldo-Gómez, J. D.; Smith, G. R.; Sansom, M. S. P. Setting up and Optimization of Membrane Protein Simulations. *Eur. Biophys. J.* **2002**, *31*, 217-227.
- (15) Kandt, C.; Ash, W. L.; Tieleman, D. P. Setting up and Running Molecular Dynamics Simulations of Membrane Proteins. *Methods* **2007**, *41*, 475-488.
- (16) Berger, O.; Edholm, O.; Jähnig, F. Molecular Dynamics Simulations of a Fluid Bilayer of Dipalmitoylphosphatidylcholine at Full Hydration, Constant Pressure, and Constant Temperature. *Biophys. J.* **1997**, *72*, 2002-2013.
- (17) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81*, 3684-3690.
- (18) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An Nlog(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089-10092.
- (19) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577-8593.
- (20) Virginia Tech Advanced Research Computing: System X, **2007-2010**. <<http://www.arc.vt.edu/arc/SystemX/index.php>>
- (21) Tieleman, D. P.; Berendsen, H. J. C. Molecular Dynamics Simulations of a fully Hydrated Dipalmitoylphosphatidylcholine Bilayer with Different Macroscopic Boundary Conditions and Parameters. *J. Chem. Phys.* **1996**, *105*, 4871-4880.
- (22) Nagle, J. F.; Tristram-Nagle, S. Structure of Lipid Bilayers. *Biochim. Biophys. Acta* **2000**, *1469*, 159-195.
- (23) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera – A Visualization System for Exploratory Research and Analysis. *J. Comp. Chem.* **2004**, *25*, 1605-1612.
- (24) Tieleman, D. P.; Sansom, M. S. P.; Berendsen, H. J. C. Alamethicin Helices in a Bilayer and in Solution: Molecular Dynamics Simulations. *Biophys. J.* **1999**, *76*, 40-49.
- (25) MATPACK C++ Numerics and Graphics Library Release 1.9.0, **1991-2005**. <<http://www.matpack.de>>

Chapter 9

Conclusions

Many different and wide-ranging combinations of biomolecular systems that have had broad applications in the study of oxidative processes and human disease have been presented in this dissertation. In Chapter 2, we used molecular dynamics (MD) simulations to identify dual modes of regulation between the enzyme monoamine oxidase B (MAO B) and the bilayer. We also identified residues and specific interactions in the active site that are important for inhibitor binding. This information would be invaluable to a knowledge-based drug design project seeking to construct a more powerful MAO B inhibitor. In Chapter 3, we used quantum mechanics (QM) to better understand the active site mechanism of the same enzyme. It seems that the barrier to ring opening of several N-cyclopropyl derivatives is not as low as was previously thought, perhaps refuting the prior assumption that the mechanism of oxidation proceeds by a single electron transfer pathway. It now seems that, at least for tertiary amines, MAO B may favor a direct hydrogen atom abstraction pathway. In Chapter 4 we examined the peculiar folding behavior of the unstructured p53 C-terminal domain, and we concluded that the secondary structure that it does adopt may be influenced by the size of the hydrophobic pocket in its chosen binding partner. In Chapter 5, we used MD simulations and non-equilibrium steered MD simulations to analyze the relationship between the protein B2 and double-stranded RNA. We identified a specific residue that, upon mutation, causes an attenuated binding capacity. On a larger scale, the methods described in the chapter could be very useful to other researchers who are trying to predict degree of binding attenuation or knockout for other protein-nucleic acid systems. In Chapter 6 we studied the dynamics of an antigenic loop in peanut proteins that varies in size with phylogeny.

The final two projects in this document steered away from pure MM or QM studies towards method development and validation. In Chapter 7, we found that the PRODRG server, which is widely used to obtain parameters for non-standard small molecules for use in MD simulation, may not be as accurate as it advertises. While bonded parameters and atom types are very accurate, we found that individual atomic charges and charge group assignments require further refinement. We suggested methods based in QM-theory to calculate better starting parameters, and MD simulations with free energy calculations to validate and refine those parameters. Finally, in Chapter 8 we described the development and application of a program used to measure the thickness and area per lipid of a bilayer system, with or without a protein embedded. Because so many drug targets are membrane-bound enzymes, it is important that the tools exist to help validate those simulations; GridMAT-MD is such a tool.

In the past 35 years, it seems computational methods have become an integral part of research in the life sciences, especially in the fields of biochemistry and drug design. Computational techniques such as MM and QM, as demonstrated by the chapters within this document, are broad in both the approach of the experiments and the scope of the application. As our computational tools continue to increase in power and in speed, and as structural information and the tools of simulation become more widely available, we are sure to see greater technological advances and growth in the popularity of applications of molecular modeling.

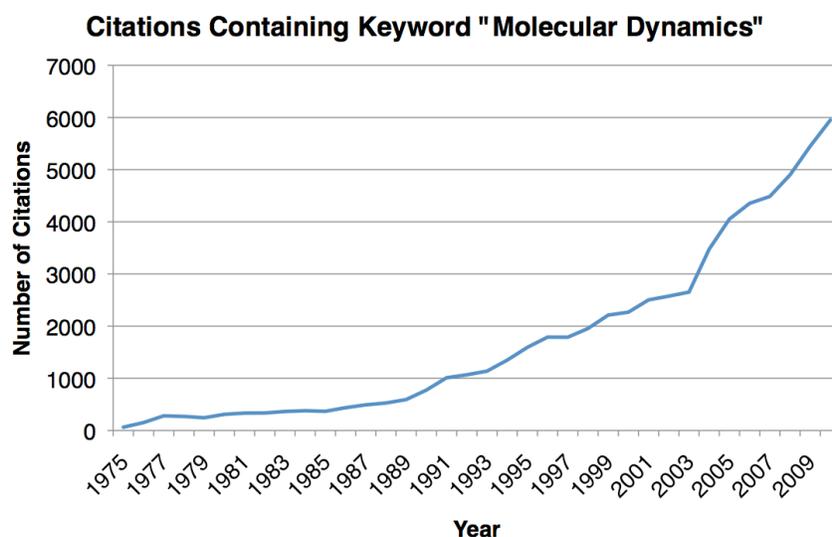


Figure 9.1. Number of citations containing the keyword “molecular dynamics” arranged by year. Data obtained from SciFinder.

Appendix A

Important *.mdp* Files

When performing molecular dynamics simulations, there are many more parameters and settings that go into the simulations than are typically described in resulting publications. In the case of the GROMACS software package, these are listed in what is called the *.mdp* file. Recorded in this Appendix A are some example *.mdp* files from the production runs of some of our simulations. They are as follows:

- 1.) Chapter 2 – MAO B in a bilayer production run
- 2.) Chapter 2 – MAO B in a bilayer steered simulation
- 3.) Chapter 4 – p53 in complex with a protein binding partner production run
- 4.) Chapter 5 – B2 protein in complex with dsRNA production run
- 5.) Chapter 5 – B2 protein in complex with dsRNA steered simulation
- 6.) Chapter 6 – Peanut protein production run
- 7.) Chapter 7 – Small molecule in condensed phase production run
- 8.) Chapter 7 – Small molecule in gaseous phase production run
- 9.) Chapter 7 – UGM and FAD complex production run

1.) Chapter 2 – MAO B in a bilayer production run

```
; 7.3.3 Run Control
integrator      = md
tinit          = 10000
dt             = 0.002
nsteps        = 5000000
comm_mode      = Linear
nstcomm        = 1
comm_grps      = Protein_FAD_1PB POPC_POPE SOL_NA+_CL-

; 7.3.8 Output Control
nstxout        = 5000000
nstvout        = 5000000
nstfout        = 5000000
nstlog         = 1000
nstenergy      = 1000
nstxtcout      = 5000
xtc_precision  = 1000
xtc_grps       = Protein_FAD_1PB POPC_POPE
energygrps     = System

; 7.3.9 Neighbor Searching
nstlist        = 5
ns_type        = grid
pbc            = xyz
rlist          = 0.8

; 7.3.10 Electrostatics
coulombtype    = PME
rcoulomb       = 0.8

; 7.3.11 VdW
vdwtype        = cut-off
rvdw           = 1.4
DispCorr       = EnerPres

; 7.3.13 Ewald
fourierspacing = 0.12
pme_order      = 4
ewald_rtol     = 1e-5

; 7.3.14 Temperature Coupling
tcoupl         = nose-hoover
tc_grps        = Protein_FAD_1PB POPC_POPE SOL_NA+_CL-
tau_t          = 0.1      0.1      0.1
ref_t          = 310      310      310

; 7.3.15 Pressure Coupling
pcoupl         = parrinello-rahman
pcoupltype     = semiisotropic
tau_p          = 1.0
compressibility = 4.5e-5 4.5e-5
ref_p          = 1.0      1.0

; 7.3.17 Velocity Generation
gen_vel        = no

; 7.3.18 Bonds
constraints     = all-bonds
constraint_algorithm = LINCS
unconstrained_start = no
lincs_order     = 4
lincs_iter      = 1
lincs_warnangle = 30
```

2.) Chapter 2 – MAO B in a bilayer steered simulation

```
; 7.3.3 Run Control
integrator      = md
tinit          = 0
dt             = 0.002
nsteps        = 250000
comm_mode     = Linear
nstcomm       = 1
comm_grps     = Protein_FAD_1PB POPC_POPE SOL_NA+_CL-

; 7.3.8 Output Control
nstxout       = 250000
nstvout       = 250000
nstfout       = 250000
nstlog        = 500
nstenergy     = 500
nstxtcout     = 250
xtc_precision = 1000
xtc_grps      = System
energygrps    = System

; 7.3.9 Neighbor Searching
nstlist       = 5
ns_type       = grid
pbc           = xyz
rlist         = 0.8

; 7.3.10 Electrostatics
coulombtype   = PME
rcoulomb      = 0.8

; 7.3.11 VdW
vdwtype       = cut-off
rvdw          = 1.4
DispCorr      = EnerPres

; 7.3.13 Ewald
fourierspacing = 0.12
pme_order     = 4
ewald_rtol    = 1e-5

; 7.3.14 Temperature Coupling
tcoupl        = nose-hoover
tc_grps       = Protein_FAD_1PB POPC_POPE SOL_NA+_CL-
tau_t         = 0.1      0.1      0.1
ref_t         = 310      310      310

; 7.3.15 Pressure Coupling
pcoupl        = parrinello-rahman
pcoupltype    = semiisotropic
tau_p         = 1.0
compressibility = 4.5e-5 4.5e-5
ref_p         = 1.0      1.0

; 7.3.17 Velocity Generation
gen_vel       = no

; 7.3.18 Bonds
constraints    = all-bonds
constraint_algorithm = LINCS
unconstrained_start = no
lincs_order   = 4
lincs_iter    = 1
lincs_warnangle = 30

; 7.3.21 COM pulling
pull          = umbrella
pull_geometry = direction
pull_start    = yes
pull_nstxout  = 10
pull_nstfout  = 10
```

```
pull_ngroups      = 1
pull_group0       = r_519
pull_group1       = pull_me2
pull_vec1         = 1.40034 0.1217 2.97393
pull_rate1        = 0.01
pull_k1           = 1000
```

3.) Chapter 4 – p53 in complex with a protein binding partner production run

```
; RUN CONTROL PARAMETERS
integrator          = md
tinit              = 0
dt                 = 0.002
nsteps             = 10000000
comm_mode          = Linear
nstcomm            = 1

; OUTPUT CONTROL OPTIONS
nstxout            = 10000000
nstvout            = 10000000
nstfout            = 0
nstlog             = 100
nstenergy          = 100
nstxtcout          = 2500
xtc_precision      = 1000
xtc_grps           = Protein_NME
energygrps         = Protein_NME SOL_NA+_CL-

; NEIGHBORSEARCHING PARAMETERS
nstlist            = 5
ns_type            = grid
pbc                = xyz
rlist              = 0.8

; OPTIONS FOR ELECTROSTATICS AND VDW
coulombtype        = PME
rcoulomb           = 0.8
fourierspacing     = 0.12
pme_order          = 4
vdw_type           = cut-off
rvdw               = 1.4

; Temperature coupling
Tcoupl             = nose-hoover
tc-grps            = Protein_NME SOL_NA+_CL-
tau_t              = 0.1      0.1
ref_t              = 310      310

; Pressure coupling
Pcoupl             = parrinello-rahman
Pcoupltype         = Isotropic
tau_p              = 1.0
compressibility     = 4.5e-5
ref_p              = 1.0

; GENERATE VELOCITIES FOR STARTUP RUN
gen_vel            = no

; OPTIONS FOR BONDS
constraints         = all-bonds
constraint-algorithm = Lincs
unconstrained-start = yes
lincs-order         = 4
lincs-iter          = 1
lincs-warnangle     = 30
```

4.) Chapter 5 – B2 protein in complex with dsRNA production run

```
; 7.3.2 Preprocessing
define                = -DPOSRES_RNA

; 7.3.3 Run Control
integrator            = md
tinit                = 0
dt                   = 0.002
nsteps               = 10000000
comm_mode            = Linear
nstcomm              = 1
comm_grps            = Protein RNA SOL_Na+_Cl-

; 7.3.8 Output Control
nstxout              = 10000000
nstvout              = 10000000
nstfout              = 10000000
nstlog               = 1000
nstenergy            = 1000
nstxtcout            = 5000
xtc_precision        = 1000
xtc_grps             = Protein RNA
energygrps           = System

; 7.3.9 Neighbor Searching
nstlist              = 5
ns_type              = grid
pbc                  = xyz
rlist                = 0.8

; 7.3.10 Electrostatics
coulombtype          = PME
rcoulomb             = 0.8

; 7.3.11 VdW
vdwtype              = cut-off
rvdw                 = 1.4
DispCorr             = EnerPres

; 7.3.13 Ewald
fourierspacing       = 0.12
pme_order            = 4
ewald_rtol           = 1e-5

; 7.3.14 Temperature Coupling
tcoupl               = nose-hoover
tc_grps              = Protein RNA SOL_Na+_Cl-
tau_t                = 0.1    0.1 0.1
ref_t                = 301    301 301

; 7.3.15 Pressure Coupling
pcoupl               = parrinello-rahman
pcoupltype           = isotropic
tau_p                = 1.0
compressibility       = 4.5e-5
ref_p                = 1.0

; 7.3.17 Velocity Generation
gen_vel              = no

; 7.3.18 Bonds
constraints           = all-bonds
constraint_algorithm = LINCS
unconstrained_start = no
lincs_order          = 4
lincs_iter           = 1
lincs_warnangle      = 30
```

5.) Chapter 5 – B2 protein in complex with dsRNA steered simulation

```
; 7.3.2 Preprocessing
define                = -DPOSRES_RNA_ALL

; 7.3.3 Run Control
integrator            = md
tinit                 = 0
dt                    = 0.002
nsteps                = 250000
comm_mode             = Linear
nstcomm               = 1
comm_grps             = Protein RNA SOL_NA+_CL-

; 7.3.8 Output Control
nstxout               = 250000
nstvout               = 250000
nstfout               = 250000
nstlog                = 100
nstenergy             = 100
nstxtcout             = 250
xtc_precision         = 1000
xtc_grps              = System
energygrps            = System

; 7.3.9 Neighbor Searching
nstlist               = 5
ns_type               = grid
pbc                   = xyz
rlist                 = 0.8

; 7.3.10 Electrostatics
coulombtype           = PME
rcoulomb              = 0.8

; 7.3.11 VdW
vdwtype               = cut-off
rvdw                  = 1.4
DispCorr              = EnerPres

; 7.3.13 Ewald
fourierspacing        = 0.12
pme_order              = 4
ewald_rtol            = 1e-5

; 7.3.14 Temperature Coupling
tcoupl                = nose-hoover
tc_grps               = Protein RNA SOL_NA+_CL-
tau_t                 = 0.1    0.1 0.1
ref_t                 = 301    301 301

; 7.3.15 Pressure Coupling
pcoupl                = parrinello-rahman
pcoupltype            = isotropic
tau_p                 = 1.0
compressibility        = 4.5e-5
ref_p                 = 1.0

; 7.3.17 Velocity Generation
gen_vel               = no

; 7.3.18 Bonds
constraints            = all-bonds
constraint_algorithm   = LINCS
unconstrained_start   = no
lincs_order            = 4
lincs_iter             = 1
lincs_warnangle        = 30

; 7.3.21 COM pulling
pull                  = umbrella
pull_geometry          = direction
```

```
pull_start          = yes
pull_nstxout        = 10
pull_nstfout        = 10
pull_ngroups        = 1
pull_group0         = RNA
pull_group1         = Protein
pull_vec1           = 1 0 0
pull_rate1          = 0.01
pull_k1             = 100000
```

6.) Chapter 6 – Peanut protein production run

```
; RUN CONTROL PARAMETERS
integrator          = md
tinit              = 0
dt                 = 0.002
nsteps             = 10000000
comm_mode          = Linear
nstcomm            = 1

; OUTPUT CONTROL OPTIONS
nstxout            = 10000000
nstvout            = 10000000
nstfout            = 0
nstlog             = 100
nstenergy          = 100
nstxtcout          = 2500
xtc_precision      = 1000
xtc_grps           = Protein
energygrps         = Protein SOL_NA+_CL-

; NEIGHBORSEARCHING PARAMETERS
nstlist            = 5
ns_type            = grid
pbc                = xyz
rlist              = 0.8

; OPTIONS FOR ELECTROSTATICS AND VDW
coulombtype        = PME
rcoulomb           = 0.8
fourierspacing     = 0.12
pme_order          = 4
vdw_type           = cut-off
rvdw               = 1.4

; Temperature coupling
Tcoupl             = nose-hoover
tc-grps            = Protein SOL_NA+_CL-
tau_t              = 0.1      0.1
ref_t              = 310      310

; Pressure coupling
Pcoupl             = parrinello-rahman
Pcoupltype         = Isotropic
tau_p              = 1.0
compressibility     = 4.5e-5
ref_p              = 1.0

; GENERATE VELOCITIES FOR STARTUP RUN
gen_vel            = no

; OPTIONS FOR BONDS
constraints         = all-bonds
constraint-algorithm = Lincs
unconstrained-start = yes
lincs-order         = 4
lincs-iter          = 1
lincs-warnangle     = 30
```

7.) Chapter 7 – Small molecule in condensed phase production run

```
; 7.3.3 Run Control
integrator      = md
tinit          = 0
dt             = 0.002
nsteps        = 10000000
comm_mode     = Linear
nstcomm       = 1
comm_grps     = System

; 7.3.8 Output Control
nstxout       = 10000000
nstvout       = 10000000
nstfout       = 10000000
nstlog        = 100
nstenergy     = 100
nstxtcout     = 500
xtc_precision = 1000
xtc_grps      = System
energygrps    = System

; 7.3.9 Neighbor Searching
nstlist       = 5
ns_type       = grid
pbc           = xyz
rlist         = 0.8

; 7.3.10 Electrostatics
coulombtype   = PME
rcoulomb      = 0.8

; 7.3.11 VdW
vdwtype       = cut-off
rvdw          = 1.4
DispCorr      = EnerPres

; 7.3.13 Ewald
fourierspacing = 0.12
pme_order     = 4
ewald_rtol    = 1e-5

; 7.3.14 Temperature Coupling
tcoupl        = nose-hoover
tc_grps       = System
tau_t         = 0.1
ref_t         = 298

; 7.3.15 Pressure Coupling
pcoupl        = parrinello-rahman
pcoupltype    = isotropic
tau_p         = 1.0
compressibility = 4.5e-5
ref_p         = 1.0

; 7.3.17 Velocity Generation
gen_vel       = no

; 7.3.18 Bonds
constraints    = all-bonds
constraint_algorithm = LINCS
unconstrained_start = no
lincs_order   = 4
lincs_iter    = 1
lincs_warnangle = 30
```

8.) Chapter 7 – Small molecule in gaseous phase production run

```
; 7.3.3 Run Control
integrator      = md
tinit          = 0
dt             = 0.002
nsteps        = 10000000
comm_mode     = Linear
nstcomm       = 1
comm_grps     = System

; 7.3.8 Output Control
nstxout       = 10000000
nstvout       = 10000000
nstfout       = 10000000
nstlog        = 100
nstenergy     = 100
nstxtcout     = 500
xtc_precision = 1000
xtc_grps      = System
energygrps    = System

; 7.3.9 Neighbor Searching
nstlist       = 5
ns_type       = grid
pbc           = xyz
rlist         = 0.8

; 7.3.10 Electrostatics
coulombtype   = cut-off
rcoulomb      = 0.8

; 7.3.11 VdW
vdwtype       = cut-off
rvdw          = 1.4
DispCorr     = EnerPres

; 7.3.14 Temperature Coupling
tcoupl       = nose-hoover
tc_grps      = System
tau_t        = 0.1
ref_t        = 298

; 7.3.15 Pressure Coupling
pcoupl       = parrinello-rahman
pcoupltype   = isotropic
tau_p        = 1.0
compressibility = 4.5e-5
ref_p        = 1.0

; 7.3.17 Velocity Generation
gen_vel      = no

; 7.3.18 Bonds
constraints   = all-bonds
constraint_algorithm = LINCS
unconstrained_start = no
lincs_order  = 4
lincs_iter   = 1
lincs_warnangle = 30
```

9.) Chapter 7 – UGM and FAD complex production run

```
; 7.3.3 Run Control
integrator      = md
tinit          = 0
dt             = 0.002
nsteps        = 10000000
comm_mode     = Linear
nstcomm       = 1
comm_grps     = Protein_FAD SOL_NA+_CL-

; 7.3.8 Output Control
nstxout       = 10000000
nstvout       = 10000000
nstfout       = 10000000
nstlog        = 500
nstenergy     = 500
nstxtcout     = 1000
xtc_precision = 1000
xtc_grps      = Protein_FAD
energygrps    = System

; 7.3.9 Neighbor Searching
nstlist       = 5
ns_type       = grid
pbc           = xyz
rlist         = 0.8

; 7.3.10 Electrostatics
coulombtype   = PME
rcoulomb      = 0.8

; 7.3.11 VdW
vdwtype       = cut-off
rvdw          = 1.4
DispCorr      = EnerPres

; 7.3.13 Ewald
fourierspacing = 0.12
pme_order     = 4
ewald_rtol    = 1e-5

; 7.3.14 Temperature Coupling
tcoupl        = nose-hoover
tc_grps       = Protein_FAD SOL_NA+_CL-
tau_t         = 0.1      0.1
ref_t         = 310      310

; 7.3.15 Pressure Coupling
pcoupl        = parrinello-rahman
pcoupltype    = isotropic
tau_p         = 1.0
compressibility = 4.5e-5
ref_p         = 1.0

; 7.3.17 Velocity Generation
gen_vel       = no

; 7.3.18 Bonds
constraints    = all-bonds
constraint_algorithm = LINCS
unconstrained_start = no
lincs_order   = 4
lincs_iter    = 1
lincs_warnangle = 30
```

Appendix B

Perl Scripts

In performing these projects, the *Perl* programming language quickly emerged as an invaluable tool for processing GROMACS output files and performing analyses for which there was no other means available. Listed in this Appendix B are several of the more useful programs that we developed for simple analyses. Within the header of each program are specific instructions on the purpose of the program and how to use it. They are as follows:

- 1.) `apl_calculator.pl` – Calculates area per lipid of a bilayer system without protein
- 2.) `average_multi.pl` – Averages the second column of multiple two-column files. For example, will average RMSD plots from different replicates.
- 3.) `bilayer_fixsol.pl` – Removes water molecules that are inserted within the hydrophobic tails of the bilayer when solvating.
- 4.) `ellipse.pl` – Calculates the surface area of an active site channel opening.
- 5.) `gen_fasta.pl` – Reads a PDB file and outputs a FASTA sequence.
- 6.) `greatest.pl` – Finds the greatest y-value in a two-column plot. Useful for determining peak force during steered simulations.
- 7.) `pull_analysis.pl` – Reads a `pullf` and `pullx` file from the same SMD run, then will map displacement of pull group (normalized to center of mass of the initial position) against force.
- 8.) `transformer.pl` – Creates a mixed POPC-POPE bilayer from a pure POPC bilayer.

```

//apl-calculator.pl//

#!/usr/bin/perl
use strict;

#
# This program measures area per lipid headgroup over a trajectory. You must first
# run g_energy to get plots of the box vector in X and Y directions. Also, delete
# the header information from the xvg files. You must also know the number of total
# lipids (the program divides it by 2). This does not work in the presence of a protein.
#

unless (@ARGV){
    die "Usage: ./\$0 <boxX.xvg> <boxY.xvg> <tot # of lipids>\n";
}

my $filenameX = $ARGV[0];
my $filenameY = $ARGV[1];
my $num_lipids = $ARGV[2]/2;

open (INPUTX, $filenameX) or die "It seems \"\$filenameX\" does not exist!\n";
my @x_vector = (<INPUTX>);
close (INPUTX);

open (INPUTY, $filenameY) or die "It seems \"\$filenameY\" does not exist!\n";
my @y_vector = (<INPUTY>);
close (INPUTY);

my @time;
my @x_value;
my @y_value;

foreach (@x_vector){
    my ($time, $value) = split(" ", $_);
    push (@time, $time);
    push (@x_value, $value);
}

foreach (@y_vector){
    my ($time, $value) = split(" ", $_);
    push (@y_value, $value);
}

my @final;
my $number = scalar(@time);

for (my $i=0; $i<$number; $i++){

    my $area = $x_value[$i]*$y_value[$i];
    my $apl = $area/$num_lipids*100;
    push (@final, $time[$i]."\t".$apl."\n");
}

open (OUTFILE, ">", "apl.xvg");
print OUTFILE @final;
close (OUTFILE);

exit();

```

```

//average_multi.pl//

#!/usr/bin/perl -w
use strict;

#
# Opens multiple .dat or .xvg files and averages the results of the equivalent lines
# in each. An equivalent line is defined as one in which the first (x) value is the
# same. The first two columns are used, any additional entries are ignored. Input is
# expected to be some generic text (data) file, the number of inputs is determined
# from the size of @ARGV
#
# Data files are assumed to be text only, with Grace-style headers (# and @)
#

unless(@ARGV) {
    die "Usage: ./\$0 <input1.xvg> <input2.xvg> ...\\n";
    exit;
}

# determine number of inputs
my $n = scalar(@ARGV);

my @big_array;

# open a series of filehandles to place each text file into an array
for (my $i=0; $i<$n; $i++) {
    open(IN, "<${ARGV[$i]}") or die "Cannot open $ARGV[$i]: !$\\n";
    my @array = <IN>;
    close(IN);

    # concatenate all inputs into one big array
    foreach $_ (@array) {
        unless ($_ =~ /^[#@]/) {
            push(@big_array, $_);
        }
    }
}

# splice out duplicate x-values
my @x_values;

foreach $_ (@big_array) {
    my @line = split(" ", $_);
    push(@x_values, $line[0]);
}

for (my $j=0; $j<scalar(@x_values); $j++) {
    for (my $k=$j+1; $k<scalar(@x_values); $k++) {
        if ($x_values[$k] == $x_values[$j]) {
            splice(@x_values, $k, 1);
        }
    }
}

# remove duplicate entry at the end
pop(@x_values);

# Now that there's a list of x-values, loop thru @big_array again, collecting
# the y-values that correspond to those x-values

foreach $_ (@x_values) {

    my @y_values;

    for (my $m=0; $m<scalar(@big_array); $m++) {
        my @line = split(" ", $big_array[$m]);
        if ($line[0] == $_) {
            push(@y_values, $line[1]);
        }
    }
}

```

```

my $sum = 0;

foreach $_ (@y_values) {
    $sum += $_;
}

# determine an average
my $average = $sum / (scalar(@y_values));

# determine the standard deviation:
# 1. for each value, subtract the average and square this difference
# 2. sum these differences, divide by number of values
# 3. take square root

my $sd_sum = 0;

foreach my $value (@y_values) {
    $sd_sum += ($value - $average)**2;
}

my $sd_dividend = $sd_sum / (scalar(@y_values));

my $std_dev = sqrt($sd_dividend);

# open output file
open(OUT, ">>output.dat") or die "Cannot open output: $!\n";
printf OUT "%d %8.3f %8.3f\n", $_, $average, $std_dev;
close(OUT);
}

```

```

//bilayer_fixsol.pl//

#!/usr/bin/perl
use strict;
my $start_time = time();

#
# This program attempts to delete some of those waters that find themselves in the middle
# of a bilayer after solvating. The input is just the newly solvated bilayer coordinate
# file and a reference atom from each lipid. It will find the highest and the lowest
# lipid reference atom positions, then narrow the margin by a cut-off on each side
# (7.5 Ang). Then it will delete all SOL molecules that fall in between those boundaries
# and write to output.gro.
#

unless (@ARGV){
    die "Usage: ./$0 <gro filename> <reference atomname>\n";
}

open (INPUT, $ARGV[0]) or die "Could not find $ARGV[0]!\n";
my @system = <INPUT>;
close (INPUT);

my $title = shift @system;
shift @system;
my $box_vectors = pop @system;

my @system_fixed;
my @reference;

foreach (@system){
    my $newline = pack("A6 A5 A7 A29", unpack("A5 A4 A6 A29", $_));
    push(@system_fixed, $newline."\n");
}

foreach (@system_fixed){
    my ($resnum, $resname, $atomname, $atomnum, $xcoor, $ycoor, $zcoor) = split(' ', $_);
    if ($atomname eq $ARGV[1]){
        push @reference, $zcoor;
    }
}

my $upper = -100;
my $lower = 100;

foreach (@reference){
    if ($_ >= $upper){
        $upper = $_;
    }

    if ($_ <= $lower){
        $lower = $_;
    }
}

$upper -= 0.75;
$lower += 0.75;
my @waters;

foreach (@system_fixed){
    my ($resnum, $resname, $atomname, $atomnum, $xcoor, $ycoor, $zcoor) = split(' ', $_);
    if ($resname eq "SOL"){
        if ($zcoor <= $upper && $zcoor >= $lower){
            push @waters, $resnum;
        }
    }
}

my %hash = map { $_, 1 } @waters;
my @unique_waters = keys %hash;
my @system_final;
my @system_final_final;

```

```

foreach (@system_fixed){
    my $flag = 0;
    my ($resnum, $resname, $atomname, $atomnum, $xcoor, $ycoor, $zcoor) = split(' ', $_);

    foreach my $water (@unique_waters){
        if ($resnum eq $water){
            $flag += 1;
        }
    }

    if ($flag == 0){
        push @system_final, $_;
    }
}

foreach (@system_final){
    my $newline = pack("A5 A4 A6 A29", unpack("A6 A5 A7 A29", $_));
    push(@system_final_final, $newline."\n");
}

unshift @system_final_final, scalar(@system_final_final)."\n";
unshift @system_final_final, $title;
push @system_final_final, $box_vectors;

open (OUTPUT, ">", "12345abcde.gro");
print OUTPUT @system_final_final;
close (OUTPUT);

system ("genconf -f 12345abcde.gro -o output.gro -renumber &>/dev/null");
system ("rm 12345abcde.gro");

my $end_time = time();
my $time = $end_time - $start_time;
print "Time taken = $time seconds\n";
exit();

```

```

//ellipse.pl//

#!/usr/bin/perl
use strict;

#
# This program is for measuring the surface area of the opening of an active site cavity,
# for example. Suppose you identify four residues that form the "mouth" of the active
# site. Use g_dist to measure the average COM distance between opposing residues. So, for
# example, if you have four residues in a square called A, B, C, and D, measure the
# distance between A and C and the distance between B and D (where those are diagonal
# from one another). Feed both of those output files to this program as described below,
# and it will treat those as opposing points in an ellipse and approximate the area.
# Remove the header lines from the .xvg files before proceeding
#

unless (@ARGV){
    die "Usage: ./\$0 <dist1.xvg> <dist2.xvg>\n";
}

my $filename1 = $ARGV[0];
my $filename2 = $ARGV[1];

open (INPUT1, $filename1) or die "It seems \"\$filename1\" does not exist!\n";
my @dist1 = (<INPUT1>);
close (INPUT1);

open (INPUT2, $filename2) or die "It seems \"\$filename2\" does not exist!\n";
my @dist2 = (<INPUT2>);
close (INPUT2);

my @time;
my @value1;
my @value2;

foreach (@dist1){
    my ($time, $value) = split(" ", $_);
    push (@time, $time);
    push (@value1, $value);
}

foreach (@dist2){
    my ($time, $value) = split(" ", $_);
    push (@value2, $value);
}

my @final;
my $number = scalar(@time);

for (my $i=0; $i<$number; $i++){
    my $area = ($value1[$i]/2)*($value2[$i]/2)*(3.14159);
    push (@final, $time[$i]."\t".$area."\n");
}

open (OUTFILE, ">", "ellipse.xvg");
print OUTFILE @final;
close (OUTFILE);

exit();

```

```

//gen_fasta.pl//

#!/usr/bin/perl
use strict;

#
# This program generates a FASTA-friendly sequence from an input pdb file
#

unless (@ARGV){
    die "Usage: perl $0 <pdb file>\n";
}

my $filename = $ARGV[0];
open (INPUT, $filename) or die "It seems \"$filename\" does not exist!\n";
my @pdb_file = (<INPUT>);
close (INPUT);

my @pdb_grep = grep /^ATOM/, @pdb_file;
my @pdb_small;
my @residues;

foreach (@pdb_grep){
    my @line = split(/,/, $_);
    my $r_num = join("", splice(@line, 22, 4));
    my $r_name = join("", splice(@line, 17, 3));
    my $new_line = (" $r_num $r_name");
    push(@pdb_small, $new_line."\n");
}

my %temp_hash = map { $_, 1} @pdb_small;
my @pdb_final = keys(%temp_hash);
@pdb_final = sort(@pdb_final);

foreach (@pdb_final){
    my @line = split(/,/, $_);
    my $r_name = join("", splice(@line, 5, 3));
    push(@residues, $r_name);
}

foreach (@residues){
    $_ =~ s/ALA/A/;
    $_ =~ s/ARG/R/;
    $_ =~ s/ASN/N/;
    $_ =~ s/ASP/D/;
    $_ =~ s/CYS/C/;
    $_ =~ s/GLU/E/;
    $_ =~ s/GLN/Q/;
    $_ =~ s/GLY/G/;
    $_ =~ s/HIS/H/;
    $_ =~ s/ILE/I/;
    $_ =~ s/LEU/L/;
    $_ =~ s/LYS/K/;
    $_ =~ s/MET/M/;
    $_ =~ s/PHE/F/;
    $_ =~ s/PRO/P/;
    $_ =~ s/SER/S/;
    $_ =~ s/THR/T/;
    $_ =~ s/TRP/W/;
    $_ =~ s/TYR/Y/;
    $_ =~ s/VAL/V/;
}

my $sequence = join('', @residues);
my $fasta_file = ">".$filename."\n".$sequence."\n";

exit();

```

```

//greatest.pl//

#!/usr/bin/perl
use strict;

#
# This program takes an XVG file with two columns, and finds the greatest value in the
# second column. This is very good for finding the maximum force of pulling on force
# curves, for example.
#

unless (@ARGV){
    die "Usage: perl $0 <xvg filename>\n";
}

my $filename = $ARGV[0];
open (INPUT, $filename) or die "It seems \"$filename\" does not exist!\n";
my @file = (<INPUT>);
close (INPUT);

my @file_grep = grep !/^#|^@/, @file;

my $max = 0;
foreach (@file_grep){
    my ($time, $force) = split(' ', $_);
    if ($force > $max){
        $max = $force;
    }
}

print "The maximum value in the second column is: $max\n";

exit();

```

```

//pull_analysis.txt//

#!/usr/bin/perl
use strict;

#
# The input to this program is a pullf and a pullx file from the same run. The program
# will figure out total displacement of the pull group, then normalize it to the center
# of mass of the initial position. Then it will pair up the distance column with the
# forces in the force file, creating one final xvg file. The integral of this file should
# be PMF.
#

unless (@ARGV){
    die "Usage: ./${0} <pullf.xvg> <pullx.xvg>\n";
}

open (INPUT1, $ARGV[0]);
my @pullf = grep !/^(#|@)/, <INPUT1>;
close (INPUT1);

open (INPUT2, $ARGV[1]);
my @pullx = grep !/^(#|@)/, <INPUT2>;
close (INPUT2);

my @pullx_vector;
push @pullx_vector, "0";

my $first_line = shift @pullx;
my ($time1, $x01, $y01, $z01, $dx1, $dy1, $dz1) = split (' ', $first_line);
my $normalized_value = ($dx1**2 + $dy1**2 + $dz1**2)**0.5;

foreach (@pullx){

    my ($time, $x0, $y0, $z0, $dx, $dy, $dz) = split (' ', $_);
    my $vector = ($dx**2 + $dy**2 + $dz**2)**0.5 - $normalized_value;
    push @pullx_vector, $vector;
}

my @final;

for (my $i=0; $i<scalar(@pullf); $i++){

    my ($time, $force) = split(' ', $pullf[$i]);
    push @final, $pullx_vector[$i]."\t".$force."\n";
}

open (OUTPUT, ">", "output.xvg");
print OUTPUT @final;
close OUTPUT;

exit();

```

```

//transformer.pl//

#!/usr/bin/perl
use strict;

#
# The input for this program is a .gro file containing a POPC lipid bilayer. You specify
# a number of lipids (fewer than the total number of lipids), and the program will choose
# that many lipids at random and transform them to POPE lipids. Useful for making a mixed
# bilayer from a pure POPC bilayer.
#

if (scalar(@ARGV) != 3){
    print "Usage: `perl $0 <filename.gro> <# POPE to make> <out_filename.gro>`\n";
    exit();
}

my $filename = $ARGV[0];
my $num_to_change = $ARGV[1];
my $out_filename = $ARGV[2];

open (INPUT, "$filename") or die "It seems \"$filename.\" does not exist!\n";
my @bilayer = (<INPUT>);
close (INPUT);

my $title = shift @bilayer;
my $num_of_atoms = shift @bilayer;
my $box_vectors = pop @bilayer;

my @num_of_residues = grep /P8/, @bilayer;
print "There are ".scalar(@num_of_residues)." POPC residues\n";

my @integers = (1..scalar(@num_of_residues));
my @changers;

for (my $i=0; $i<$num_to_change; $i++){
    my $pos = int(rand(scalar(@integers)));
    my $changer = splice(@integers, $pos, 1);
    push @changers, $changer;
}

my @sorted = sort(@changers);
print "You are trying to change ".scalar(@sorted)." residues from POPC to POPE\n";

my @bilayer_unpacked;
my @bilayer_changed;

foreach (@bilayer){
    my $newline = pack("A6 A5 A7 A29", unpack("A5 A4 A6 A29", $_));
    push @bilayer_unpacked, $newline;
}

foreach my $old (@bilayer_unpacked){
    my $new_line;
    my $flag;
    my ($resid, $resname, $atomname, $atomid, $bx, $by, $bz) = split(' ', $old);

    foreach my $new (@changers){
        if ($resid eq $new){
            if ($atomname eq "C1"){
                $new_line = ($resid." POPE H1 ".$atomid." ".$bx." ".$by."
".$bz."\n");
                $flag = 1;
            } elsif ($atomname eq "C2"){
                $new_line = ($resid." POPE H2 ".$atomid." ".$bx." ".$by."
".$bz."\n");
                $flag = 1;
            } elsif ($atomname eq "C3"){
                $new_line = ($resid." POPE H3 ".$atomid." ".$bx." ".$by."
".$bz."\n");
                $flag = 1;
            } else {

```

```

    $new_line = ($resid." POPE ".$atomname." ".$atomid." ".$bx."
".$by." ".$bz."\n");
    $flag = 1;
    }
    }
    if ($flag == 1){
        # do nothing
    } else {
        $new_line = ($resid." ".$resname." ".$atomname." ".$atomid." ".$bx." ".$by."
".$bz."\n");
    }

    $flag = 0;
    push @bilayer_changed, $new_line;
}

my @bilayer_final;
my $final_line;

foreach (@bilayer_changed){
    $final_line = pack("A5 A5 A5 A5 A8 A8 A8", split(' ', $_));
    push @bilayer_final, $final_line."\n";
}

push @bilayer_final, $box_vectors;
unshift @bilayer_final, $num_of_atoms;
unshift @bilayer_final, $title;

open (OUTPUT, ">", "123456abcdef.gro");
print OUTPUT @bilayer_final;
close (OUTPUT);

system("editconf -f 123456abcdef.gro -o ".$out_filename." &>/dev/null");
system("rm 123456abcdef.gro");

exit();

```