Fall 2007

Molecular Dynamics Simulation Studies of DNA and Proteins: Force Field Parameter Development of Small Ligands and Convergence Analysis for Simulations of Biomolecules

Anne Loccisano

MOLECULAR DYNAMICS SIMULATION STUDIES OF DNA AND PROTEINS: FORCE FIELD PARAMETER DEVELOPMENT FOR SMALL LIGANDS AND CONVERGENCE ANALYSIS FOR SIMULATIONS OF BIOMOLECULES

A Dissertation

Submitted to the Bayer School

of Natural and Environmental Sciences

Duquesne University

In partial fulfillment of the requirements for

the degree of Doctor of Philosophy

by

Anne Elizabeth Loccisano

December 2007

MOLECULAR DYNAMICS SIMULATION STUDIES OF DNA AND PROTEINS: FORCE FIELD PARAMETER DEVELOPMENT FOR SMALL LIGANDS AND CONVERGENCE ANALYSIS FOR SIMULATIONS OF BIOMOLECULES

By

Anne Elizabeth Loccisano

Approved December 2007

_____

Jeffrey D. Evanseck, Ph.D.
Dissertation Advisor
Professor of Chemistry and Biochemistry
Duquesne University

_____

Charles T. Dameron, Ph.D.
Committee Member
Associate Professor of Chemistry and Biochemistry
Duquesne University/St. Francis University

_____

Steven M. Firestine, Ph.D.
External Reviewer
Assistant Professor of Pharmaceutical Sciences
Wayne State University

_____

Jeffry D. Madura, Ph.D.
Chairman, Department of Chemistry and Biochemistry
Professor of Chemistry and Biochemistry
Duquesne University

_____

David W. Seybert, Ph.D.
Dean, Bayer School of Natural and Environmental Sciences
Professor of Chemistry and Biochemistry
Duquesne University

ABSTRACT

MOLECULAR DYNAMICS SIMULATION STUDIES OF DNA AND PROTEINS: FORCE FIELD PARAMETER DEVELOPMENT FOR SMALL LIGANDS AND CONVERGENCE ANALYSIS FOR SIMULATIONS OF BIOMOLECULES

By

Anne Elizabeth Loccisano

December 2007

Dissertation Supervised by: Dr. Jeffrey D. Evanseck

In the first part of this dissertation, CHARMM force field parameters for DNA minor groove-binding polyamides were developed. The parameterization involved the subdivision of the polyamides into model compounds, which were calibrated against MP2/6-31G(d) data. To test the new parameters, fourteen 10 ns molecular dynamics crystal simulations have been carried out on a DNA/polyamide complex at low (113K) and high (300K) temperatures. Of the 18 helical parameters examined, only one (stagger) is found to be statistically significant from the crystal structure with a t-test at the 95% confidence level. For the high temperature, stagger is non-significant at the 97% confidence level, which underscores the importance of running multiple trajectories. It is observed that when the simulations are run at 300K, the DNA fragment begins to distort; however, better sampling is achieved. Competition between water and polyamides for hydrogen bonding to DNA is found to explain weak or unpredictable binding.

In the second part, force field parameters for retinoids were developed. The retinoids were divided into model compounds and calibrated against MP2/6-31G(d) data. To test the parameters, five molecular dynamics crystal simulations of reported x-ray structures of protein/retinoid complexes were performed. The structural and geometric analysis of these simulations compares well to experiment, and some dynamics that could be important to ligand binding were discovered. The new parameters can now be used in simulations of retinoid-binding proteins to better understand these systems and in drug design to make new retinoids with therapeutic and anticancer potential.

The last part explores the convergence of structural parameters in biomolecular systems. A simple statistical test was applied to the different parameters from a few long and many short simulations to observe which strategy is best. For the protein, both the long and short simulations gave similar results with respect to convergence. For the DNA, it was found that fraying effects penetrate four base pairs in from the ends of the helix. Structural parameters converge more quickly for the middle four bases than for all bases, and the long simulations yielded better results with respect to convergence than the short simulations.

To Ben and my family

ACKNOWLEDGEMENTS

I am truly fortunate to have had all the wonderful people in my life who have helped and influenced me during my graduate career. This was the most difficult task I have taken on thus far in my life, and I needed all the help I could get.

I want to first thank my graduate advisor, Dr. Jeffrey Evanseck, for everything he has done for me. He has been an inspiration to me, and I have learned so much more from him than I ever expected I would as a graduate student. He has supported, guided, and pushed me as a graduate student, and I am so grateful for his time and patience with me.

Thank you to my dissertation committee members, Dr. Jeff Madura, Dr. Charles Dameron, and Dr. Steven Firestine (my outside reader), for all of their advice, ideas, guidance, and just everyday conversation. Dr. Madura's classes showed me that there is no reason to fear quantum mechanics, and all of these professors have inspired, challenged, and pushed me to achieve my best.

I would like to thank our collaborators, Dr. Alex MacKerell, Dr. Steven Firestine, and Dr. John Kern. They read the long documents I sent them, they offered their guidance and advice on my research projects, they were patient in answering all my questions (no matter how ridiculous), and I am grateful for all they taught me. Dr. Kern taught me so much about statistics (and he was so patient about it!) and the S-plus program, and I gained a real appreciation for all of it.

I want to thank Dr. Paul Johnson for all of his guidance and support as my teaching supervisor. I couldn't have asked for a better supervisor—he has confidence in

his TAs, and he gave us the freedom to run our organic labs in the ways that worked for us. Dr. Johnson gave me confidence in my teaching abilities, and I really enjoyed teaching my organic labs under his guidance.

Ian, Sue, and everyone who works in the stockroom—these are the people who are responsible for making my teaching labs run so smoothly. I want to thank them for answering my lab/stockroom questions, always having equipment and chemicals ready for my organic labs, and running to the stockroom whenever my students needed anything.

Sandy, Amy, Mary Jo, Julie, Andrea, Dianne, and all the janitors and maintenance men—I don't know what I would have done without all of these people! They have all helped me out in some way or another, whether it was answering stupid questions; showing me how not to break the photocopier, fax machine, or printer; lending me the vacuum cleaner, or fixing the thermostat in the lab. They have all helped me to avoid major crises on more than one occasion.

Carolina—She was so patient with all my questions about health insurance, the library, all the paperwork I had to fill out, and everything else concerning my life at this school. Carolina always went out of her way to help me out, and she listened when I had a problem that I though was life or death (like parking). I want to thank her for all of her help and for just being a friend. Also, thanks to Mary Ann—she was also a great help with administrative matters while she was here.

Thank you to the past and present members of the Evanseck and Madura research group members who have helped me, supported me, and are just my friends. Orlando Acevedo, Amy Waligorski, Jason DeChancie, Tugba Kukkucal, Sarah Mueller Stein,

Melinda Harrison, the late Heather Altman, Kristine Baker, Ed Franklin, Saichand Pakkala, Jennifer Makowski, Tom Dick, Pranav Dalal, Peter Krouskop, Martine Indarte, and Shankar Manapelli—all of them have helped me out, made me laugh, or were just there to listen at one time or another, and I am grateful to all of them for what they taught me and for their friendship. I'd also like to give a shout out to those who did not support me.

Thanks goes to the other past and present grad students who have worked in the department with me who will always be my friends. Eliza Ruben—she came to our group for a few years, and she became a good friend to me. She was always there for advice, complaining on various topics, and just conversation. I'm truly grateful for her friendship and that she could be a part of the Evanseck group for a while. Brian Kail, John Williams, Shayna Smith, Jonathan Lekse, Tyler Collins, RJ Mycka, Lakshmi Menon, Rachelle Palchesko, Courtney Sparacino, Kelly Papariella, Rosalynn Quinones, Chris Kabana, Will Eckenhoff, Bobby Rager, and Mitch Fedak—these people were always there to listen and commiserate with. They have all willingly put themselves through the same pain of graduate school or they simply witnessed the distress from the sidelines. All have given me advice, support, friendship, or lunch at one time or another, and they reminded me that I was not alone in feeling overwhelmed, frustrated, or ready to break someone's arms. I'd also like to thank George Lusnak, who I taught several organic lab sections with. I think George was probably the best co-TA—he knows how to make all the experiments work quickly and get the students out of the lab, and best of all, his stories are great. I was never bored in the organic lab when I taught with George.

I want to thank my friends for their constant support and understanding of this whole process and the chaos that is my life. Jessica Bellian, Honore Sullivan-Galli, Rayeann Gretz, Nicole Falace, Kendra Brozich, Erin Black, Kelly Bolden, Angela Hughes, Sara Silvestri, Julie Mackey, Kristine Baker, Jen Makowski, Shayna Smith, and Len Vuocolo—they know me better than anyone, and they are the best friends anyone could ask for. They have encouraged me, pushed me, helped out in so many ways, and most of all, they listen to my stories. They are always there to remind me that I am loved and appreciated.

I want to thank my parents, brother, grandparents, and the rest of my family for all their love and patience. Mom, Dad, and Aaron—I would have never been able to get this far or pull this off without their unwavering support, advice, and love (they have also supported my efforts to be the eternal student; I know, I have to get a job now). The discipline that my parents instilled in me enabled me to finish what I started here. My dad always told me that your education is the one thing that no one can ever take away from you, and that has been one of my guiding principles the whole time I have been in school. I figured I better get a lot of it in case I lost everything else.

Above all, I want to thank my husband, Ben. He has been so unbelievably supportive and understanding of everything I went through as a graduate student. He, like my parents, supported the student cause—he encouraged me to finish my degree even at the times I wanted to quit, and it meant that I had no money. He believed in me when I didn't. He used to cook rice for me and bring it to me at the lab when I had to work really late—what more can I ask for? He did everything he could to help me out, from housework to job applications. He was my inspiration to finish this degree, and

some days, he was my only sanity. His constant love, support, advice, belief in me, and everything else he does for me reminds me every day how lucky I am to have found someone who understands me in the way that he does. I do not have words to describe how truly fortunate I am to have him in my life, how much he means to me, or how much I love him. Part of this degree is his.

Table of Contents

Page

Abstract……...iv

Dedication……...vi

Acknowledgements……vii

List of Figures……...xiv

List of Tables……xx

Chapter 1 Introduction 1

1.1 Historical Perspective on Simulation 1

1.2 Force Field Development 2

1.3 Project Overviews 6

1.4 DNA Minor Groove-Binding Polyamides 7

1.5 Retinoids and π-Conjugated Systems 17

1.6 Convergence Analysis of Biomolecular Simulations 24

References 32

Chapter 2 Methods 56

2.1 Molecular Dynamics 56

2.2 Analysis of Trajectories 71

2.3 Quantum Mechanical Methods 76

2.4 Description of Calculations 90

References 109

Chapter 3 Force Field Parameters for DNA Minor Groove-Binding Pyrrole-Imidazole Polyamides and Significance of Multiple Simulations 117

3.1 Parameterization of DNA minor groove-binding polyamides 117

3.2 DNA/polyamide Simulations in the Crystal Environment 134

3.3 Conclusions 180

References 185

Chapter 4 Force Field Parameters for the Simulation of Conjugated Dienes with a Focus on Retinoids 192

4.1 Parameterization of Retinoids 192

4.2 Comparison to Available Experimental Data 208

4.3 Retinoid Crystal Simulations 217

4.4 Retinoid/Protein Complex Simulations 221

4.6 Conclusions 287

References 290

Chapter 5 Convergence Testing in Biomolecular Simulations 300

5.1 Potential Scale Reduction Test 300

5.2 Protein Convergence 303

5.3 DNA Convergence 311

5.4 Conclusions 325

References 330

Appendix A 333

Appendix B 407

List of Figures

Page

Figure 1.1 3-D representation of polyamides bound 2:1 in the minor groove of DNA. .. 10

Figure 1.2 Schematic diagram of Im-Py-Hp polyamide fragment. 13

Figure 1.3 Structures of four retinoids 17

Figure 3.1 Model compounds for polyamides..... 118

Figure 3.2 Structure (top) and the three torsion energy plots for N-methylpyrrole. 122

Figure 3.3 Structure (top) and the three torsion energy plots for N-methylimidazole. . 124

Figure 3.4 Structure (top left) and the four torsion energy plots for hydroxypyrrole ... 127

Figure 3.5 Structure and the two torsion energy plots for the cationic tail. 128

Figure 3.6 Structure and the two torsion energy plots for the γ linker. 131

Figure 3.7 Structure and the two torsion energy plots for the β-alanine linker..... 133

Figure 3.8 Structure of netropsin and RMSD plot for the netropsin simulation 135

Figure 3.9 Time series for one dihedral of the amidino group tail..... 136

Figure 3.10 Standard reference frame and helical parameters 138

Figure 3.11 Starting structure for DNA/Dervan polyamide simulations 141

Figure 3.12 Sample output from 3DNA for a DNA 10-mer for six helical parameters. 143

Figure 3.13 Time series plots of t-values over time for the stagger helical parameter. 146

Figure 3.14 Time series plots of t-values over time for the buckle helical parameter .. 147

Figure 3.15 Principal component analysis plots of stagger..... 148

Figure 3.16 Principal component analysis plots of buckle..... 150

Figure 3.17 RMSD vs. time plots for the 113 K simulations and for the 300 K simulations compared to the starting structures for the simulations. 151

Figure 3.18 T-test for all eight low temperature simulations. 154

Figure 3.19 T-test for all six high temperature simulations 155

Figure 3.20 T-test for three new high temperature simulations 156

Figure 3.21 T-test for combined high and low temperature simulations 158

Figure 3.22 Comparison of x-ray structure and high temperature simulated structure. 159

Figure 3.23 Comparison of x-ray structure and low temperature simulated structure.. 161

Figure 3.24 Probability density curve and histogram bar plot for minor groove width of low temperature simualtions. 162

Figure 3.25 Probability density curve and histogram bar plot for minor groove width of high temperature simualtions. 163

Figure 3.26 Polyamide distances measured (dashed lines) in the x-ray structure and from the low and high temperature crystal simulations..... 165

Figure 3.27 Plot of distance from water to thymine O2 over the 10 ns time span of the simulation for low temperature simulation 167

Figure 3.28 Starting structure for DNA/hydroxypyrrole polyamide simulation..... 169

Figure 3.29 Polyamide distances measured (dashed lines) in the x-ray structure and from the hydroxypyrrole polyamide crystal simulations. 171

Figure 3.30 Starting structure for DNA/netropsin simulation..... 172

Figure 3.31 RMSD vs. time of DNA/netropsin complex. 174

Figure 3.32 Hydrogen bond distances (numbered) from netropsin to DNA..... 175

Figure 3.33 RMSD vs. time of DNA/distamycin complex 176

Figure 3.34 Hydrogen bond distances from distamycin to DNA (numbered) 177

Figure 3.35 RMSD vs. time of DNA/netropsin complex using AMBER parameters .. 179

Figure 4.1 Model compounds and patch residues for retinoids..... 194

Figure 4.2 Previous 1,3,5-hexatriene carbon atom types 195

Figure 4.3 Torsion energy plots for 1,3-butadiene and 1,3-pentadiene..... 197

Figure 4.4 Torsion energy plots for 2-methyl-1,3-butadiene, 4-methyl-1,3-pentadiene, and 2-methyl-1,3-pentadiene 200

Figure 4.5 Torsion energy plot for N-3-dimethyl-2-butenamide and *p*-acetamide phenol 201

Figure 4.6 Torsion energy surfaces for 1,6,6-trimethyl-2-ethene-cyclohexene 202

Figure 4.7 Torsion energy surfaces for 2-propenoic acid, 2-propenol, and 2-propenal . 204

Figure 4.8 Torsion energy surfaces for 2-butene (left) and 1,3,5-hexatriene (right)..... 206

Figure 4.9 Torsion energy plot for deprotonated and protonated Schiff base 207

Figure 4.10 The left structure is that of retinal 214

Figure 4.11 The structure on the right is that of retinoic acid 215

Figure 4.12 Comparison of monoclinic crystal and CHARMM forms..... 216

Figure 4.13 Examples of overlayed x-ray and simulation average structures..... 218

Figure 4.14 Examples of overlayed x-ray and simulation average structures..... 220

Figure 4.15 Example of overlayed x-ray and simulation average structures 221

Figure 4.16 The four retinoids built from the model compounds for use in simulations. 222

Figure 4.17 Retinoic acid complexed with CRABPII 223

Figure 4.18 RMSD plots of the retinoid acid/CRABPII complex..... 224

Figure 4.19 RMSD over time of different parts of the retinoic acid ligand. 225

Figure 4.20 Overlay of the average structure of the ligand from the simulation 227

Figure 4.21 Distance vs. time plots of Arg132 and Tyr134 to retinoic acid. 229

Figure 4.22 Distance from Arg111 to different water oxygens..... 231

Figure 4.23 Two of the nonpolar ligand-protein contacts (retinoic acid/CRABPII)..... 232

Figure 4.24 Two of the protein-protein distances (retinoic acid/CRABPII). 233

Figure 4.25 Retinol complexed with zebrafish CRBP 235

Figure 4.26 RMSD plots of the retinol/CRBP complex 237

Figure 4.27 Overlay of x-ray (red) and simulation average structure (blue)..... 238

Figure 4.28 Polar protein/ligand contacts (retinol/CRBP). 240

Figure 4.29 Two of the contacts from the retinol ligand chain to the protein. 241

Figure 4.30 Two of the contacts of the retinol ligand ring to the protein..... 242

Figure 4.31 Water in binding cavity of zebrafish CRBP..... 244

Figure 4.32 Starting structures for retinoic acid/CRABPII complex 245

Figure 4.33 Fenretinide/RBP complex 246

Figure 4.34 RMSD plots of the fenretinide/RBP complex..... 248

Figure 4.35 RMSDs for different parts of the ligand 249

Figure 4.36 Overlay of x-ray structure of ligand and simulation average structure of ligand..... 250

Figure 4.37 Distance from hydroxy group of the phenol ring of the ligand to various water molecules. 251

Figure 4.38 Distance over time for sidechain carbons of Leu64 to ligand..... 253

Figure 4.39 Distances over time for resides that move away from the ligand during the 1.5-2.0 ns time period. 254

Figure 4.40 Overlay of snapshots of the fenretinide ligand. 255

Figure 4.41 Distance from Leu35 to ligand..... 257

Figure 4.42 Overlays of the ligand/protein complex..... 259

Figure 4.43 Snapshot from 500 ps (before distance increase)..... 260

Figure 4.44 Snapshot from 1600 ps (after distance increase) 260

Figure 4.45 Sensory rhodopsin II complexed with retinal 262

Figure 4.46 RMSD plots of the protein-ligand complex..... 263

Figure 4.47 Overlay of the x-ray ligand (red) with the simulation average structure (blue). 264

Figure 4.48 Schiff base linkage over time. 265

Figure 4.49 Cl^- ion distances to nearby residues (retinal/SRII)..... 267

Figure 4.50 Distances for Lys205, Asp75, Asp201, water 401, and water 402. 269

Figure 4.51 Distances from retinal ligand to aromatic residues. 270

Figure 4.52 Protein-protein contacts near the ligand (retinal/SRII). 271

Figure 4.53 Protein water distances near the ligand (retinal/SRII). 273

Figure 4.54 Three of the protein-ligand contacts measured from the simulation (retinal/SRII). 274

Figure 4.55 Retinal complexed with bacteriorhodopsin..... 276

Figure 4.56 RMSD plots for retinal/BR complex 278

Figure 4.57 Overlays of x-ray structure (red) and simulation average structure (blue)..279

Figure 4.58 Schiff base linkages for retinal/BR complex. 280

Figure 4.59 Protein-protein contacts measured from BR/retinal simulation..... 282

Figure 4.60 Four of the ligand-protein distances measured from the BR/retinal simulation..... 284

Figure 4.61 Water-ligand distances measured from the BR/retinal simulation. 286

Figure 5.1 Examples of time series plots for hydrogen bonds…….....305

Figure 5.2 Time series plots for radius of gyration and helix-helix distance……..306

Figure 5.3 PCA plot for radius of gyration……. 308

Figure 5.4 Examples of time series plots for short-time crambin simulations……310

Figure 5.5 Time series plots for propeller and stretch for long DNA simulations……..313

Figure 5.6 RMSD plots for long DNA simulations…….315

Figure 5.7 Time series plots for twist and y-displacement……..318

Figure 5.8 Time series plots for opening and shift……..320

Figure 5.9 Time series plots for shift and opening from short simulations…….322

List of Tables

Page

Table 3.1 RMSDs of complexes, DNA, and polyamides 142

Table 3.2 RMSDs of complexes, DNA, and polyamides 152

Table 4.1 Experimental, QM, and MM comparison for 1,3-butadiene. 209

Table 4.2 Experimental, QM, and MM comparison for 2-methyl-1,3-butadiene. 210

Table 4.3 Experimental, QM, and CHARMM comparison of 2-propenoic acid. 211

Table 4.4 Experimental, QM, and MM comparison of 1,3,5-hexatriene 212

Table 4.5 Experimental, QM, and MM comparison of *p*-acetamide phenol[b] 213

Table 4.6 RMSD of ligands for triclinic form of retinoic acid 218

Table 4.7 RMSD of ligands compared to x-ray structure for retinal 219

Table 4.8 RMSD of ligands for monoclinic form of retinoic acid 220

Table 4.9 Protein-ligand distances (CRABPII/retinoic acid complex) 228

Table 4.10 Protein-ligand distances (CRBP/retinol complex) 239

Table 4.11 Protein-ligand distances (RBP/fenretinide complex) 252

Table 4.12 Protein-ligand and protein-protein distances (SRII/retinal complex) 266

Table 4.13 Protein-ligand and protein-protein distances (BR/retinal complex) 281

Table 5.1 PSRT results (long-time crambin simulations) 304

Table 5.2 PSRT results (short-time crambin simulations) 309

Table 5.3 Parameter averages for long and short crambin simulations 311

Table 5.4 PSRT results (long-time DNA simulations; 12 bases) 312

Table 5.5 PSRT results (long-time DNA simulations; 4 bases) 317

Table 5.6 PSRT results (short-time DNA simulations; 12 bases) 319

Table 5.7 PSRT results (short-time DNA simulations; 4 bases) 321

Table 5.8 Parameter averages for long and short DNA simulations (12 and 4 bases) ... 324

Chapter 1

Introduction

1.1 Historical Perspective on Simulation

The first molecular dynamics simulation of a biomolecular system was published in 1977.[1] The simulation was of bovine pancreatic trypsin inhibitor (BPTI), and although the system was small (58 residues; 885 atoms), the simulation was performed in vacuum, the force field was primitive, and the simulation only lasted 9.2 ps, it showed that proteins exhibit substantial fluctuations on the picosecond timescale.[1] The first molecular dynamics simulations of DNA followed in 1983 and were performed on a 12- and 24-base pair fragment in vacuum and without accounting for electrostatic forces.[2] Since then, a wide range of protein and nucleic acid motions and structure features have been discovered and explored through molecular dynamics simulations.[3-11] Advances in computer hardware, software, force fields, improvement in simulation methodology, and also the tremendous progress in experimental structure determination have allowed the field of MD simulation to grow at a rapid rate. Since the first BPTI simulation in 1977 in vacuum on a short timescale, these advances now enable large (10^4-10^6 atoms) and complicated systems[3, 12] (such as a membrane-bound protein with explicit lipid and

water molecules) to be simulated on longer timescales (hundreds of nanoseconds to microseconds) with a more realistic representation of the solvent and surroundings.

Molecular dynamics simulations can provide the methodology to model, in ultimate detail, the internal atomic motions and fluctuations of a biomolecular system as a function of time. MD simulations allow for observation of phenomena and properties of systems that cannot be observed with experimental techniques. Although MD simulations have been widely used to give insight into the underlying natural dynamics, observe functionally important motions, make predictions, and facilitate experimental structure determination, the dynamics of many biomolecular systems are poorly or not understood.[12-14] In addition, in order to compute and extract accurate information from the modeled system, the models used must be as realistic as possible (e.g., the force field used must be parameterized for the molecules of interest). Also, the various properties of interest that are calculated need to have reached satisfactory levels of convergence; the properties must be stabilized if analysis is to be performed on them.[15, 16]

1.2 Force Field Development

In molecular dynamics and molecular mechanics, the intra- and intermolecular forces between the atoms are described by an empirical force field (the CHARMM force field, which is used in this dissertation, is discussed in Chapter 2). The mathematical terms in empirical force fields that are used to describe the intra- and intermolecular forces in biomolecules are relatively simple, and force fields assume that an atom is the smallest particle in the system of interest, rather than considering nuclei and electrons explicitly as in quantum mechanics.[3, 17] Both of these factors

allow for the computational demands of biomolecular simulation studies to be fulfilled. However, the force field parameters must be developed correctly in order to obtain accurate and reliable results and predictions.

The quality and accuracy obtained from a calculation performed with a particular force field is obviously based on the methods and target data used to optimize the parameters. Several considerations should be taken into account when parameterizing a force field for a new class of compounds.[18] The first is the issue of transferability; this refers to the ability to take parameters optimized for a specific set of target data and then apply those parameters to molecules that are not included in the target data. In a transferable force field, the dihedral parameters that are optimized for rotation about the C-C bond of ethane could then be used for butane. In a nontransferable force field, the dihedrals present in butane but not in ethane (C-C-C-C) must be optimized specifically by using target data for butane. However, how transferable parameters are depends on the similarity of the molecules, and what parameters are actually transferable and which are not is not well-defined.[18] Any new compounds to be used with the force field must be tested in order to ensure that they are treated correctly and the required accuracy is obtained. Generally, the extent of transferability is considered to be minimal, and new parameters must be generated for each new class of molecules to be used with the force field.[18]

Since new parameters must usually be generated and optimized, a second consideration is the selection of target data that is to be used to optimize the new parameters. The form of the CHARMM force field is discussed in Chapter 2; therefore, the terms will not be described and shown here. The parameters needed to

extend the force field for a new class of compounds will fit into the different terms of the force field, and these are generally optimized using different sources. The internal parameters such as equilibrium bond lengths, equilibrium angle values, and dihedral multiplicity are often optimized to reproduce gas-phase geometric data obtained from quantum mechanical calculations, electron diffraction, or microwave experiments. The internal force constants, such as the bond and angle force constants, are usually optimized using vibrational spectra, which contain individual frequencies and their assignments. The external parameter optimization is usually more difficult than the internal parameter optimization; the amount of target data relative to the amount of needed parameters is decreased compared to the internal parameters.[18] Upon the development of the OPLS force field, optimization of van der Waals terms improved.[19] Heats of vaporization of pure solvents were computed through Monte Carlo simulations, and this data was then used to refine the van der Waals parameters.[19] This same approach has also been applied to the CHARMM and AMBER force fields.[20, 21] Van der Waals parameters have also been optimized using the calculated heats of sublimation of crystals.[22] The approaches for optimization of electrostatic parameters are mostly dominated by the reproduction of target data from QM calculations.[18] One method is based on optimizing partial atomic charges to reproduce the electrostatic potential (ESP) around a molecule calculated with QM methods. A popular variation of this method is called the RESP method,[23] where the charges on atoms that are minimally exposed to solvent are restrained. The goal of both methods is to produce the partial atomic charges that reproduce the electrostatic field created by the molecule.[24]

When parameterizing the force field for a new set of compounds, it is important to remember that the force field is "empirical". The choice of method and target data and even the form of the potential energy function is the decision of the developer of the force field.[18] Even if the same form of the potential energy function is used, if developers have used different target data, the quality and results of the force field can differ. A great amount of correlation exists between all the different, individual parameters of the force field, and therefore, a number of different combinations of parameters can reproduce a given set of target data.[18] Therefore, the optimization process is extremely important to the quality and results of the force field. Automatic parameterization procedures have been attempted;[25-27] however, a significant amount of manual work is generally required.[18]

A third consideration is the procedure used in which to optimize the force field. When a force field is selected for use, the information that one wishes to extract must be considered. For example, if one is interested in examining the atomic details of water interactions with protein residues, the proper force field to use would be an all-atom force field (as opposed to an extended atom force field where hydrogens are not explicitly represented but treated as part of the atom to which they are bound) designed specifically for biomolecules that allows for explicit representation of water molecules. In this case, the AMBER or CHARMM force fields would be a suitable choice; however, they may not be of sufficient accuracy for a particular study.[18] If the study mentioned above also involved different but structurally similar ligands being bound to the protein, the AMBER or CHARMM force field must be extended to treat the new class of compounds. The optimization

method used must be consistent with the method used for the original development of the force field. Chapters 3 and 4 of this dissertation contain the details of how the CHARMM force field was parameterized and tested for two new sets of compounds, polyamides and retinoids.

1.3 Project Overviews

This dissertation focuses on the three issues mentioned above: force field parameterization, gaining insight into dynamics of the systems using the new force field parameters, and analyzing biomolecular systems for convergence. The first part of the force field parameterization focuses on the polyamides, which are a class of DNA minor groove binders.[28] The molecules have potential as DNA sequence-specific recognition agents; however, why some show poor binding affinity for their target sites and why some sites are more difficult than others to target is not well understood. Force field parameters for the polyamides were created, and through testing the parameters with molecular dynamics simulations in the crystal environment, a new conceptual idea as to why polyamides are weak binders emerged as well as an interesting phenomenon concerning the sampling of DNA helical parameters. The second part of the force field parameterization focuses on retinoids. Attempts to model π-conjugated systems have been difficult to implement[29-31] or the compounds have not been explicitly parameterized.[26, 32] New parameters were made for conjugated systems with a focus on retinoids. Retinoids are an attractive class of compounds in anticancer therapy now; however, toxicity limits their use.[33, 34] Through testing the parameters with crystal simulations, interesting details were uncovered with respect to ligand motions that could be important in ligand binding

and dynamics. These motions enable better understanding of how the ligand behaves in the binding site of the protein, and this can aid in the future design of less toxic retinoids for use as anticancer agents. The last part of this dissertation focuses on the convergence of structural/geometric parameters in two biomolecular systems, DNA and a small protein. Multiple trajectories are commonly run in order to improve sampling of conformational space,[35-37] and this approach was used in this study. A set of a few long simulations and a set of many short simulations were run, the parameters of interest were measured, and a statistical test was applied to the different parameters from the sets of simulations to observe which simulation strategy works best.

1.4 DNA Minor Groove-Binding Polyamides

The sequence specific recognition of DNA by small, organic molecules is an area of intense research due to the ability of these agents to interfere with critical biochemical processes involving DNA. Pioneering work by Dervan,[38-57] Dickerson,[58-63] and Lown[64-70] on analogues of the natural products distamycin and netropsin (which are made up of three and two N-methylpyrrole (Py) rings, respectively) has led to the development of compounds that are capable of recognizing any base pair combination found in DNA. These compounds, termed polyamides or lexitropsins, have functional groups that are complementary to the distinct pattern and position of hydrogen bond donors and acceptors found in the minor groove of DNA.[28, 39, 41, 45, 71-73] The polyamides are made up of five-membered (and recently, six-membered) heterocyclic amino acids linked by amide bonds that can be combined as antiparallel ring pairs in the minor groove of DNA to recognize

predetermined sequences of DNA, with affinities and specificities comparable to those of DNA-binding proteins (i.e., K_d < 1nM).[41]

The heterocycle pairs used in polyamides are pieced together with the intervening amide bonds in order to recognize a predetermined DNA sequence based on hydrogen bond donors and acceptors, molecular shape, and electrostatic potential of the minor groove. The Dervan group has established a code describing this relationship.[40, 74-77] It was first discovered that a Py/Py pair could distinguish an AT or TA pair from CG or GC pairs (AT/TA pairs are fairly symmetrical compared to the unsymmetrical nature of GC/CG pairs due to the exocyclic NH2 group of G); shortly after, it was found that an imidazole (Im)/Py pair could distinguish GC from the other three base pairs (due to a linear hydrogen bond between the Im nitrogen and the exocyclic NH2 group of G). It took more time to distinguish AT from TA; it was finally discovered that the Py/hydroxypyrrole (Hp) pair could distinguish AT from TA, due to the hydrogen bond formed between the hydroxyl group and the O2 of thymine. Thus, the basic Dervan pairing code is as follows: Im/Py pair binds GC, a Py/Im pair binds CG, a Py/Py pair binds AT or TA, and an Hp/Py pair and a Py/Hp pair distinguish TA and AT, respectively.[40, 41, 74] All polyamides contain a positively charged group at the tail end (dimethylaminopropyamine (Dp)), which interacts with the negatively charged phosphate backbone and increases affinity.

Polyamides can bind to DNA in either 1:1 or 2:1 stoichiometries; however, less is understood about polyamide recognition in the 1:1 mode.[78] It is known that in the 1:1 binding mode, polyamides can form bifurcated hydrogen bonds between bases on either one of the two strands in DNA. Also, because DNA sequences containing a

run of purine bases has a particularly narrow minor groove[79] and cannot accommodate two polyamides in the groove, the 1:1 motif shows promise in targeting these types of sequences.[78, 80]

In the 2:1 binding mode, which is the most common binding mode, the heterocycles of two polyamide molecules are paired and stack on top of each other when bound to the minor groove of DNA. In this mode, one polyamide is capable of reading one strand of DNA, while the other polyamide recognizes the second strand. In addition to forming hydrogen bonds to the DNA, the interaction is stabilized by the DNA-polyamide van der Waals interactions and π-stacking interactions formed between the ring pairs of the polyamides. Figure 1.1 shows a 3-D representation of how two antiparallel polyamides bind in the minor groove of DNA (left) and a schematic diagram of how the paired heterocycles recognize the hydrogen bond donor-acceptor pattern of the minor groove (right).

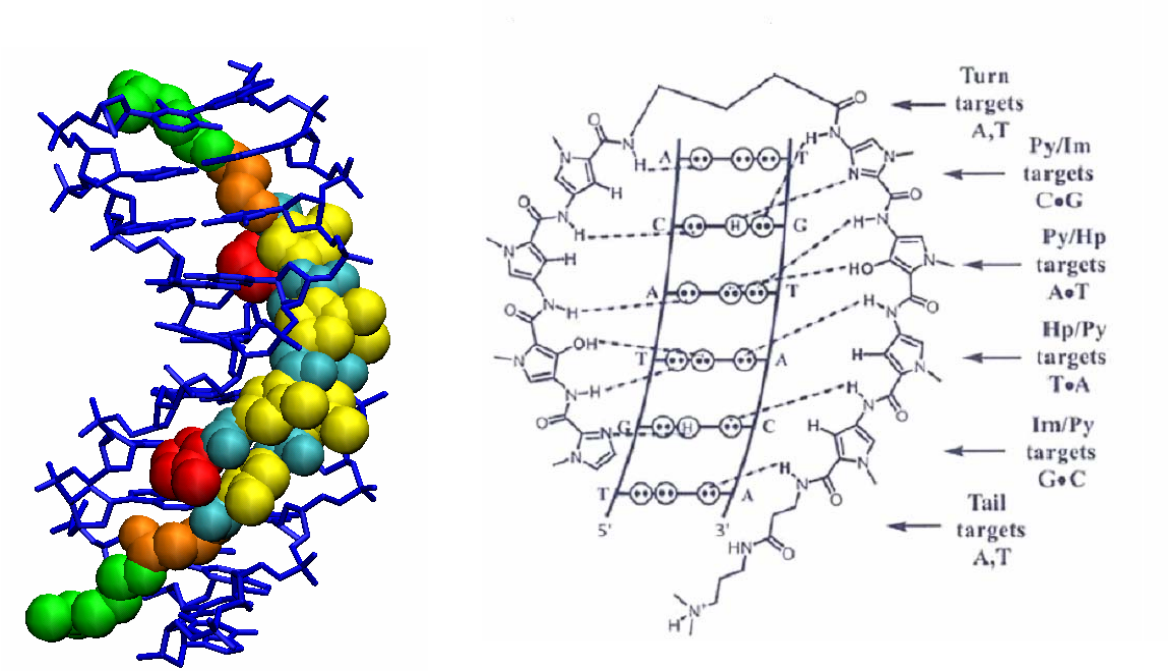


Figure 1.1 On the left: 3-D representation of polyamides bound 2:1 in the minor groove of DNA (PDB[81] id 1CVY[73]). DNA is shown as blue sticks, and the polyamides are shown in van der Waals representation (imidazoles in red, pyrroles in yellow, amide bonds in cyan, β-alanine linker in orange, and cationic tail in green). The structure contains the sequence 5'-CCAGATCTGG-3' complexed with two polyamides of sequence ImPyPyPy-β-Dp. On the right: Schematic diagram of how the paired heterocycles recognize the hydrogen bond donors and acceptor of the minor groove. The "ladder" in the middle is the DNA binding site, the dashed lines are the hydrogen bonds, the circles with the dots are the lone pairs on O2 of T and C and N3 of A and G, and circles with an H are the Hs of the exocyclic NH2 of G. The tail is made up of two parts: β-alanine, which targets A, T bases and the cationic Dp, which interacts with the phosphate backbone (figure adapted from Dervan, et al., 2005[71]).

In the 2:1 binding mode, the heterocycles are usually paired by the covalent linkage of two antiparallel polyamide strands; the covalent linkage results in increased affinity and specifity (100-fold higher affinity than their unlinked counterparts).[28] The standard motif for polyamides is a hairpin containing eight heterocycles, where two strands containing four heterocycles each are linked by a γ-aminobutyric acid linker. Called a γ-turn, it shows preference for binding AT/TA pairs over GC/CG pairs (due to steric clashes with the exocyclic NH2 group of G[82]),

and it connects the carboxylic end of one polyamide to the amino end of the other polyamide (thus, the polyamides are antiparallel). The γ-turn stabilizes the register of the heterocycle pairings with the DNA bases, which is most likely the reason for the higher affinity and specificity.[83] Hairpin polyamides align themselves with the DNA so that the amino end of the polyamide is oriented toward the 5' direction of the adjacent DNA strand, and the carboxylic end oriented toward the 3' direction of the adjacent DNA strand.[84]

Although the normal eight-ring hairpin motif binds with high affinity and specificity, this motif binds only six base pairs of DNA. In order for the polyamides to be useful in biology or medicine, specifically recognizing a very long stretch of DNA is important; longer sequences would be expected to occur less frequently in large genome.[28, 71] However, expanding the binding site size simply by adding extra heterocycles to the polyamides has been shown to decrease affinity and specificity because the curvature of the polyamides is tighter than that of the curvature of the DNA groove.[85] Actually, beyond five heterocycles, the polyamide shape is no longer complementary to the DNA groove.[47] In order to allow more conformational freedom in the polyamides, a flexible β-alanine residue is used. The β-alanine residue is used as a replacement for Py, so β/Py and β/Im pairs function as Py/Py and Py/Im pairs and simultaneously relaxe the curvature of the polyamide, restoring complementarity to the DNA groove.[86] Polyamides with β-alanine have been used to target successfully up to 16 base pairs with high affinity and specificity. [52]

Although most polyamide research has focused on the five-membered heterocycles, other heterocycles have recently been shown to bind to DNA.[28, 71]

These heterocycles are based on the benzimidazole (Bi) ring system, which consists of a 6-5 bicyclic ring structure. Different substituents can be added to the six-membered ring, and polyamides with the benzimidazole derivatives have a curvature that is more complementary to the DNA than the traditional five-membered ring polyamides.[87, 88] The benzimidazole derivatives (imidazopyridine (Ip) and hydroxybenzimidazole (Hz)) have been incorporated into the eight-ring hairpin polyamides and have been found to be as effective as the five-membered ring polyamides at recognizing the minor groove of DNA; the hydrogen bonding contacts between the DNA and polyamide are preserved.[55] Of particular interest are the Hz/Py and Py/Hz pairing. This pair can replace Hp/Py and Py/Hp pairs, and because Hp-containing polyamides have been found to degrade over time and the Hz polyamides are chemically stable, the Hz/Py pair is a candidate to replace the Hp/Py pair.[28] However, because the benzimidazole system is more hydrophobic and has a greater surface area, this may alter the DNA-polyamide van der Waals interactions and the π-stacking interactions between the polyamide ring pairs.[28]

Polyamides utilized in this dissertation are composed of three most common five-membered ring heterocycle types joined by an amide functional group. The NH group of the amide bonds is critical for binding as it forms specific hydrogen bonds with the bases when bound to the DNA. The three heterocycles, N-methylpyrrole (Py), N-methylimidazole (Im), and N-methyl-4-hydroxypyrrole (Hp) and the amide linkages are shown in Figure 1.2.

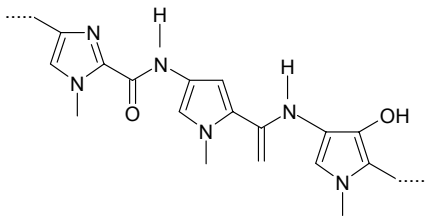


Figure 1.2 Schematic diagram of Im-Py-Hp polyamide fragment.

The Dervan pairing rules have been used to design literally hundreds of synthetic ligands that do bind their predetermined DNA target sequences with the high affinity and specificity comparable to DNA binding proteins. The polyamides have been shown to affect both transcriptional activation (by inhibiting repressor proteins and thus activating a particular gene)[43, 89-93] and repression (by inhibiting assembly of transcriptional proteins on DNA and thus downregulating a particular gene) in vitro and in cell cultures.[39, 42, 69, 94-98] Polyamides labeled with fluorescent dyes have also been shown to be taken up by cells, although nuclear uptake is dependent on cell type.[38, 44, 99] However, even though the polyamides bind to B-DNA and have been found to perform the functions mentioned above, there are limitations regarding targeted DNA sequences. Sequence dependent structural variations of the DNA minor groove affect the energetics of binding, a number of DNA sequences remain difficult to target (especially purine tracts), and binding affinity and specificity at some sequences is weak and/or unpredictable. Variation in the DNA minor groove width, curvature, flexibility, hydration, and the relative positions of hydrogen bond donors/acceptors could all possibly influence polyamide binding and affinity for a particular sequence.[71] Because molecular dynamics simulations allow for an atomistic view and examination of the DNA minor groove, this is an ideal tool for better understanding of the interactions of the polyamides with the DNA.

Computational studies are important for the description on how distamycin A, netropsin, and related polyamides bind to DNA.[72, 76, 78, 83, 86, 100-121] For example, molecular dynamics simulations have been used to aid in the refinement of NMR structures of DNA/distamycin and DNA/polyamide complexes.[72, 78, 83, 100, 102-104, 121] Free energy perturbation studies have been utilized to understand binding affinities of DNA/polyamide complexes,[103, 109] and unrestrained molecular dynamics have been employed to understand features of the complex such as hydrogen bonding, hydration, and structural distortions on the DNA induced by the polyamides.[108, 110-112, 114-120] Finally, quantum mechanical studies have been applied to understand the electronic and geometric properties of novel polyamides, and to interpret the results of experimental binding studies.[76, 101, 106]

The quality of the force fields implemented to treat polyamides varies greatly between studies. Close examination of the force field parameters utilized underscores the need for parameter development. For example, treatments of DNA complexes with an isopropyl-thiazole polyamide,[100] lexitropsin,[103] CPI-lexitropsin,[102] distamycin A,[104, 107-109, 121] netropsin,[109, 114, 117, 118, 122] 2-imidazoledistamycin,[107] carbocyclic analogs of netropsin and distamycin,[110, 111] carbocyclic derivatives of distamycin with a chlorambucil moiety,[112] ImHpPyPy-β-Dp,[115, 116, 119] ImPyPy-γ-PyPyDp,[112] ImPy-β-Im-β-ImPy-β-Dp,[78] Ac-ImPyPy-γ-PyPyPy-Gly-Dp,[72] and ImPyPy-γ-PyPyPy-Gly-Dp[72] have used default force field parameters that are best guesses or extrapolations from similar functionality. Force field development based upon the recruitment of similar terms and parameters may be successful in some cases;[18] however, in the case of polyamides made up of Im, Py, and Hp, serious errors

in the torsional terms, multiplicity, phase shift, and force constant magnitude have been found. The simulation results could lead to artificial behavior that could impact the interpretation of structure, dynamics, and properties of polyamides and their complexes.

Zhang and coworkers created polyamide parameters in the AMBER force field[21] to determine the structure of a DNA/cyclic polyamide complex by NMR.[105] Energy minimization and molecular dynamics with the AMBER force field were used to aid in structure refinement; their results for the structure indicate distortion of the DNA in the polyamide binding site and altered stacking of the ligand rings relative to noncyclic polyamides.[105] A number of specific bonds, angles, and torsional terms were calibrated against B3LYP/6-31G(d) data. However, a majority of the terms were derived by comparison with other functionalities, and no comparison was made back to experimental structures. It is well known that density functional methods, such as B3LYP, do not account for the effects of dispersion.[123-131] Dispersion effects are important in the development of polyamide force fields because polyamides are cresent-shaped, which complements the shape of the narrow minor groove and improves binding by promoting van der Waals interactions.[132] Also, in the case of the 2:1 binders, the rings of the polyamides must stack on top of one another in the minor groove, which is stabilized mainly by nonbond interactions.[132]

Pang developed force field parameters for AMBER[21] to study the formation of DNA/polyamide complexes.[120] The results suggest that three and four ring polyamides have the same tendency for forming the same dimer conformations in water as they do in their ternary complexes with DNA and that the “pre-dimerization”

of polyamides in water assists in generating a tighter and more selective binding to DNA.[120] The majority of bonds, angle, and dihedral terms were calibrated against HF/6-31G(d), and no comparison was made back to any experimental structures. In a separate study, Topham and Smith made parameters for the CHARMM force field for distamycin and netropsin.[113] Netropsin and distamycin were divided into smaller model compounds, and potential energy surfaces for specific torsions were generated using HF/6-31G(d). The energy-minimized model of netropsin had an RMSD of 0.28 Å with the crystal structure. It is known that Hartree-Fock does not account for the effects of electron correlation.[133]

Wellenzohn, et al. performed simulation studies with the AMBER force field[21] on DNA/netropsin[114, 118] and other DNA/polyamide complexes.[21, 115-117, 119] The netropsin molecule and other polyamides were not explicitly parameterized for the force field. Only the partial charges on the atoms of the molecules were explicitly derived; the other parameters for these molecules were taken from analogous existing functionality and molecules in the force field. Generally, the extent of transferability of force field parameters is considered to be minimal, and new parameters must be generated for each new class of molecules to be used with the force field.[18] No comparison back to experiment was reported.

Given the problem of interest in polyamide/DNA simulation and the lack of accurate force field parameters to perform molecular dynamics simulations, we have developed parameters for the polyamides for the CHARMM force field.[20] Parameterization is accomplished by first dividing the polyamides into six model compounds and carrying out energy minimization using MP2/6-31G(d), which

includes the effects of electron correlation and dispersive interactions. The bond, angle, torsional, and improper terms were modified to match the quantum mechanical results. The parameters were tested by performing molecular dynamics crystal simulations of the reported x-ray crystal structures of a netropsin crystal, a 2:1 polyamide/DNA complex, a netropsin/DNA complex, a distamycin/DNA complex, and DNA without bound polyamides. The new parameters can be used for improved simulations of polyamides and their nucleic acid complexes.

1.5 Retinoids and π-Conjugated Systems

Retinoids make up a large group of naturally occurring and synthetic compounds related to retinol, otherwise known as vitamin A.[134 135 136] Physiologically occuring retinoids consist of a β-ionene ring, a chain of conjugated double bonds (that includes the ring), and a polar functional group at the end of the chain. The figure below illustrates the general retinoid skeleton with two naturally occurring retinoids, retinol (**1**) and retinal (**2**), and two synthetic retinoids, axerophthene (**3**), and fenretinide (**4**).

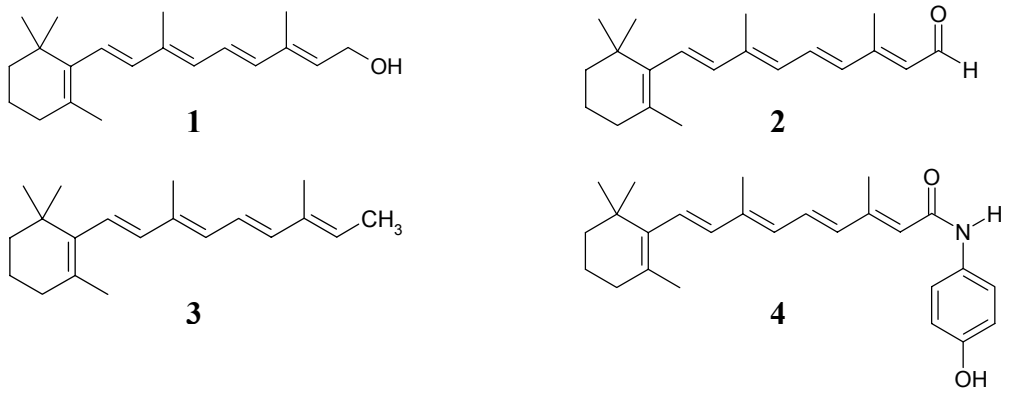


Figure 1.3 Structures of four retinoids. **1**, retinol; **2**, retinal; **3**, axerophthene; **4**, fenretinide.

Retinoids are essential in a number of vertebrate physiological processes; they play vital roles in cell growth,[137] reproduction (spermatogenesis, conception, and placental formation),[138] embryogenesis,[138] vision,[136, 139] learning and memory,[140] and resistance to and recovery from infection.[136] Retinol (also known as vitamin A) and its active metabolites (the most active is retinoic acid) also control the differentiation, proliferation, and apoptosis of many different types of cells from conception to death,[138, 141] including epithelial cells in the digestive tract, respiratory system, skin, bone, and the central nervous system.[136, 142, 143] Retinol cannot be synthesized *de novo* by the human body, so it is an essential part of the diet.[134] Dietary retinol is taken in as retinyl esters (obtained from animal products primarily as retinyl palmitate) or as β-carotene (obtained from vegetables).[144] The retinyl esters and β-carotene are then metabolized into the various retinoids found in the body, including retinol, retinoic acid, and retinal. Retinyl palmitate is thought to be hydrolyzed to form retinol and a carboxylic acid in the intestine before being transported to the liver. In the liver, which is the main storage site of vitamin A, retinol is esterified again to maintain the storage levels. β-carotene is thought to be cleaved to form two retinal molecules in the intestine. The retinal can then be reduced by retinaldehyde reductase to yield retinoic acid or it can be esterified to retinyl esters for storage.[33, 138]

The stored retinyl esters in the liver are hydrolyzed to all-*trans* retinol, which is the major circulating retinoid, for release into blood circulation.[144] The carrier protein for all-*trans* retinol in the blood to various target tissues is binding protein (RBP), which controls the plasma levels of retinol. Retinoids have low solubility in their uncomplexed form in aqueous medium so they must be protected, solubilized,

and transported in the blood and in the cell by specific retinoid binding proteins.[135, 145-147] In circulation, holo RBP is bound to transthyretin (TTR), which consists of four subunits. Formation of the complex is thought to prevent filtration of the relatively small RBP through the kidneys and is hypothesized to be the complex that interacts with cell surface receptors that mediate the uptake of retinol.[148, 149]

Once inside the cells of target tissues, retinol can be metabolized by various enzymes to retinal and retinoic acid or converted back to the ester form for storage; all forms (alcohol, acid, aldehyde and ester) can be present in the cell and are interconvertible.[150] Cytoplasmic retinol-binding proteins (CRBP I and CRBP II) and cytoplasmic retinoic acid-binding proteins (CRABP I and CRABP II) bind retinol and retinoic acid with high affinity, respectively, inside the cell.[151] Other than the obvious binding of retinoids in the cell, the functions of these cytoplasmic retinoid-binding proteins remain to be defined; however, they are thought to modulate intracellular retinoid metabolism by influencing the amount of ligand available to nuclear receptors.[135, 151, 152] The regulation of retinoid metabolism in the cell is crucial because a shortage or excess or retinoic acid can affect retinoid-signaling pathways and is known to result in developmental defects.[33, 150, 153] However, it is known that retinol bound to CRBP I is the substrate of enzymes that synthesize retinyl esters or oxidize retinol to retinoic acid. Retinoic acid bound to CRABP I or II is then delivered to the nuclear receptors, which direct and regulate the expression of specific retinoic-acid inducible genes.[143] CRBPs and CRABPs show a specific tissue distribution in adult and developing animals, which suggests that these proteins each serve different functions.[151] Visual tissues contain other specific binding proteins for

11-*cis* retinal and 11-*cis* retinol (known to function only in vision), cellular retinaldehyde-binding protein (CRALBP) and interphotoreceptor retinol binding protein (IRBP).[138, 154]

Two classes of nuclear receptors mediate the biological activities of retinoids, retinoic acid receptors (RARs) and retinoid X receptors (RXRs), both of which have three subtypes (α, β, and γ).[34, 134, 140] RARs bind and are activated by all-*trans* retinoic acid and 9-*cis* retinoic acid with similar affinity but bind 13-*cis* retinoic acid only weakly; RXRs bind and are activated by only 9-*cis* retinoic acid.[34] RARs and RXRs are transcription factors that bind to specific DNA sequence-retinoic acid response elements as RAR-RXR heterodimers and RXR homodimers, acting as ligand-dependent transcriptional regulators for retinoic acid responsive genes.[34, 155, 156]

Because retinoids play essential roles in cell differentiation, growth, proliferation, and apoptosis in all stages of life in vertebrates, their pharmaceutical and chemopreventative properties have been exploited. The many effects that retinoids have on cells have made this class of compounds an attractive candidate for use in cancer therapy and other diseases.[34, 155, 157, 158] Both naturally occurring and synthetic retinoids are used to treat various skin disorders (acne, psoriasis, and ageing),[155] they have been in used to augment treatments of other diseases, such as type II diabetes,[159] and they show promise for the treatment of several cancers,[142, 158, 160, 161] HIV infection,[162] and late-onset Alzheimer's disease.[140] These agents have targeted the RARs and RXRs; however, the scope of retinoid therapy is limited due to high toxicity (only a few retinoids are used routinely for treatment of skin disorders[163-165]),[34] and the design of less toxic retinoids is necessary for further

clinical use. Because vitamin A and its derivatives play a multitude of roles in all stages of life through activation of widely expressed RARs and RXRs, administration of additional retinoids (natural or synthetic) causes toxic side effects that are characteristic of excess levels of vitamin A.[33] Current efforts in retinoid development involve increasing the selectivity of synthetic retinoids to decrease the toxic side effects.[155, 166]

Because of their biological importance and potential theraputic uses, it is of interest to understand the interactions and dynamics of retinoids with the various proteins that bind them at the molecular level. Molecular dynamics simulations are a powerful tool for elucidating the atomistic details of biomolecular systems; however, the empirical force field parameters used must be optimized so that they accurately treat the class of compounds that one is attempting to model. Previous attempts have been made to treat conjugated systems properly, which are fundamental to retinoids. The first attempts were made independently by Warshel and Karplus[29] and Allinger and Sprague,[30] where both involved a combined quantum mechanical/molecular mechanical approach.

The method of Warshel and Karplus was based on the formal separation of σ and π electrons, where the σ electrons were treated using an empirical force field (the consistent force field, or CFF[167]), and π electrons with Pariser-Parr-Pople semiempirical theory[168, 169] corrected for orbital overlap. The total energy of a molecule was then expressed as the sum of σ and π electron contributions. The goal of the method was to treat the equilibrium conformations and vibrational spectra of the ground and excited states of conjugated molecules. Experimental data from

ethylene, butadiene, benzene, and propylene were used to fit the σ and π parameters used in their method.[29] The method was tested against the ground and excited states of several conjugated molecules, including 1,3-butadiene, 1,3-cyclohexadiene, and 1,3,5-hexatriene. The results were in good agreement with experimental geometries and vibrational frequencies; however, it was never generally used and thus never made widely available because at that time it required excessive computer resources, and it was not easily extended for large systems of interest.

Allinger also developed a combined quantum and molecular mechanical approach that separated the σ and π electrons.[30] Allinger treated the σ electrons with the MMP1 force field,[170] and the π electrons were treated with a variable-electronegativity self-consistent field (VESCF) calculation. The method was tested against the experimental geometries (x-ray crystallographic and electron diffraction structures) and heats of formation for a representative group of 65 compounds. The computed results yielded satisfactory agreement with experiment over the range of compounds with the largest difference between experiment and computation being 2.0 kcal/mol for the heats of formation. The program was never widely used by the scientific community due to the same issues limiting the Warshel and Karplus method.

A more recent method by Treboux for calculating geometries and vibrational frequencies of ground state and excited conjugated hydrocarbons also utilized the combined quantum mechanics and molecular mechanics methods.[31] This method again partitioned the total energy into σ and π parts. In order to reduce the computational time involved with the iterative SCF algorithm utilized in the last two

methods described, this method makes use of a Heisenberg Hamiltonian for the π electrons.[31] The geometries and energy barriers calculated from various conjugated systems were in good agreement with experimental data; however, the authors made note of the fact that it could be practically applied to only 20 carbon atoms at the time the method was devised, which was in 1995.[31]

The most recent approach for treating a general class of large biomolecules, which include conjugated systems, was based completely on molecular mechanics. Both CHARMM and AMBER force fields were designed to be general purpose force fields to describe a wide variety of organic molecules, specifically, moieties that are commonly found in pharmaceuticals.[20, 26, 32] Neither CHARMM nor AMBER were parameterized explicitly for conjugated systems.

The limitations of force fields to treat conjugated systems, such as retinoids, have been recognized. Numerous molecular dynamics studies involving retinal-binding proteins have been performed, primarily with the goal of understanding different aspects of the bacteriorhodopsin or rhodopsin photocycle.[171-174 175-185] Some studies involved the parameterization of the retinal ligand using quantum mechanical calculations; however, the quantum mechanical calculations used to parameterize the retinal were performed using Hartree-Fock[176, 177, 186] or B3LYP,[180, 183, 187, 188] which do not account for the effects of electron correlation and dispersion interactions, respectively. Many other studies used parameters for the retinal ligand that were simply adapted from other atoms or molecules in the force field being used in the study.[171-175, 178, 179, 181]

In order to understand the interactions and dynamics of retinoid binding proteins and design new retinoid ligands with therapeutic potential using molecular dynamics simulations, this class of compounds has been explicitly parameterized for this dissertation. New force field parameters have been developed for the CHARMM force field for retinoids, which properly account for the alternating single and double bonds present in the conjugated systems of these compounds. The new force field has been evaluated against x-ray crystallographic data and quantum mechanical structures and energies. Through testing the parameters with crystal simulations, some ligand motions were also discovered that could be important in ligand binding and dynamics or interaction of the protein/ligand complex with other molecules. The observed motions aid in understanding more about how the ligands move internally within the protein, which can be used in design of new ligands. The parameters can now be used in simulations of retinoid binding proteins to better understand their structure, dynamics, and how they interact with other proteins and for the design of new retinoids that show promise as therapeutic and anticancer agents.

1.6 Convergence Analysis of Biomolecular Simulations

Understanding DNA and protein dynamics is critical not only for expanding our knowledge of DNA and protein structure in itself, but for understanding DNA and protein interactions and dynamics with other proteins, nucleic acids, and ligands. Therefore, it is crucial to build a knowledge base of DNA and protein dynamics, which is not currently available, especially for DNA. The energy landscapes of biomolecules are rough and rugged, containing many accessible local minima (corresponding to similar but slightly different conformations) separated by energy

barriers, and transitions between these minima take time.[15, 35, 36, 189, 190] On the current time scale of MD simulations (hundreds of nanoseconds), the simulation can get stuck in the many minima on the potential energy surface, and thus not allow for complete sampling of all possible conformations of even a small (8-10 base pairs) DNA duplex or small protein (46 residues).[191] Slow motions, such as large conformational changes, ion channel gating motions, or other functionally relevant motions, often occur on timescales that not easily accessible and thus are not well-characterized with simulations (microseconds to milliseconds).[37, 191] Furthermore, simulated biomolecular systems commonly contain a complex mixture of nucleic acids or proteins, water, and counterions (such as sodium, chlorine, or magnesium), and in some systems, lipid bilayers, which cause the convergence to occur more slowly, resulting from less efficient sampling.[192-196]

Sampling of minima is essential for the conformational flexibility of biomolecules, and therefore, to their function.[35, 197, 198] Thus, if the biomolecular system of interest is not properly sampled, the properties measured from the simulation will not be representative of its true behavior, and predictions made could be incorrect. The results and predictions from simulation studies depend on proper sampling of the molecule's conformational space.

Simulations of biomolecular systems must address the issue of the ergodic hypothesis, which a problem for any system with multiple minima on its potential energy surface.[35] The system may not cross all of the energy barriers sufficiently in the time that the simulation is run to sample properly all states accessible to it. The ergodic hypothesis states that average of a property over the entire simulation

trajectory is equal to an average over all states accessible to the system.[15, 35] The hypothesis is difficult to prove for any system; however, it is easier and possible to evaluate a necessary criterion for the ergodic hypothesis to be valid: At equilibrium, independent trajectories over an ergodic system must be self-averaging.[15] This means that a property measured from two or more independent trajectories should be equal if the system dynamics are ergodic. In other words, if two independent simulations do not converge to equal values, the system is nonergodic on the timescale that the simulations were run.[15] Obviously, different properties of the system will exhibit different convergence behavior (some will converge more quickly than others), so multiple properties should be examined in order to address ergodicity and convergence.[35, 199]

Because of the complexity of biomolecular systems and the need for adequate sampling, one of the primary issues with MD simulations is convergence: How long does the simulation need to be run in order to extract reliable information? Acquiring reliable and accurate information on dynamics from MD simulations requires us to know how much sampling is required for the convergence of all of the properties of interest. Attaining convergence is obviously going depend on the system under study and the properties of interest, and it is a difficult question to answer. For effective sampling and therefore ergodicity, the simulations must correctly sample the regions of phase space that are accessible to the system under the given conditions (e.g., temperature and pressure). If only a single long simulation is run, the system can get stuck in one of the many minima and thus not sample effectively. Multiple shorter trajectories have been found to be more effective at sampling phase space than a

single long one,[35-37, 200-204] and this is the approach that is utilized in this dissertation for the convergence studies.

Rigid convergence of a system can never be proved, because there is no guarantee that the past dynamics of the system will predict the future dynamics, and in principle, new events or conformations of a system may be discovered with more extensive sampling.[32, 37, 191, 192] However, efforts can be made to ensure that the various properties calculated from simulations have reached a level of satisfactory convergence, and different approaches have been used to do this. Some methods are relatively simple and commonly used, such as monitoring the RMSD of the simulation structures over time with respect to the starting structure and calculating correlation times among different properties.[32, 196, 205-208] However, the RMSD is not always a good indicator of convergence;[191, 209] we have found in this study that DNA simulations started from different atomic coordinates result in RMSDs that converge at different rates or fluctuate at different points when the simulation has run for a considerable amount of time (>100 ns). In other words, the RMSD failed to settle down to a constant value; the system appears to be moving between various substates (see Chapter 5). This is not surprising for a flexible biomolecule like DNA, but it indicates that this method cannot determine when the simulation has converged. Also, if correlation times (time over which a particular property value takes to become independent from the previous values) longer than the overall trajectory length exist for the system, then the proper correlation time will not be observed; the correlation times will also depend on what part of the protein (a loop vs. a helix) or DNA (entire fragment vs. selected base pairs) is being examined.[37, 191, 209] Another

commonly used method is block averaging.[32, 195, 210] With block averaging, the total trajectory is broken up into independent, equally sized blocks, and the averages are then calculated and compared over the blocks. However, in order obtain independent blocks, the relaxation time of the property of interest must determined to figure out what length of block can be considered independent.[32] This, as with correlation times, will present a problem if the trajectory has not been run sufficiently long enough for the property to relax.

Other more advanced methods, such as principal components analysis (PCA),[198, 206, 211-216] cluster population analysis, [37, 191, 199, 217-219] or energy-based methods[189, 220, 221] have also been used to diagnose sampling and convergence. PCA, also called essential dynamics,[211, 222] has been to applied biomolecular simulations to visualize large-scale motions.[198, 206, 211-216, 223, 224 225, 226] It is assumed that the dynamics of the large-scale motions analyzed by PCA are converged on the timescale of the simulation.[37] However, this may not be the case; several studies have found that the motion visualized from PCA were not consistent among individual simulations.[37, 227-230] Although these studies showed that portions of a single trajectory or that individual simulations yield different results when PCA is applied, they do show that the simulation times examined are not long enough to sample properly the structure and fluctuations of the systems studied.

Cluster population analysis[37, 191, 199, 217-219] has been used to examine the relative populations of substates on the potential energy surface of the simulated system. This method aims to classify the structures generated from trajectories into clusters based on the RMSD between the trajectory structures and a reference

structure. In practice, sampling is limited so there are statistical variations among the cluster probability distributions; the magnitude of the variations is then considered a measure of the degree of convergence among the simulations. However, interpreting the cluster population variance can be difficult,[37] and as a result, the whole procedure requires time and processing of data. Despite the time investment, this method has been used, and results suggest the same as those of PCA: the simulation times used were not long enough to produce ergodic behavior of the systems studied.[37, 191]

The energy-base method (also called the "ergodic measure"[189] and not limited to energy-based quantities) developed by Straub, Thirumalai, and Mountain[189, 220, 221] examines the fluctuations of some quantity averaged over independent simulations, and the timescale required for that quantity to be ergodic is determined. However, a thorough sampling of the conformational space of the system's structures is required for convergence, and many structures are close in energy but not structurally similar, and energy-based methods will not be sensitive to this.[191, 209]

In order to understand more about sampling and convergence in simulations of DNA and protein systems, we have approached the problem using several long-time simulations (4 simulations of 150 ns each for DNA; 10 simulations of 45 ns each for the protein crambin) and many short-time simulations (20 simulations of 2 ns each for both the DNA and crambin) and a simple statistical test. The statistical test used is called the potential scale reduction test,[231] and it is a popular convergence diagnostic used by statisticians in order to monitor convergence in multiple Markov chain Monto Carlo simulations.[232-235] Approximate convergence of a particular parameter measured from the simulations is assumed when the variance between the

different simulations for a particular parameter is no larger than the variance seen in that particular parameter from an individual simulation.[235] This approach enables us to know which methodology is better (several long simulations or many short simulations) for achieving convergence of several geometric parameters. Eighteen helical parameters[236] were measured from the DNA simulations and 16 parameters, including radius of gyration and hydrogen bonding distances, were measured from the crambin simulations. This approach will also give insight as to which parameters may take longer to stabilize over multiple simulations and how long they take to stabilize (effectively telling us how long we need to run the simulations in order to achieve convergence of a particular parameter thus obtain reliable results). To our knowledge, the potential scale reduction test has never been applied to biomolecular systems in order to assess convergence. Like other approaches, this test cannot diagnose absolute convergence, but the approach is very simple. All that is required are the values of the properties of interest over time; the statistical test (see Chapter 5 for details and formula) can easily be implemented in any statistical program.

In addition to gaining knowledge of the convergence of different properties, we have also discovered interesting and important information about which base pairs of the simulated DNA fragment should be included in the analysis of DNA simulations. For example, we have found that the effects of fraying (a well-known phenomenon where the end base pairs open and close rapidly)[237-240] penetrate beyond the end base pairs of the DNA. An understanding of sampling and convergence in biomolecular simulations will enable us to better model these systems and thus use

computational techniques to examine structural and dynamic features that cannot be observed by experiment.

References

(1) McCammon, J. A.; Gelin, B. R.; Karplus, M., Dynamics of Folded Proteins, *Nature*, 1977, 267, 585-590.

(2) Levitt, M. In *Computer Simulation of DNA Double-Helix Dynamics*, Cold Spring Harbor Symposia on Quantitative Biology, 1983; 'Ed.'^'Eds.' Cold Spring Harbor Laboratory: 1983; p^pp.

(3) Schlick, T., *Molecular Modeling and Simulation: An Interdisciplinary Guide*. Springer-Verlag: New York, 2000; 21.

(4) Daura, X.; Jaun, B.; Seebach, D.; van Gunsteren, W. F.; Mark, A. E., Reversible Peptide Folding in Solution by MD simulation, *J. Mol. Biol.*, 1998, 280, 925-932.

(5) Zagrovic, B.; Sorin, E. J.; Pande, V. S., Beta-hairpin Folding Simulations in Atomistic Detail Using an Implicit Solvent Model, *J. Mol. Biol.*, 2001, 313, 151-169.

(6) Young, M. A.; Beveridge, D. L., Molecular Dynamics Simulations of an Oligonucleotide Duplex with Adenine Tracts Phased by a Full Helix Turn, *J. Mol. Biol.*, 1998, 281, 675-687.

(7) Duan, Y.; Kollman, P. A., Pathways to a Protein Folding Intermediate Observed in a 1-microsecond simulation in Aqueous Solution, *Science*, 1998, 23, 740-744.

(8) Israilev, S.; Crofts, A. R.; Berry, E. A.; Schulten, K., Steered Molecular Dynamics Simulation of the Rieske Subunit Motion in the Cytochrome bc1 Complex, *Biophys. J.*, 1999, 77, 1753-1768.

(9) Jensen, M. O.; Tajkhorshid, E.; Schulten, K., The Mechanism of Glycerol Conduction in Aquaglyceroporins, *Structure*, 2001, 9, 1083-1093.

(10) Shirts, M.; Pande, V. S., Screen Savers of the World Unite!, *Science*, 2000, 290, 1903-1904.

(11) Almond, A.; Blundell, C. D.; Higman, V. A.; MacKerell, A. D.; Day, A. J., Using Molecular Dynamics Simulations to Provide New Insights into Protein Structure on the Nanosecond Timescale: Comparison with Experimental Data and

Biological Inferences for the Hyaluronan-Binding Link Module of TSG-6, *Journal of Chemical Theory and Computation*, 2007, 3, 1-16.

(12) Karplus, M.; McCammon, J. A., Molecular Dynamics Simulations of Biomolecules, *Nat. Struc. Biol.*, 2002, 9, 646-652.

(13) Daggett, V., Long Timescale Simulations, *Curr. Opin. Struc. Biol.*, 2000, 10, 160-164.

(14) Hansson, T.; Oostenbrink, C.; van Gunsteren, W. F., Molecular Dynamics Simulations, *Curr. Opin. Struc. Biol.*, 2002, 12, 190-196.

(15) Berne, B. J.; Straub, J. E., Novel Methods of Sampling Phase Space in the Simulation of Biological Systems, *Curr. Opin. Struc. Biol.*, 1997, 7, 181-189.

(16) Lei, H.; Duan, Y., Improved Sampling Methods for Molecular Simulation, *Curr. Opin. Struc. Biol.*, 2007, 17, 187-191.

(17) Leach, A. R., Molecular Dynamics Simulation Methods. In *Molecular Modelling: Principles and Applications*, Pearson Education Limited: Essex, England, 2001; 353-409.

(18) MacKerell, A. D., Jr., Atomistic Models and Force Fields. In *Computational Biochemistry and Biophysics*, Marcel Dekker, Inc.: New York, 2001; 7-38.

(19) Jorgensen, W. L., Theoretical Studies of Medium Effects on Conformational Equilibria, *J. Phys. Chem.*, 1983, 87, 5304-5314.

(20) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B. R., W.E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M., All-atom empirical potential for molecular modeling and dynamics studies of proteins, *Journal of Physical Chemistry*, 1998, 102, 3586.

(21) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeye, 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.

(22) Warshel, A.; Lifson, S., Consistent Force Field Calculations. II. Crystal Structures, Sublimation Energies, Molecular and Lattice Vibrations, Molecular Conformations, and Enthalpies of Alkanes, *J. Chem. Phys.*, 1970, 53, 582-594.

(23) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A., A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The RESP Model, *J. Phys. Chem.*, 1993, 97, 10269-10280.

(24) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollman, P. A., Application of RESP Charges to Calculate Conformational Energies, Hydrogen Bond Energies, and Free Energies of Solvation, *J. Am. Chem. Soc.*, 1993, 115, 9620-9631.

(25) Thacher, T. S.; Hagler, A. T.; Rabitz, H., Application of Sensitivity Analysis of the Establishment of Intermolecular Potential Functions, *J. Am. Chem. Soc.*, 1991, 113, 2020-2033.

(26) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Development and Testing of a General AMBER force field, *J. Comp. Chem.*, 2004, 25, 1157-1174.

(27) Maple, J. R.; Dinur, U.; Hagler, A. T., Derivation of Force Fields for Molecular Mechanics and Dynamics from ab initio Energy Surfaces, *Proc. Natl. Acad. Sci.*, 1988, 85, 5350-5354.

(28) Dervan, P. B.; Poulin-Kerstien, A. T.; Fechter, E. J.; Edelson, B. S., Regulation of Gene Expression by Synthetic DNA-Binding Ligands, *Topics in Current Chemistry*, 2005, 253, 1-31.

(29) Warshel, A.; Karplus, M., Calculation of Ground and Excited State Potential Surfaces of Conjugated Molecules. I. Formulation and Parameterization, *Journal of the American Chemical Society*, 1972, 94, 5612-5625.

(30) Allinger, N. L.; Sprague, J. T., Calculation of the Structures of Hydrocarbons Containing Delocalized Systems by the Molecular Mechanics Method, *J. Am. Chem. Soc.*, 1973, 95, 3893.

(31) Treboux, G.; Maynau, D.; Malrieu, J. P., Combining Molecular Mechanics with Quantum Treatments for Large Conjugated Hydrocarbons. 2. A Geometry-Dependent Heisenberg Hamiltonian, *J. Phys. Chem.*, 1995, 99, 6417-6423.

(32) MacKerell, A. D., Jr., Nucleic Acid Simulations. In *Computational Biochemistry and Biophysics*, Marcel Dekker, Inc.: New York, 2001; 441-463.

(33) Moise, A. R.; Noy, N.; Palczewski, K.; Blaner, W. S., Delivery of Retinoid-Based Therapies to Target Tissues, *Biochemistry*, 2007, 46, 4449-4458.

(34) Kagechika, H.; Shudo, K., Synthetic Retinoids: Recent Developments Concerning Structure and Clinical Utility, *J. Med. Chem.*, 2005.

(35) Caves, L. S. D.; Evanseck, J. D.; Karplus, M., Locally Accessible Conformations of Proteins: Multiple Molecular Dynamics Simulations of Crambin, *Protein Science*, 1998, 7, 649-666.

(36) Schulze, B. G.; Evanseck, J. D., Cooperative Role of Arg45 and His64 in the Spectroscopic A3 State of Carbonmonoxy Myoglobin: Molecular Dynamics Simulations, Multivariate Analysis, and Quantum Mechanical Computations, *J. Am. Chem. Soc.*, 1999, 121, 6444-6454.

(37) Grossfield, A.; Feller, S. E.; Pitman, M. C., Convergence of Molecular Dynamics Simulations of Membrane Proteins, *Proteins: Structure, Functions, and BioInformatics*, 2007, 67, 31-40.

(38) Best, T. P.; Edelson, B. S.; Nickols, N. G.; Dervan, P. B., Nuclear Localization of Pyrrole-Imidazole Polyamide-fluorescein Conjugates in Cell Culture, *Proc. Natl. Acad. Sci. USA*, 2003, 100, 12063-12068.

(39) Chiang, S.-Y.; Burli, R. W.; Benz, C. C.; Gawron, L.; Scott, G. K.; Dervan, P. B.; Beerman, T. A., Targeting the Ets Binding Site of the HER2/neu Promoter with Pyrrole-Imidazole Polyamides, *J.Biol. Chem.*, 2000, 275, 24246-24254.

(40) Dervan, P. B.; Burli, R. W., Sequence-Specific DNA Recognition by Polyamides, *Curr. Opin. Chem. Biol.*, 1999, 3, 688-693.

(41) Dervan, P. B., Molecular Recognition of DNA by Small Molecules, *Bioorg. & Med. Chem.*, 2001, 9, 2215-2235.

(42) Dickinson, L. A.; Gulizia, R. J.; Trauger, J. W.; Baird, E. E.; Mosier, D. E.; Gottesfeld, J. M.; Dervan, P. B., Inhibition of RNA Polymerase II Transcription in Human Cells by Synthetic DNA-Binding Ligands, *Proc. Natl. Acad. Sci. USA*, 1998, 95, 12890-12895.

(43) Dickinson, L. A.; Trauger, J. W.; Baird, E. E.; Ghazal, P.; Dervan, P. B.; Gottesfeld, J. M., Anti-Repression of RNA Polymerase II Transcription by Pyrrole-Imidazole Polyamides, *Biochemistry*, 1999, 38, 10801-10807.

(44) Belitsky, J. M.; Leslie, S. J.; Arora, P. S.; Beerman, T. A.; Dervan, P. B., Cellular Uptake of N-Methylpyrrole/N-Methylimidazole Polyamide Dye Conjugates, *Bioorg. & Med. Chem.*, 2002, 10, 3313-3318.

(45) Ellervik, U.; Wang, C. C. C.; Dervan, P. B., Hydroxybenzamide/Pyrrole Pair Distinguishes TA from AT Base Pairs in the Minor Groove of DNA, *J. Am. Chem. Soc.*, 2000, 122, 9354-9360.

(46) Gearhart, M. D.; Dickinson, L. A.; Ehley, J.; Melander, C.; Dervan, P. B.; Wright, P. E.; Gottesfeld, J. M., Inhibition of DNA Binding by Human Estrogen-Related Receptor 2 and Estrogen Receptor a with Minor Groove Binding Polyamides, *Biochemistry*, 2005, 44, 4196-4203.

(47) Kielkopf, C. L.; Baird, E. E.; Dervan, P. B.; Rees, D. C., Structural Basis for GC Recognition in the DNA Minor Groove, *Nat. Struct. Biol.*, 1998, 5, 104.

(48) Kielkopf, C. L.; White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B.; Rees, D. C., A Structural Basis for Recognition of AT and TA Base Pairs in the Minor Groove of B-DNA, *Science*, 1998, 282, 111-115.

(49) Rucker, V. C.; Melander, C.; Dervan, P. B., Influence of β-alanine on Hairpin Polyamide Orientation in the DNA Minor Groove, *Helv. Chim. Acta*, 2003, 86, 1839-1851.

(50) Swalley, S. E.; Baird, E. E.; Dervan, P. B., Discrimination of 5'-GGGG-3', 5'-GCGC-3', and 5'-GGCC-3' Sequences in the Minor Groove of DNA by Eight-Ring Hairpin Polyamides, *J. Am. Chem. Soc.*, 1997, 119, 6953-6961.

(51) Swalley, S. E.; Baird, E. E.; Dervan, P. B., Effects of γ-Turn and β-Tail Amino Acids on Sequence-Specific Recognition of DNA by Hairpin Polyamides, *J. Am. Chem. Soc.*, 1999, 121, 1113-1120.

(52) Trauger, J. W.; Baird, E. E.; Dervan, P. B., Recognition of 16 Base Pairs in the Minor Groove of DNA by a Pyrrole-Imidazole Polyamide Dimer, *J. Am. Chem. Soc.*, 1998, 120, 3534-3535.

(53) White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B., Recognition of the four Watson-Crick Base Pairs in the DNA Minor Groove by Synthetic Ligands, *Nature*, 1998, 391, 468-471.

(54) Yang, F.; Belitsky, J. M.; Villanueva, R. A.; Dervan, P. B.; Roth, M. J., Inhibition of Moloney Murine Leukemia Virus Integration Using Polyamides Targeting the Long-Terminal Repeat Sequences, *Biochemistry*, 2003, 42, 6249-6258.

(55) Doss, R. M.; Marques, M. A.; Foister, S.; Chenoweth, D. M.; Dervan, P. B., Programmable Oligomers for Minor Groove DNA Recognition, *J. Am. Chem. Soc.*, 2006, 128, 9074-9079.

(56) Hsu, C.-H.; Phillips, J. W.; Trauger, J. W.; Farkas, M. E.; Belitsky, J. M.; Heckel, A.; Olenyuk, B. Z.; Puckett, J. W.; Wang, C. C. C.; Dervan, P. B., Completion of a programmable DNA-binding small molecule library, *Tetrahedron*, 2007, 63, 6146-6151.

(57) Nickols, N. G.; Jacobs, C. S.; Farkas, M. E.; Dervan, P. B., Improved nuclear localization of DNA-binding polyamides, *Nucleic Acids Research*, 2007, 35, 363-370.

(58) Filipowsky, M. E.; Kopka, M. L.; Brazil-Zilson, M.; Lown, J. W.; Dickerson, R. E., Linked Lexitropsins and the in Vitro Inhibition of HIV-1 Reverse Transcriptase RNA-Directed DNA Polymerization: A Novel Induced-Fit of 3,5 m-Pyridyl Bisdistamycin to Enzyme-Associated Template-Primer, *Biochemistry*, 1996, 35, 15397-15410.

(59) Kopka, M. L.; Yoon, C.; Goodsell, D. S.; Pjura, P.; Dickerson, R. E., The Molecular Origin of DNA-Drug Specificity in Netropsin and Distamycin, *Proc. Natl. Acad. Sci. USA*, 1985, 82, 1376-1380.

(60) Kopka, M. L.; Goodsell, D. S.; Han, G. W.; Chiu, T. K.; Lown, J. W.; Dickerson, R. E., Defining GC-Specificity in the Minor Groove: Side-by-side binding of the di-imidazole lexitropsin to CATGGCCATG, *Structure*, 1997, 5, 1033-1046.

(61) O'Hare, C. C.; Mack, D.; Tandon, M.; Sharma, S. K.; Lown, J. W.; Kopka, M. L.; Dickerson, R. E.; Hartley, J. A., DNA Sequence Recognition in the Minor Groove

by Crosslinked Polyamides: The Effect of N-terminal Head Group and Linker Length on Binding Affinity and Specificity, *Proc. Natl. Acad. Sci. USA*, 2002, 99, 72-77.

(62) Walker, W. L.; Landaw, E. M.; Dickerson, R. E.; Goodsell, D. S., Estimation of the DNA Sequence Discriminatory Ability of Hairpin-Linked Lexitropsins, *Proc. Natl. Acad. Sci. USA*, 1997, 94, 5634-5639.

(63) Walker, W. L.; Landaw, E. M.; Dickerson, R. E.; Goodsell, D. S., The Theoretical Limits of DNA Sequence Discrimination by Linked Polyamides, *Proc. Natl. Acad. Sci. USA*, 1998, 95, 4315-4320.

(64) Dwyer, T. J.; Geierstanger, B. H.; Bathini, Y.; Lown, J. W.; Wemmer, D. E., Design and Binding of a Distamycin A Analog to d(CGCAAGTTGGC).d(GCCAACTTCGC): Synthesis, NMR Studies, and Implications for the Design of Sequence-Specific Minor Groove Binding Oligopeptides, *J. Am. Chem. Soc.*, 1992, 114, 5911.

(65) Geierstanger, B. H.; Dwyer, T. J.; Bathini, Y.; Lown, J. W.; Wemmer, D. E., NMR Characterization of a Heterocomplex Formed by Distamycin and Its Analog 2-ImD with d(CGCAAGTTGGC):d(GCCAACTTGCG): Preference for the 1:1:1 2-ImD:Dst:DNA Complex over the 2:1 2-ImD:DNA and the 2:1 Dst:DNA Complexes, *J. Am. Chem. Soc.*, 1993, 115, 4474-4482.

(66) Lee, M.; Krowicki, K.; Hartley, J. A.; Pon, R. T.; Lown, J. W., Molecular Recognition between Oligopeptides and Nucleic Acids: Influence of van der Waals Contacts in Determining the 3'-Terminus of DNA Sequences Read by Monocationic Lexitropsins, *J. Am. Chem. Soc.*, 1988, 110, 3641-3649.

(67) Lown, J. W.; Krowicki, K.; Bhat, U. G.; Skorobogaty, A.; Ward, B.; Dabrowiak, J. C., Molecular Recognition between Oligopeptides and Nucleic Acids: Novel Imidazole-Containing Oligopeptides Related to Netropsin that Exhibit Altered DNA Sequence Specificity, *Biochemistry*, 1986, 25, 7408.

(68) Lown, J. W.; Krowicki, K.; Balzarini, J.; Newman, R. A.; De Clercq, E., Novel linked antiviral and antitumor agents related to netropsin and distamycin: Synthesis and biological evaluation, *J. Med. Chem.*, 1989, 32, 2368-2375.

(69) Sharma, S. K.; Billaud, J.-N.; Tandon, J.; Billet, O.; Choi, S.; Kopka, M. L.; Phillips, T. R.; Lown, J. W., Inhibition of feline immunodeficiency virus (FIV) replication by DNA binding polyamides, *Bioorg. & Med. Chem. Lett.*, 2002, 12, 2007-2010.

(70) Wang, W.; Lown, J. W., Anti-HIV-I Activity of Linked Lexitropsins, *J. Med. Chem.*, 1992, 35, 2890-2897.

(71) Dervan, P. B.; Doss, R. M.; Marques, M. A., Programmable DNA Binding Oligomers for Control of Transcription, *Current Medicinal Chemistry*, 2005, 5, 373-387.

(72) Hawkins, C. A.; Baird, E. E.; Dervan, P. B.; Wemmer, D. E., Analysis of Hairpin Polyamide Complexes Having DNA Binding Sites in Close Proximity, *J. Am. Chem. Soc.*, 2002, 124, 12689-12696.

(73) Kielkopf, C. L.; Bremer, R. E.; White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B.; Rees, D. C., Structural Effects of DNA Sequence on TA Recognition by Hydroxypyrrole/Pyrrole Pairs in the Minor Groove, *J. Mol. Biol.*, 2000, 295, 557-567.

(74) Dervan, P. B.; Edelson, B. S., Recognition of the DNA Minor Groove by Pyrrole-Imidazole Polyamides, *Curr. Opin. Struc. Biol.*, 2003, 13, 284-299.

(75) Edelson, B. S.; Best, T. P.; Olenyuk, B. Z.; Nickols, N. G.; Doss, R. M.; Foister, S.; Heckel, A.; Dervan, P. B., Influence of Structural Variation on Nuclear Localization of DNA-Binding Polyamide-Fluorophore Conjugates, *Nucleic Acids Research*, 2004, 32, 2802-2818.

(76) Marques, M. A.; Doss, R. M.; Foister, S.; Dervan, P. B., Expanding the Repertoire of Heterocycle Ring Pairs for Programmable Minor Groove DNA Recognition, *J. Am. Chem. Soc.*, 2004, 126, 10339-10349.

(77) Rucker, V. C.; Dunn, A. R.; Sharma, S. K.; Dervan, P. B.; Gray, H. B., Mechanism of Sequence-Specific Fluorescent Detection of DNA by N-Methyl-Imidazole, N-Methyl-pyrrole, and b-alanine linked Polyamides, *J. Phys. Chem. B*, 2004, 108, 7490-7494.

(78) Urbach, A. R.; Love, J. J.; Ross, S. A.; Dervan, P. B., Structure of a β-alanine-linked Polyamide Bound to a Full Helical Turn of Purine Tract DNA in the 1:1 Motif, *J. Mol. Biol.*, 2002, 320, 55-71.

(79) Hunter, C. A., Sequence-Dependent DNA Structure: The Role of Base Stacking Interactions, *J. Mol. Biol.*, 1993, 230, 1025-1054.

(80) Urbach, A. R.; Dervan, P. B., Toward Rules for 1:1 Polyamide:DNA Recognition, *Proc. Natl. Acad. Sci. USA*, 2001, 98, 4343-4348.

(81) 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 Research*, 2000, 28, 235-242.

(82) Woods, C. R.; Ishii, T.; Boger, D. L., Synthesis and DNA Binding Properties of Iminodiacetic Acid-Linked Polyamides: Characterization of Cooperative Extended 2:1 Side-by-Side Parallel Binding, *J. Am. Chem. Soc.*, 2002, 124, 10676-10682.

(83) de Clairac, R. P. L.; Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E., NMR Characterization of Hairpin Polyamide Complexes with the Minor Groove of DNA, *J. Am. Chem. Soc.*, 1997, 119, 7909-7916.

(84) White, S.; Baird, E. E.; Dervan, P. B., Orientation Preferences of Pyrrole-Imidazole Polyamides in the Minor Groove of DNA, *J. Am. Chem. Soc.*, 1997, 119, 8756-8765.

(85) Kelly, J. J.; Baird, E. E.; Dervan, P. B., Binding Site Size Limit of the 2:1 Pyrrole-Imidazole Polyamide-DNA Motif, *Proc. Natl. Acad. Sci. USA*, 1996, 93, 6981-6985.

(86) Turner, J. M.; Swalley, S. E.; Baird, E. E.; Dervan, P. B., Aliphatic/Aromatic Amino Acid Pairings for Polyamide Recognition in the Minor Groove of DNA, *J. Am. Chem. Soc.*, 1998, 120, 6219.

(87) Briehn, C. A.; Weyermann, P.; Dervan, P. B., Alternative Heterocycles for DNA Recognition: The Benzimidazole/Imidazole Pair, *Chem. Eur. J.*, 2003, 9, 2110-2122.

(88) Renneberg, D.; Dervan, P. B., Imidazopyridine/Pyrrole and Hydroxybenzimidazole/Pyrrole Pairs for DNA Minor Groove Recognition, *J. Am. Chem. Soc.*, 2003, 125, 5707-5716.

(89) Arora, P. S.; Ansari, A. Z.; Best, T. P.; Ptashne, M.; Dervan, P. B., Design of Artificial Transcriptional Activators with Rigid Poly-L-proline Linkers, *J. Am. Chem. Soc.*, 2002, 124, 13067-13071.

(90) Arndt, H.-D.; Hauschild, K. E.; Sullivan, D. P.; Lake, K.; Dervan, P. B.; Ansari, A. Z., Toward Artificial Development Regulators, *J. Am. Chem. Soc.*, 2003, 125, 13322-13323.

(91) Mapp, A. K.; Ansari, A. Z.; Ptashne, M.; Dervan, P. B., Activation of Gene Expression by Small Molecule Transcription Factors, *Proc. Natl. Acad. Sci. USA*, 2000, 97, 3930-3935.

(92) Coull, J. J.; He, G.; Melander, C.; Rucker, V. C.; Dervan, P. B.; Margolis, D. M., Targeted Depression of the Human Immunodeficiency Virus Type I Long Terminal Repeat by Pyrrole-Imidazole Polyamides, *J. Virology*, 2002, 76, 12349-12354.

(93) Ansari, A. Z.; Mapp, A. K.; Nguyen, D. H.; Dervan, P. B.; Ptashne, M., Towards a Minimal Motif for Artificial Transcription Activators, *Chem. Biol.*, 2001, 8, 583-592.

(94) Gottesfeld, J. M.; Neely, L.; Trauger, J. W.; Baird, E. E.; Dervan, P. B., Regulation of Gene Expression by Small Molecules, *Nature*, 1997, 387, 202-205.

(95) Wurtz, N. R.; Pomerantz, J. L.; Baltimore, D.; Dervan, P. B., Inhibition of DNA Binding by NF-κB with Pyrrole-Imidazole Polyamides, *Biochemistry*, 2002, 41, 7604-7609.

(96) Takahashi, R.; Bando, T.; Sugiyama, H., Specific Alkylation of Human Telemere Repeats by Hairpin Pyrrole-Imidazole Polyamide, *Bioorg. & Med. Chem.*, 2003, 11, 2503-2509.

(97) Fechter, E. J.; Dervan, P. B., Allosteric Inhibition of Protein-DNA Complexes by Polyamide-Intercalator Complexes, *J. Am. Chem. Soc.*, 2003, 125, 8476-8485.

(98) Nguyen-Hackley, D. H.; Ramm, E.; Taylor, C. M.; Joung, J. K.; Dervan, P. B.; Pabo, C. O., Allosteric Inhibition of Zinc-Finger Binding in the Major Groove of DNA by Minor-Groove Binding Ligands, *Biochemistry*, 2004, 43, 3880-3890.

(99) Chenoweth, D. M.; Viger, A.; Dervan, P. B., Fluorescent Sequence-Specific dsDNA Binding Oligomers, *J. Am. Chem. Soc.*, 2007, 129, 2216-2217.

(100) Anthony, N. G.; Johnston, B. F.; Khalaf, A. I.; Mackay, S. P.; Parkinson, J. A.; Suckling, C. J.; Waigh, R. D., Short Lexitropsin that Recognizes the DNA Minor Groove at 5'-ACTAGT-3': Understanding the Role of Isopropyl-thiazole, *J. Am. Chem. Soc.*, 2004, 126, 11338-11349.

(101) Foister, S.; Marques, M. A.; Doss, R. M.; Dervan, P. B., Shape Selective Recognition of TA Base Pairs by Hairpin Polyamides Containing N-Terminal 3-Methoxy (and 3-Chloro) Thiophene Residues, *Bioorg. & Med. Chem.*, 2003, 11, 4333-4340.

(102) Fregeau, N. L.; Wang, Y.; Pon, R. T.; Wylie, W. A.; Lown, J. W., Characterization of a CPI-Lexitropsin Conjugate-Oligonucleotide Covalent Complex by 1H NMR and Restrained Molecular Dynamics Simulations, *J. Am. Chem. Soc.*, 1995, 117, 8917-8925.

(103) Singh, M. P.; Kumar, S.; Joseph, T.; Pon, R. T.; Lown, J. W., A 1H-NMR Study of the DNA Binding Characteristics of Thioformyldistamycin, an Amide Isosteric Lexitropsin, *Biochemistry*, 1992, 31, 6453-6461.

(104) Pelton, J. G.; Wemmer, D. E., Structural Modeling of the Distamycin A-d(CGCCAATTCGC)2 Complex Using 2D NMR and Molecular Mechanics, *Biochemistry*, 1988, 27, 8088-8096.

(105) Zhang, Q.; Dwyer, T. J.; Tsui, V.; Case, D. A.; Cho, J.; Dervan, P. B.; Wemmer, D. E., NMR Structure of a Cyclic Polyamide-DNA Complex, *J. Am. Chem. Soc.*, 2004, 126, 7958-7966.

(106) Marques, M. A.; Doss, R. M.; Urbach, A. R.; Dervan, P. B., Toward an Understanding of the Chemical Etiology for DNA Minor-Groove Recognition by Polyamides, *Helv. Chim. Acta*, 2002, 85, 4485-4517.

(107) Singh, S. B.; Wemmer, D. E.; Kollman, P. A., Relative Binding Affinities of Distamycin and its Analog to d(CGCAAGTTGGC) d(GCCAACTTGCG): Comparison of Simulation Results with Experiment, *Proc. Natl. Acad. Sci. USA*, 1994, 91, 7673-7677.

(108) Boehncke, K.; Nonella, M.; Schulten, K., Molecular Dynamics Investigation of the Interaction between DNA and Distamycin, *Biochemistry*, 1991, 30, 5465-5475.

(109) Dolenc, J.; Oostenbrink, C.; Koller, J.; van Gunsteren, W. F., Molecular Dynamics Simulations and Free Energy Calculations of netropsin and distamycin binding to an AAAAA DNA binding site, *Nucleic Acids Research*, 2005, 33, 725-733.

(110) Bielawski, K.; Bielawska, A.; Bartulewicz, D.; Rozanski, A., Molecular Modeling of the Interaction of Carbocyclic Analogues of Netropsin and Distamycin with d(CGCGAATTCGCG)2, *Acta Biochimica Polonica*, 2000, 47, 855-866.

(111) Bartulewicz, D.; Markowska, A.; Wolczynski, S.; Dabrowska, M.; Rozanski, A., Molecular modelling, synthesis, and antitumor activity of carbocyclic analogues of netropsin and distamycin--new carriers of alkylating agents, *Acta Biochimica Polonica*, 2000, 47, 23-35.

(112) Bartulewicz, D.; Bielawski, K.; Bielawska, A.; Rozanski, A., Synthesis, molecular modelling, and antiproliferative and cytotoxic effects of carbocyclic derivatives of distamycin with chlorambucil moiety, *Eur. J. Med. Chem.*, 2001, 36, 461-467.

(113) Topham, C. M.; Smith, J. C., Molecular Mechanics Potential Functions for Drug Design: DNA Minor Groove Binders, *J. Chim. Phys.*, 1997, 94, 1313-1338.

(114) Wellenzohn, B.; Winger, R. H.; Hallbrucker, A.; Mayer, E.; Liedl, K. R., Simulation of EcoRI Dodecamer Netropsin Complex Confirms Class I Complexation Mode, *J. Am. Chem. Soc.*, 2000, 122, 3927-3931.

(115) Wellenzohn, B.; Flader, W.; Winger, R. H.; Hallbrucker, A.; Mayer, E.; Liedl, K. R., Structural Flexibility of the d(CCAGTACTGG)$_2$ B-DNA Decamer and Its Complex with Two Polyamides, *J. Phys. Chem. B*, 2001, 105, 3135-3142.

(116) Wellenzohn, B.; Flader, W.; Winger, R. H.; Hallbrucker, A.; Mayer, E.; Liedl, K. R., Complex of B-DNA with Polyamides Freezes DNA Backbone Flexibility, *J. Am. Chem. Soc.*, 2001, 123, 5044-5049.

(117) Wellenzohn, B.; Flader, W.; Winger, R. H.; Hallbrucker, A.; Mayer, E.; Liedl, K. R., Significance of Ligand Tails for Interaction with the Minor Groove of B-DNA, *Biophys. J.*, 2001, 81, 1588-1599.

(118) Wellenzohn, B.; Flader, W.; Winger, R. H.; Hallbrucker, A.; Mayer, E.; Liedl, K. R., Influence of Netropsin's Charges on the Minor Groove Width of d(CGCGAATTCGCG)2, *Biopolymers*, 2002, 61, 276-286.

(119) Wellenzohn, B.; Loferer, M. J.; Trieb, M.; Rauch, C.; Winger, R. H.; Mayer, E.; Liedl, K. R., Hydration of Hydroxypyrrole Influences Binding of ImHpPyPy-β-Dp Polyamide to DNA, *J. Am. Chem. Soc.*, 2003, 125, 1088-1095.

(120) Pang, Y.-P., Nonbonded Bivalence Approach to Cell-Permeable Molecules that Target DNA Sequences, *Bioorg. & Med. Chem.*, 2004, 12, 3063-3068.

(121) Pelton, J. G.; Wemmer, D. E., Structural Characterization of a 2:1 Distamycin A
d(CGCAAATTGGC) Complex by Two-Dimensional NMR, *Proc. Natl. Acad. Sci. USA*, 1989, 86, 5723-5727.

(122) Bren, U.; Hodoscek, M.; Koller, J., Development and Validation of Empirical Force Field Parameters for Netropsin, *J. Chem. Inf. Model.*, 2005, 45, 1546-1552.

(123) Kristyan, S.; Pulay, P., Can (semi) local density functional theory account for the London dispersion forces?, *Chemical Physics Letters*, 1994, 229, 175-180.

(124) Kleis, J.; Schroder, E., van der Waals interaction of simple, parallel polymers, *J. Chem. Phys.*, 2005, 122.

(125) Langreth, D. C.; Dion, M.; Rydberg, H.; Schroder, E.; Hyldgaard, P.; Lundqvist, B. I., van der Waals Density Functional Theory with Applications, *Int. J. Quantum Chem.*, 2005, 101, 599.

(126) Perez-Jorda, J. M.; Becke, A. D., A Density-Functional Study of van der Waals forces: rare gas diatomics, *Chemical Physics Letters*, 1995, 233, 134-137.

(127) Dobson, J. F.; Wang, J., Testing the local density approximation with energy-versus-separation curves of jellium slab pairs, *Phys. Rev. B*, 2004, 69, 235104.

(128) Zimmerli, U.; Parrinello, M.; Koumoutsakos, P., Dispersion Corrections to Density Functionals for Water Aromatic Interactions, *J. Chem. Phys.*, 2004, 120, 2693.

(129) Hult, E.; Andersson, Y.; Lundqvist, B. I., Density Functional for van der Waals Forces at Surfaces, *Physical Review Letters*, 1996, 77, 2029.

(130) Cybulski, S. M.; Severson, C. E., Critical Examination of the Supermolecule Density Functional Theory Calculations of Intermolecular Interactions, *J. Chem. Phys.*, 2005, 122, 014117.

(131) Hyla-Kryspin, I.; Haufe, G.; Grimme, S., Weak hydrogen bridges: A systematic theoretical study on the nature and strength of C-H---F-C Interactions, *Chem. Eur. J.*, 2004, 10, 3411-3422.

(132) Mrksich, M.; Wade, W. S.; Dwyer, T. J.; Geierstanger, B. H.; Wemmer, D. E.; Dervan, P. B., Antiparallel Side-by-Side Dimeric Motif for Sequence-Specific Recognition in the Minor Groove of DNA by the Designed Peptide 1-methylimidazole-2-carboxamide netropsin, *Proc. Natl. Acad. Sci. USA*, 1992, 89, 7586-7590.

(133) Szabo, A.; Ostlund, N. S., *Modern Quantum Chemistry: Introduction fo Advanced Electronic Structure Theory*. Dover Publications: Mineola, N.Y., 1996.

(134) Gudas, L. J., Retinoids and Vertebrate Development, *J. Biol. Chem.*, 1994, 269, 15399-15402.

(135) Zanotti, G.; Berni, R., Retinoids in Mammals: A Crystallographic Perspective, *Croatica Chemica Acta*, 2002, 75, 835-845.

(136) Napoli, J. L., Interactions of retinoid binding proteins and enzymes in retinoid metabolism, *Biochimica et Biophysica Acta*, 1999, 1440, 139-162.

(137) Calderone, V.; Folli, C.; Marchesani, A.; Berni, R.; Zanotti, G., Identification and Structural Analysis of a Zebrafish Apo and Holo Cellular Retinol-Binding Protein, *J. Mol. Biol.*, 2002, 321, 527-535.

(138) Napoli, J. L., Biochemical Pathways of Retinoid Transport, Metabolism, and Signal Transduction, *Clinical Immunology and Immunopathology*, 1996, 80, S52-S62.

(139) Folli, C.; Calderone, V.; Ottonello, S.; Bolchi, A.; Zanotti, G.; Stoppini, M.; Berni, R., Identification, retinoid binding, and x-ray analysis of a human retinol-binding protein, *Proc. Natl. Acad. Sci. USA*, 2001, 98, 3710-3715.

(140) Goodman, A. B., Retinoid Receptors, Transporters, and Metabolizers as Therapeutic Targets in Late Onset Alzheimer Disesase, *J. Cellular Physiology*, 2006, 209, 598-603.

(141) Cowan, S. W.; Newcomer, M. E.; Jones, T. A., Crystallographic studies on a family of cellular lipophilic transport proteins. Refinement of P2 myelin protein and the structure determination and refinement of cellular retinol-binding protein in complex with all-trans-retinol., *J. Mol. Biol.*, 1993, 230, 1225.

(142) Ulukaya, E.; Wood, E. J., Fenretinide and its relation to cancer, *Cancer Treatment Reviews*, 1999, 25, 229.

(143) Borghi, R.; Vene, R.; Arena, G.; Schubert, D.; Albini, A.; Tosetti, F., Transient Motion of cytoplasmic and nuclear retinoid receptors expression in differentiating human teratocarcinoma NT2 cells, *Journal of Neurochemistry*, 2003, 84, 94-104.

(144) Newcomer, M. E., Retinoid-binding proteins: Structural determinants important for function, *FASEB J.*, 1995, 9, 229-239.

(145) Zanotti, G.; Marcello, M.; Malpeli, G.; Folli, C.; Sartori, G.; Berni, R., Crystallographic Studies on Complexes Between Retinoids and Plasma Retinol-Binding Protein, *J. Biol. Chem.*, 1994, 269, 29613-29620.

(146) Zanotti, G.; D'Acunto, M. R.; Malpeli, G.; Folli, C.; Berni, R., Crystal Structure of the Transthyretin-retinoic acid complex, *Eur. J. Biochem.*, 1995, 234, 563-569.

(147) Thompson, J. R.; Bratt, J. M.; Banaszak, L. J., Crystal Structure of Cellular Retinoic Acid Binding protein I shows increased access to the binding cavity due to formation of an intermolecular β-sheet, *J. Mol. Biol.*, 1995, 252, 433-446.

(148) Newcomer, M. E.; Ong, D. E., Plasma retinol binding protein: structure and function of the prototypic lipocalin, *Biochimica et Biophysica Acta*, 2000, 1482, 57-64.

(149) Raghu, P.; Sivakumar, B., Interactions amongst plasma retinol-binding protein, transthyretin, and their ligands: implications in vitamin A homeostasis and transthyretin amyloidosis, *Biochimica et Biophysica Acta*, 2004, 1703, 1-9.

(150) Chaudhuri, B. N.; Kleywegt, G. J.; Hermite, B.-L.; Bergfors, T.; Senn, H.; Le Motte, P.; Partouche, O.; Jones, T. A., Structures of cellular retinoic acid binding proteins I and II in complex with synthetic retinoids., *Acta. Cryst.*, 1999, D55, 1850.

(151) Li, E.; Norris, A. W., Structure/Function of Cytoplasmic vitamin A-binding proteins, *Annu. Rev. Nutr.*, 1996, 16, 205-234.

(152) Li, E., Structure and function of cytoplasmic retinoid binding proteins, *Molecular and Cellular Biology*, 1999, 192, 105-108.

(153) Summerbell, D., The effect of local application of retinoic acid to the anterior margin of the developing chick limb, *J. Embryol. Exp. Morph.*, 1983, 78, 269-289.

(154) Crabb, J. W.; Goldflam, S.; Harris, S. E.; Saari, J. C., Cloning of the cDNAs Encoding the Cellular Retinaldehyde-binding protein from Bovine and Human Retina and Comparison of the protein structures, *J. Biol. Chem.*, 1988, 263, 18688-18692.

(155) Vivat-Hannah, V.; Zusi, F. C., Retinoids as Therapeutic Agents: Today and Tomorrow, *Mini-Reviews in Medicinal Chemistry*, 2005, 5, 755-760.

(156) Torrisi, R.; Descensi, A.; Formelli, F.; Camerini, T.; De Palo, G., Chemoprevention of Breast Cancer with Fenretinide, *Drugs*, 2001, 61, 909-918.

(157) Zusi, F. C.; Lorenzi, M. V.; Vivat-Hannah, V., Selective Retinoids and Rexinoids in Cancer Therapy and Chemoprevention, *Drug Discovery Today*, 2002, 7, 1165-1174.

(158) Altucci, L.; Gronemeyer, H., The Promise of Retinoids to Fight Against Cancer, *Nature Reviews Cancer*, 2001, 1, 181-193.

(159) Faul, M. M.; Grese, T. A., Selective RXR Modulators for the treatment of type II diabetes, *Curr. Opin. Drug Discov. Devel.*, 2002, 5, 974-985.

(160) Bonnani, B.; Lazzeroni, M.; Veronesi, U., Synthetic Retinoid Fenretinide in Breast Cancer Chemoprevention, *Expert Review of Anticancer Therapy*, 2007, 7, 423-432.

(161) Altucci, L.; Leibowitz, M. D.; Ogilvie, K. M.; de Lera, A. R.; Gronemeyer, H., RAR and RXR Modulation in cancer and metabolic disease, *Nature Reviews Drug Discovery*, 2007, 6, 798-810.

(162) Maciaszek, J. W.; Coniglio, S. J.; Talmage, D. A.; Viglianti, G. A., Retinoid-Induced Repression of Human Immunodeficiency Virus Type I Core Promotor Activity Inhibits Virus Replication, *J. Virology*, 1998, 72, 5862-5869.

(163) Taylor, G. A.; Shalita, A. R., Retinoid Therapy of acne and sebocyte-related disorders, *Basic and Clinical Dermatology*, 2007, 39, 103-123.

(164) Van de Kerkhof, P. C. M.; Verfaille, C. J., Retinoids and retinoic acid metabolism blocking agents in psoriasis, *Basic and Clinical Dermatology*, 2007, 39, 125-152.

(165) Rittie, L.; Fisher, G. J.; Griffiths, C. E. M., Anti-aging effects of retinoids and mechanisms of action, *Basic and Clinical Dermatology*, 2007, 39, 77-101.

(166) de Lera, A. R.; Bourguet, W.; Altucci, L.; Gronemeyer, H., Design of selective nuclear receptor modulators: RAR and RXR as a case study, *Nature Reviews Drug Discovery*, 2007, 6, 811-820.

(167) Lifson, S.; Warshel, A., Consistent Force Field for Calculations of Conformations, Vibrational Spectra, and Enthalpies of Cycloalkane and n-Alkane Molecules, *J. Chem. Phys.*, 1968, 49, 5116.

(168) Pariser, R.; Parr, R. G., A Semi-Empirical Theory of the Electronic Spectra and Electronic Structure of Complex Unsaturated Molecules. II, *J. Chem. Phys.*, 1953, 21, 767-776.

(169) Pople, J. A., Electron Interaction in Unsaturated Hydrocarbons, *Trans. Faraday Soc.*, 1953, 49, 1375-1385.

(170) Kao, J.; Allinger, N. L., Conformational Analysis. 122. Heats of Formation of Conjugated Hydrocarbons by the Force Field Method., *J. Am. Chem. Soc.*, 1977, 99, (4), 975.

(171) Bashford, D.; Gerwert, K., Electrostatic Calculations of the pKa values of ionizable groups in bacteriorhodopsin, *J. Mol. Biol.*, 1992, 224, 473-486.

(172) Zhou, F.; Windemuth, A.; Schulten, K., Molecular Dynamics Study of the Proton Pump Cycle of Bacteriorhodopsin, *Biochemistry*, 1993, 32, 2291-2306.

(173) Humphrey, W.; Logunov, I.; Schulten, K.; Sheves, M., Molecular Dynamics Study of Bacteriorhodopsin and Artificial Pigments, *Biochemistry*, 1994, 33, 3668-3678.

(174) Humphrey, W.; Xu, D.; Sheves, M.; Schulten, K., Molecular Dynamics Study of the Early Intermediates in the Bacteriorhodopsin Photocycle, *J. Phys. Chem.*, 1995, 99, 14549-14560.

(175) Xu, D.; Sheves, M.; Schulten, K., Molecular Dynamics Study of the M12 Intermediate of Bacteriorhodopsin, *Biophys. J.*, 1995, 69, 2745-2760.

(176) Nina, M.; Roux, B.; Smith, J. C., Functional Interactions in Bacteriorhodopsin: A Theoretical Analysis of Retinal Hydrogen Bonding with Water, *Biophys. J.*, 1995, 68, 25-39.

(177) Baudry, J.; Tajkhorshid, E.; Molnar, F.; Phillips, J.; Schulten, K., Molecular Dynamics Study of Bacteriorhodopsin and the Purple Membrane, *J. Phys. Chem. B*, 2001, 105, 905-918.

(178) Rajamani, R.; Gao, J., Combined QM/MM study of the opsin shift in bacteriorhodopsin, *J. Comp. Chem.*, 2002, 23, 96-105.

(179) Rohrig, U. F.; Guidoni, L.; Rothlisberger, U., Early Steps of the Intramolecular Signal Transduction in Rhodopsin Explored by Molecular Dynamics Simulations, *Biochemistry*, 2002, 41, 10799-10809.

(180) Hayashi, S.; Tajkhorshid, E.; Schulten, K., Structural Changes during the Formation of Early Intermediates in the Bacteriorhodopsin Photocycle, *Biophys. J.*, 2002, 83, 1281-1297.

(181) Ulmschneider, M. B.; Tieleman, D. P.; Sansom, M. S. P., Interactions of a Transmembrane Helix and a Membrane: Comparative Simulations of Bacteriorhodopsin Helix A, *J. Phys. Chem. B*, 2004, 108, 10149-10159.

(182) Huber, T.; Botelho, A. V.; Beyer, K.; Brown, M. F., Membrane Model for the G-Protein-Coupled Receptor Rhodopsin: Hydrophobic Interface and Dynamical Structure, *Biophys. J.*, 2004, 86, 2078-2100.

(183) Lemaitre, V.; Yeagle, P. L.; Watts, A., Molecular Dynamics Simulations of Retinal in Rhodopsin: From the Dark-Adapted State towards Lumirhodopsin, *Biochemistry*, 2005, 44, 12667-12680.

(184) Sato, Y.; Hata, M.; Neya, S.; Hoshino, T., Computational Analysis of the Proton Translocation from Asp96 to Schiff Base in Bacteriorhodopsin, *J. Phys. Chem. B*, 2006, 110, 22804-22812.

(185) Saam, J.; Tajkhorshid, E.; Hayashi, S.; Schulten, K., Molecular Dynamics Investigation of Primary Photoinduced Events in the Activation of Rhodopsin, *Biophys. J.*, 2002, 83, 3097-3112.

(186) Roux, B.; Nina, M.; Pomes, R.; Smith, J. C., Thermodynamic Stability of Water Molecules in the Bacteriorhodopsin Proton Channel: A Molecular Dynamics Free Energy Perturbation Study, *Biophys. J.*, 1996, 71, 670-681.

(187) Tajkhorshid, E.; Baudry, J.; Schulten, K.; Suhai, S., Molecular Dynamics Study of the Nature and Origin of Retinal's Twisted Structure in Bacteriorhodopsin, *Biophys. J.*, 2000, 78, 683-693.

(188) Sugihara, M.; Buss, V.; Entel, P.; Hafner, J., The Nature of the Complex Counterion of the Chromophore in Rhodopsin, *J. Phys. Chem. B*, 2004, 108, 3673-3680.

(189) Straub, J. E.; Rashkin, A. B.; Thirumalai, D., Dynamics in Rugged Energy Landscapes with Applications to the S-Peptide and Ribonuclease A, *J. Am. Chem. Soc.*, 1994, 116, 2049-2063.

(190) Schulze, B. G.; Grubmuller, H.; Evanseck, J. D., Functional Significance of Hierarchical Tiers in Carbonmonoxy Myoglobin: Conformational Substates and Transitions Studied by Conformational Flooding Simulations, *J. Am. Chem. Soc.*, 2000, 122, 8700-8711.

(191) Lyman, E.; Zuckerman, D. M., Ensemble-Based Convergence Analysis of Biomolecular Trajectories, *Biophysical Journal*, 2006, 91, 164-172.

(192) Beveridge, D. L.; Barreiro, G.; Byun, K. S.; Case, D. A.; Cheatham, T. E., III; Dixit, S. B.; Giudice, E.; Lankas, F.; Lavery, R.; Maddocks, J. H.; Osman, R.; Seibert, E.; Sklenar, H.; Stoll, G.; Thayer, K. M.; Varnai, P.; Young, M. A., Molecular Dynamics Simulations of the 136 Unique Tetranucleotide Sequences of DNA Oligonucleotides. I. Research Design and Results on d(CpG) Steps, *Biophysical Journal*, 2004, 87, 3799-3813.

(193) Beveridge, D. L.; Dixit, S. B.; Barreiro, G.; Thayer, K. M., Molecular Dynamics Simulations of DNA Curvature and Flexibility: Helix Phasing and Premelting, *Biopolymers*, 2004, 73, 380-403.

(194) Cheatham, T. E., III, Simulation and modeling of nucleic acid structure, dynamics and interactions, *Curr. Opin. Struc. Biol.*, 2004, 14, 360-367.

(195) de Vries, A. H.; Chandrasekhar, I.; van Gunsteren, W. F.; Hunenberger, P. H., Molecular Dynamics Simulations of Phospholipid Bilayers: Influence of Artificial Periodicity, System Size, and Simulation Time, *J. Phys. Chem. B*, 2005, 109, 11643-11652.

(196) Ponomarev, S. Y.; Thayer, K. M.; Beveridge, D. L., Ion Motions in Molecular Dynamics Simulations on DNA, *Proc. Natl. Acad. Sci.*, 2004, 101, 14771-14775.

(197) Norberg, J.; Nilsson, L., Conformational Free Energy of ApApA from Molecular Dynamics Simulations, *J. Phys. Chem.*, 1996, 100, 2550-2554.

(198) Noy, A.; Perez, A.; Lankas, F.; Luque, J.; Orozco, M., Relative Flexibility of DNA and RNA: A Molecular Dynamics Study, *J. Mol. Biol.*, 2004, 343, 627-638.

(199) Smith, L. J.; Daura, X.; van Gunsteren, W. F., Assessing Equilibration and Convergence in Biomolecular Simulations, *Proteins: Structure, Functions, and Genetics*, 2002, 48, 487-496.

(200) Sarzynska, J.; Kulinski, T.; Nilsson, L., Conformational Dynamics of a 5S rRNA Hairpin Domain Containing Loop D and a Single Nucleotide Bulge, *Biophys. J.*, 2000, 79, 1213-1227.

(201) Auffinger, P.; Louise-May, S.; Westhof, E., Multiple molecular dynamics simulations of the anticodon loop of tRNAAsp in aqueous solution with counterions, *J. Am. Chem. Soc.*, 1995, 117, 6720-6726.

(202) Auffinger, P.; Westhof, E., H-bond stability in the tRNAAsp anticodon hairpin: 3 ns of multiple molecular dynamics simulations, *Biophysical Journal*, 1996, 71, 940-954.

(203) Auffinger, P.; Westhof, E., RNA Hydration: Three Nanoseconds of Multiple Molecular Dynamics Simulations of the Solvated $tRNA^{Asp}$ Anticodon Hairpin, *J. Mol. Biol.*, 1997, 269, 326-341.

(204) Mihailescu, D.; Reed, J.; Smith, J. C., Convergence in peptide folding simulation: Multiple trajectories of a potential AIDS pharmacophore, *Biopolymers*, 2003, 70, (2), 121-133.

(205) Genest, D., How Long Does DNA Keep the Memory of Its Conformation? A Time-Dependent Canonical Correlation Analysis of Molecular Dynamics Simulation, *Biopolymers*, 1996, 38, 389-399.

(206) Zhang, W.; Wu, C.; Duan, Y., Convergence of Replica Exchange Molecular Dynamics, *J. Chem. Phys.*, 2005, 123, 154105-154101-154105-154109.

(207) Stella, L.; Melchionna, S., Equilibration and Sampling in Molecular Dynamics Simulations of Biomolecules, *J. Chem. Phys.*, 1998, 109, 10115-10117.

(208) Babin, V.; Baucom, J.; Darden, T. A.; Sagui, C., Molecular Dynamics Simulations of DNA with Polarizable Force Fields: Convergence of an Ideal B-DNA Structure to the Crystallographic Structure, *J. Phys. Chem. B*, 2006, 110, 11571-11581.

(209) Lyman, E.; Zuckerman, D. M., On the Structural Convergence of Biomolecular Simulations by Determination of the Effective Sample Size, *J. Phys. Chem. B*, 2007.

(210) Feig, M.; Pettitt, B. M., Structural Equilibrium of DNA Represented with Different Force Fields, *Biophys. J.*, 1998, 75, 134-149.

(211) Amadei, A.; Linssen, A. B.; Berendsen, H. J., Essential dynamics of proteins, *Proteins*, 1993, 17, (4), 412-425.

(212) Arcangeli, C.; Bizzarri, A. R.; Cannistraro, S., Molecular dynamics simulation and essential dynamics study of mutated plastocyanin: structural, dynamical and functional effects of a disulfide bridge insertion at the protein surface, *Biophysical Chemistry*, 2001, 92, (3), 183-199.

(213) Chau, P. L.; van Aalten, D. M. F.; Bywater, R. P.; Findlay, J. B. C., Functional concerted motions in the bovine serum retinol-binding protein, *Journal of Computer-Aided Molecular Design*, 1999, 13, (1), 11-20.

(214) Chen, C.; Xiao, Y.; Zhang, L., A directed essential dynamics simulation of peptide folding, *Biophysical Journal*, 2005, 88, (5), 3276-3285.

(215) Jha, S.; Coveney, P. V.; Laughton, C. A., Force Field Validation for Nucleic Acid Simulations: Comparing Energies and Dynamics of a DNA Dodecamer, *J. Comp. Chem.*, 2005, 26, 1617-1627.

(216) Perez, A.; Blas, J. R.; Rueda, M.; Lopez-Bes, J. M.; de la Cruz, X.; Orozco, M., Exploring the Essential Dynamics of B-DNA, *Journal of Chemical Theory and Computation*, 2005, 1, 790-800.

(217) Daura, X.; van Gunsteren, W. F.; Mark, A. E., Folding-Unfolding Thermodynamics of a β-Heptapeptide From Equilibrium Simulations, *Proteins: Structure, Functions, and Genetics*, 1999, 34, 269-280.

(218) Karpen, M. E.; Tobias, D. J.; Brooks, C. L., III, Statistical Clustering Techniques for the Analysis of Long Molecular Dynamics Trajectories: Analysis of 2.2ns Trajectories of YPGDV, *Biochemistry*, 1993, 32, 412-420.

(219) Elmer, S. P.; Pande, V. S., Foldamer Simulations: Novel Computational Methods and Applications to Poly-Phenylacetylene Oligomers, *J. Chem. Phys.*, 2004, 121, 12760-12771.

(220) Thirumalai, D.; Mountain, R. D.; Kirkpatrick, T. R., Ergodic Behavior in Supercooled Liquids and in Glasses, *Phys. Rev. A*, 1989, 39, 3563-3574.

(221) Mountain, R. D.; Thirumalai, D., Measures of Effective Ergodic Convergence in Liquids, *J. Phys. Chem.*, 1989, 93, 6975-6979.

(222) de Groot, B. L.; Daura, X.; Mark, A. E.; Grubmuller, H., Essential Dynamics of Reversible Peptide Folding: Memory-free Conformational Dynamics Governed by Internal Hydrogen Bonds, *Journal of Molecular Biology*, 2001, 309, (1), 299-313.

(223) de Groot, B. L.; Hayward, S.; van Aalten, D. M.; Amadei, A.; Berendsen, H. J., Domain motions in bacteriophage T4 lysozyme: a comparison between molecular dynamics and crystallographic data, *Proteins*, 1998, 31, (2), 116-127.

(224) de Groot, B. L.; Vriend, G.; Berendsen, H. J., Conformational changes in the chaperonin GroEL: new insights into the allosteric mechanism, *Journal of molecular biology*, 1999, 286, (4), 1241-1249.

(225) Kazmierkiewicz, R.; Czaplewski, C.; Lammek, B.; Ciarkowski, J., Essential dynamics/factor analysis for the interpretation of molecular dynamics trajectories, *Journal of Computer-Aided Molecular Design*, 1999, 13, (1), 21-33.

(226) Lee, M. C.; Deng, J.; Briggs, J. M.; Duan, Y., Large-Scale Conformational Dynamics of the HIV-1 Integrase Core Domain and its Catalytic Loop Mutants, *Biophysical Journal*, 2005, 88, 3133-3146.

(227) Faraldo-Gomez, J. D.; Forrest, L. R.; Baaden, M.; Bond, P. J.; Domene, C.; Patargias, G.; Cuthbertson, J.; Sansom, M. S. P., Conformational Sampling and Dynamics of Membrane Proteins from 10-Nanosecond Computer Simulations, *Proteins: Structure, Functions, and BioInformatics*, 2004, 57, 783-791.

(228) Balsera, M. A.; Wriggers, W.; Oono, Y.; Schulten, K., Principal Component Analysis and Long Time Protein Dynamics, *Journal of Physical Chemistry*, 1996, 100, 2567-2572.

(229) Hess, B., Similarities between principal components of protein dynamics and random diffusion, *Physical Review E: Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, 2000, 62, (6-B), 8438-8448.

(230) Hess, B., Convergence of sampling in protein simulations, *Physical Review E*, 2002, 65, 031910-031911-031910-031910.

(231) Gelman, A.; Rubin, D. B., Inference from Iterative Simulation Using Multiple Sequences, *Statistical Science*, 1992, 7, 457-472.

(232) Brooks, S. P.; Roberts, G. O., Convergence Assessment Techniques for Markov Chain Monte Carlo, *Statistics and Computing*, 1998, 8, 319-335.

(233) Chen, M.-H.; Shao, Q.-M.; Ibrahim, J. G., *Monte Carlo Methods in Bayesian Computation*. Springer-Verlag: 2000.

(234) Cowles, M. K.; Carlin, B. P., Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review, *J. Am. Stat. Assoc.*, 1996, 91, 883-904.

(235) Gelman, A., Inference and monitoring convergence. In *Markov Chain Monte Carlo in Practice*, Chapman and Hall/CRC Press: 1996; 131-141.

(236) Dickerson, R. E., Definitions and Nomenclature of Nucleic Acid Structure Components, *Nucleic Acids Research*, 1989, 17, 177-199.

(237) Chou, S.-H.; Wemmer, D. E.; Hare, D. R.; Reid, B. R., Sequence-Specific Recognition of DNA: NMR studies of the Imino Protons of a Synthetic RNA Polymerase Promoter, *Biochemistry*, 1984, 23, 2257-2262.

(238) Dornberger, U.; Leijon, M.; Fritzsche, H., High Base Pair Opening Rates in Tracts of GC Base Pairs, *J. Biol. Chem.*, 1999, 274, 6957-6962.

(239) Gorenstein, D. G., Conformation and Dynamics of DNA and Protein-DNA Complexes by ^{31}P NMR, *Chemical Reviews*, 1994, 94, 1315-1338.

(240) Pardi, A.; Tinoco, I., Kinetics for Exchange of Imino Protons in Deoxyribonucleic Acid, Ribonucleic Acid, and Hybrid Oligonucleotide Helices, *Biochemistry*, 1982, 21, 4686-4693.

Chapter 2

Methods

This chapter discusses the background, methods, and procedures used for computations. The theoretical background of molecular dynamics is described in Section 2.1, the statistical tests used for analysis are described in Section 2.2, the theoretical background of quantum mechanical methods used is described in Section 2.3, and descriptions of the computations performed are in Section 2.4.

2.1 Molecular Dynamics

Molecular dynamics simulation provides a method for examination of molecular motion at the level of the atom and thus is a useful tool in chemistry and physics.[1] MD is a classical simulation method, meaning that the motions of the particles in the system obey Newton's laws of motion and that electronic degrees of freedom are not explicitly described. Thus, one of the most important limitations of MD is that it cannot be used to describe phenomena involving bond breaking/formation events, excited states, or isomerizations.[1] MD simulations enable the time dependence of properties of a molecular system (structural, dynamic, and thermodynamic) to be studied by numerically

solving Newton's equations of motion. A simple description of a simulation can be summarized by the following four steps:[2]

1. Initial coordinates are obtained (in the case of biomolecules, these are usually obtained from the Brookhaven Protein Data Bank[3]or from the Rutgers Nucleic Acid Database[4]. Initial velocities are then assigned to the initial coordinates (usually assigned from some sort of distribution—Gaussian, Maxwell, etc.).

2. Compute the forces on acting on every particle in the system. This is the most time-consuming part of most simulations because the contribution of to the force on every particle due to all of its neighbors must be considered. If the system contains N particles, and if no techniques are employed in order to speed up the evaluation of short- or long-range forces (use of nonbond cutoffs), the force evaluation scales as N^2.

3. Integrate Newton's equations of motion. After all forces are computed, Newton's equations of motion are integrated by breaking the calculation down into small time steps. Integration of Newton's equations is described in further detail below. Steps 2 and 3 make up the bulk of the simulation; they repeated until the time evolution of the system has been computed for the specified amount of time.

4. Determine desired averages and properties of the system. When the calculation is complete, the desired structural, dynamic, and thermodynamic properties can then be computed.

2.1.1 CHARMM Force Field. In molecular dynamics and molecular mechanics, the intra- and intermolecular forces between the atoms are described by an empirical force field. Empirical force fields use atomistic models, where atoms are the smallest particles in the system, rather than the electrons and nuclei used in quantum mechanics.

The empirical force field, as known as a potential energy function, allows for the potential energy (V) of the system to be calculated as a function of the 3-D structure of the system (R). The potential energy function consists of harmonic approximations of intramolecular (bond stretching, angle bending, and torsion interaction) terms and intermolecular (Coulombic and Van der Waals) terms (Eqn.2.1).

$$V(R)_{total} = V(R)_{intramolecular} + V(R)_{intermolecular} \qquad \text{Eqn. 2.1}$$

The harmonic approximations used in the force field are sufficient for biomolecular simulations. Simulations of biomolecules are typically performed at temperatures where bond lengths and angles usually stay near their equilibrium values. Also, since there is no bond breakage/formation occurring in an MD simulation, a harmonic potential can accurately describe any bond and angle distortions.

The CHARMM force field, which is used in this dissertation, has the following form (Eqn.2.2).

$$V(R) = \sum_{bonds} K_b(b-b_0)^2 + \sum_{UB} K_{UB}(S-S_0)^2 + \sum_{angle} K_\theta(\theta-\theta_0)^2 +$$
$$\sum_{dihedrals} K_\chi(1+\cos(n\chi-\delta)) + \sum_{impropers} K_{imp}(\phi-\phi_0)^2 + \qquad \text{Eqn. 2.2}$$
$$\sum_{nonbonded} \varepsilon\left[\left(\frac{R\min_{ij}}{r_{ij}}\right)^{12} - \left(\frac{R\min_{ij}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{\varepsilon_1 r_{ij}}$$

The first five terms are the intramolecular terms, and the last two terms are the intermolecular terms. The parameters K_b, K_{UB}, K_θ, K_χ, and K_{imp} are the force constants and are usually determined experimentally or from quantum mechanical calculations.[5] The terms in parentheses (e.g., b-b_0) describe the deviation (b) of the parameter value from the equilibrium value (b_0). For example, (b-b_0) describes the deviation of the

current bond length from the equilibrium bond length (the other terms are analogous). The UB term is the Urey-Bradley term which describes 1,3-nonbonded interactions. In the dihedral term, χ is the torsion angle value, δ is the phase shift, and n is the multiplicity. The first intermolecular term is the Lennard-Jones term, where $R_{min,ij}$ is the minimum interaction radius (this is dependent on the atoms interacting and is determined by experiment), ε is the Lennard-Jones well depth, $1/r^6$ describes the attractive interactions, and $1/r^{12}$ describes the repulsive interactions. The second intermolecular term is the Coulombic term, where q_i and q_j are the partial charges on the interacting atoms, r_{ij} is the distance between the interacting atoms, and ε_1 is the dielectric constant, which is generally treated as equal to 1 (permittivity of vacuum).

The 3-D structure (usually Cartesian coordinates obtained from the PDB or NDB in the case of biomolecules) and all parameters required in the force field are needed in order to begin the simulation. Once these are known, the energy of the system can then be calculated using Eqn.2.2. The first derivative of the energy with respect to the positions of the atoms yields the forces acting on the atoms, which is then used for the MD simulation (Eqn.2.3),

$$\frac{\partial V(x_1, x_2, \ldots x_i \ldots x_N)}{\partial x_i} = F(x_1, x_2, \ldots x_i \ldots x_N) \qquad \text{Eqn. 2.3}$$

where x are the coordinates (x_i is the coordinate for the ith particle and N is the number of particles in the system).

There are currently a variety of existing force fields, and choice for which to use depends on the nature of the system. In this dissertation, the CHARMM force field is used for MD simulations, which is designed for use with biomolecular systems.[6-8]

2.1.2 Integration of Newton's equations. Once the 3-D coordinates are obtained, velocities are assigned to each particle in the system (these are usually assigned from a Gaussian or Maxwell distribution). The motion of the system is then simulated by integrating Newton's second law of motion,

$$F = ma \qquad \text{Eqn. 2.4}$$

where F is the force acting on a particle (this is obtained from the potential energy function in Eqn.2.2), m is the mass of the particle, and a is its acceleration. The force can also be expressed as the gradient of the potential energy.

$$F = -\nabla V(R) \qquad \text{Eqn. 2.5}$$

Since acceleration is the second derivative of position with respect to time, this can be substituted in and Eqns. 2.4 and 2.5 can be combined to yield

$$-\frac{dV}{dR} = m\frac{d^2R}{dt^2} \qquad \text{Eqn. 2.6}$$

Newton's equation of motion can then relate the derivative of the potential energy to the changes in position as a function of time.

The potential energy is a function of all coordinates of all particles in the system. Numerical treatment is accomplished by using finite difference methods, which break the integration down into small timesteps (Δt). The timestep for the simulation must be chosen so that the algorithm will accurately describe the fastest degrees of freedom (highest frequency motions) of the system. For biomolecular systems, the fastest motions are the stretching vibrations of hydrogens covalently bound to heavy atoms (this occurs on a timescale of ~1 fs; therefore, the integration is only accurate for a timestep of ~1-2 fs). In order to use a timestep of 2 fs, the SHAKE algorithm[9] is applied to covalent

hydrogen bonds in molecular dynamics simulations. SHAKE holds the covalent hydrogen bonds (e.g. C-H) constrained, eliminating the high frequency motions, and thus allows for a larger timestep. From the force, the accelerations of the particles can be calculated, and those are then combined with the positions and velocities at time t to calculate the new positions and velocities at time $t + \Delta t$. The forces on the particles in the new positions are then calculated, which gives new positions and velocities at time $t + 2\Delta t$, etc.

There are several algorithms which are commonly used for integrating Newton's equations using finite difference methods. All of them assume that the positions, velocities, and accelerations of the particles in the system can be approximated by a Taylor series expansion. The most commonly used algorithm in simulations of biomolecules is the Verlet algorithm, which is based on two Taylor series expansions, a forward expansion ($t + \Delta t$) and a backward ($t - \Delta t$) expansion:

$$r_{n+1} = r_n + v_n \Delta t + \frac{1}{2}\left(\frac{F_n}{m}\right)\Delta t^2 + O(\Delta t^3) \qquad \text{Eqn. 2.7}$$

$$r_{n-1} = r_n + v_n \Delta t + \frac{1}{2}\left(\frac{F_n}{m}\right)\Delta t^2 - O(\Delta t^3) \qquad \text{Eqn. 2.8}$$

where r_n indicates the position at step n, $n+1$ indicates the position at the next step, and $O(\Delta t^n)$ is the term of order Δt^n or smaller. These two expansions can be summed together, which gives an algorithm for propagating the positions:

$$r_{n+1} = 2r_n - r_{n-1} + \left(\frac{F_n}{m}\right)\Delta t^2 + O(\Delta t^4) \qquad \text{Eqn. 2.9}$$

The Verlet algorithm is carried out in two steps: use the current position r_n to calculate the current force F_n and then use the current and previous positions r_n and r_{n+1} together

with the current force F_n to obtain the position in the next step, r_{n+1}. These two steps are repeated for every timestep for every particle in the system.

Because the velocities do not explicitly appear in the Verlet algorithm for computing the new positions, subtracting Eqn. 2.8 from Eqn. 2.7 yields an algorithm for propagating the velocities:

$$v_n = \frac{r_{n+1} - r_{n+1}}{2\Delta t} + O(\Delta t^2) \qquad \text{Eqn. 2.10}$$

In this dissertation, the leap-frog integrator is used. The leap-frog integrator is a variation on the Verlet integrator; the modification was made to improve upon the large velocity errors of the velocity propagation of the Verlet algorithm. In the leap-frog integrator, velocities are determined at the mid-point of the position evaluation. The algorithm is written as:

$$r_{n+1} = r_n + v_{n+1/2}\Delta t \qquad \text{Eqn. 2.11}$$

$$v_{n+1/2} = v_{n-1/2} + \left(\frac{F_n}{m}\right)\Delta t \qquad \text{Eqn. 2.12}$$

where $v_{n\pm1/2}$ is the velocity at the mid-step time, $t \pm (1/2)\Delta t$. The leap-frog integrator involves three steps: use the current position r_n to calculate the current force F_n, use the current force and previous mid-step velocity $v_{n-1/2}$ to calculate the next mid-step velocity $v_{n+1/2}$, and finally use the current position and the next mid-step velocity to obtain the position in the next step r_{n+1}.

2.1.3 Nonbond Calculations. The expensive part of an MD simulation is the evaluation of nonbond forces. The number of bond, angle, and dihedral terms in the

force field are proportional to the number of particles in the system (the number of internal coordinates is *3N-6*, which is linear in *N*); however, the number of nonbond terms increases on the order of N^2 (nonbond forces must be calculated between every pair of atoms in system). The non-bond attractive and repulsive terms of the Lennard-Jones potential decay quickly with distance; at 2.5σ (σ is the distance at which the force between the two interacting particles is zero) the Lennard-Jones potential has only 1% of its value at σ.[10] However, charge-charge interactions do not decay as rapidly with distance; the electrostatic energy is ~1/r for two point charges. In both instances, various algorithms are employed in order to reduce computational time.[1, 11]

2.1.3.1. Use of Cutoffs. The most common way of reducing computational time and effort in the calculation of nonbond interactions is the use of nonbond cutoffs. These methods generally work by explicitly calculating the interaction energies between all pairs of particles within a specified distance (this is known as the cutoff distance and is usually set to 10-12 Å for biomolecular systems) and approximating the contributions from the particles beyond the cutoff distance.

Abruptly truncating interactions at the cutoff distance causes discontinuities in the potential energy and in the force near the cutoff distance.[10] This is obviously a problem, and methods have been developed to reduce the discontinuities. Two commonly used algorithms are known as switching and shifting functions. Both methods set the nonbond energies at zero beyond some distance value; however, the values for interactions at less than the cutoff distance, *r*, are treated differently. For a switching function *S(r)*, *S(r)* is a polynomial of *r* that alters the nonbond energies smoothly and gradually over a buffer region $[a,b]$ so that the energy at $b = 0$ and the energies at $r \leq a$ are not changed (the

buffer region is typically 1-2 Å).[11] Shifting functions alter the nonbond energies more gradually over a larger region than switching function. This method underestimates short-range forces and alters the nonbond energies over the region $r \leq b$, rather than the smaller region that is affected in switching functions.[11]

2.1.3.2. Ewald Summation. Because long-range electrostatic interactions do not decay as rapidly as other nonbond interactions, long-range electrostatics can be problematic in MD. Until recently, cutoffs were applied to electrostatics (just like they were to other, more rapidly decaying nonbond interactions) in order to cut back on computational time. However, it was noticed that when cutoffs were used for electrostatic interactions, artifacts were introduced to the system, especially when the system contained many localized charges, as in the case of DNA.[12, 13] For example, it was found that even when explicit counterions were used in DNA simulations with periodic boundary conditions and cutoffs, substantial distortions were observed for the DNA structure.[14] In simulations of polypeptides, the size of the cutoff distance was found to affect the stability of an α-helix.[15]

Because of the artifacts caused by cutoffs, use of the Ewald summation, which is a more rigorous treatment of electrostatics, is required for highly charged systems, such as DNA.[12, 16] Studies have shown that use of the Ewald summation has led to more stable trajectories of DNA, RNA, and proteins.[17-20] The Ewald summation is a technique for calculating the electrostatic energy of a system using periodic boundary conditions (simulated system is placed in a unit cell and is considered to have infinite images of itself in space). The method uses a combination of real and reciprocal space sums of interaction energies in order to evaluate long-range electrostatic energies. In

crystallography, the reciprocal lattice is an orthogonal system related to the orthogonal system associated with the atoms of the unit cell, which is the real space lattice. In the Ewald sum, two terms arise from the calculation of electrostatic energy; one term from pairs of atoms in the direct lattice (central simulation cell) and the other term from atom pairs corresponding to interactions with images of the central cell atoms. The Ewald sum is a periodic method; it includes a complete representation of electrostatic energies from an infinite number of images. Because it is a periodic method, the Ewald sum avoids the truncation effects in energy forces observed when using cutoff methods (and also requires use of periodic boundary conditions).

In the Ewald method, a particle interacts with all other particles in the simulation cell and with all of their images in the periodic system. There is thus a contribution to the total energy from interactions in the central simulation box along with the interactions between the central box and the periodic image boxes. The electrostatic contribution to the potential energy for all pairs of charges in the central simulation cell is written

$$V = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \qquad \text{Eqn. 2.13}$$

where q_i is the partial charge on particle i, ε_0 is the dielectric constant, N is the number of particles in the system, r_{ij} is the minimum distance between particles i and j, and ½ is in front to avoid double-counting. The charge-charge contribution between the particles in the central cell and all the images of all the particles in the periodic system is written as

$$V = \frac{1}{2}\sum_{\mathbf{n}}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{q_i q_j}{4\pi\varepsilon_0 \left|r_{ij}+\mathbf{n}\right|} \qquad \text{Eqn. 2.14}$$

where **n** is a vector whose components are integral multiples of the length, L, of the central simulation box (**n** = n_xL, n_yL, n_zL). The sum in Eqn. 2.14 is very slow to converge, so the Ewald sum converts the summation to two series, both of which converge much more quickly. The split must take into account the slow decay of Coulomb interactions when r is a large distance and also the variation Coulomb interactions when r is a short distance. The split is of the form

$$\frac{1}{r} = \frac{f(r)}{r} + \frac{1 - f(r)}{r} \qquad \text{Eqn. 2.15}$$

where *f(r)* must be a function to take both problems with Coulomb interaction into account. The first term is for real space and is short-ranged, and the second term is for reciprocal space and is a long-ranged term (this term can be Fourier-transformed). The Ewald sum uses the complementary error function *(erfc(r))* for *f(r)*.

$$erfc(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty \exp(-t^2) dt \qquad \text{Eqn. 2.16}$$

In the Ewald sum, each charge is considered to be surrounded by a neutralizing charge distribution which is of equal magnitude but opposite in sign to the charge. This is normally a Gaussian distribution of the form

$$\rho_i(r) = \frac{q_i \alpha^3}{\pi^{3/2}} \exp(-\alpha^2 r^2) \qquad \text{Eqn. 2.17}$$

The sum is now the sum of the interactions between charges and the neutralizing Gaussians. This is the real space part of the Ewald sum.

$$V = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{|\mathbf{n}|=0}{}' \frac{q_i q_j}{4\pi\varepsilon_0}\frac{erfc(\alpha\left|r_{ij}+\mathbf{n}\right|}{\left|r_{ij}+\mathbf{n}\right|} \qquad \text{Eqn. 2.18}$$

The summation with the prime indicates that the series does not include the interaction $i = j$ at $\mathbf{n} = 0$. The new summation that uses the error function now converges quickly (and beyond some cutoff distance, the value is considered negligible). The rate of convergence depends on α, the width of the neutralizing Gaussians, and the wider the Gaussian, the faster the sum converges.

Another charge distribution is then added to the system to counteract the first neutralizing distribution. This is the reciprocal space sum (**k** are reciprocal vectors) and is a Fourier series.

$$V = \frac{1}{2}\sum_{k\neq 0}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{\pi L^3}\frac{q_i q_j}{4\pi\varepsilon_0}\frac{4\pi^2}{k^2}\exp\left(-\frac{k^2}{4\alpha^2}\right)\cos(\mathbf{k}\cdot\mathbf{r}_{ij}) \qquad \text{Eqn. 2.19}$$

The reciprocal space sum, like the real space sum, also converges much more quickly than the original sum. However, unlike the real space sum, the reciprocal space sum converges more quickly for a small α (narrower Gaussian). The number of terms that must be included in the reciprocal sum increases as α increases, which results in more time required for evaluation.

The sum of the neutralizing Gaussians in real space includes the interaction of each Gaussian with itself. Therefore, a term must be subtracted from the Ewald sum:

$$V = -\frac{\alpha}{\sqrt{\pi}}\sum_{k=1}^{N}\frac{q_k^2}{4\pi\varepsilon_0} \qquad \text{Eqn. 2.20}$$

The sums described are convergent; however, they are convergent only for a system with a net charge of zero. If the system has a non-zero net charge, then an extra term must be added.

$$V = \frac{2\pi}{3L^3}\left|\sum_{i=1}^{N}\frac{q_i}{4\pi\varepsilon_0}\mathbf{r}_i\right|^2 \qquad \text{Eqn. 2.21}$$

The full Ewald summation is shown in Eqn. 2.22.

$$V = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left\{\begin{array}{l}\sum_{|\mathbf{n}=0|}^{\infty}{}' \frac{q_i q_j}{4\pi\varepsilon_0}\frac{erfc(\alpha\left|r_{ij}+\mathbf{n}\right|}{\left|r_{ij}+\mathbf{n}\right|} \\ +\sum_{k\neq 0}\frac{1}{\pi L^3}\frac{q_i q_j}{4\pi\varepsilon_0}\frac{4\pi^2}{k^2}\exp\left(-\frac{k^2}{4\alpha^2}\right)\cos(\mathbf{k}\cdot\mathbf{r}_{ij}) \\ -\frac{\alpha}{\sqrt{\pi}}\sum_{k=1}^{N}\frac{q_k^2}{4\pi\varepsilon_0}+\frac{2\pi}{3L^3}\left|\sum_{i=1}^{N}\frac{q_i}{4\pi\varepsilon_0}\mathbf{r}_k\right|^2\end{array}\right\} \qquad \text{Eqn. 2.22}$$

The Ewald method is computationally expensive to implement (it scales as N^2); however, methods have been devised to cut back on computational time.[14, 16] By optimizing α, the relative rates of convergence of the real and reciprocal space sums can be adjusted to suit the system under study. The optimal balance between the sums will enable the Ewald sum to scale as $N^{3/2}$, which is still a considerable amount of computational time for large biomolecular systems. The particle mesh Ewald sum (PME) developed by Darden and coworkers reduces the Ewald sum to *N log(N)*.[16] In the PME method, the trigonometric function values of the Fourier series present in the reciprocal space sum are evaluated by a smooth approximation of the potential over a grid, or mesh. The smoothing function is known as the Euler spline, which expresses the value of the trigonometric function at the actual charge coordinates in terms of the charge value at

neighboring grid points. The resulting sums over the grid points are then evaluated efficiently by a fast Fourier transform. This dissertation uses the PME for all simulations; in the case of net-charged systems, a neutralizing background is used by default in the standard Ewald calculation.[21]

The Ewald sum was developed by P.P.Ewald in the 1920s,[22] long before the advent of computers, in order to compute the electrostatic energy of crystal. In the 1970s, it was applied to computer simulations of particles for precisely the same reason: to calculate efficiently the electrostatic energy of a periodic system (effectively, a crystal).

2.1.4. Boundaries. The treatment of boundaries and their effects is important to MD simulations because it allows macroscopic properties to be simulated using a relatively small number of particles. To be realistic, biomolecular simulations have to include some sort of description of a solvent environment. This modeling of bulk solvent is accomplished through the use of boundary conditions. There are various approaches to doing this, which include continuum boundary conditions, finite boundary conditions, and periodic boundary conditions.

With continuum boundary conditions, the solute (protein or DNA) is treated as a macroscopic object surrounded by a continuum which represents the solvent (water molecules are represented explicitly). With finite boundary conditions, the solute is surrounded by a layer of explicit water molecules while the bulk solvent molecules are modeled by some form of boundary potential at the solvent/vacuum border. [1, 12] These methods have the advantage of decreasing the number of particles that are explicitly represented and thus decrease computational expense; however, they have drawbacks. With continuum solvent methods, explicit interaction of the solute with water molecules

at the atomic level cannot be observed (e.g., interaction of specific water molecules with amino acids in a protein).[12, 23] With finite boundary conditions, water molecules cannot diffuse naturally due to the finite size of the system, and movements of water molecules can only be observed in the interior parts of the protein, such as binding cavities.[12, 23]

Periodic boundary conditions allow simulations to be performed using a relatively small number of particles so that the particles in the system experience forces as if they were in bulk solvent. The simulated system (the solute and the solvent) is placed in a unit cell and is considered to have an infinite number of images in space. This replication of the unit cell forms an infinite lattice in three dimensions. In three dimensions, the atoms of the simulated system are contained within a unit cell analogous to the unit cell of a crystal. During the simulation, only the coordinates of the unit cell need are included. If an atom of the unit cell leaves the unit cell by crossing the boundary, then an image atom enters to replace it; therefore, the number of particles is conserved. In order to save computational time in evaluation of forces, the minimum image convention is customarily applied. With the minimum image convention, each atom only sees one image of every other atom in the system. The forces are calculated with only with closest periodic image.

The choice of geometry for the periodic cell is important. The cubic cell is the simplest periodic cell; however, it is appropriate to use a periodic cell that reflects the underlying geometry of the system. For example, a rectangular cell is probably best suited for a solvated DNA system, while a more spherical cell is better for a globular or spherical protein.

The simulations in this dissertation are all performed with periodic boundary conditions in the crystalline environment (crystal simulations allow for better comparison back to experiment when the experimental structure is a crystal structure[24]). The periodic cells and the dimensions used for the systems were those reported by the authors of the experimental crystal structures.

2.2 Analysis of Trajectories

2.2.1 RMSD. For a quantitative comparison of one structure to another (such as a simulation-generated structure to an experimental structure), some sort of similarity measurement is required. A commonly used method is to calculate the root-mean-square deviation (RMSD) between two structures or conformations. This is done by first superimposing the structures on top of one another so as to eliminate deviations caused by translation or rotation and then calculating the RMSD using Eqn. 2.23.

$$RMSD = \sqrt{\frac{1}{N}\sum_{k=1}^{N}(r_k^i - r_k^j)^2} \qquad \text{Eqn. 2.23}$$

where N is the number of atoms in the structure, k is the index over those atoms, and the r_k's are the coordinates of atom k in conformations i and j. The summation in Eqn. 2.23 can include any set of atoms of the structure under consideration; it can include all the atoms in the structure or just a subset, such as the atoms of a ligand bound to a protein. In the analysis of biomolecular simulations, the summation usually includes only the heavy atoms (non-hydrogen atoms) and in the case of proteins and DNA, sometimes it only includes the alpha carbons of the protein backbone or the phosphate groups of the DNA backbone. RMSD is usually in the units of Å.

2.2.2. T-tests. The t-test[25] assess whether the means of two groups are statistically different from each other. The null hypothesis (H_0) of the test is that the means of the populations are equal. There are two forms of the t-test; one form assumes that the population standard deviation is the same for both sets of measurements being compared, and the other form assumes that the population standard deviations are different. In practice, both forms give about the same results (the numerical result of the test will fluctuate, but the overall significance of the test usually will not). In this dissertation, both forms of the t-test were used in order to compare the properties extracted from simulations to the experimental structure that was used as the starting point for the simulations. Eqns. 2.24 and 2.25 show both forms of the t-test.

$$t_{calc} = \frac{\overline{x_1} - \overline{x_2}}{s_{pooled}} \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \qquad \text{Eqn. 2.24}$$

$$t_{calc} = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{(s_1^{\,2} / n_1) + (s_2^{\,2} / n_2)}} \qquad \text{Eqn. 2.25}$$

Eqn. 2.24 is the form of the t-test that assumes that the population standard deviations are the same, and Eqn. 2.25 is the form that assumes they are different. In both equations, *s* is the standard deviation for each population considered, *n* is the number of samples used, and *x* is the mean of each population. In Eqn. 2.24, s_{pooled} is the pooled standard deviation (this makes use of both data sets) is calculated using Eqn. 2.26

$$s_{pooled} = \sqrt{\frac{s_1^{\,2}(n_1 - 1) + s_2^{\,2}(n_2 - 1) + \ldots + s_k^2(n_k - 1)}{n_1 + n_2 + \ldots + n_k - k}} \qquad \text{Eqn. 2.26}$$

where *s* is the standard deviation of each population considered, *n* is the number of samples, and *k* is the number of populations. When used in the t-test, s_{pooled} (Eqn. 2.26)

involves two data sets (the t-test compares the means of two groups at a time, $k = 2$); however, Eqn. 2.26 can be used to pool as many standard deviations as needed.

Once the t_{calc} value is calculated, this number is compared with the t_{table} number at the specified degrees of freedom and confidence level (the 95% confidence level is normally used). If t_{calc} is greater than t_{table} at the 95% confidence level, the results are considered to be statistically significantly different. The number of degrees of freedom is determined using Eqn. 2.27.

$$dof = n_1 + n_2 - 2 \qquad \text{Eqn. 2.27}$$

In this dissertation, the *dof* was ~100 for the properties being analyzed. At the 95% confidence interval for 100 and greater *dof*, the t_{table} value is 1.96, and this will be used throughout the document, unless otherwise noted. The specific details of what properties the t-test was applied to and how it was used (number of samples used, etc.) will be discussed with the simulation results.

2.2.3. Principal Components Analysis. Principal components analysis (PCA)[26-29] is a multivariate analysis technique that is used to identify patterns in data and express the original data in such a way as to highlight similarities and differences. PCA transforms data to a new coordinate system such the greatest variance by any projection of the data comes to lie on the first coordinate (first principal component), the second variance lies on the second coordinate, and so on. PCA can be used to reduce the dimensionality of a data set by retaining those characteristics of the data that contribute most to the variance of the data set. The lower-order principal components are kept, while the higher-order ones are ignored because the lower order ones contribute most to the variance.

The procedure by which PCA works is typically described using an example data set. Suppose that this data set consists of property with *p* measurements over *n* conformations for multiple MD simulations. This data set is matrix **X** with *p* columns and *n* rows. The purpose of performing PCA in this dissertation was to see the relationship of a particular property among the all conformations from different simulations and their relationship to an experimental structure that was used to initialize the simulations (i.e., which conformations among the simulations are similar or different to each other and the experimental structure). PCA plots were also used to select starting structures for additional simulations. The data set is multidimensional, and the whole point of PCA is to reduce dimensionality in order to examine the relationships of the different conformations from the simulations. First, for PCA to work properly, the mean of each dimension of the data set must be subtracted from each dimension (this produces a data set whose mean is zero). Next, a covariance matrix **C** is constructed. Covariance is measured between two dimensions, and it is used to find out how much the dimensions of the data vary from the mean with respect to each other (variance is the analogous measure in one dimension). The covariance between two dimensions *a* and *b*, would be calculated by Eqn. 2.28.

$$\mathrm{cov}(a,b) = \frac{\sum_{i=1}^{n}(a_i - \bar{a})(b_i - \bar{b})}{(n-1)} \qquad \text{Eqn. 2.28}$$

If the data set contains more than two dimensions, obviously more than one covariance measurement can be calculated. For an *n* dimensional data set, **C** will be an *n* x *n* matrix, and each element is the result of calculating the covariance between two separate

dimensions (Eqn. 2.28). The diagonal of **C** will contain the covariance between one of the dimensions and itself, which are just the variances for that dimension. Also, since cov*(a,b)* = cov*(b,a)*, **C** is symmetric about the diagonal.

The next step is to diagonalize **C**. By doing this, a matrix of eigenvectors **E** and their corresponding eigenvalues. For an *n* x *n* covariance matrix, *n* eigenvectors will be obtained. All the eigenvectors of **C** are orthogonal to each other (as with any matrix), and this is important because it means that we will be able to express the original data in terms of these orthogonal eigenvectors rather in terms of the usual x, y, and z axes (these will not show how each data point relates to the rest of the data). The eigenvalues that are associated with each eigenvector indicate the magnitude of the variances in the direction of their corresponding eigenvectors.

Once the eigenvectors and eigenvalues are found from **C**, the next step is to sort them in order by eigenvalue, highest to lowest. The sorting can be visualized by using a Scree plot, which simply shows which components contribute most to the variance of the data set (the percentage that each component contributes to the variance of the data set is plotted versus the component number). Usually, only the first 1 or 2 components will contribute most to the variance, and the lower components can be ignored. Some information will be lost when components are ignored; however, if the eigenvalues are small, not much information is lost, and the dimensionality of the data set is reduced (if the data set has *n* dimensions, and only the first *p* eigenvectors are retained, the final data set will only contain *p* dimensions). Another matrix $\mathbf{E_1}$ is then formed from the chosen eigenvectors.

The final step in PCA is to project the original data set onto the new set given by the chosen eigenvectors. The new components are the principal components (PCs), and while the original data contained *n* dimensions and may have been correlated, the new data set contains only *p* dimensions (if p eigenvectors were retained), and the PCs are uncorrelated. The final data set (which is made up of the PCs) is computed by the following equation,

$$M = E_1^{\mathrm{T}} \times X^{\mathrm{T}} \qquad \text{Eqn. 2.29}$$

where **M** is the final data matrix, $\mathbf{E_1}^{\mathbf{T}}$ is the transposed matrix of the chosen eigenvectors, and $\mathbf{X}^{\mathbf{T}}$ is the transpose of the original data matrix with the means subtracted out. To get the data in a plot-friendly format, it helps to transpose **M**. The original data is now solely in terms of the chosen eigenvectors, and if not all eigenvectors were used, the dimensions of the data have been reduced. The PCs can now be plotted against one another in order to visualize the projections of the data onto the eigenvectors with the largest variances.

In this dissertation, PCA was applied to DNA/polyamide complex data sets. The S-plus program[30] was used to perform PCA.

2.3 Quantum Mechanical Methods

2.3.1 Schrödinger Equation. The time-independent solution to the Schrödinger equation for a molecular system can be expressed by the following equation

$$H\Psi = E\Psi \qquad \text{Eqn. 2.30}$$

where Ψ is the wavefunction, E is the energy of the system, and H is the Hamiltonian operator, which is equal to

$$H = \frac{-h^2}{8\pi^2 m}\nabla^2 + V \qquad \text{Eqn. 2.31}$$

where h is Planck's constant, m is the mass of the particle, ∇ is the Laplacian operator, and V is the potential energy. The Hamiltonian is made up of the sum of the kinetic and potential energy contributions of the nuclei and electrons in the system. The kinetic energy is a summation of ∇^2 over all the particles in the molecule (first part of Eqn. 2.31):

$$KE = \frac{-h^2}{8\pi^2}\sum_k \frac{1}{m_k}\left(\frac{\partial^2}{\partial x_k^2} + \frac{\partial^2}{\partial y_k^2} + \frac{\partial^2}{\partial z_k^2}\right) \qquad \text{Eqn. 2.32}$$

The potential energy component is the Coulomb repulsion between each pair of charged particles:

$$V = \frac{1}{4\pi\varepsilon_0}\sum_j \sum_{k<j} \frac{e_j e_k}{\Delta r_{jk}} \qquad \text{Eqn. 2.33}$$

Δr_{jk} is the distance between the two particles j and k, and e_j and e_k are the charges on the particles. For an electron, the charge is $-e$, and for a nucleus, the charge is Ze (Z is the atomic number of the atom).

Therefore, we can write:

$$V = \frac{1}{4\pi\varepsilon_0}\left(-\sum_i \sum_I \left(\frac{Z_I e^2}{\Delta r_{iI}}\right) + \sum_i \sum_{j<i}\left(\frac{e^2}{\Delta r_{ij}}\right) + \sum_i \sum_{J<I}\left(\frac{Z_I Z_J e^2}{\Delta R_{IJ}}\right)\right) \qquad \text{Eqn. 2.34}$$

where the first term is for the nuclei-electron attraction, the second term is for electron-electron repulsion, and the third term is for nuclear-nuclear repulsion.

Under the Born-Oppenheimer approximation,[31] nuclear and electronic motions can be separated, which simplifies the Schrödinger equation. This approximation is reasonable because the mass of a nucleus is thousands of times greater than that of an electron, and nuclei move very slowly with respect to the electrons (electrons move instantaneously in response to changes in nuclear position). Therefore, the nuclei are fixed relative to the electrons, and the motion of the electrons can be considered to occur in a fixed potential produced by the nuclei.

$$H = KE^{elec}(\vec{r}) + KE^{nuclear}(\vec{R}) + V^{nuclear-elec}(\vec{R},\vec{r}) + V^{elec}(\vec{r}) + V^{nuclear}(\vec{R}) \qquad \text{Eqn. 2.35}$$

Where $\vec{r}$ and $\vec{R}$ designate the positions of the electrons and nuclei, respectively. The Born-Oppenheimer approximation says that the two parts (electronic and nuclear) can be solved independently, so a Hamiltonian that omits the nuclear kinetic energy term can be written:

$$H^{elec} = -\frac{1}{2}\sum_i\left(\frac{\partial^2}{\partial x_i^2}+\frac{\partial^2}{\partial y_i^2}+\frac{\partial^2}{\partial z_i^2}\right) - \sum_i\sum_I\left(\frac{Z_I}{\left|\vec{R}_I-\vec{r}_i\right|}\right) + \sum_i\sum_{j<i}\left(\frac{1}{\left|\vec{r}_i-\vec{r}_j\right|}\right) + \sum_I\sum_{J<I}\left(\frac{Z_I Z_J}{\left|\vec{R}_I-\vec{R}_J\right|}\right) \qquad \text{Eqn. 2.36}$$

Using this electronic Hamiltonian, a Schrödinger equation is obtained for describing the motion and energy of the electrons in a fixed field of nuclei, where *n* is the number of electrons.

$$H^{elec}\Psi^{n}(\vec{r},\vec{R}) = E^{n}(\vec{R})\Psi^{n}(\vec{r},\vec{R}) \qquad \text{Eqn. 2.37}$$

The Schrödinger equation cannot be solved exactly with the exception of one-electron systems (such as the hydrogen atom).[32, 33] The electron-electron repulsion term in Eqn. 2.36 makes it impossible to find an analytical solution to the Schrödinger equation. Therefore, some sort of approximation method, such as perturbation theory, is necessary to obtain a solution.

The product of the wavefunction Ψ with its complex conjugate (Ψ* Ψ) is interpreted the probability of finding the particle at some point in space. The square of Ψ thus gives the electron density at any given point. Therefore, Ψ must be normalized; if the probability of finding the particle is integrated over all space, the result must be 1 (the particle has to be somewhere).

$$\int \Psi^{*}\Psi d\tau = 1 \qquad \text{Eqn. 2.38}$$

$d\tau$ indicates integration over all space. Aside from the normality requirement, the wavefunction has the requirement of being antisymmetric, which means that Ψ changes sign when two identical particles are interchanged (this follows from the fact that electrons are indistinguishable from one another, and there is a fundamental symmetry that the wavefunction must obey in describing the behavior of many electrons).

The solutions (the wavefunctions, Ψ) to the Schrödinger equation (Eqn. 2.37) describe some stationary state of the system (the solutions are called stationary-state wave functions because they are independent of time). The wavefunction depends on both the position of the particle (x, y, z coordinates) and their spin states. Solving Eqn. 2.37 gives Ψ^{n} and E^{n}, where the Ψs are the eigenfunctions of the Hamiltonian operator

and the E^ns are the eigenvalues (the E^ns are the allowed energy values). $E_0{}^n$ is the energy of the ground-state energy of the system, while all the E^ns above that are the energies of the excited states.

2.3.2 Hartree-Fock Theory. Hartree-Fock (HF) is one of the most common approximations for determining the ground-state wavefunction and ground-state energies for a many-electron system.[32-34] There is no correct solution for a many-electron system; there must be some way to determine if one wavefunction is better than another one. Hartree-Fock utilizes the variational principle, which says that for a time-independent Hamiltonian operator, any trial wavefunction ϕ will have an energy expectation value that is greater than or equal to the true ground state wavefunction corresponding to the given Hamiltonian.

$$\int \frac{\phi^* \hat{H} \phi d\tau}{\phi^* \phi d\tau} \geq E_{gs} \qquad \text{Eqn. 2.39}$$

Therefore, the Hartree-Fock energy is an upper bound to the true ground state energy of a given system.

The next approximation made by Hartree-Fock theory is that the trial wavefunctions to be used are approximated by a single Slater determinant. A Slater determinant is the simplest form of an orbital wavefunction that satisfies the asymmetry requirement. Because asymmetry must be considered, electron spin has to be taken into account. Electron spin can be conveniently treated by combining spatial orbitals with the spin functions α and β, resulting in a spin orbital that a function of location (spatial part) and spin (spin part). Eqn. 2.40 is an example of a Slater determinant

$$\Psi = \frac{1}{\sqrt{N!}} \begin{pmatrix} \phi_1(1)\alpha & \phi_1(1)\beta \ldots & \phi_n(1)\beta \\ \phi_1(2)\alpha & \phi_1(2)\beta \ldots & \phi_n(2)\beta \\ \ldots & \ldots & \ldots \\ \phi_1(n)\alpha & \phi_1(n)\beta \ldots & \phi_n(n)\beta \end{pmatrix} \qquad \text{Eqn. 2.40}$$

where $\frac{1}{\sqrt{N!}}$ is a factor that ensures that the wavefunction is normalized, $\phi_i(n)$ is the spatial orbital for the atom *n*, and α and β are the spin functions.

The Hartree-Fock equation for each spatial orbital is

$$F_i\phi_i = \varepsilon_i\phi_i \qquad \text{Eqn. 2.41}$$

where ϕ_i is the wavefunction, ε_i is the energy of the orbital, and F_i is the Fock operator. The Fock operator has the form

$$F_i = -\frac{1}{2}\nabla_i^2 - \sum_{A=1}^{M}\frac{Z_A}{r_{ia}} + v^{HF}(i) \qquad \text{Eqn. 2.42}$$

where $v^{HF}(i)$ is the average potential felt by the *i*th electron due to the other electrons; this average potential is approximation to the third term of Eqn. 2.36.

When a Hartree-Fock calculation is started, neither F_i nor ϕ_i are known. An initial guess is made for the ϕ_i, which in turn allows calculation of F_i. By using the variational method, the ϕ_i and F_i are optimized in order to approximate ground state energy for the system. This iterative procedure is known as the self-consistent field (SCF) method.

The largest drawback of Hartree-Fock theory is that it does not include effects of electron correlation. The electrons are considered to be moving in an average potential of the other electrons; thus, electrons are not specifically influenced by the presence of individual neighboring electrons. However, despite the lack of electron correlation, HF

does allow for quantum calculations to be applied to larger and more realistic systems, because it is less computationally demanding. HF is also a useful starting point for more advanced quantum mechanical methods.

2.3.3. Perturbation Theory. Because Hartree-Fock does not include effects of electron correlation, it will represent many properties incorrectly, such as bond formation/breakage.[34, 35] Methods that go beyond Hartree-Fock to include the effects of electron correlation are needed; these are known as post-Hartree-Fock or post-SCF methods.

One method of treating electron correlation is Møller-Plesset perturbation theory. This method estimates the electron correlation energy by splitting the Hamiltonian operator into two parts. The perturbation Hamiltonian is the difference between the exact and Hartree-Fock Hamiltonians. The true Hamiltonian H is expressed as the sum of a zeroeth order Hamiltonian H_0 (for which a set of molecular orbitals can be obtained and therefore can be solved exactly) and a perturbation (λV), which is assumed to be small in comparison to H_0.

$$H = H_0 + \lambda V \qquad \text{Eqn. 2.43}$$

The eigenfunctions of the true Hamiltonian operator are Ψ_i with corresponding eigenvalues E_i. The eigenfunctions of the zeroeth order Hamiltonian are Ψ_i^0 with corresponding energies E_i^0. The ground state wavefunction and energy are therefore Ψ_0^0 and E_0^0. The parameter λ can vary from 0 to 1; when $\lambda = 0$, the H is equal to the zeroeth order

Hamiltonian, but when $\lambda = 1$, H is equal to its true value. The eigenfunctions Ψ_i and eigenvalues E_i of H can then be expressed as powers of λ:

$$\Psi_i = \Psi_i^0 + \lambda\Psi_i^1 + \lambda^2\Psi_i^2 + \lambda^3\Psi_i^3 + \ldots \qquad \text{Eqn. 2.44}$$

$$E_i = E_i^0 + \lambda E_i^1 + \lambda^2 E_i^2 + \lambda^3 E_i^3 + \ldots \qquad \text{Eqn. 2.45}$$

The perturbed wavefunction and energy are then substituted into the full Schrodinger equation ($H\Psi_i = E_i\Psi_i$) and like powers of λ are equated to give the zeroeth (Eqn. 2.46), first (Eqn. 2.47), and second orders (Eqn. 2.48) of perturbation:

$$H_0\Psi_i^0 = E_i^0\Psi_i^0 \qquad \text{Eqn. 2.46}$$

$$(H_0 - E_i^0)\Psi_i^1 = (E_i^1 - V)\Psi_i^0 \qquad \text{Eqn. 2.47}$$

$$(H_0 - E_i^0)\Psi_i^2 = E_i^2\Psi_i^0 + E_i^1\Psi_i^1 - V\Psi_i^1 \qquad \text{Eqn. 2.48}$$

Solving these equations yields the following for E_i and Ψ_i:

$$E_i = E_i^0 + \lambda H_{ii}^1 + \lambda^2 \sum_{j \neq i} \frac{H_{ij}^1 H_{ji}^1}{E_i^0 - E_j^0} + \ldots \qquad \text{Eqn. 2.49}$$

$$\Psi_i = \Psi_i^0 + \lambda \sum_{j \neq 1} \frac{H_{ji}^1}{E_i^0 - E_j^0}\Psi_j^0 + \ldots \qquad \text{Eqn. 2.50}$$

Where $H_{ij}^1 = \int \Psi_i^0 V \Psi_j^0 d\tau$. The first-order correction to the energy can be calculated by using the unperturbed wavefunction and the perturbed Hamiltonian:

$$E_i^1 = H_{ii}^1 = \int \Psi_i^0 V \Psi_i^0 d\tau \qquad \text{Eqn. 2.51}$$

The higher-order terms are developed similarly; for example, the second-order correction is given by:

$$E_i^2 = \sum_{j \neq i} \frac{\left(\int \Psi_i^0 V \Psi_j^0 d\tau \right)^2}{E_i^0 - E_j^0} \qquad \text{Eqn. 2.52}$$

So far, these are the results of general perturbation theory. In the particular case of Møller-Plesset theory, the Hartree-Fock wavefunction is used as the zeroeth order approximation, and the difference between the true electron-electron repulsions and the averaged ones from HF is treated as the perturbation (the perturbation Hamiltonian is the difference between the exact and Hartree-Fock Hamiltonians). The zeroeth order Hamiltonian is a sum of Fock operators, and the zeroeth order energy is sum of orbital energies.

$$H_0 = \sum_i F_i \qquad \text{Eqn. 2.53}$$

$$E_i^0 = \sum_i E_i \qquad \text{Eqn. 2.54}$$

The first-order energy is given by Eqn. 2.54, and adding E_i^0 to E_i^1 gives the full Hartree-Fock energy. Therefore, to get any treatment of electron correlation, the second-order perturbation must be incorporated. The second-order energy and the first-order correction to the wavefunction are given by:

$$E_i^2 = \sum_j \frac{\left| \int \Psi_i^0 V \Psi_j d\tau \right|^2}{E_i^0 - E_j} \qquad \text{Eqn. 2.55}$$

$$\Psi_i^1 = \sum_j \frac{\int \Psi_i^0 V \Psi_j^0 d\tau}{E_i^0 - E_j} \Psi_j \qquad \text{Eqn. 2.56}$$

where the sum is over all excited states, denoted as Ψ_j and E_j. The E_i^2 numerator is always positive because it is squared, while the denominator is negative because Ψ_i^0 is the ground state and Ψ_j is an excited state. Therefore, the second-order energy correction is always negative or zero. Higher order corrections may be either positive or negative.

The higher orders of Møller-Plesset theory perturbation are denoted as MP*n*, where *n* is the order of perturbation. The MP1 energy is the same as the Hartree-Fock energy; MP2 includes the effects of double excitation and is the most practical treatment for electron correlation.[34, 35] Third, fourth, and higher orders of perturbation are derived similarly to second-order. The terms quickly become more mathematically complicated and computationally expensive. For *n* basis functions, HF scales as n^4 and MP2 scales as n^5; for higher orders of perturbation, the calculations become more costly. MP2 calculations require more computational effort; however, MP2 yields a significant correction to geometries and energies over the corresponding HF calculations.[34, 35]

2.3.4. Density Functional Theory. The central idea in density functional theory (DFT) is that a relationship exists between the total electronic energy and the overall electron density. DFT methods are based on the ideas of Hohenberg and Kohn,[36] who stated that the ground state electronic energy of a molecule can be expressed exactly as a functional of the electron density of the molecule. The term functional refers to a function of a function, which in this case means that the total energy has a functional dependence on electron density, which in turn is dependent on the coordinates of the electrons of the system. The approximate functionals used in DFT methods use the following separation of the total electronic energy.

$$E = E^T + E^V + E^J + E^{XC} \qquad \text{Eqn. 2.57}$$

E^T is the kinetic energy term arising from electron motion, E^V is the potential energy term and includes nuclear-electron attraction and nuclear-nuclear repulsion, E^J is electron-electron repulsion term (also described as the Coulomb self-interaction of the electron density), and E^{XC} is the exchange-correlation term and includes the contributions due to electron exchange and correlation. All terms of Eqn. 2.57 with the exception the nuclear-nuclear repulsion are functions of the electron density, ρ.

Eqn. 2.57 follows the ideas of Kohn and Sham,[37] who suggested that the sum of the kinetic energy of the electrons and the contributions from electronic interactions should be expressed as a sum of three terms: the kinetic energy (which is defined as the kinetic energy of a system with non-interacting electrons), the electron-electron repulsion energy, and the electron exchange-correlation energy. In practice, Kohn-Sham DFT calculations are performed using an iterative approach, analogous to the SCF approach used in HF calculations. An initial guess is made for ρ, which allows for derivation of a set of orbitals that leads to an improved value for ρ. The improved ρ value is then used in the second iteration and so on until convergence is reached.[33]

The exchange term in Eqn. 2.57 represents the effects of electron exchange and correlation on the total energy of the system, which is an advantage over HF, which does not treat effects of electron correlation. However, there is no known exact expression for E^{XC} (Hohenberg and Kohn showed that E^{XC} is dependent entirely on ρ; however, the theorem does not provide the form of this functional).[33, 36, 37] A number of approximate expressions for E^{XC} have been developed, which in turn lead to a variety of methods. In

practice, E^{XC} is usually divided into separate parts, which are the exchange and correlation parts.

$$E^{XC}(\rho) = E^{X}(\rho) + E^{C}(\rho) \qquad \text{Eqn. 2.58}$$

All terms are again functionals of ρ. The functionals on the right side of the equation are the exchange and correlation functionals, respectively. Both parts can be one of two types, a local functional (depends only on ρ) or a gradient-corrected functional (depends on ρ and the gradient of $\rho, \nabla\rho$).

The local exchange functional is almost always defined as

$$E_{LDA}{}^{X} = -\frac{3}{2}\left(\frac{3}{4\pi}\right)^{1/3} \int \rho^{4/3} d^{3}\vec{r} \qquad \text{Eqn. 2.59}$$

where $\vec{r}$ represents the coordinates of the electrons. This is known as the local density approximation (LDA), and it states that in regions of the system where charge density varies slowly, the exchange correlation energy can be considered the same as that of a uniform electron gas of the same charge density.[35] The LDA approximation is generally not of high enough accuracy to be useful for determining structural properties or dissociation energies of molecules.[33, 35] However, the quality of the results can be improved by adding correction terms to Eqn. 2.59 that depend on the gradient of ρ.

The functional developed by Becke in 1988[38, 39] makes a correction to the LDA approximation that improves upon many of its deficiencies. The following gradient-corrected exchange functional is Becke's form, which is now commonly used,

$$E_{Becke}{}^{X} = E_{LDA}{}^{X} - \gamma \int \frac{\rho^{4/3} x^{2}}{(1 + 6\gamma \sinh^{-1} x)} d^{3}\vec{r} \qquad \text{Eqn. 2.60}$$

where $x^2 = \rho^{-4/3}|\nabla\rho|$, and γ is a parameter chosen to fit the known exchange energies of inert gas atoms (Becke defines γ = 0.0042 Hartrees).[38]

As for the exchange functional, corrections (local and gradient) exist for the correlation functional. A commonly used method is that of Lee, Yang, and Parr,[40]

$$E_{LYP}^{C} = -a\int \frac{1}{1+d\rho^{-1/3}}\{r + b\rho^{-2/3}[c_F\rho^{5/3} - 2t_W + \left(\frac{1}{9}t_W + \frac{1}{18}\nabla^2\rho\right)e^{-cr^{-1/3}}]\}d\vec{r} \qquad \text{Eqn. 2.61}$$

where *a*, *b*, *c*, and *d* are constants and $t_W(\vec{r}) = \sum_{i=1}^{N}\frac{|\nabla\rho_i(\vec{r})|^2}{\rho_i(\vec{r})} - \frac{1}{8}\nabla^2\rho$. Usually, DFT methods are formed by combining an exchange functional with a correlation functional. For example, the BLYP functional is formed by combining Becke's gradient-corrected exchange functional with Lee, Yang, and Parr's gradient-corrected correlation functional.

The key feature of DFT methods is the way in which electron exchange and correlation effects are directly incorporated; correlation effects are truly only considered in more complex, post-HF methods, such as MP2. Because DFT methods include both exchange and correlation effects, higher accuracy is achieved compared to Hartree-Fock, which does not include electron correlation effects. Despite the improvements in DFT in recent years, there are still difficulties in using density functional theory to describe properly certain intermolecular interactions, in particular, van der Waals (dispersion) forces. The poor treatment of dispersion by DFT renders this method unsuitable for the treatment of systems that are dominated by dispersion forces, such as noble gases.[33]

2.3.5. Basis Sets. A basis set is a set of functions used to create the molecular orbitals of a molecule. The functions that make up the basis set are known as basis

functions, and they are expanded as linear combinations to make up the basis set.[33] The basis functions are usually centered on the atomic nuclei, so they represent atomic orbitals. The accuracy of quantum mechanical calculations depends largely on the basis set chosen. The basis set specifies which and how many basis functions are used to describe a molecule.

An individual molecular orbital can be defined as

$$\phi_i = \sum_{\mu=1}^{N} c_{\mu i} \chi_\mu \qquad \text{Eqn. 2.62}$$

where the coefficients $c_{\mu i}$ are known as the molecular orbital expansion coefficients, which are determined computationally and χ_μ are the basis functions (these are chosen to be normalized). Gaussian03,[41] which was used in this dissertation for quantum mechanical calculations, uses Gaussian-type atomic functions as basis functions. Gaussian functions have the general form

$$g(\alpha, \vec{r}) = cx^n y^m z^l e^{-\alpha r^2} \qquad \text{Eqn. 2.63}$$

where $\vec{r}$ is made up of x, y, and z, and α is a constant that determines the size, or spread, of the function (a large α value gives a function that does not spread very far; a small α value will give a large spread). In a Gaussian function, $e^{-\alpha r^2}$ is multiplied by powers of x, y, z, and normalization constant, so that when the square of the Gaussian function is integrated over all space, the result is 1. The parameter c depends on α, l, m, and n. Linear combinations of primitive Gaussian functions of the form in Eqn. 2.63 are used to form the basis functions. These are known as contracted Gaussians and have the form

$$\chi_\mu = \sum_p d_{\mu p} g_p \qquad \text{Eqn. 2.64}$$

where $d_{\mu p}$ are fixed constants within a given basis set. The basis functions are combined to yield the linear expansion that forms the molecular orbital (Eqn. 2.65).

$$\phi_i = \sum_\mu c_{\mu i} \chi_\mu = \sum_\mu c_{\mu i} \left(\sum_p d_{\mu p} g_p \right) \qquad \text{Eqn. 2.65}$$

In this dissertation, the 6-31G(d) basis set was used in all quantum mechanical calculations. This is a combination of split-valence and polarized basis sets. Valence electrons principally take part in chemical bonding; therefore, valence orbitals are represented by more than one basis function, which in turn are composed of a linear combination of Gaussian primitive functions. Split-valence basis sets have two or more sizes of basis function for each valence orbital. Split-valence basis sets allow orbitals to change size but not shape. Polarized basis sets overcome this by adding orbitals to each atom with angular momentum beyond what is required for the ground state (e.g., d functions are added to carbon, p functions can be added to hydrogen). In the 6-31G(d) basis set, one basis function comprised of six Gaussian primitives is applied to the core atomic orbitals, and the valence orbitals are split into two basis functions. The first one is composed of three Gaussian primitives and the second by one Gaussian primitive. In this specific basis set, d orbitals are added to heavy atoms.

2.4 Description of Calculations

2.4.1. General Methods. Calculations were conducted using the computational resources at the Center for Computational Sciences[42, 43] at Duquesne University and the

Pittsburgh Supercomputing Center. Electronic structure calculations were carried out with the Gaussian03 program[41] with the second-order Møller-Plesset (MP2) method[44] and the 6-31G(d) basis set.[45, 46] Molecular mechanics calculations were performed using the CHARMM program[47] and the CHARMM force field for proteins[8, 48] and nucleic acids.[6, 7] The water model in all simulations was TIP3P.[49] Crystal simulations employed the CRYSTAL[50] facility in CHARMM. Molecular dynamics simulations were performed in the NPT[51] ensemble using the Leap-Frog integrator and periodic boundary conditions with the SHAKE[9] algorithm applied to covalent hydrogen bonds, which allowed for a 0.002 ps timestep. A dielectric constant of 1.0 was used in all simulations. Electrostatics were treated using the particle-mesh Ewald method.[16] In the case of net-charged systems, a neutralizing background is used by default in the standard Ewald calculation.[21] Ewald calculations were performed using the specified real space cutoffs with the Lennard-Jones interactions truncated at the same distance. Kappa (screening parameter) was set to 0.36; the order parameter was set to 6.[6] Nonbond pair lists were maintained to 14 Å, nonbond interactions were truncated at 12 Å, and the image cutoff was kept at 14.0 Å. Nonbond lists, hydrogen bond lists, and image lists were updated heuristically. The volumes of the systems were monitored to ensure that there were only minimal fluctuations in the volume over time. Snapshots of the systems were saved every ten picoseconds. Simulation details for each specific system are given in the following sections

2.4.2. Quantum Mechanical Calculations (polyamide model compounds). Each compound was energy minimized using MP2/6-31G(d). For comparison, the structures were also energy minimized using HF and B3LYP with the 6-31G(d) basis set.

Frequency calculations were also performed at the MP2/6-31G(d) level of theory. In order to obtain the potential energy surfaces (PES) for selected torsion angles, torsion energies were calculated by holding the selected torsion angle fixed at different increments (30º increments from –180º to 180º) and geometry optimizing all other degrees of freedom at the three levels of theory mentioned above.

2.4.3. Netropsin Crystal Simulation. The starting structure for the simulation was an x-ray structure solved by Berman, et al.[52] The crystal was of the C2/c space group (monoclinic crystal; unit cell dimensions are a = 21.7 Å, b = 6.37 Å, c = 42.7 Å; $\alpha = \gamma =$ 90.0º, β = 107.0º). The netropsin molecule was built using the CHARMM program,[47] and the HBUILD algorithm[53] was used to add hydrogens to netropsin. The structure also contained five crystallographic waters and one sulfate ion, which were retained for the simulation. No additional water or ions were added to the system. The system was then energy minimized using the steepest descent algorithm (SD) for 500 steps and for 100 steps using the ABNR algorithm to an energy convergence of 0.001 kcal/mol. Dynamics were then performed for 5 ns.

2.4.4. DNA with Bound Polyamides. The initial structure for the simulations was the x-ray crystal structure PDB[3] id 1CVY.[54] The crystal contained a DNA decamer (5'-CCAGATCTGG-3') with two polyamides bound antiparallel in the minor groove (polyamide sequence was ImPyPyPy-β-Dp). The polyamides recognize the six base pair target site 5'-AGATCT-3'. In addition to the two DNA strands and two polyamides, the structure also contained 52 crystallographic waters. The structure was solved to 2.15 Å resolution, and the crystal is of the $P2_12_12_1$ space group (orthorhombic crystal; unit cell dimensions are a = 36.4 Å, b = 38.7 Å, c = 47.2 Å; $\alpha = \beta = \gamma =$ 90.0º). The DNA-

polyamide complex was generated in CHARMM using the revised nucleic acids parameters.[6, 7] The HBUILD algorithm[53] was used to add hydrogens to the DNA and polyamide molecules, and the hydrogens were then energy minimized for 50 steps using the steepest descent (SD) algorithm. The 52 crystallographic waters were also added to the system. The simulation was performed on the asymmetric unit of the crystal; this contained all the atoms of the DNA duplex and the two polyamides. Water was added to fill the vacuum spaces in the crystal by generating the primary and image atoms of the DNA, polyamide, and crystallographic waters and overlaying them with a waterbox that was larger than the dimensions of the asymmetric unit (waterbox dimensions were 59.9 Å x 59.9 Å x 63.8 Å). The waterbox was centered over the primary atoms, and any water molecules within 2.5 Å of the DNA or polyamides were deleted. The waterbox was then overlayed three times with the original atoms and rotated 90º each time in order to obtain the correct density for the crystal (the first time the waterbox was rotated 90º degrees about the x-axis, the second time the waterbox was rotated 90º about the y-axis, and the third time the waterbox was rotated 90º about the z-axis). Any overlapping waters with the original atoms were deleted. An image update and image centering was then performed in order to delete the appropriate amount of waters for the size and symmetry of the unit cell. The waters that moved after the image update and centering were deleted. No ions were identified or reported in the published x-ray crystal structure, and none were added. The water, DNA, and polyamide were energy minimized for 500 steps of steepest descent to an energy convergence of 0.001 kcal/mol. The water was then equilibrated for 20 ps using the NVT ensemble with the DNA and polyamide constrained (RMSD of waters converged to ~5.6 Å). Following relaxation of water, the constraints

on the DNA and polyamide were released, and all atoms were allowed to relax for 20 ps in the NVT ensemble (RMSD of whole complex and waters converged to ~5.8 Å; average RMSD of DNA/polyamide complex was 0.65 Å). All atoms were then subjected to dynamics for 10 ns at 300 K. Many studies involving DNA crystal simulations are performed at 300 K.[7, 19, 24, 55, 56] Two more simulations were also run at 300 K; however, the systems were perturbed by rotating the waterbox –90º about the y-axis (rotated 90º about the x- and z-axes) and –90º about the z-axis (rotated 90º about the x- and y-axes) during the reoverlaying procedure (perturbing the systems in this way gave different initial coordinates for the solvent atoms and thus allowed for better sampling of conformational space[57-61]). The systems were prepared in the same way as described above, and data was then collected over 10 ns in the NPT ensemble. Because the x-ray crystal data was collected at –160ºC (113 K), four simulations were run at 113 K. The systems were set up and run in the same way as the simulations run at 300 K (rotations of the waterbox, equilibration, and data collection were done in the same way as the 300 K crystal simulations).

In order to improve sampling of conformational space, four more simulations were run at 113 K and three more simulations were run at 300 K. Principal component analysis (PCA) was performed, and points (structures) that were close to the x-ray value on the PCA plots were used as the starting structures for the additional simulations. In order to perturb the systems, each simulation was started from different initial velocities.

A simulation was also performed with a hydroxypyrrole-containing polyamide. The initial structure for the simulations was the x-ray crystal structure PDB[3] id 1CVX.[54] The crystal contained the DNA decamer (5'-CCAGATCTGG-3') with two polyamides

bound antiparallel in the minor groove (polyamide sequence was ImPyHpPy-β-Dp). As with the other simulated Dervan polyamides, the polyamides recognize the six base pair target site 5'-AGATCT-3'. In addition to the two DNA strands and two polyamides, the structure also contained 30 crystallographic waters, which were retained for the simulation. The structure was solved to 2.27 Å resolution, and the crystal is of the $P3_1$ space group (trigonal crystal; unit cell dimensions are a = 31.2 Å, b = 31.2 Å, c = 46.1 Å; $\alpha = \beta = 90.0^{\circ}$, $\gamma = 120.0°$). The system was set up and generated similarly to the other DNA/polyamide complex simulations, and the simulation was run for 10 ns at 103 K.

2.4.5. DNA/Netropsin Complex. The initial structure for the simulation was the x-ray crystal structure PDB[3] id 121D.[62] The crystal contained a DNA dodecamer (5'-CGCAAATTTGCG-3') with the drug netropsin in the minor groove. Netropsin has a binding preference for AT sequences over GC sequences, and in this structure, netropsin covers five of the six AT base pairs of the DNA fragment. In addition to the two DNA strands and netropsin, the structure also contained 47 crystallographic waters, which were retained for the simulation. The structure was solved to 2.2 Å resolution, and the crystal is of the $P2_12_12_1$ space group (orthorhombic crystal; unit cell dimensions are a = 25.65 Å, b = 42.03 Å, c = 65.33 Å; $\alpha = \beta = \gamma = 90.0^{\circ}$). The system was set up and generated similarly to the DNA/polyamide complex simulations, and the simulation was run for 10 ns at 300 K.

2.4.6. DNA/Distamycin Complex. The initial structure for the simulation was the x-ray crystal structure PDB[3] id 121D.[63] The crystal contained a DNA dodecamer (5'-CGCAAATTTGCG-3') with the drug netropsin in the minor groove. Like netropsin, distamycin has a binding preference for AT sequences over GC sequences, and in this

structure, distamycin covers five of the six AT base pairs of the DNA fragment. In addition to the two DNA strands and distamycin, the structure also contained 75 crystallographic waters, which were retained for the simulation. The structure was solved to 2.2 Å resolution, and the crystal is of the $P2_12_12_1$ space group (orthorhombic crystal; unit cell dimensions are a = 25.20 Å, b = 41.07 Å, c = 64.65 Å; $\alpha = \beta = \gamma = 90.0^{o}$). The system was set up and generated similarly to the DNA/polyamide and DNA/netropsin complex simulations, and the simulation was run for 10 ns at 288 K.

2.4.7. Netropsin and DNA/Complex Simulations Using AMBER parameters. Two tests were performed using previously described parameters intended for use with the AMBER[64] force field.[65, 66] The first test was a simulation of a netropsin crystal; the starting structure was the same used for the netropsin crystal simulation with the new CHARMM parameters, which is that by Berman, et al.[52] The simulation was prepared and run for 5 ns in the same way as the netropsin simulation using the CHARMM parameters. The second test was a crystal simulation of a DNA/netropsin complex. The starting structure for the simulation was the same as that used for the new CHARMM parameters (PDB[3] id 121D[62]). The system was set up and run identically to the CHARMM DNA/netropsin simulation. The system contained the same amount of water and was started from the same initial velocities as the CHARMM simulation and run for 10 ns at 300 K.

2.4.8. Quantum Mechanical Calculations (retinoid model compounds). Optimization and frequency calculations were carried out on the model compounds at the MP2/6-31G(d) level of theory. In order to obtain the potential energy surfaces (PES) for selected torsion angles, torsion energies were calculated by holding the selected torsion

angle fixed at different increments (20º increments from –180º to 180º) and geometry optimizing all other degrees of freedom at the three levels of theory mentioned above.

2.4.9. Small Retinoid Crystal Simulations. For the retinoid crystal simulations, the unit cell was built using the CHARMM program.[47] No water or ions were reported in the x-ray structures, and none were added. The systems were energy minimized using the steepest descent algorithm (SD) for 500 steps and for 1000 steps using the ABNR algorithm to an energy convergence of 0.001 kcal/mol. All simulations were carried out at the temperature in which the crystal structure was determined, unless noted otherwise.

2.4.9.1. All-Trans Retinal Crystal. The initial structure for the simulation was the x-ray crystal structure CSD[67] id TRETAL01[68] (space group $P2_1$ (monoclinic), unit cell dimensions were a = 15.27 Å, b = 8.26 Å, c = 14.94 Å; $\alpha = \gamma = 90.0°$, $\beta = 104.7°$; four retinal molecules were reported for the unit cell, which was what was used for the simulation). Five molecular dynamics simulations started from different velocities were then performed for 1 ns at 273 K.

2.4.9.2. All-Trans Retinoic Acid Crystal (Monoclinic). The initial structure for the simulation was the x-ray crystal structure CSD[67] id VITAAC01[69] (space group P21 (monoclinic), unit cell dimensions were a = 8.078 Å, b = 34.103 Å, c = 7.387 Å; $\alpha = \beta = 90.0°$, $\gamma = 118.76°$). The retinoic acid molecule parameterized for the force field is the deprotonated form; however, the crystal structure of the ligand contained the protonated form. A patch analogous to the existing patches for neutral aspartic acid or neutral glutamic acid was made in order to protonate the retinoic acid. Five molecular dynamics simulations started from different velocities were then performed for 1 ns at 393 K.

2.4.9.3. All-Trans Retinoic Acid Crystal (Triclinic). The initial structure for the simulation was the x-ray crystal structure CSD[67] id TRETAL10[70] (space group P1' (triclinic), unit cell dimensions were a = 8.04 Å, b = 28.49 Å, c = 5.996 Å; α = 50.58 °,β = 71.38°, γ = 95.7°; two molecules of retinoic acid were found in the unit cell, which was what was used for the simulation). Five molecular dynamics simulations started from different velocities were then performed for 1 ns at 123 K.

2.4.10. Protein/Retinoid Complex Crystal Simulations.

2.4.10.1. Retinoic Acid Complexed with cellular retinoic acid binding protein II (CRABP II). The initial structure for the simulation was the x-ray crystal structure PDB[3] id 1CBS[71] (resolution 1.80 Å; space group $P2_12_12_1$ (orthorhombic), unit cell dimensions were a = 45.65 Å, b = 47.56 Å, c = 77.61 Å; $\alpha = \beta = \gamma = 90.0^0$). The retinoid-protein complex was generated in CHARMM using the revised protein parameters,[48] the HBUILD algorithm[53] in CHARMM was used to add hydrogens to the protein and ligand molecules, and the hydrogens were then energy minimized for 50 steps using the steepest descent (SD) algorithm (energy tolerance was 0.001 kcal/mol). The 100 crystallographic waters present in the x-ray structure were added, and three sodium ions were added to neutralize the system (no ions were reported in the crystal structure). The sodium ions were placed randomly, and they were all placed at least 10 Å away from the protein to avoid bad contacts. The crystal was then built using the CRYSTAL[50] facility in CHARMM. Water was added to fill the vacuum spaces in the crystal by generating the primary and image atoms of the protein, ligand, and crystallographic waters and overlaying them with a waterbox that was larger than the dimensions of the asymmetric unit (waterbox dimensions were 59.9 Å × 59.9 Å × 63.8 Å). The waterbox was centered

over the primary atoms, and any water molecules within 2.5 Å of the protein or ligand were deleted. The waterbox was then overlayed three times with the original atoms and rotated 90º each time in order to obtain the correct density for the crystal (the first time the waterbox was rotated 90º degrees about the x-axis, the second time the waterbox was rotated 90º about the y-axis, and the third time the waterbox was rotated 90º about the z-axis). Any overlapping waters with the original atoms were deleted (overlapping atoms were any atoms within 2.8 Å of the original atoms). An image update and image centering was then performed in order to delete the appropriate amount of waters for the size and symmetry of the unit cell. The waters that moved after the image update and centering were deleted. The water, ions, protein, and ligand were energy minimized for 500 steps of steepest descent to an energy convergence of 0.001 kcal/mol. The solvent was then equilibrated for 20 ps using the NVT ensemble with the protein and retinoic acid constrained. Following relaxation of water, the constraints on the protein/ligand complex were released, and all atoms were allowed to relax for 20 ps in the NVT ensemble (RMSD of whole complex and solvent converged to ~14 Å; average RMSD of protein/retinoic acid complex after equilibration was 0.89 Å). All atoms were then subjected to dynamics for 2 ns at 277 K.

2.4.10.2. Retinol Complexed with Cellular Retinol Binding Protein (CRBP) from Zebrafish. The initial structure for the simulation was the x-ray crystal structure PDB[3] id 1KQW[72] (resolution 1.38 Å; space group I4 (tetragonal), unit cell dimensions were a = 89.02 Å, b = 89.02 Å, c = 38.27 Å; $\alpha = \beta = \gamma = 90.0^0$). The retinoid-protein complex was generated in CHARMM in the same way as the retinoic acid/CRABP II system. The crystal structure was determined at pH 5.60, so the His residues were protonated for the

simulation. The 194 crystallographic waters present in the x-ray structure were added, and three chlorine ions were added to neutralize the system (no ions were reported in the crystal structure). The chloride ions were placed randomly and at least 10 Å away from the protein to avoid bad contacts. The crystal was then built, and water was added to fill the vacuum spaces in the crystal in the same way as for the retinoic acid/CRABP II system. Equilibration and dynamics were performed on the retinol/CRBP complex in the same way as for the retinoic acid/CRABP II complex. All atoms were then subjected to dynamics for 2 ns at 100 K.

2.4.10.3. Fenretinide Complexed with retinol binding protein (RBP). The initial structure for the simulation was the x-ray crystal structure PDB[3] id 1FEL[73] (resolution 1.80 Å; space group $P2_12_12_1$ (orthorhombic), unit cell dimensions were a = 46.63 Å, b = 49.26 Å, c = 76.64 Å; $\alpha = \beta = \gamma = 90.0^0$). The last 8 residues (176-183) are unresolved for this x-ray structure, so the missing residues were generated with homology modeling using the MOE homology program[74, 75] based on a segment matching procedure.[76] The residues generated with the homology modeling were then energy minimized with steepest descent to an energy convergence of 0.001 kcal/mol, while holding the rest of the protein/ligand complex fixed. The retinoid-protein complex was then generated in CHARMM in the same way as the retinoic acid/CRABP II and retinol/CRBP systems. Bovine RBP contains three disulfide bridges. The crystal structure was determined at pH 5.0-5.3, so the His residues were protonated for the simulation. The 163 crystallographic waters present in the x-ray structure were added, and one sodium ion was added to neutralize the system, which was placed randomly (no ions were reported in the crystal structure). The crystal was then built, and water was added to fill the vacuum

spaces in the crystal in the same way as for the retinoic acid/CRABP II and the retinol/CRBP systems. Equilibration and dynamics were performed on the fenretinide/RBP complex in the same way as for the other systems. All atoms were then subjected to dynamics for 3 ns at 275 K.

2.4.10.4. Retinal Complexed with Sensory Rhodopsin II (SRII). The initial structure for the simulation was the x-ray crystal structure PDB[3] id 1H68[77] (resolution 2.10 Å; space group $C222_1$ (orthorhombic), unit cell dimensions were a = 84.83 Å, b = 128.74 Å, c = 50.74 Å; $\alpha = \beta = \gamma = 90.0^0$). The asymmetric unit of the x-ray structure contains one chain, and this is how SRII is thought to exist biologically; however, the oligomeric state of SRII is unknown. The first residue and the last 20 residues (220-239) are unresolved for this x-ray structure, so the missing residues were generated with homology modeling using the MOE homology program.[74, 75] The residues generated with the homology modeling were then energy minimized with steepest descent to an energy convergence of 0.001 kcal/mol while holding the rest of the protein/ligand complex fixed. The retinoid-protein complex was then generated in CHARMM in the same way as the other protein/ligand systems. The crystal structure was determined at pH 4.6 so the His residues were protonated for the simulation. The 27 crystallographic waters and one chloride ion present in the x-ray structure were added, and three sodium ions were added to neutralize the system. The sodium ions were randomly placed at least 10 Å away from the protein to avoid bad contacts. The crystal was then built, and water was added to fill the vacuum spaces in the crystal in the same way as for the other protein/ligand systems. Equilibration and dynamics were performed on the retinal/SRII complex similar to the other systems. The solvent was equilibrated for 30 ps in the NVT

ensemble, and the whole system was then equilibrated for 30 ps in the NVT ensemble (average RMSD of retinal/SRII complex after equilibration was 1.67 Å). All atoms were then subjected to dynamics for 2 ns at 100 K.

2.4.10.5. Retinal Complexed with Bacteriorhodopsin (BR). The initial structure for the simulation was the x-ray crystal structure PDB[3] id 1PXR[78] (resolution 1.70 Å; space group $P12_11$ (monoclinic), unit cell dimensions were a = 44.97 Å, b = 108.84 Å, c = 55.98 Å; $\alpha = \gamma = 90.0^0$, $\beta = 113.60^o$). The asymmetric unit of this x-ray structure contains two chains; however, native BR exists in the purple membrane of *Halobacterium salinarum* as a trimeric structure.[79] Also, this structure is a P50A mutant. The first five residues and the last 18 residues (232-249) are unresolved in both chains for this x-ray structure, so the missing residues were generated with homology modeling using the MOE homology program.[74, 75] The residues generated with the homology modeling were then energy minimized with steepest descent to an energy convergence of 0.001 kcal/mol, while holding the rest of the protein/ligand complex fixed. The retinoid-protein complex was then generated in CHARMM in the same way as the other protein/ligand systems. The crystal structure was determined at pH 3.5 so the His residues were protonated for the simulation. Only solvent-exposed Glu and Asp residues were protonated. The 124 crystallographic waters present in the x-ray structure were added, and ten chloride ions were added to neutralize the system. The ions were randomly placed ~10 Å away from the protein to avoid any bad contacts. The crystal was then built, and water was added to fill the vacuum spaces in the crystal in the same way as for the other protein/ligand systems. Equilibration and dynamics were performed on the retinal/BR complex similarly to the other systems. The solvent was equilibrated

for 30 ps in the NVT ensemble, and the whole system was then equilibrated for 30 ps in the NVT ensemble. All atoms were then subjected to dynamics for 3.5 ns at 110 K.

2.4.11. Crambin Simulations (Convergence Analysis). Molecular dynamics simulations were performed using the CHARMM program[47] and the CHARMM force field.[6] Calculations were conducted using the computational resources at the Pittsburgh Supercomputing Center. The water model in all simulations was TIP3P.[49] Crystal simulations employed the CRYSTAL[50] facility in CHARMM. Molecular dynamics simulations were performed in the NPT[51] ensemble using the Leap-Frog integrator and periodic boundary conditions with the SHAKE[9] algorithm applied to covalent hydrogen bonds, which allowed for a 0.002 ps timestep. A dielectric constant of 1.0 was used in all simulations. Electrostatics were treated using the particle-mesh Ewald method.[16] Ewald calculations were performed using the specified real space cutoffs with the Lennard-Jones interactions truncated at the same distance. Kappa (screening parameter) was set to 0.36; the order parameter was set to 6. Nonbond pair lists were maintained to 14 Å, nonbond interactions were truncated at 12 Å, and the image cutoff was kept at 14.0 Å. Nonbond lists, hydrogen bond lists, and image lists were updated heuristically. The volumes of the systems were monitored to ensure that there were only minimal fluctuations in the volume over time. Snapshots of the systems were saved every ten picoseconds.

2.4.11.1. Long-time protein simulations. The initial protein (crambin) structure for all simulations was PDB[3] id 1EJG.[80] The crystal contained only the 46-residue protein crambin; no water or ions were reported in the structure. The structure was solved to 0.54 Å resolution, and the crystal was of the $P2_1$ space group (monoclinic crystal; unit cell

dimensions were a = 40.82 Å, b = 18.50 Å, c = 22.37 Å; $\alpha = \gamma = 90.0^{o}$, $\beta = 90.47^{o}$). The protein was generated in CHARMM using the revised protein backbone parameters.[48] The HBUILD algorithm[53] was used to add hydrogens to the protein molecule, and the hydrogens were then energy minimized for 50 steps using the steepest descent (SD) algorithm. The simulation was performed on the asymmetric unit of the crystal, which contained only the protein. Water was added to fill the vacuum spaces in the crystal in the same way as was done for the DNA. No ions were added to the system, and the final system had a charge of 0. The protein and water were energy minimized for 1000 steps of steepest descent to an energy convergence of 0.001 kcal/mol. The water was then equilibrated for 20 ps using the NVT ensemble with the protein constrained. Following relaxation of water, the constraints on the protein were released, and all atoms were allowed to relax for 20 ps in the NVT ensemble. All atoms were then subjected to dynamics for 45 ns at 100 K, which was the temperature at which the x-ray crystal structure was determined. Nine more simulations were also run at 100 K; however, the systems were perturbed by assigning different initial velocities to the atoms in each simulation. The systems were prepared in the same way as described above, and data was then collected over 45 ns in the NPT ensemble. Snapshots were saved every 10 ps. The total simulation time over all ten simulations was 450 ns.

2.4.11.2. Short-time protein simulations. The systems were set up in the same way as for the long-time simulations (same starting coordinates and protocol). Each simulation was started from different initial velocities in order to introduce a perturbation. Snapshots were collected every 10 ps. Ten 2 ns simulation were run. The first 2 ns of the long-time simulations were used along with the 10 short simulations for

data analysis of short-time simulations to give a total of 20 short simulations (total simulation time was 40 ns).

2.4.12. DNA Simulations (Convergence Analysis). Molecular dynamics simulations were performed using the CHARMM program[47] and the CHARMM force field.[6] Calculations were conducted using the computational resources at the Pittsburgh Supercomputing Center. The water model in all simulations was TIP3P.[49] Crystal simulations employed the CRYSTAL[50] facility in CHARMM. Molecular dynamics simulations were performed in the NPT[51] ensemble using the Leap-Frog integrator and periodic boundary conditions with the SHAKE[9] algorithm applied to covalent hydrogen bonds, which allowed for a 0.002 ps timestep. A dielectric constant of 1.0 was used in all simulations. Electrostatics were treated using the particle-mesh Ewald method.[16] In the case of charged systems, electrostatics were treated using the modified particle mesh Ewald method for net-charged systems.[21] This modified Ewald summation is implemented in versions 25 and higher of the CHARMM program. Ewald calculations were performed using the specified real space cutoffs with the Lennard-Jones interactions truncated at the same distance. Kappa (screening parameter) was set to 0.36; the order parameter was set to 6. Nonbond pair lists were maintained to 14 Å, nonbond interactions were truncated at 12 Å, and the image cutoff was kept at 14.0 Å. Nonbond lists, hydrogen bond lists, and image lists were updated heuristically. The volumes of the systems were monitored to ensure that there were only minimal fluctuations in the volume over time. Snapshots of the systems were saved every 10 ps.

2.4.12.1. Long-time DNA Simulations. The initial DNA structure for all simulations was the x-ray crystal structure PDB[3] id 119D.[81] The crystal contained a DNA

dodecamer (5'-CGTAGATCTACG-3') along with 1 Mg^{+2} ion, which was retained for the simulations, and 137 crystallographic waters, which were also retained. The structure was solved to 2.25 Å resolution, and the crystal was of the C2 space group (monoclinic crystal; unit cell dimensions were a = 64.83 Å, b = 35.36 Å, c = 25.35 Å; $\alpha = \gamma = 90.0^{o}$, $\beta = 92.24^{o}$). The DNA dodecamer was generated in CHARMM using the revised nucleic acids parameters.[6, 7] The HBUILD algorithm[53] was used to add hydrogens to the DNA molecule, and the hydrogens were then energy minimized for 50 steps using the steepest descent (SD) algorithm. The 137 crystallographic waters were then added to the system. The simulation was performed on the asymmetric unit of the crystal; this contained all the atoms of the DNA duplex and the one Mg^{2+} ion. Water was added to fill the vacuum spaces in the crystal by generating the primary and image atoms of the DNA, Mg^{2+} ion, and crystallographic waters and then overlaying them with a waterbox that was larger than the dimensions of the asymmetric unit (waterbox dimensions were 59.9 Å x 59.9 Å x 63.8 Å). The waterbox was centered over the primary atoms, and any water molecules within 2.5 Å of the DNA or magnesium ion were deleted. The waterbox was then overlayed three times with the original atoms and rotated 90º each time in order to obtain the correct density for the crystal (the first time the waterbox was rotated 90º degrees about the x-axis, the second time the waterbox was rotated 90º about the y-axis, and the third time the waterbox was rotated 90º about the z-axis). Any overlapping waters with the original atoms were deleted. An image update and image centering was then performed in order to delete the appropriate amount of waters for the size and symmetry of the unit cell. The waters that moved after the image update and centering were deleted. Only the Mg^{2+} ion that was identified and reported in the published x-ray crystal

structure was used in the system; no additional ions were added. The water, DNA, and magnesium ion were energy minimized for 500 steps of steepest descent to an energy convergence of 0.001 kcal/mol. The water and Mg^{2+} ion were then equilibrated for 20 ps using the NVT ensemble with the DNA constrained (RMSD of waters converged to ~5.6 Å). Following relaxation of water, the constraints on the DNA were released, and all atoms were allowed to relax for 20 ps in the NVT ensemble (RMSD of DNA and waters converged to ~5.8 Å; average RMSD of DNA only was 0.65 Å). All atoms were then subjected to dynamics for 150 ns at 288 K, which was the temperature at which the x-ray crystal structure was determined. Many studies involving DNA crystal simulations are performed at 300 K.[7, 19, 24, 55, 56] Three more simulations were also run at 288 K; however, the systems were perturbed by rotating the waterbox –90º about the y-axis (rotated 90º about the x- and z-axes) and –90º about the z-axis (rotated 90º about the x- and y-axes) during the reoverlaying procedure (perturbing the systems in this way gave different initial coordinates for the solvent atoms and thus allowed for better sampling of conformational space[57, 58, 61]). The systems were prepared in the same way as described above, and data was then collected over 150 ns in the NPT ensemble. The total simulation time over all four simulations was 600 ns.

2.4.12.2. Short-time DNA Simulations. The short DNA simulations were set up similarly to the long-time simulations (same starting DNA structure, same protocol used). Each system was starting from different initial solvent coordinates in order to introduce a perturbation in each simulation. This was achieved by rotating the waterbox in different degrees about the x-, y-, and z-axes. A total of 20 simulations were run for 2 ns each

(total of 40 ns). Snapshots were saved every 10 ps. Listed below are the rotations of the waterbox about each axis for the 20 simulations.

Table 2.1 Rotations about x, y, and z axes for short DNA simulations

Simulation	x	y	z
1	20	20	-20
2	25	25	-25
3	30	30	-30
4	35	35	-35
5	40	40	-40
6	45	45	-45
7	50	50	-50
8	55	55	-55
9	60	60	-60
10	65	65	-65
11	70	70	-70
12	75	75	-75
13	80	80	-80
14	85	85	-85
15	100	100	-100
16	105	105	-105
17	110	110	-110
18	115	115	-115
19	120	120	-120
20	125	125	-125

References

(1) Leach, A. R., Molecular Dynamics Simulation Methods. In *Molecular Modelling: Principles and Applications*, Pearson Education Limited: Essex, England, 2001; 353-409.

(2) Frenkel, D.; Smit, B., Molecular Dynamics Simulations. In *Understanding Molecular Simulation: From Algorithms to Applications*, Academic Press: San Diego, 2002; 63-107.

(3) 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 Research*, 2000, 28, 235-242.

(4) Berman, H. M.; Olson, W. K.; Beveridge, D. L.; Westbrook, J.; Gelbin, A.; Demeny, T.; Hsieh, S.-H.; Srinivasan, A. R.; Schneider, B., The Nucleic Acid Database: A Comprehensive Relational Database of Three-Dimensional Structures of Nucleic Acids, *Biophysical Journal*, 1992, 63, 751-759.

(5) MacKerell, A. D., Jr., Atomistic Models and Force Fields. In *Computational Biochemistry and Biophysics*, Marcel Dekker, Inc.: New York, 2001; 7-38.

(6) Foloppe, N.; MacKerell, A. D., Jr., All-Atom Empirical Force Field for Nucleic Acids: I. Parameterization Optimization Based on Small Molecule and Condensed Phase Macromolecular Target Data, *J. Comp. Chem.*, 2000, 21, 86-104.

(7) MacKerell, A. D., Jr.; Banavali, N., All-Atom Empirical Force Field for Nucleic Acids: II. Application to Molecular Dynamics Simulations of DNA and RNA in Solution, *J. Comp. Chem.*, 2000, 21, 105-120.

(8) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M., All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins, *J. Phys. Chem. B*, 1998, 102, (18), 3586-3616.

(9) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C., Numerical Integration of the Cartesian equations of motions of a system with constraints: Molecular Dynamics of n-alkanes, *J. Comp. Phys.*, 1977, 23, 327-341.

(10) Leach, A. R., Computer Simulation Methods. In *Molecular Modelling: Principles and Applications*, Pearson Education Limited: Essex, England, 2001; 303-352.

(11) Schlick, T., *Molecular Modeling and Simulation: An Interdisciplinary Guide*. Springer-Verlag: New York, 2000; 21.

(12) Darden, T. A., Treatment of Long-Range Forces and Potential. In *Computational Biochemistry and Biophysics*, Marcel Dekker, Inc.: New York, 2001; 91-114.

(13) Cramer, C. J., Molecular Mechanics. In *Essentials of Computational Chemistry: Theories and Models*, John Wiley & Sons, Ltd.: West Sussex, England, 2002; 17-61.

(14) Darden, T. A.; Perera, L.; Li, L.; Pedersen, L. G., New Tricks for modelers from the crystallography toolkit: the particle mesh Ewald algorithm and its use in nucleic acid simulations, 1999.

(15) Schreiber, H.; Steinhauser, O., Cutoff Size Does Strongly Influence Molecular Dynamics Results on Solvated Polypeptides, *Biochemistry*, 1992, 31, 5856-5860.

(16) Darden, T. A.; York, D.; Pedersen, L. G., Particle Mesh Ewald: an N log(N) method for Ewald Sums in Large Systems, *J. Chem. Phys.*, 1993, 98, (12), 10089-10092.

(17) Cheatham, T. E., III; Miller, J. L.; Fox, T.; Darden, T. A.; Kollman, P. A., Molecular dynamics simulations on solvated biomolecular systems--the particle mesh Ewald method leads to stable trajectories of DNA, RNA, and proteins, *J. Am. Chem. Soc.*, 1995, 117, 4193-4194.

(18) Lee, H.; Darden, T. A.; Pedersen, L. G., Molecular Dynamics Simulation Studies of a High Resolution Z-DNA Crystal, *J. Chem. Phys.*, 1995, 102, 3830-3834.

(19) Lee, H.; Darden, T. A.; Pedersen, L. G., Accurate crystal molecular dynamics simulations using particle mesh Ewald-RNA dinucleotides--ApU and GpC, *Chem. Phys. Lett.*, 1995, 243, 229-235.

(20) York, D.; Yang, W.; Lee, H.; Darden, T. A.; Pedersen, L. G., Toward the Accurate Modeling of DNA: the Importance of Long-Range Electrostatics, 1995.

(21) Bogusz, S.; Cheatham, T. E., III; Brooks, B. R., Removal of Pressure and Free Energy Artifacts in Charged Periodic Systems via Net Charge Corrections to the Ewald Potential, *J. Chem. Phys.*, 1998, 108, 7070-7084.

(22) Ewald, P. P., Die Berechnung optischer und elektrostatischer Gitterpotentiale, *Annalen der Physik*, 1921, 369, 253-287.

(23) Roux, B., Implicit Solvent Models. In *Computational Biochemistry and Biophysics*, Marcel Dekker, Inc.: New York, 2001; 133-149.

(24) Babin, V.; Baucom, J.; Darden, T. A.; Sagui, C., Molecular Dynamics Simulations of DNA with Polarizable Force Fields: Convergence of an Ideal B-DNA Structure to the Crystallographic Structure, *J. Phys. Chem. B*, 2006, 110, 11571-11581.

(25) Colton, T., *Statistics in Medicine*. Little, Brown & Co., Inc.: Boston, MA, 1974.

(26) Hotelling, H., Analysis of a Complex of Statistical Variables in Principal Components, *Journal of Educational Psychology*, 1933, 24, 417-441.

(27) Jolliffe, I. T., *Principal Components Analysis*. Springer-Verlag: New York, 2002.

(28) Manly, B. F. J., *Multivariate Statistical Methods: A Primer*. Chapman and Hall: 2005.

(29) Pearson, K., On Lines and Planes of Closest Fit to Systems of Points in Space, *Philosophical Magazine*, 1901, 2, 559-572.

(30) *S-Plus*, Insightful Corp.: Seattle, WA, 2007.

(31) Born, M.; Oppenheimer, R., Zur Quantentheorie der Molekeln (On the Quantum Theory of Molecules), *Annalen der Physik*, 1927, 84, 457-484.

(32) McQuarrie, D. A.; Simon, J. D., *Physical Chemistry: A Molecular Approach*. University Science Books: Sausalito, CA, 1997.

(33) Foresman, J. B.; Frisch, A., *Exploring Chemistry with Electronic Structure Methods*. Gaussian, Inc.: Pittsburgh, 1996.

(34) Hinchliffe, A., *Molecular Modelling for Beginners*. John Wiley & Sons: England, 2003.

(35) Ratner, M. A.; Schatz, G. C., *Introduction to Quantum Mechanics in Chemistry*. Prentice Hall: New Jersey, 2001.

(36) Hohenberg, P.; Kohn, W., Inhomogeneous Electron Gas, *Phys. Rev. B*, 1964, 136, 864.

(37) Kohn, W.; Sham, L. J., Self-Consistent Equations Including Exchange and Correlation Effects, *Phys. Rev. A*, 1965, 10, 1133-1138.

(38) Becke, A. D., Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior, *Phys. Rev. A*, 1988, 38, 3098-3100.

(39) Becke, A. D., Density-Functional Thermochemistry. I. The Effect of the Exchange-Only Gradient Correction, *J. Chem. Phys.*, 1992, 96, 2155-2160.

(40) Lee, C.; Yang, W.; Parr, R. G., Development of the Colle-Salvetti Correlation Energy Formula into a Functional of the Electron Density, *Phys. Rev. B*, 1988, 37, 785-789.

(41) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A.; 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.; Haseqawa, 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. *Gaussian03, Revision B.03*, Gaussian, Inc.: Pittsburgh, PA, 2003.

(42) *NSF CHE-0354052*; p^pp.

(43) *NSF AAB/PSC CHE-030008P*; p^pp.

(44) Møller, C.; Plesset, M. S., Note on an Approximation Treatment for Many-Electron Systems, *Physical Review*, 1934, 46, 618.

(45) Hehre, W. J.; Ditchfield, R.; Pople, J. A., Self-Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian-type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules, *J. Chem. Phys.*, 1972, 56, 2257-2261.

(46) Ditchfield, R.; Hehre, W. J.; Pople, J. A., Self-Consistent Molecular Orbital Methods. IX. An Extended Gaussian-type Basis for Molecular Orbital Studies of Organic Molecules, *J. Chem. Phys.*, 1971, 54, 724-728.

(47) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M., CHARMM: a Program for Macromolecular Energy, Minimization, and Dynamics Calculations, *J. Comp. Chem.*, 1983, 4, (2), 187.

(48) MacKerell, A. D.; Feig, M.; Brooks, C. L., Improved Treatment of the Protein Backbone in Empirical Force Fields, *J. Am. Chem. Soc.*, 2004, 126, 698.

(49) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of Simple Potential Functions for Simulating Liquid Water, *J. Chem. Phys.*, 1983, 79, 926.

(50) Field, M.; Karplus, M. *CRYSTAL: Program for Crystal Calculations in CHARMM*, Harvard University: Cambridge, MA, 1992.

(51) Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, R. W., Constant Pressure Molecular Dynamics Simulations: The Langevin Piston Method, *J. Chem. Phys.*, 1995, 103, 4613-4621.

(52) Berman, H. M.; Neidle, S.; Zimmer, C.; Thrum, H., Netropsin, A DNA-Binding Oligopeptide: Structural and Binding Studies, *Biochimica et Biophysica Acta*, 1979, 561, 124-131.

(53) Brunger, A. T.; Karplus, M., Polar Hydrogen Positions in Proteins: Empirical Energy Placement and Neutron Diffraction Comparison, *Proteins: Structure, Functions, and Genetics*, 1988, 4, 148-156.

(54) Kielkopf, C. L.; Bremer, R. E.; White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B.; Rees, D. C., Structural Effects of DNA Sequence on TA Recognition by Hydroxypyrrole/Pyrrole Pairs in the Minor Groove, *J. Mol. Biol.*, 2000, 295, 557-567.

(55) Baucom, J.; Transue, T.; Fuentes-Cabrera, M.; Krahn, J. M.; Darden, T. A.; Sagui, C., Molecular Dynamics Simulations of the d(CCAACGTTGG)2 Decamer in Crystal Environment: Comparison of Atomic Point-Charge, Extra-Point, and Polarizable Force Fields, *J. Chem. Phys.*, 2004, 121, 6998-7008.

(56) Bevan, D. R.; Li, L.; Pedersen, L. G.; Darden, T. A., Molecular Dynamics Simulations of the d(CCAACGTTGG)$_2$ Decamer: Influence of the Crystal Environment, *Biophys. J.*, 2000, 78, 668-682.

(57) Schulze, B. G.; Evanseck, J. D., Cooperative Role of Arg45 and His64 in the Spectroscopic A3 State of Carbonmonoxy Myoglobin: Molecular Dynamics Simulations, Multivariate Analysis, and Quantum Mechanical Computations, *J. Am. Chem. Soc.*, 1999, 121, 6444-6454.

(58) Schulze, B. G.; Grubmuller, H.; Evanseck, J. D., Functional Significance of Hierarchical Tiers in Carbonmonoxy Myoglobin: Conformational Substates and Transitions Studied by Conformational Flooding Simulations, *J. Am. Chem. Soc.*, 2000, 122, 8700-8711.

(59) Grossfield, A.; Feller, S. E.; Pitman, M. C., Convergence of Molecular Dynamics Simulations of Membrane Proteins, *Proteins: Structure, Functions, and BioInformatics*, 2007, 67, 31-40.

(60) Auffinger, P.; Louise-May, S.; Westhof, E., Multiple molecular dynamics simulations of the anticodon loop of tRNAAsp in aqueous solution with counterions, *J. Am. Chem. Soc.*, 1995, 117, 6720-6726.

(61) Caves, L. S. D.; Evanseck, J. D.; Karplus, M., Locally Accessible Conformations of Proteins: Multiple Molecular Dynamics Simulations of Crambin, *Protein Science*, 1998, 7, 649-666.

(62) Tabernero, L.; Verdaguer, N.; Coll, M.; Fita, I.; van der Marel, G. A.; van Boom, J. H.; Rich, A.; Aymami, J., Molecular Structure of the A-tract DNA Dodecamer d(CGCAAATTTGCG) Complexed with the Minor Groove Binding Drug Netropsin, *Biochemistry*, 1993, 32, 8403-8410.

(63) Coll, M.; Frederick, C. A.; Wang, A. H.-J.; Rich, A., A Bifurcated Hydrogen-Bonded Conformation in the Base Pairs of the DNA Dodecamer d(GCGAAATTTGCG) and Its Complex with Distamycin, *Proc. Natl. Acad. Sci. USA*, 1987, 84, 8385-8389.

(64) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; 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.

(65) Wellenzohn, B.; Winger, R. H.; Hallbrucker, A.; Mayer, E.; Liedl, K. R., Simulation of EcoRI Dodecamer Netropsin Complex Confirms Class I Complexation Mode, *J. Am. Chem. Soc.*, 2000, 122, 3927-3931.

(66) Wellenzohn, B.; Flader, W.; Winger, R. H.; Hallbrucker, A.; Mayer, E.; Liedl, K. R., Influence of Netropsin's Charges on the Minor Groove Width of d(CGCGAATTCGCG)2, *Biopolymers*, 2002, 61, 276-286.

(67) Allen, F. H., The Cambridge Structural Database: a quarter of a million crystal structures and rising, *Acta Cryst.*, 2002, B58, 380-388.

(68) Hamanaka, T.; Mitsui, T.; Ashida, T.; Kakudo, M., The Crystal Structure of All-trans Retinal, *Acta Crystallographica*, 1972, B28, 214-222.

(69) Stam, C. H., The Crystal Structure of a Monoclinic Modification and the Refinement of a Triclinic Modification of Vitamin A Acid (Retinoic Acid), $C_{20}H_{28}O_2$, *Acta Crystallographica*, 1972, B28, 2936-2945.

(70) Stam, C. H.; MacGillavry, C. H., The Crystal Structure of the Triclinic Modification of Vitamin-A Acid, *Acta Crystallographica*, 1963, 16, 62-68.

(71) Kleywegt, G. J.; Bergfors, T.; Senn, H.; Le Motte, P.; Gsell, B.; Shudo, K.; Jones, T. A., Crystal Structures of Cellular Retinoic Acid Binding Proteins I and II in Complex with all-trans retinoic acid and a synthetic retinoid, *Structure*, 1994, 2, 1241.

(72) Calderone, V.; Folli, C.; Marchesani, A.; Berni, R.; Zanotti, G., Identification and Structural Analysis of a Zebrafish Apo and Holo Cellular Retinol-Binding Protein, *J. Mol. Biol.*, 2002, 321, 527-535.

(73) Zanotti, G.; Marcello, M.; Malpeli, G.; Folli, C.; Sartori, G.; Berni, R., Crystallographic Studies on Complexes Between Retinoids and Plasma Retinol-Binding Protein, *J. Biol. Chem.*, 1994, 269, 29613-29620.

(74) *MOE*, Chemical Computing Group: Quebec, Canada, 2004.

(75) Kelly, K. Multiple sequence and structural alignment in MOE. http://www.chemcomp.com/Journal_of_CCG/Features/align.htm

(76) Levitt, M., Accurate modeling of protein conformation by automatic segment matching, *J. Mol. Biol.*, 1992, 226, 507-533.

(77) Royant, A.; Nollert, P.; Edman, K.; Neutze, R.; Landau, E. M.; Pebay-Peyroula, E.; Navarro, J., X-ray Structure of Sensory Rhodopsin II at 2.1-A resolution, *Proc. Natl. Acad. Sci. USA*, 2001, 98, 10131-10136.

(78) Faham, S.; Yang, D.; Bare, E.; Yohannan, S.; Whitelegge, J. P.; Bowie, J. U., Side-Chain Contributions to Membrane Protein Structure and Stability, *J. Mol. Biol.*, 2004, 335, 297-305.

(79) Wang, J. M.; Link, S.; Heyes, C. D.; El-Sayed, M., Comparison of the Dynamics of the Primary Events of Bacteriorhodopsin in Its Trimeric and Monomeric States, *Biophys. J.*, 2002, 83, 1557-1566.

(80) Jelsch, C.; Teeter, M. M.; Lamzin, V.; Pichon-Pesme, V.; Blessing, R. H.; Lecomte, C., Accurate Protein Crystallography at ultra-high resolution: valence electron distribution in crambin, *Proc. Natl. Acad. Sci.*, 2000, 97, 3171-3176.

(81) Leonard, G. A.; Hunter, W. N., Crystal and Molecular Structure of d(CGTAGATCTACG), *J. Mol. Biol.*, 1993, 234, 198-208.

Chapter 3

Force Field Parameters for DNA Minor Groove-Binding Pyrrole-Imidazole Polyamides and Significance of Multiple Simulations

This chapter describes the parameterization procedure for the DNA minor groove-binding polyamides, the results of the crystal simulations performed to test the parameters, and conclusions made from the study.

3.1 Parameterization of DNA minor groove-binding polyamides

Parameterization of the polyamides began by first dividing the polyamides into six model compounds and carrying out energy minimization using MP2/6-31G(d), which includes the effects of electron correlation and dispersive interactions. The bond, angle, torsional, and improper terms were modified to match the quantum mechanical results. The parameters were evaluated by performing molecular dynamics crystal simulations of the reported x-ray crystal structures of a netropsin crystal, 2:1 Dervan polyamide/DNA complexes, a netropsin/DNA complex, and a distamycin/DNA complex. The new parameters can be used for improved simulations of polyamides and their nucleic acid complexes.

Parameterization was performed by first dividing the molecule into components, otherwise known as model compounds.[1-15] The models, shown in Figure 3.1, consist of

three commonly used heterocycles for DNA recognition (N-methylpyrrole, N-methylimidazole, and N-methylhydroxypyrrole), β-alanine used for enhancing flexibility of the molecule, a cationic tail (dimethylaminopropylamide) used for phosphate binding, and γ-aminobutyric acid as a linker in hairpin polyamides. Since heterocycles and linkers are joined together in the polyamide by amide bonds, compounds that contain terminal amide groups to represent the linkage to another heterocycle or linker in the polyamide sequence were chosen.

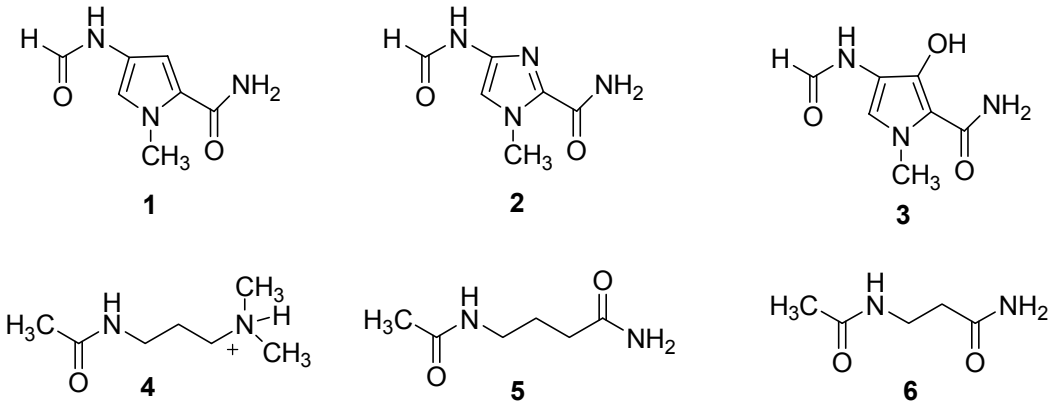


Figure 3.1 Model compounds for polyamides (some hydrogens not shown for clarity).

Each model compound was energy minimized using MP2/6-31G(d). Model compound **1** was also minimized using CCSD/6-31G(d) to verify the MP2 results. Comparison of MP2 and CCSD energy minimized structures revealed that the error for bonds was 0.003 ± 0.005 Å, 0.2 ± 0.1º for angles, and 0.4 ± 0.3º for dihedral angles. The comparison indicated that the MP2/6-31G(d) level of theory is appropriate for force field parameterization, a result in agreement with others for parameterization of polymers.[16, 17] For comparison, the structures were also energy minimized using HF and B3LYP with the 6-31G(d) basis set. Initial topology and parameters were then created from existing

compounds in the CHARMM force field.[18] Indole was the starting point for the pyrrole-based rings (**1** and **3**), histidine was the starting point for the imidazole ring (**2**), and N-methylacetamide was the starting point for the tail and linker compounds (**4**, **5**, **6**).

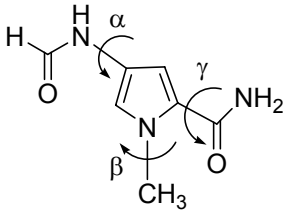
The initial topology and parameters were then modified to match the quantum mechanical results. The minimized bond lengths and angle values obtained from the MP2 optimized structures were used as the equilibrium bond lengths and equilibrium angle values for the new CHARMM parameters for the associated bond and angle types. Each model compound was energy minimized with the CHARMM force field with steepest descent and Newton-Raphson (tolerance gradient was 10^{-6} kcal/mol/Å for each minimization). The parameters where then adjusted to minimize the error with the MP2/6-31G(d) structure. The average error between CHARMM and MP2 is 0.02 ± 0.02 Å for bond distances and 2 ± 2º for angles.

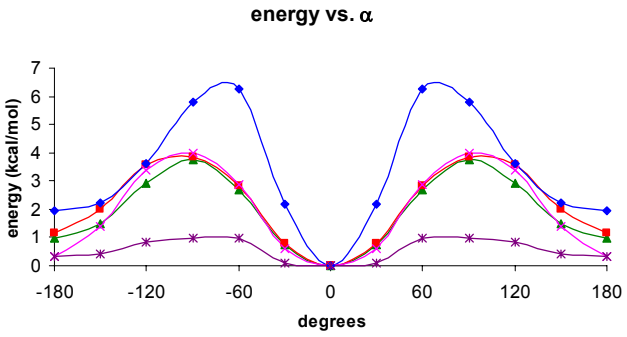
In addition to bonds and angles, torsion terms were also parameterized. For each model compound, the potential energy surfaces (PES) for selected torsion angles (see Figures 3.2-3.7 for selected torsion angles; the bonds with the arrows were the bonds which were rotated about) were calculated by holding the selected torsion angle fixed at different increments (30º increments from –180º to 180º) and geometry optimizing all other degrees of freedom. The torsion angles selected were for rotations about the bonds connecting the amide groups to the heterocycle, the bond connecting the methyl group to the heterocycle, and the carbon-carbon single bonds of the linker and tail model compounds. The torsion angles selected are necessary for the polyamide to optimize contacts with the bases of the minor groove of DNA. The CHARMM dihedral angle parameters force constant ($K\chi$), multiplicity (n), and phase shift (δ) were then modified to

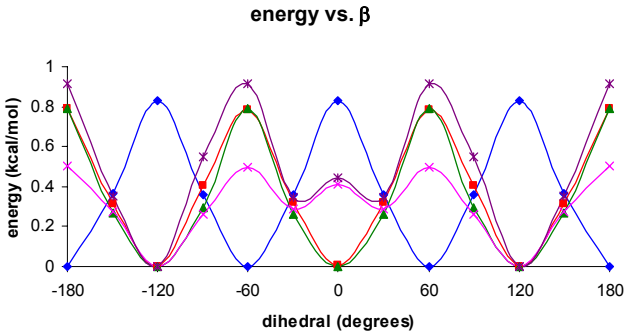
match the quantum mechanical results from the potential energy surfaces. Each model compound was energy minimized in CHARMM (steepest descent and Newton-Raphson; tolerance gradient was 10^{-6} kcal/mol/Å) with the selected torsion angle held fixed, and the potential energy surface of the dihedral was calculated. If the surface did not match the MP2 results, the multiplicity, phase shift, or force constant were adjusted as necessary. After the results from the CHARMM compounds matched those obtained from the QM calculations, the model compounds were pieced together to make the entire polyamide molecule for use in molecular mechanics energy minimization or molecular dynamics simulations.

For N-methylpyrrole, (**1**) three dihedrals were selected for parameterization, which are the rotation of the methyl group and the rotation of the two amide groups on the ring. The most flexible points on the polyamide are most likely along the backbone where the heterocycles are connected to the amide groups. The polyamide will have to rotate around those bonds to optimize interactions with the DNA, and the methyl groups on the ring have free rotation, which will interact with the DNA through van der Waals contacts, so they will also have to rotate to optimize their interactions with the DNA. The heterocycles themselves are aromatic, so the rings stay planar. The initial CHARMM parameters for three important dihedrals borrowed from similar functionality did not agree with the MP2/6-31G(d) potential energy surface (Figure 3.2). For example, the methyl rotation (β) is 180° out of phase, both amide groups have energy barriers that are too high (~3.4 kcal/mol too high for α and ~4.1 kcal/mol too high for γ), and the second amide rotation (γ) does not give the lowest energy at 0°. The energies obtained from HF/6-31G(d) and B3LYP/6-31G(d) do not necessarily match those obtained with MP2/6-

31G(d). For the methyl rotation, B3LYP gives a barrier at 0°, whereas MP2 gives a minimum. HF shows the correct phase for methyl rotation with similar barriers to MP2. As shown for the α torsional plot, HF barriers are 34% lower than MP2.







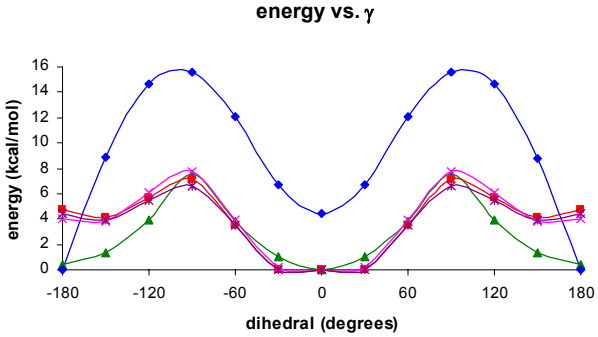


Figure 3.2 Structure (top) and the three torsion energy plots for N-methylpyrrole. Blue diamond (♦), initial CHARMM; red square (+), MP2/6-31G(d); green triangle (▲), final CHARMM; pink x (X), B3LYP/6-31G(d); purple star (*), HF/6-31G(d).

Similarly, for N-methylimidazole (**2**), three dihedrals were selected for parameterization. The chosen dihedrals are again for the rotation of the amide groups

and the methyl group on the imidazole ring. The results for the imidazole torsions are shown in Figure 3.3. The initial CHARMM parameters for N-methylimidazole borrowed from similar functionality are also in poor agreement with the MP2/6-31G(d) results. The methyl rotation is out of phase by 60º, and the energy barriers are ~1.5 kcal/mol too high. Both amide group rotations do not show the correct behavior. The first amide group rotation (α) should have a maximum at 180º and –180º; instead the maxima are at 90º and –90º. The second amide group rotation (γ) has energy barriers that are ~6.3 kcal/mol too high, and there are minima at 180º and –180º, which the MP2 does not show. The HF and B3LYP results agree well with MP2 for both the amide group rotations; however, for the methyl rotation, the B3LYP is 30º out of phase.

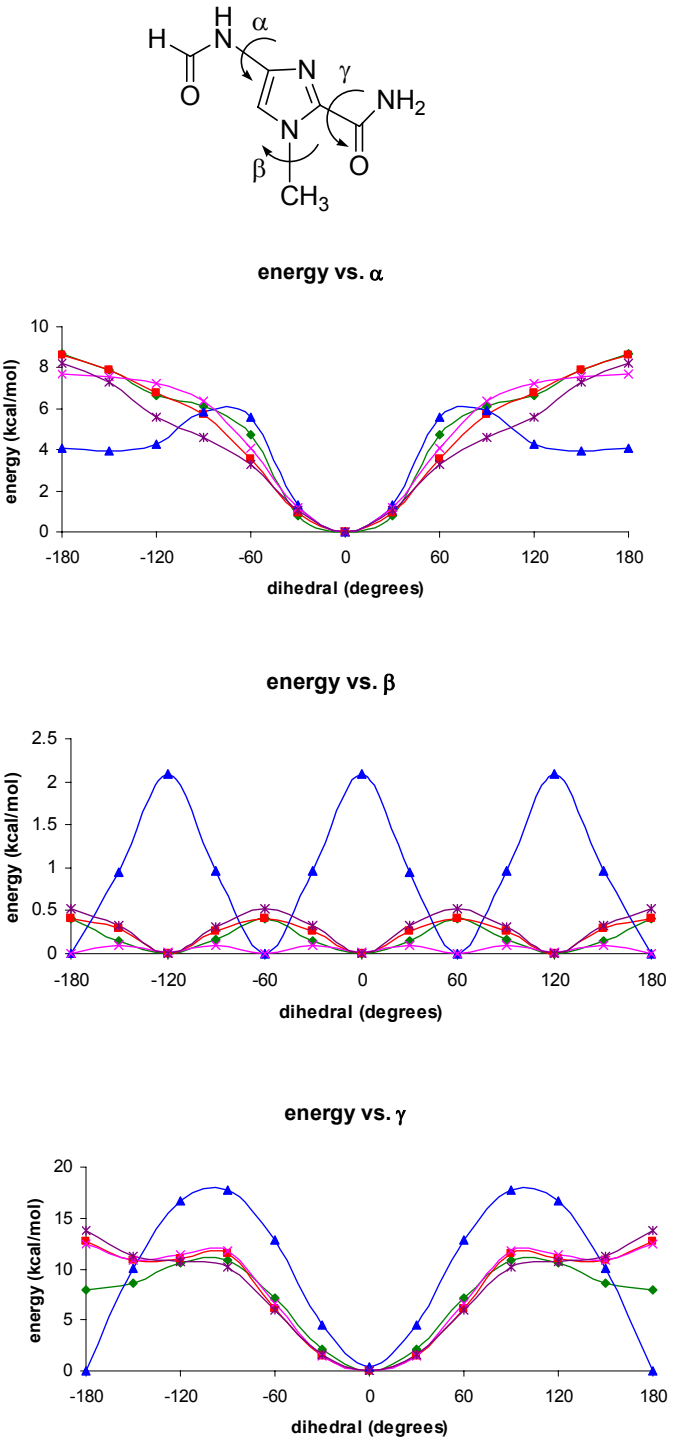


Figure 3.3 Structure (top) and the three torsion energy plots for N-methylimidazole. Blue diamond (♦), initial CHARMM; red square (+), MP2/6-31G(d); green triangle (▲), final CHARMM; pink x (X), B3LYP/6-31G(d); purple star (*), HF/6-31G(d).

Four torsions were chosen for hydroxypyrrole (**3**), the two amide group rotations, the methyl group rotation, and the OH group rotation. The OH group rotation was chosen for this model compound because the H of the hydroxyl group hydrogen bonds with the O2 of thymine, so this group will have to rotate to optimize interaction with the DNA. The results for hydroxypyrrole are shown in Figure 3.4. The initial CHARMM parameters are in poor agreement with the MP2/6-31G(d) results; all of the selected torsion angles are out of phase, the energy barriers are too high for the methyl rotation (~1.3 kcal/mol) and the OH group rotation (~17.0 kcal/mol), and all but the methyl rotation show the incorrect multiplicity. The final CHARMM parameters give correct methyl rotation barriers and show the right multiplicity, the rotation of the OH group on the pyrrole ring shows the correct maximum, the rotation of the first amide group (α) shows the correct minimum and multiplicity, and the rotation of the second amide group shows the correct barrier heights. The height of the small maximum at 0º is ~2.5 kcal/mol higher than that of the MP2 results; however, this is the best agreement that could be obtained in order to obtain simultaneous agreement for the other dihedrals of the hydroxypyrrole model compound. The second amide torsion plot (γ) shows a minimum at 180º and a small maximum at 0º, whereas the other heterocycles show the global minimum at 0º for this torsion surface. The MP2 calculations are carried out in the gas phase, so the amide group rotates 180º for the carbonyl oxygen to form a hydrogen bond with the H of the OH group, lowering the energy. The other heterocycles do not have the OH group; therefore, the minimum is at 0º for the others. Both HF and B3LYP are in reasonable agreement with MP2 for δ; they both give the correct phase, multiplicity, and a maximum at 0º. HF is 0.65 kcal/mol lower than MP2, and B3LYP is 0.52 kcal/mol

lower than MP2. HF shows good agreement with MP2 for β and γ. For β, HF shows the correct phase and multiplicity and is 0.28 kcal/mol higher than MP2. For γ, HF again shows the correct phase and multiplicity and is 0.39 kcal/mol higher than MP2. However, HF shows poor agreement with MP2 for α; HF gives a maximum at 150º, and a minimum at 180º, while MP2 gives a maximum at 180º. B3LYP shows reasonable agreement with MP2 for α and β. For α, B3LYP gives the correct phase and multiplicity, and the maximum at 180º is 0.92 kcal/mol lower than MP2. For β, B3LYP shows the correct phase and multiplicity, and the barriers are ~0.32 kcal/mol lower than MP2. For γ, B3LYP shows the correct phase and multiplicity; however, the barriers at 90º and -90º are 3.41 kcal/mol higher than MP2.

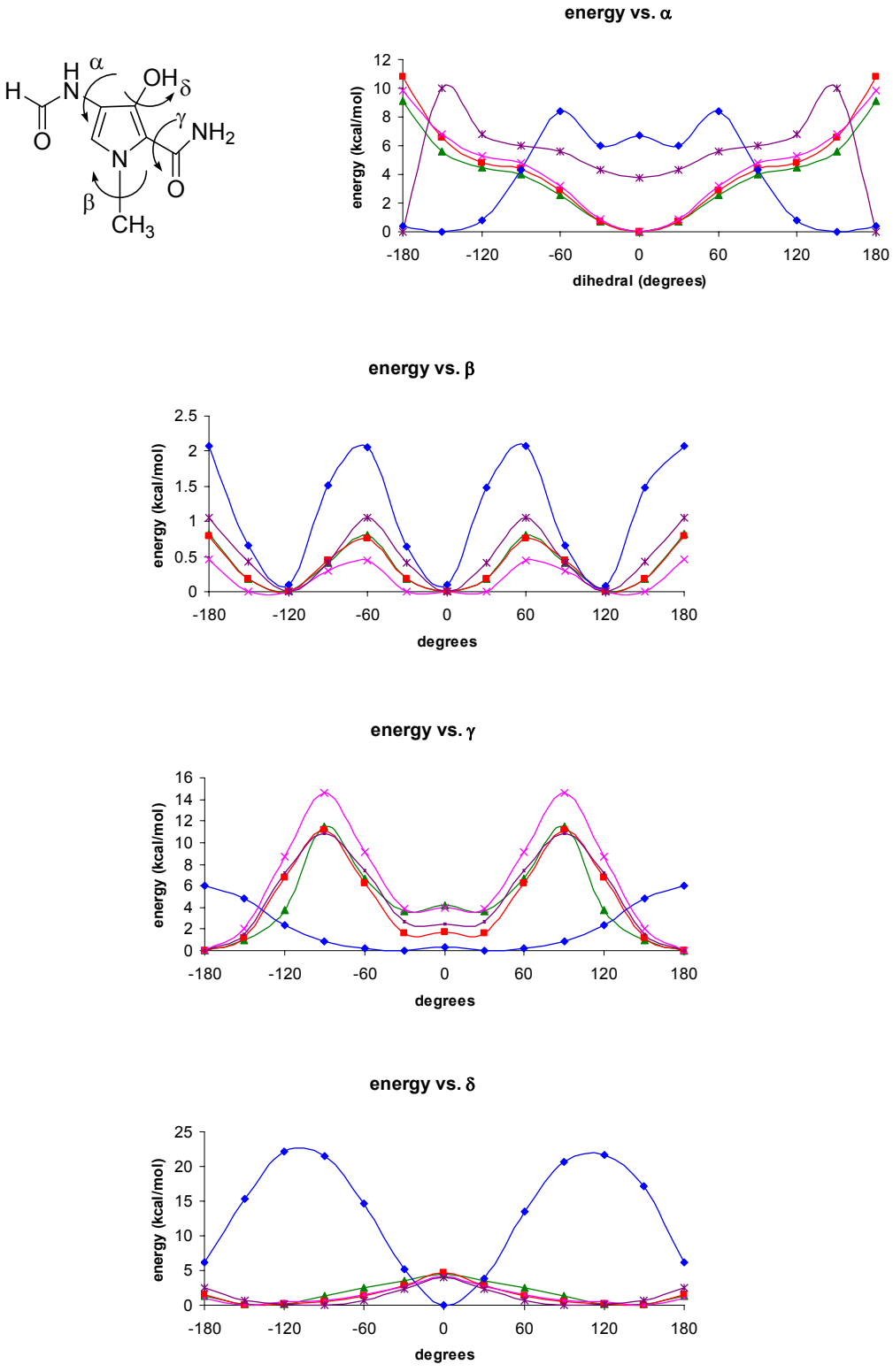


Figure 3.4 Structure (top left) and the four torsion energy plots for hydroxypyrrole. Blue diamond (♦), initial CHARMM; red square (+), MP2/6-31G(d); green triangle (▲), final CHARMM; pink x (X), B3LYP/6-31G(d); purple star (*), HF/6-31G(d).

Similar analyses were carried out for the cationic tail (**4**), the γ-aminobutyric acid linker, and the β-alanine linker. For the cationic tail, two dihedral angles were selected for rotation, which are for the rotation of the amide group and the rotation of the dimethyl amino group. The torsion plots are shown in Figure 3.5.

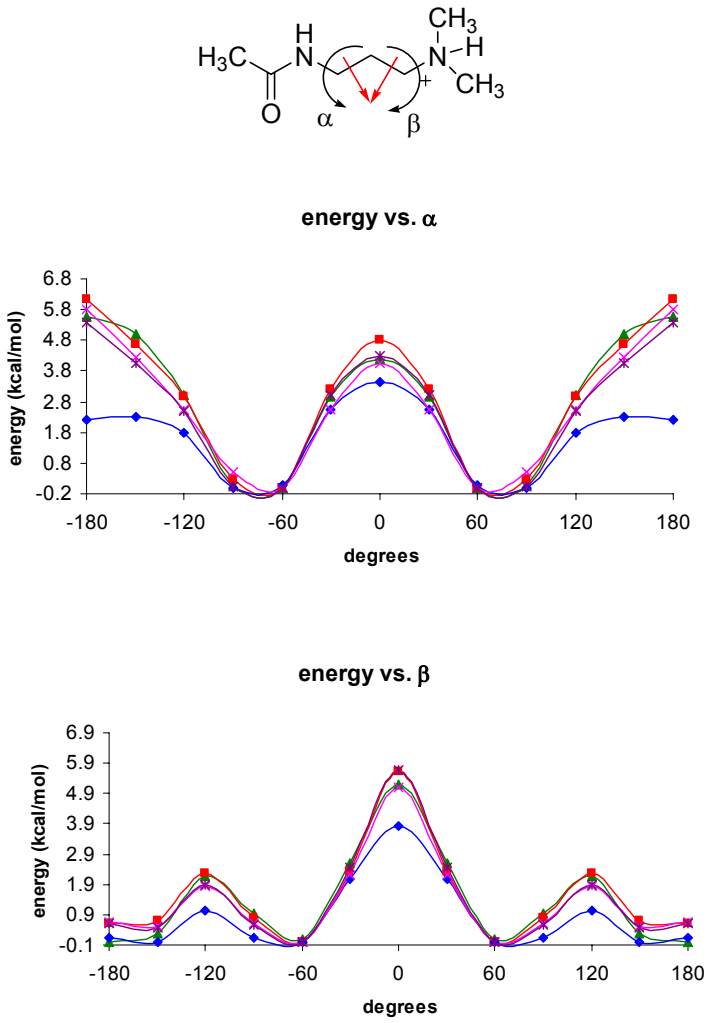


Figure 3.5 Structure (top) and the two torsion energy plots for the dimethylaminopropyl cationic tail. Blue diamond (♦), initial CHARMM; red square (+), MP2/6-31G(d); green triangle (▲), final CHARMM; pink x (X), B3LYP/6-31G(d); purple star (*), HF/6-31G(d).

From the torsion energy plots for the cationic tail, it can be seen that the initial CHARMM parameters for this model compounds are in better agreement with the MP2

results than the initial heterocycle parameters. The MP2 calculations were performed with two dihedrals constrained at 180º, which are marked by the straight red arrows. The molecule was initially optimized with no constraints; however, in the gas phase, the molecule folded up on itself into a ball, so the constraints were necessary. The initial CHARMM parameters for the tail show the correct phase and number of minima and maxima; however, the energy barriers for both torsion angles are initially too low (1.34 kcal/mol for α, and 1.78 kcal/mol for β). Initially, constraints were placed on the same dihedrals in CHARMM when making the new parameters; however, the energy barriers were too high, and the constraints had to be removed. The height of the energy barriers in the final CHARMM parameters is in good agreement with the calculated MP2 energy barriers (0.6 kcal/mol lower than MP2 for α and 0.43 kcal/mol lower for β). The HF and B3LYP results are in reasonable agreement with the MP2 results for α. HF is 0.5 kcal/mol lower than MP2 and B3LYP is 0.74 kcal/mol lower than MP2. For β, the agreement is better with HF being 0.05 kcal/mol higher than MP2 and B3LYP being 0.52 kcal/mol lower than MP2.

Two dihedrals were selected for rotation for the γ-aminobutyric acid linker (**5**) and for the β-alanine linker (**6**). For the γ-aminobutyric acid linker, the two dihedrals selected are for rotation of the amide groups. The results are shown in Figure 3.6. The initial CHARMM parameters for the γ-linker are in slight agreement with MP2, so they did not require significant modification. As with the cationic tail model compound, the MP2 calculations were performed with two dihedrals constrained at 180º (the constrained dihedrals are marked by the straight red arrows) for the same reason they were place on the cationic tail. Again, constraints were placed on the same dihedrals in CHARMM

initially, but the energy barriers were too high and could not be lowered until the constraints were removed. The plots show the correct phase and multiplicity. However, the one energy barrier is ~1.15 kcal/mol too high (α), and the other ~0.97 kcal/mol too low (β). The final CHARMM gives energy barriers that are good agreement with MP2; 6.24 kcal/mol for α, which is 0.19 kcal/mol higher than MP2, and 0.22 kcal/mol higher than MP2 for β. The energies for 120º and 150º for β do not follow the MP2 energies, but this is the best agreement that could be obtained in order to have both torsions agree well with MP2. The HF and B3LYP results are in reasonable agreement with the MP2 results for both torsions. For α, HF is 0.73 kcal/mol higher than MP2, and B3LYP is 0.34 kcal/mol higher than MP2. For β, the agreement is better; HF is 0.34 kcal/mol higher than MP2, and B3LYP is 0.09 kcal/mol higher than MP2.

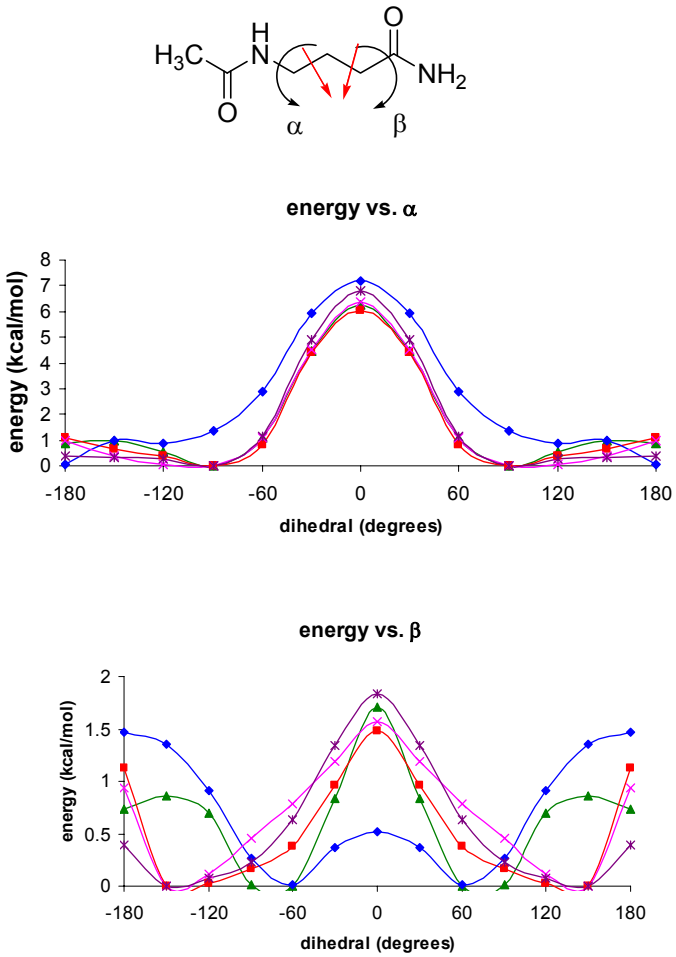


Figure 3.6 Structure (top) and the two torsion energy plots for the γ-aminobutyric acid linker. Blue diamond (♦), initial CHARMM; red square (+), MP2/6-31G(d); green triangle (▲), final CHARMM; pink x (X), B3LYP/6-31G(d); purple star (*), HF/6-31G(d).

As with the γ-aminobutyric acid linker, the two dihedrals selected for the β-alanine linker are for rotation of the amide groups. The results are shown in Figure 3.7. As for the tail and γ-linker model compounds, the torsion energy plots for the β-alanine linker show that the initial CHARMM parameters were in better agreement with MP2 than were the initial heterocycle parameters. One dihedral was constrained at 180º (marked by the straight red arrow) for the MP2 calculations for the same reasons as the cationic tail and the γ-linker. When the structure was energy minimized with CHARMM,

no constraints were used because the energy barriers were too high. The energy barrier is ~5.9 kcal/mol too high for α and the barrier is ~0.94 kcal/mol too low for β; however, they are in phase and show the correct multiplicity. With adjustment of force constants, the final CHARMM parameters agree well with the MP2 results. The energy barrier for β is ~0.10 kcal/mol higher than MP2, and α is 2.85 kcal/mol higher than MP2; however, this better than the initial CHARMM parameters and is the best agreement that could be obtained and still have both torsions agree well with the MP2 results. HF and B3LYP are in good agreement with MP2 for both dihedrals. HF is 0.15 kcal/mol lower than MP2 and B3LYP is 0.49 kcal/mol lower than MP2 for α. For β, HF is 0.41 kcal/mol higher than MP2 and B3LYP is 0.39 kcal/mol lower than MP2.

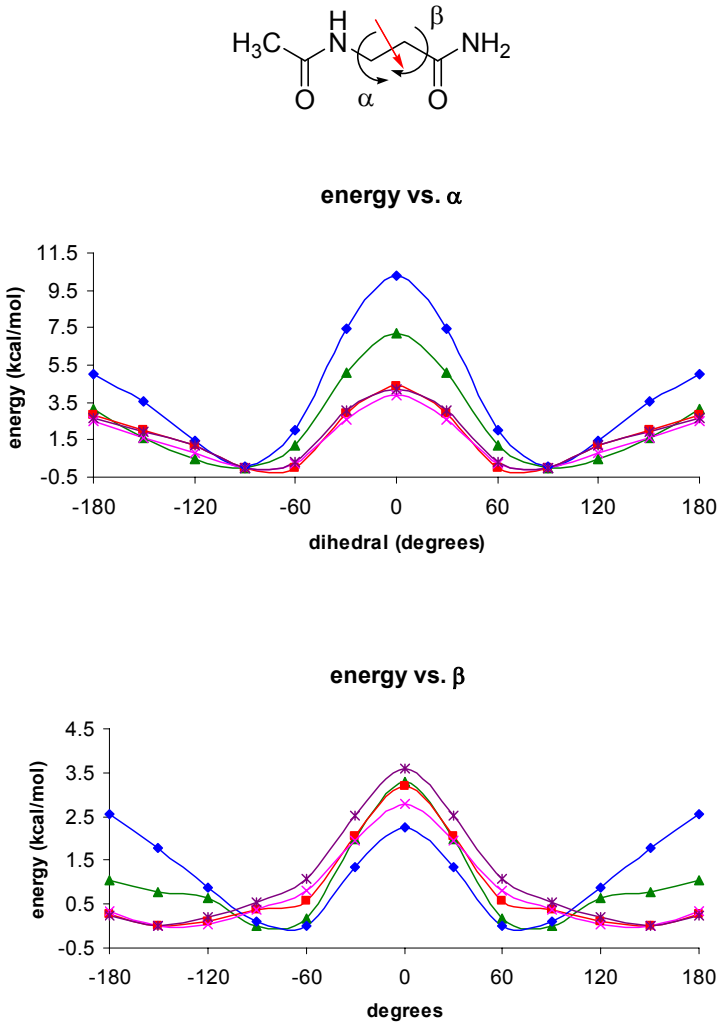


Figure 3.7 Structure (top) and the two torsion energy plots for the □-alanine linker. Blue diamond (♦), initial CHARMM; red square (+), MP2/6-31G(d); green triangle (▲), final CHARMM; pink x (X), B3LYP/6-31G(d); purple star (*), HF/6-31G(d).

The same trend of systematic errors in multiplicity, phase shift, and force constant is observed for all six model compounds. Because HF and B3LYP do not necessarily show the same behavior as MP2, electron correlation and dispersion effects are probably important for this class of compounds and are necessary in polyamide parameter development. HF and B3LYP are therefore not appropriate methods to use for parameterization of these compounds.

Final CHARMM parameters for the model compounds produce structures and energies in good agreement with MP2/6-31G(d) results. Overall, the error between CHARMM and MP2 geometry optimized structures is 0.02 ± 0.02 Å for bonds, 2 ± 2° for angles, and 5 ± 9° for dihedrals. The error in barrier heights between CHARMM and MP2 is 0.38 ± 0.40 kcal/mol.

3.2 DNA/polyamide Simulations in the Crystal Environment

Simulations were performed in the crystal environment in order to validate the new force field parameters for polyamides. There are several high-resolution crystal structures of DNA with bound polyamides available,[1-3, 19] and performing the simulations in the crystal environment allows for a direct comparison back to experiment. Simulations were analyzed by comparing the calculated DNA helical parameters and DNA/polyamide distances back to the starting x-ray structure using the t-test.

3.2.1 Netropsin Crystal Simulation. As an initial test of the parameters, a 5 ns crystal simulation of netropsin (CCSD[20] id NETRSN[21]; R-factor 6.7%) was performed. The system contained one netropsin molecule, five water molecules, and one sulfate ion. The average RMSD of the simulation was 0.42 ± 0.09 Å, as shown in Figure 3.8 (top right). The low RMSD indicates that the parameters are in good agreement with the experimental crystal structure.

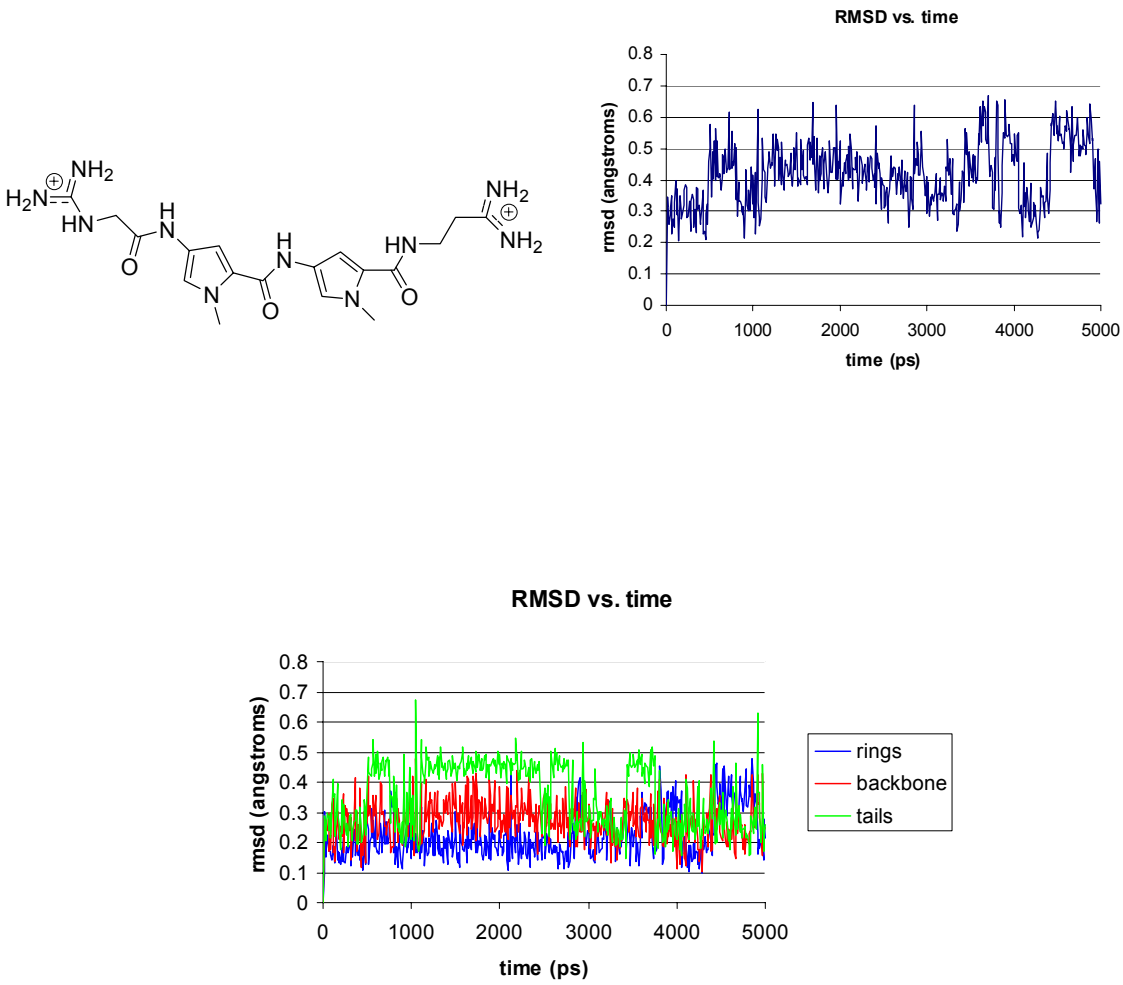


Figure 3.8 Structure of netropsin (top left) and RMSD vs. time plot (top right) for the netropsin simulation (hydrogens not included in RMSD). The bottom plot is the RMSD vs. time for different parts of the netropsin molecule (rings in blue, amide backbone in red, and tails in green).

Exploration of the differences between the simulated and experimentally derived structures was done by examining the RMSDs of the pyrrole rings, amide backbone, and the cationic tail. It was found that the tail region moves more than any other region during the simulation (RMSD of 0.4 ± 0.1 Å). The backbone has the next largest contribution with an average RMSD of 0.27 ± 0.07 Å, and the rings contribute the least with an average RMSD of 0.22 ± 0.08 Å (Figure 3.8, bottom). This is reasonable because the rings are aromatic and planar so there will not be as much movement within

the rings, the backbone has a little more freedom to move, but not as much as the tail. Rotation is allowed where the amide groups are attached to the rings, and the tails consist primarily of C-C single bonds so free rotation is allowed. The high RMSD of the tail with the amidino group at the end involves a rotation around the carbon-carbon bonds from ~ 40° to 80° (Figure 3.9).

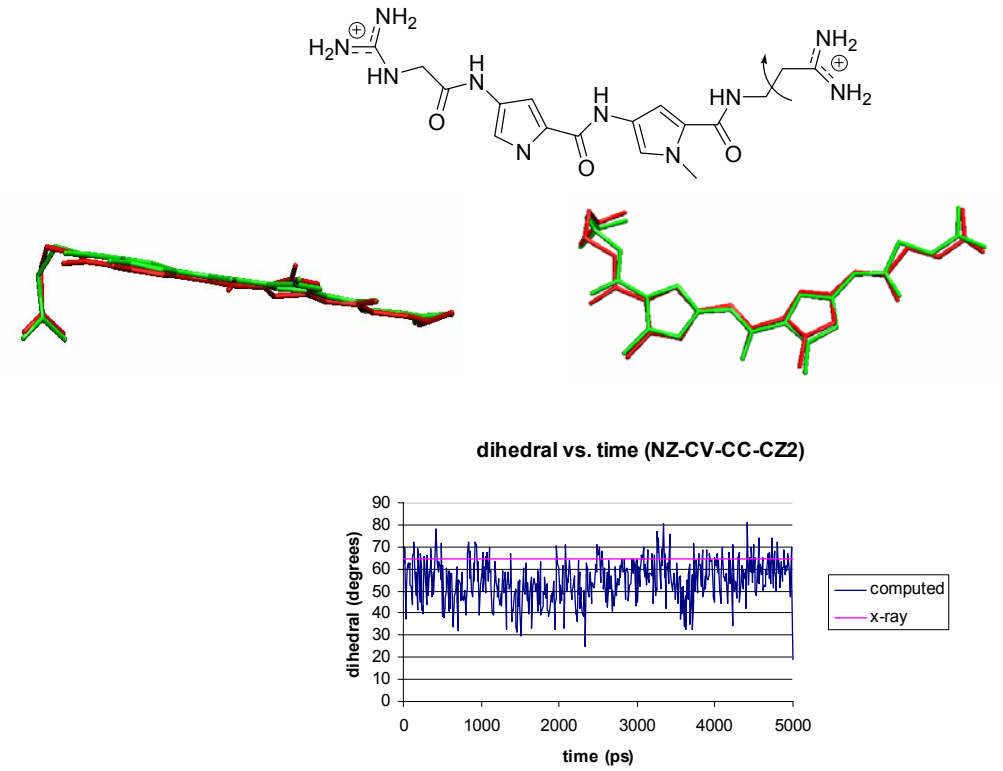


Figure 3.9 Time series (bottom) for one dihedral of the amidino group tail (indicated by arrow on netropsin structure on the top). This end of the molecule appears to contribute most to the RMSD due to rotation about carbon-carbon bonds. The overlay of the x-ray (red) and average structures (green) shows that the simulated structure retains the same orientation of the amidino group.

The x-ray value for the dihedral shown is 65°; the simulation yields 54.7 ± 9.7°. The overlay of the x-ray and average structures shows that even though the dihedrals of the tail show deviation, on average, the tail retains the same orientation as the x-ray structure.

3.2.2. Nucleic Acid Helical Parameters. The next tests of the developed parameters were crystal simulations of DNA/polyamide complexes. An important part of nucleic acid simulation analysis is the examination of helical parameters. Helical parameters describe how the bases and base pairs are oriented with respect to each other and/or the global helical axis of the DNA.[22-25] Because they describe the relative orientation of the bases and base pairs, and a ligand that binds to a nucleic acid will affect these orientations, a successful force field for ligands that bind to nucleic acids should be able to reproduce these parameters accurately.[26] As part of the validation of the new parameters, the helical parameters of the simulated DNA were compared the values of the starting x-ray crystal structure using a t-test at the 95% confidence level.

A complete set of helical parameters (definitions, names, and spatial arrangements of bases and base pairs) was established at a 1988 conference,[25] a standard reference frame has also been adopted,[23] which is used by the Nucleic Acids Database (NDB)[27] for calculating helical parameters. Various computer programs[22, 28-35] have been developed to calculate nucleic acid helical parameters, and although the algorithms used are different, this has been found to have a limited effect on the computed parameters.[24, 36] However, the choice of reference frame was found to have a profound effect on the computed parameters.[24] Figure 3.10 shows the standard reference frame and helical parameters that are calculated by most helical parameter analysis programs.

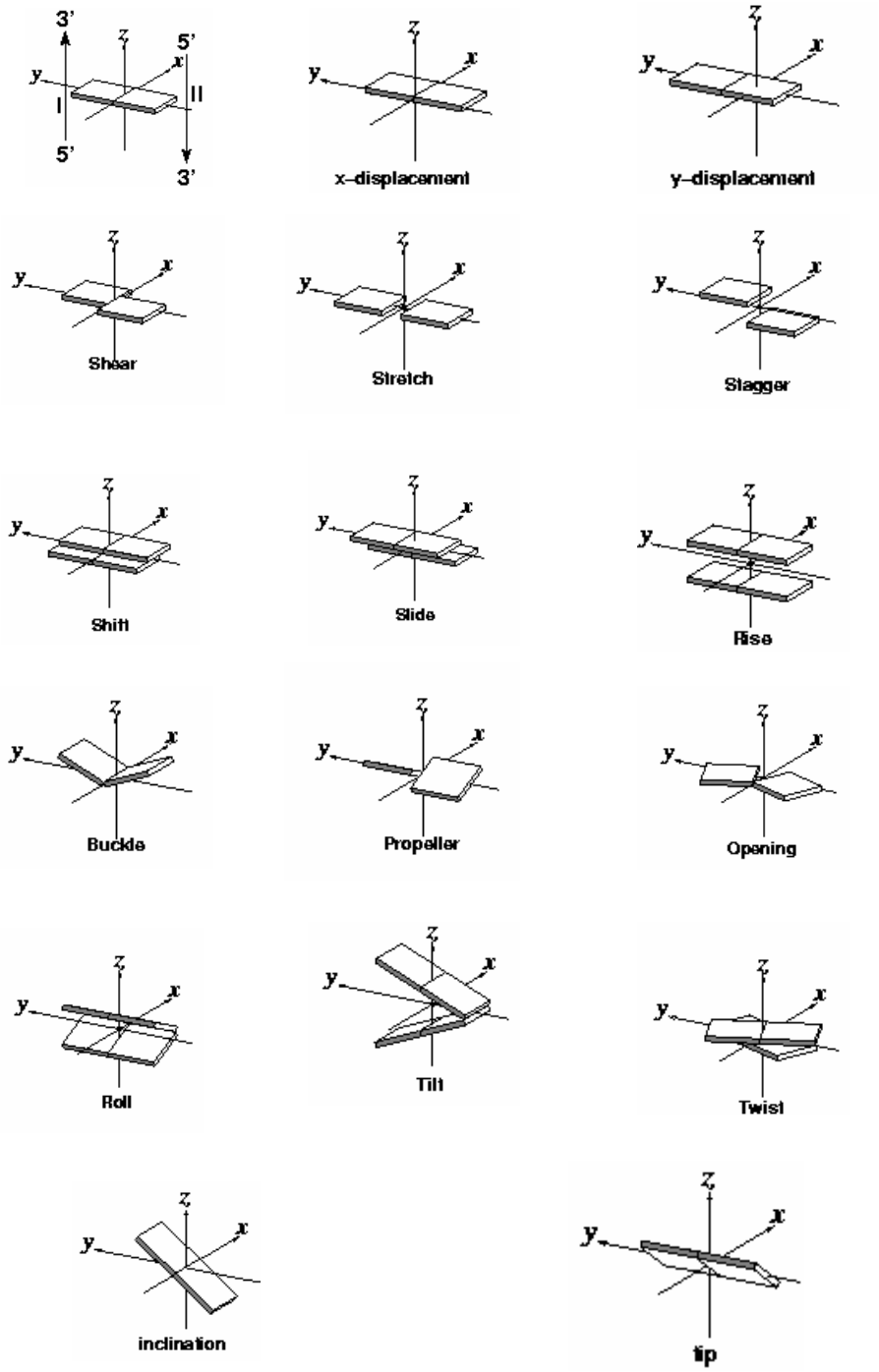


Figure 3.10 Standard reference frame (top left) and helical parameters. In the reference frame, the x-axis points away from the minor groove. Translations are shown in the upper half of the figure (x-displacment through rise) and rotations are shown in the bottom half of the figure (buckle through tip). Each small tile represents a DNA base; two of them together represent a base pair.

The helical parameters can be divided into translational motions and rotational motions; they can also be divided further into three families: base pair-axis parameters, intrabase pair parameters, and interbase pair parameters.

There are two different ways of describing nucleic acid helical parameters, which are termed global and local.[37, 38] In the global approach, the helical parameters calculated describe the overall arrangement of the bases and base pairs with respect to a global helical axis, which is an axis that runs vertically down a double-stranded helix and is usually taken to be linear. The difficulty with a global approach is that a linear helical axis is not an appropriate assumption for many irregular nucleic acid conformations (e.g., a structure with a bend). In the local approach, the helical parameters calculated describe the orientations of the bases/base pairs within the local framework of two successive base pairs along the nucleic acid fragment; a local helical axis is calculated for each base pair step rather than a global helical axis for the entire fragment. There has been no final decision on calculating global versus local helical parameters; both have their advantages and disadvantages.[37] Local parameter algorithms have the advantage that they avoid the difficulty of defining a global helical axis, and they also yield parameters that depend only the conformation of the given base pair step. Global parameters depend on the conformation of the whole fragment and involve the determination of the global helical axis. However, global parameters have the advantage of distinguishing between the different helical conformations (e.g., A-DNA vs. B-DNA). If there is curvature in the fragment, the location and extent can be identified more easily with global parameters.[37] The various available computer algorithms utilize one of the two approaches, but as mentioned, neither method has been found to be truly superior to another.[37] In this

dissertation, the 3DNA program,[28] which uses the local approach, was used for calculation of helical parameters.

3.2.3. DNA/Dervan polyamide complex simulations. Adequate sampling of conformational space is a difficult task, especially for large biomolecules such as DNA in which the potential energy surface contains many local minima.[39-41] In this study, we employ multiple trajectories because other studies have reported that this method improves sampling of conformational space.[42-45] For this system, four initial trajectories were performed at the temperature at which the starting x-ray structure was determined (113 K), and three initial trajectories were performed at 300 K. The theoretical melting temperature for this DNA fragment is 305 K (2°C for AT pairs and 4°C for GC pairs), which will be increased a few degrees by the polyamides in the minor groove, so the melting temperature should be above the simulation temperature.

RMSD. Shown in Figure 3.11 is the RMSD, relative to the starting x-ray crystal structure (PDB[46] id 1CVY[3]); left side of Figure 3.11), of the DNA/polyamide complexes from the seven initial simulations done at the low temperature (113 K) and high temperature (300 K).

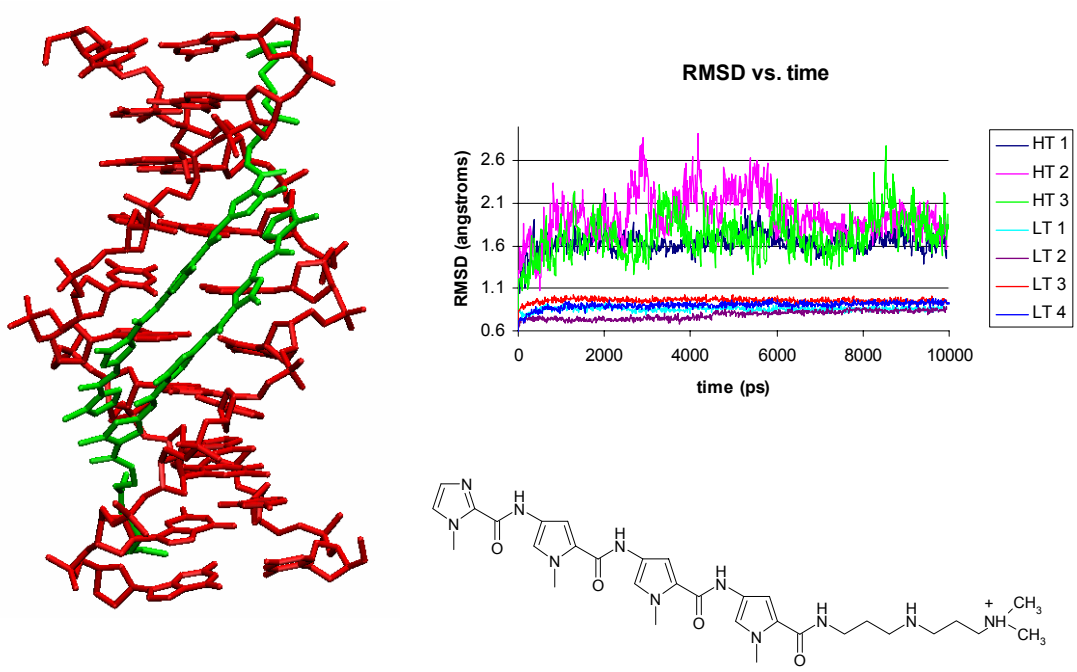


Figure 3.11 Starting structure for DNA/Dervan polyamide simulations (PDB[46] id 1CVY[3]; left). The DNA is in red, and the polyamides are in green. The structure of the polyamides bound in the minor groove is shown on the bottom right (there are two of them bound antiparallel in the minor groove).The plot on the top right is the RMSD vs. time for the 113 K simulations (bottom) and for the 300 K simulations (top) compared to the starting x-ray crystal structure.

The RMS deviation converges quickly to 0.8-1 Å for all of the low temperature simulations, with the exception of one low temperature simulation (LT2), which converges with the rest of the low temperature simulations at ~ 5 ns. As expected for the 300 K simulations, the RMSD shows more fluctuation than for the lower temperature simulations. The RMSD stays below 3 Å, but there is great amount of fluctuation up to ~6 ns of simulation time. After 6 ns, the RMS deviations start to converge to around 1.7 Å, with the exception of one simulation (HT3), which shows deviations greater than 2.5 Å. Table 3.1 lists the average RMSDs of the complex, DNA, and polyamides separately.

Table 3.1 RMSDs of complexes, DNA, and polyamides

		RMSD (Å)		
simulation	Temp (K)	complex	DNA only	polyamides only
HT1	300.0 ± 0.83	1.64 ± 0.13	1.15 ± 0.15	2.52 ± 0.33
HT2	300.0 ± 0.84	1.95 ± 0.27	1.87 ± 0.28	1.68 ± 0.24
HT3	300.0 ± 0.83	1.71 ± 0.24	1.65 ± 0.24	1.42 ± 0.25
LT1	113.0 ± 0.29	0.96 ± 0.04	0.89 ± 0.04	1.13 ± 0.06
LT2	113.0 ± 0.29	0.86 ± 0.04	0.73 ± 0.04	1.01 ± 0.05
LT3	113.0 ± 0.30	0.79 ± 0.05	0.58 ± 0.04	0.95 ± 0.15
LT4	113.0 ± 0.31	0.90 ± 0.05	0.83 ± 0.05	1.00 ± 0.05

For all of the low temperature simulations and one high temperature simulation, the RMSD for the complex is influenced specifically from the tails of the polyamide, which contain C-C single bonds with free rotation. In the remaining two high temperature simulations, the end base pairs of the DNA exhibit some fraying, which is the most likely cause of higher RMSD contribution of the DNA.

Statistical Analysis of Helical Parameters. Helical parameters were computed using the 3DNA program.[28] The helical parameters were computed and then averaged from snapshots from the simulations. A t-test was then performed between the computed and x-ray helical parameters at the 95% confidence level (see Chapter 2, Section 2.2.2 for a description and formulas for the t-test). The output from 3DNA gives a value for each base or base pair in the DNA fragment (depending on what parameter is being measured); therefore, there is an average and standard deviation associated with each parameter for each snapshot. Figure 3.12 shows a sample output from 3DNA.

	bp	Shear	Stretch	Stagger	Buckle	Propeller	Opening
1	C-G	0.71	-0.26	-0.27	9.75	-18.27	-3.72
2	C-G	0.67	-0.12	-0.01	6.20	-14.32	1.72
3	A-T	-0.13	0.08	-0.50	-15.35	-25.94	26.90
4	G-C	-0.21	-0.08	-0.19	-21.65	-8.44	2.00
5	A-T	0.03	-0.23	0.30	-5.25	-4.01	-3.27
6	T-A	0.11	-0.18	-0.08	-5.73	-10.23	0.31
7	C-G	0.10	-0.12	0.05	-15.78	-3.56	-0.07
8	T-A	0.17	0.53	-0.27	14.11	-18.42	24.29
9	G-C	-0.41	-0.32	-0.35	-6.05	-5.31	-2.27
10	G-C	-0.64	0.16	-0.67	-10.19	-7.20	5.08
	ave.	0.04	-0.05	-0.20	-4.99	-11.57	5.10
	s.d.	0.42	0.25	0.28	11.73	7.43	11.14

Figure 3.12 Sample output from 3DNA for a DNA 10-mer for six helical parameters.

An average of each helical parameter was taken over all snapshots used for a simulation. In order to take the standard deviation from each snapshot in account, a pooled standard deviation was performed for each helical parameter. For each 10 ns simulation, 100 snapshots, or samples, (one every 100 ps) were evaluated using 3DNA, and an overall average and a pooled standard deviation (Eqn.3.1) was calculated for each helical parameter.

$$s_{pooled} = \sqrt{\frac{s_1^{\,2}(n_1-1) + s_2^{\,2}(n_2-1) + \ldots + s_k^2(n_k-1)}{n_1 + n_2 + \ldots + n_k - k}} \qquad \text{Eqn. 3.1}$$

n =10 for each snapshot (n =9 in the case of base pair step parameters), s is the standard deviation for each snapshot taken from a simulation, and k =100 (100 snapshots used).

After the pooled standard deviation was calculated, the t-test comparing the simulation averages to the x-ray values was performed for each helical parameter.

$$t_{calc} = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{(s_1^{\,2}/n_1) + (s_2^{\,2}/n_2)}} \qquad \text{Eqn. 3.2}$$

x_1 is the simulation average, x_2 is the x-ray average, s_1 is the simulation standard deviation (this number is the number obtained from the pooled standard deviation), s_2 is the x-ray standard deviation (obtained from the 3DNA output for the x-ray structure), n_1 is the number of samples for the simulation, and n_2 is the number of samples for the x-ray structure. For the simulations, n_1 = 1000 (100 snapshots * 10 values) for individual base pair parameters, or n_1 = 900 (100 snapshots * 9 values) for base pair step parameters (the DNA/Dervan polyamide systems contained 10 base pairs). For the DNA/netropsin complex, the DNA/distamycin complex, and DNA with no bound polyamides, n_1 = 1200 for individual base pairs, and n_1 = 1100 for base pair steps (these systems contained 12 base pairs). The number of samples (n_2) for the x-ray structure was set to 10 (1 structure * 10 values); 9 was used for the helical parameters that involve base pair steps rather than individual base pairs (in the case of the DNA/netropsin complex, DNA/distamycin complex, and DNA with no bound polyamides, 12 and 11 were used).

T-test results for Helical Parameters. According to the t-test, there is good agreement between the helical parameters calculated from both the low and high temperature simulations and the x-ray structure, where 17 out of the 18 helical parameters measured are not statistically significant according to the t-test. The only helical parameter that was significantly different in all seven individual simulations is stagger at the 95% confidence level. Tables of the t-test results for the individual simulations can be found in Appendix A.

Because stagger was the only helical parameter that was found to be statistically significant in all seven individual simulations, t-tests were performed on the combined sets of simulations (one t-test performed on the low temperature simulations and one t-

test performed on the high temperature simulations) to see if the combined t-test result would be non-significant. The t-tests were performed using Eqn. 3.3,

$$t_{calc} = \frac{\overline{x_1} - \overline{x_2}}{s_{pooled}} \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \qquad \text{Eqn. 3.3}$$

where x_1 is the simulation average, x_2 is the x-ray value, s_{pooled} is the pooled standard deviation of the simulations and x-ray standard deviation, n_1 is the number of samples used from the simulations, and n_2 is the number of samples used for the x-ray structure. In order to see if the combined t-test results and how the t-values fluctuated over time, the t-test was performed in the following way. Rather than taking the average value of stagger over the ten bases of the DNA fragment for each snapshot, all ten values from the 100 snapshots were used for the t-test for each simulation. Because this t-test was performed over multiple simulations, the t-test for the high temperature simulations used 300 snapshots, and the t-test for the low temperature simulations used 400 snapshots. For comparison to the x-ray structure, the x-ray values of stagger for each base were compared to the corresponding bases from the simulations rather using the average over the ten bases. The snapshots from the simulations were ordered chronologically, the stagger value for each base at each snapshot was averaged over all simulations considered, and a t-test was then performed between each snapshot and the x-ray value. Because the stagger values were averaged over all simulations considered, there were 100 snapshots and thus a total of 100 t-tests performed. The resulting time series plots showed how the t-values for stagger evolve with time.

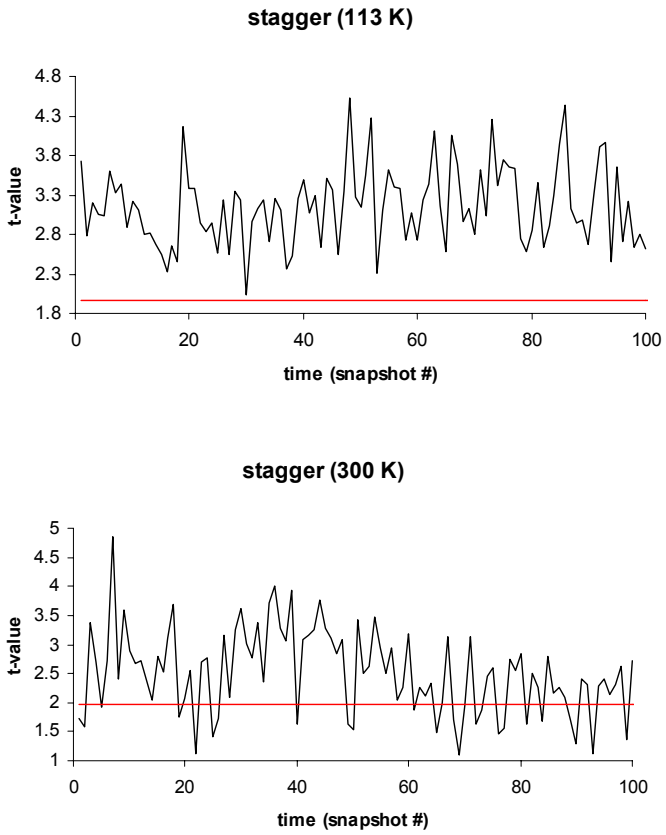


Figure 3.13 Time series plots of t-values over time for the stagger helical parameter. The time series for the low temperature simulations is on the top; the time series for the high temperature simulations is on the bottom. The red lines denote the t-table value (1.96) at the 95% confidence level.

Stagger is also observed to be statistically significant when all low temperature simulations are examined together and when all high temperature simulations are examined together. In the time series plot for the low temperature simulations, all 100 t-tests performed yield a t-value greater than the t-table value (1.96) at the 95% confidence level. The average of the t-values is 3.16; stagger is statistically significant when all low temperature simulations are considered together. For the high temperature simulations, the average of the t-values is 2.51, so stagger is still statistically significant when all high temperature simulations are considered together. For comparison, the t-test results of a helical parameter (buckle) that did not exhibit any statistical significance in the individual

simulations is shown in Figure 3.14. Over the four low temperature simulations, all t-values are less than 1.96, and the average t-value is 0.93. For the high temperature simulations, all t-values are again less then 1.96, and the average t-value is 0.64.

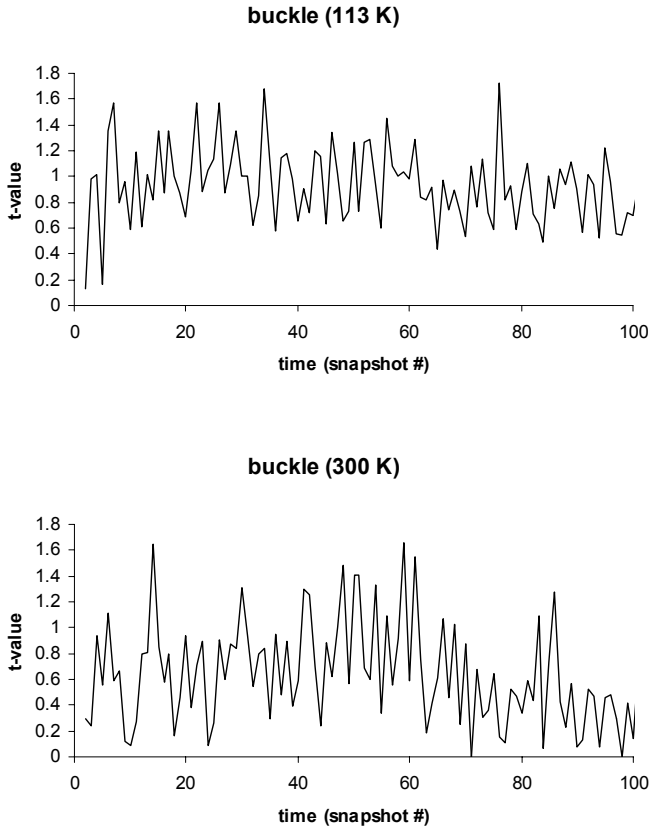


Figure 3.14 Time series plots of t-values over time for the buckle helical parameter. The time series for the low temperature simulations is on the top; the time series for the high temperature simulations is on the bottom. None of the individual or combined simulations exhibited significance with respect to buckle.

Principal Components Analysis (PCA). Neither the high nor the low temperature simulations appear to be sampling the x-ray value for the stagger parameter properly. In order to examine where the simulations are sampling with respect to the x-ray value of stagger and which simulations sample the x-ray value best (or not at all), PCA was performed on the stagger parameter for both the high and low temperature simulations (Figure 3.15).

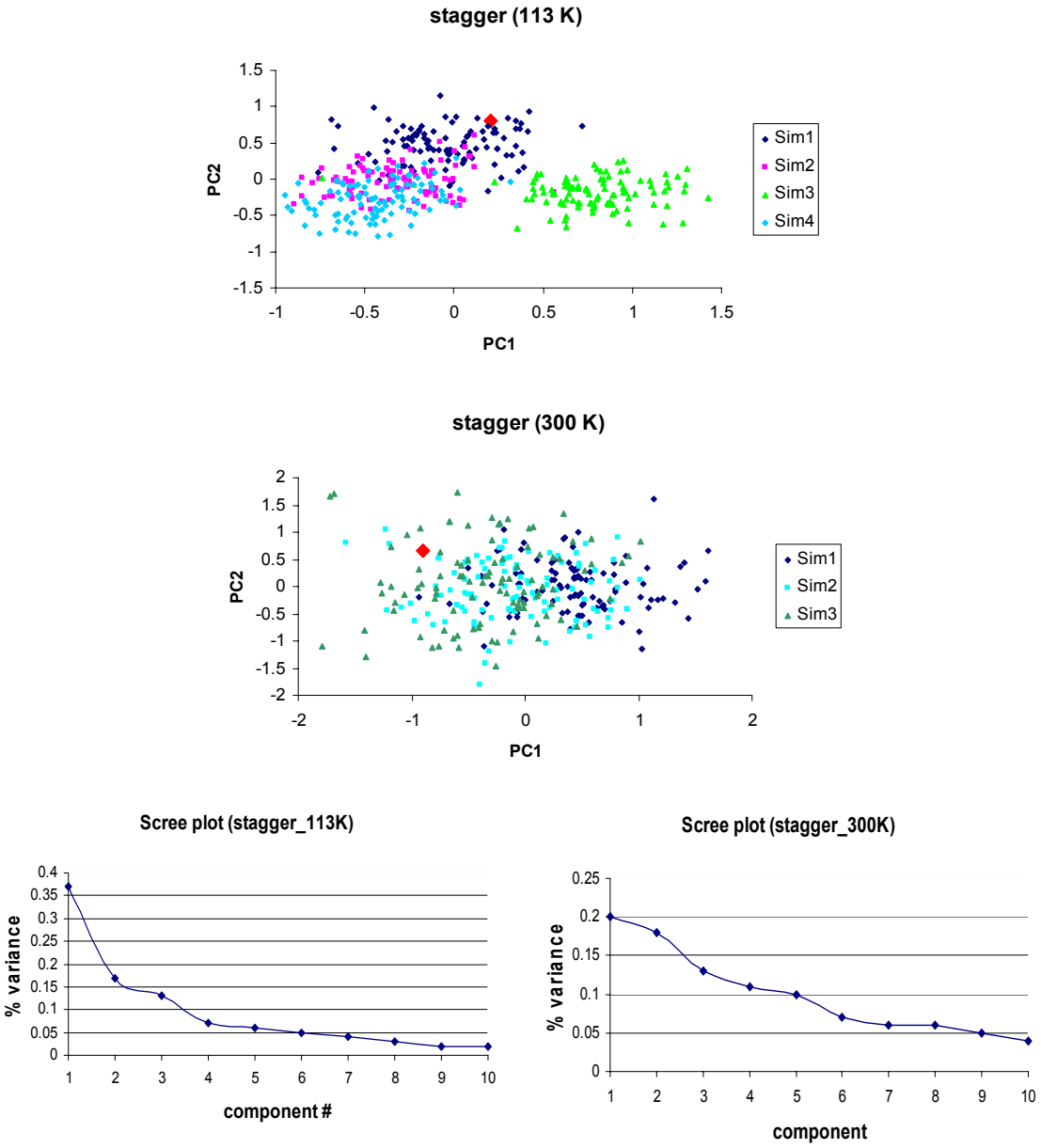


Figure 3.15 Principal component analysis plots of stagger (low temperature simulations, top; high temperature simulations, middle). The different colors represent the different simulations. The red diamond on each PCA plot represents the x-ray value. The Scree plots associated with the PCA plots are shown at the bottom (low T, left; high T, right).

Upon examination of the PCA plots, the x-ray point falls outside of the regions where the simulations are sampling. Three of the low temperature simulations (green, light blue,

and most of the pink) are not even in the region of where the x-ray value falls; the dark blue simulation is sampling right on the edge of where the x-ray value falls. The high temperature simulations seem to better sample the region of conformational space where the x-ray value falls for stagger; however, the x-ray value lies on the edge of the sampling area. The PCA plots confirm that the simulations simply are not sampling enough conformational space to sample the x-ray value.

For comparison, PCA was also carried out for the buckle parameter, which did not show statistical significant using the t-test. For the low temperature simulations (Figure 3.16, top) the x-ray point falls in the middle of the region around where each of the simulations sample; the simulations average out to sample the x-ray value. For the high temperature simulations (Figure 3.16, middle), the x-ray point falls in the middle of the region where the simulations sample. The PCA plots show that x-ray value of buckle is sampled efficiently by both sets of simulations, unlike stagger.

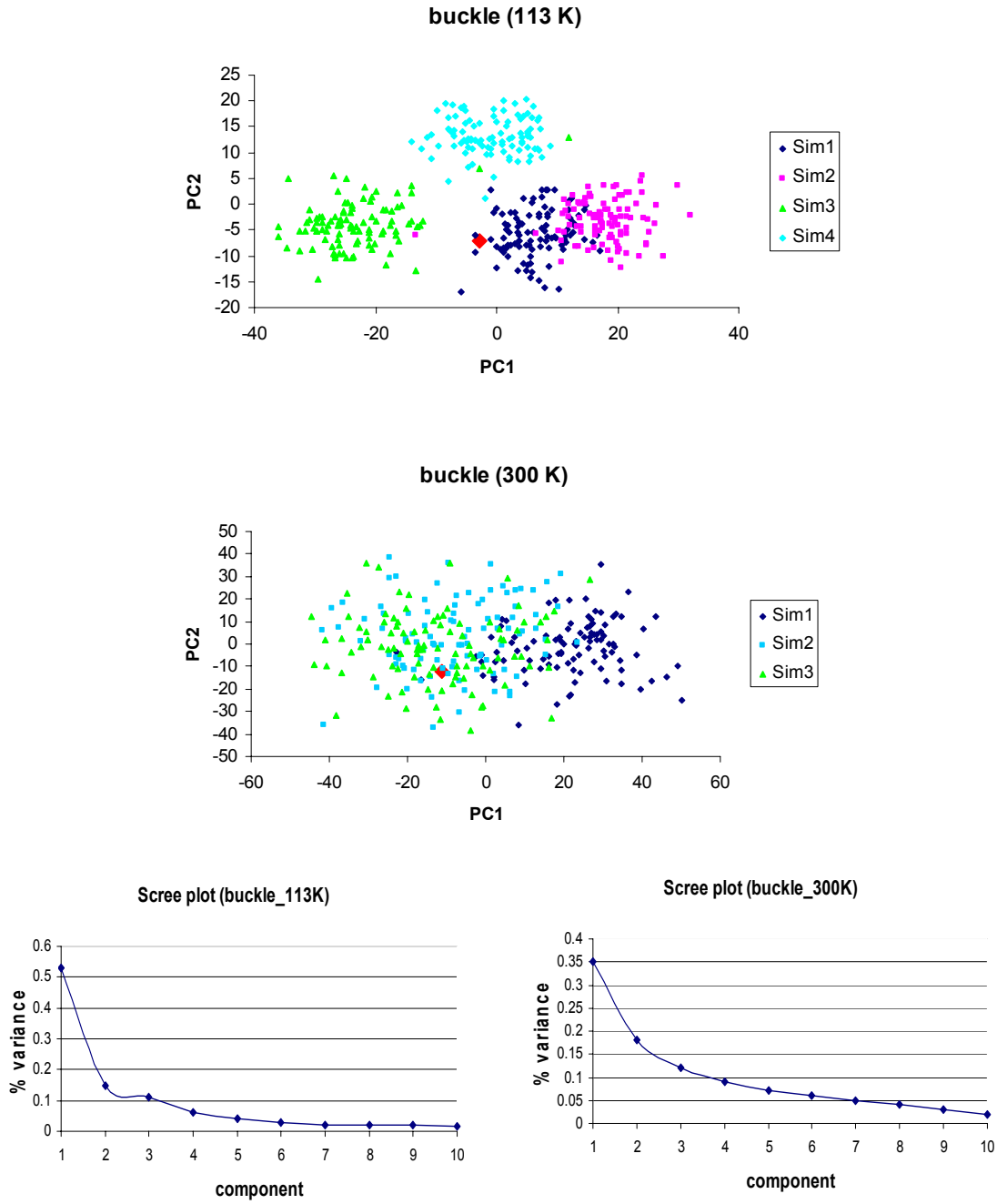


Figure 3.16 Principal component analysis plots of buckle (low temperature simulations, top; high temperature simulations, middle). The different colors represent the different simulations. The red diamond on the PCA plots represents the x-ray value. The Scree plots associated with the PCA plots are shown at the bottom.

Additional Simulations. To help improve sampling of conformational space of the stagger parameter, additional simulations were performed (four new simulations at the lower temperature (113 K) and three new simulations at the higher temperature (300 K)). The simulations were started from structures that were in regions not well sampled by the simulations on the PCA plots for stagger (each started from different initial velocities). Shown in Figure 3.17 is the RMSD vs. time for the DNA/polyamide complexes for the additional simulations.

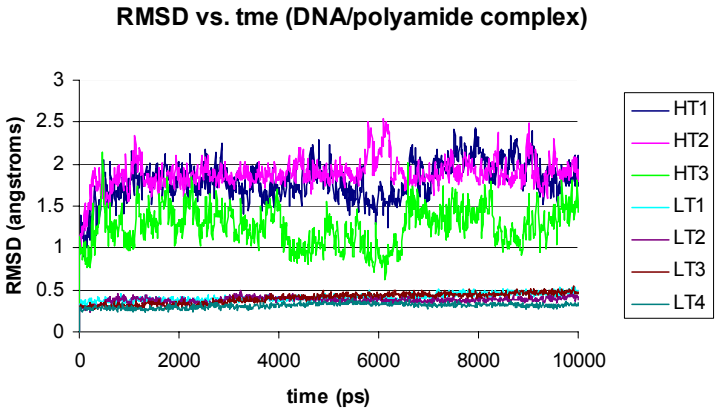


Figure 3.17 RMSD vs. time plots for the 113 K simulations (bottom) and for the 300 K simulations (top) compared to the starting structures for the simulations.

The RMS deviation converges quickly to ~0.5 Å for all of the low temperature simulations at ~ 4 ns. As expected for the 300 K simulations, the RMSD shows more fluctuation than for the lower temperature simulations. There is a great amount of fluctuation during the simulations, with HT3 having the most fluctuation. However, the RMSD does not rise above ~2.5 Å for any of the simulations, and HT3 has the lowest RMSD at ~1-1.5 Å. Table 3.2 lists the average RMSDs of the complex, DNA, and polyamides separately.

Table 3.2 RMSDs of complexes, DNA, and polyamides

		RMSD (Å)		
simulation	Temp (K)	complex	DNA only	polyamides only
HT1	300.0 ± 0.82	1.80 ± 0.24	1.76 ± 0.22	1.56 ± 0.57
HT2	300.0 ± 0.85	1.88 ± 0.18	1.81 ± 0.18	1.80 ± 0.25
HT3	300.0 ± 0.83	1.26 ± 0.23	1.29 ± 0.25	0.86 ± 0.16
LT1	113.0 ± 0.30	0.36 ± 0.04	0.42 ± 0.05	0.28 ± 0.04
LT2	113.0 ± 0.28	0.38 ± 0.04	0.38 ± 0.04	0.30 ± 0.03
LT3	113.0 ± 0.32	0.40 ± 0.06	0.42 ± 0.07	0.27 ± 0.05
LT4	113.0 ± 0.31	0.31 ± 0.03	0.31 ± 0.03	0.25 ± 0.03

For two of the high temperature simulations and three of the low temperature simulations, the DNA contributes most to the RMSD of the complex. Some fraying of the end base pairs was observed, which is most likely the cause of the higher contribution of the DNA. In the remaining simulations, the polyamides contribute more to the RMSD which is due to the tails of the polyamides (free rotation is allowed about the C-C single bonds of the tails).

Low-Temperature Simulations. A total of eight simulations were performed at 113 K (four original and four new). All individual simulations with the exception of one showed that stagger is statistically significantly different from the x-ray structure (all other helical parameters are non-significant; see Appendix A for tables of helical parameter and t-values). When all eight simulations were combined together, the stagger parameter is still statistically significant according to the t-test (Figure 3.18, top). All t-values are greater than 1.96; the average t-value is 3.11 (this is a slight improvement over the four initial simulations, which had an average t-value of 3.16). A PCA plot of the stagger parameter shows that the four new simulations do improve sampling; more simulation points are in the region where the x-ray value (red diamond) lies. However,

some simulations are still not sampling even near where the x-ray value lies (e.g., the green, light blue, and pink simulations). Even when these three simulations were not considered in the analysis, the t-test still yielded statistical significance (average t-value was 3.04; plot not shown). In the case of the low temperature simulations, even the additional sampling is not sufficient because according to the t-test, stagger is still statistically significant from the x-ray structure at the 95% confidence level.

High-Temperature Simulations. A total of six simulations were performed at 300 K (three original and three new). In the original three individual simulations, stagger is the only statistically significant helical parameter; in the three new simulations, none of the helical parameters are statistically significant. When the six simulations were considered together, the t-test yields statistical significance at the 95% confidence level, but not at the 97% confidence level ($t\text{-table}_{97\%}$ =2.24) (Figure 3.19; top). The average t-value is 2.03, which is just above 1.96; however, this value is less than 2.24. The additional high temperature simulations did improve the average t-value (the three original simulations yielded an average t-value of 2.51), and the improvement observed is greater than that for the low temperature simulations. A PCA plot of stagger for the six simulations shows that the additional simulations improve the sampling of stagger (Figure 3.19, middle).

Because the six simulations together produced an average t-value that was close to the t-table value at the 95% confidence interval, only the three new high temperature simulations were then considered together for the t-test. With only the three new simulations, the t-test result for stagger is overall non-significance (Figure 3.20, top) at the 95% confidence level (average t-value is 1.64). Upon examination of a PCA plot of

stagger for only the three new simulations (Figure 3.20, middle), the simulations do sample the x-ray value more efficiently than the first three simulations and even when all six simulations are averaged together. For the three new simulations, all of the simulations appear to sample values that fall in the region of the x-ray value.

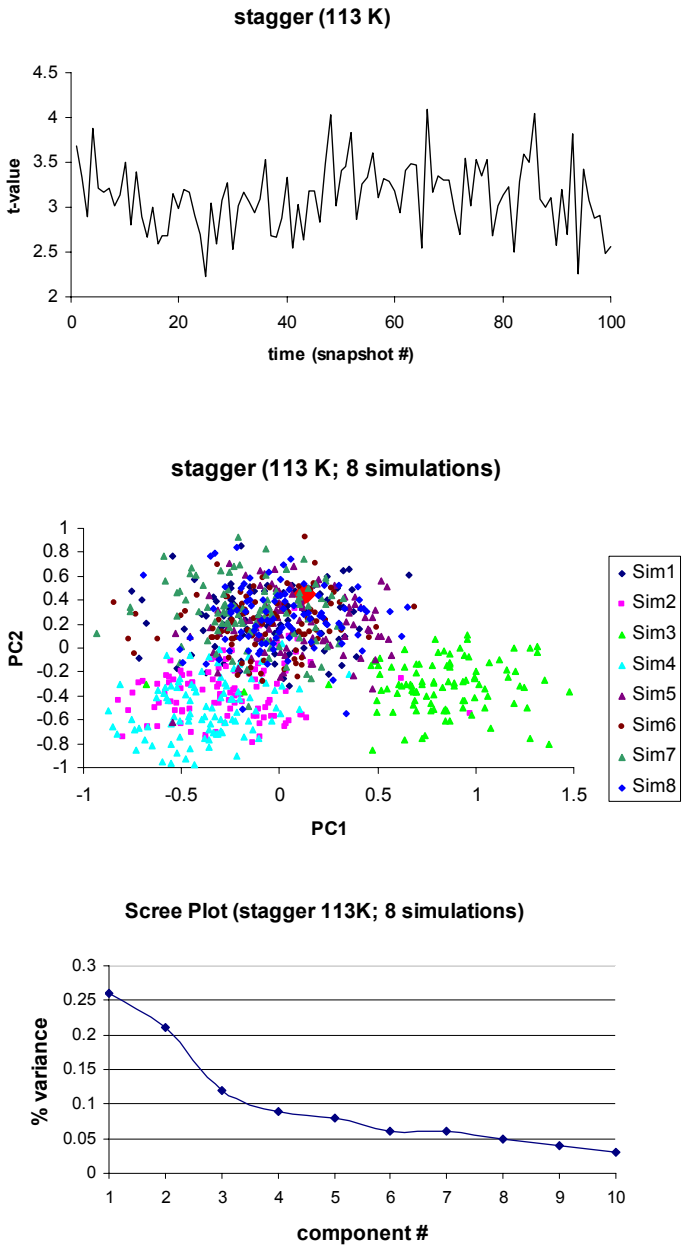


Figure 3.18 T-test for all eight low temperature simulations (top); PCA plot of all eight simulations (x-ray value is shown as a red diamond, middle); Scree plot corresponding to PCA plot (bottom).

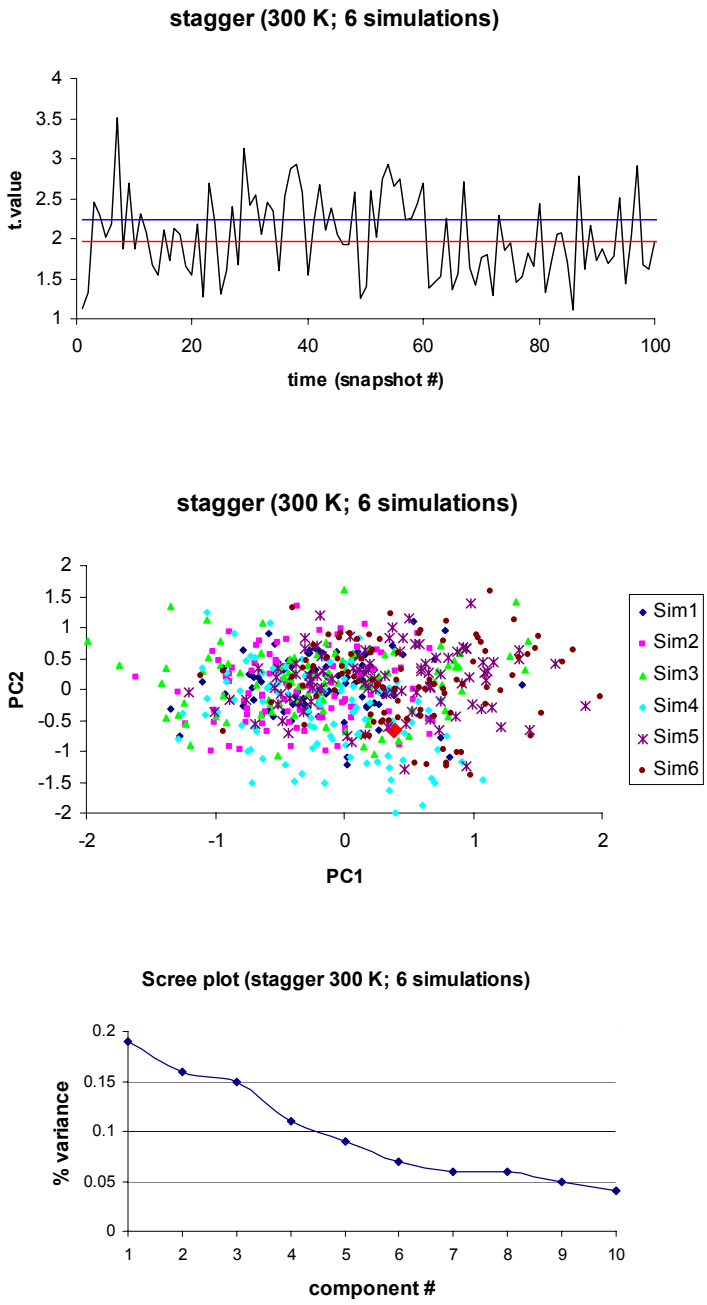


Figure 3.19 T-test for all six high temperature simulations (top). The t-table value at the 95% CL (1.96) is marked by the red line, and the t-table value at the 97% CL (2.24) is marked by the blue line. PCA plot of all six simulations (x-ray value is shown as a red diamond, middle); Scree plot corresponding to PCA plot (bottom).

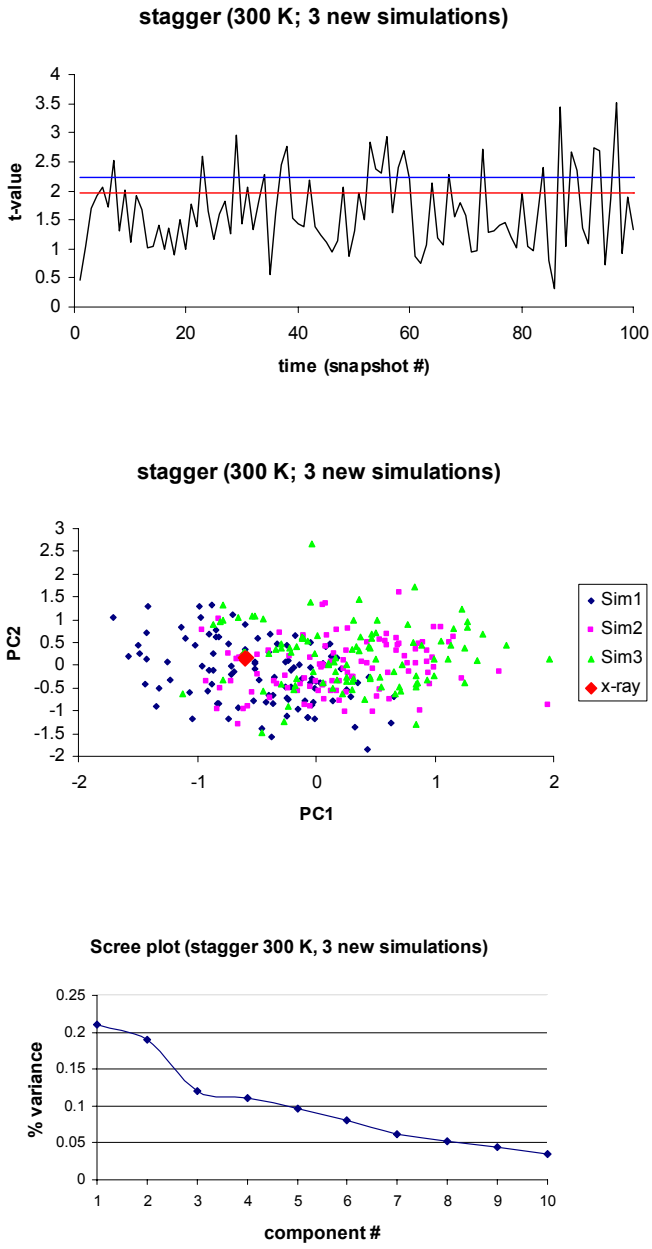


Figure 3.20 T-test for three new high temperature simulations (top). The t-table value at the 95% CL (1.96) is marked by the red line, and the t-table value at the 97% CL (2.24) is marked by the blue line. PCA plot of three new simulations (x-ray value is shown as a red diamond, middle); Scree plot corresponding to PCA plot (bottom).

The additional simulations improve the sampling in both the low and high temperature simulations. However, the low temperature average t-value is only improved upon

marginally (3.16 vs. 3.11), whereas in the high temperature simulations, a larger improvement is observed (2.51 vs. 2.03). From the average t-value and the PCA plot, the additional sampling with the low temperature simulations is not sufficient to obtain non-statistical significance for the stagger parameter. The high temperature simulations do achieve non-significance. All six simulations together show statistical significance at the 95% confidence level but not at the 97% confidence level, which is already an improvement over the initial three simulations. When only the three new simulations are considered, stagger is not statistically significant at the 97% or 95% confidence level. As expected, the temperature at which the simulations are run influence the sampling of stagger for this particular system, and the higher temperature is the obvious choice for better sampling.

Although the high and low temperature simulations yielded a statistically significant t-value when considered by temperature, all 14 (8 low temperature and 6 high temperature) simulations gave a non-significant t-value for all helical parameters when considered together. The figure below shows the value of the t-test over time for all 14 simulations. The average t-value was 1.04. None of the other helical parameters were statistically significant over all 14 high or low temperature simulations (see Appendix A for a table of t-values).

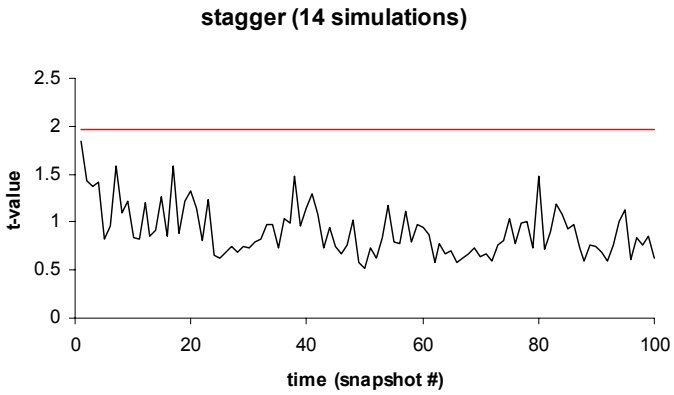


Figure 3.21 T-test for combined high and low temperature simulations. The t-table value at the 95% CL (1.96) is marked by the red line.

Although the t-test for stagger displays non-significant results when the 14 simulations are combined, the high temperature simulations are obviously contributing more to the non-significance. The additional high temperature simulations showed more improvement in the t-test results than the low temperature simulations. The low temperature simulations did not sample enough to get the x-ray value; however, the additional high temperature simulations do improve sampling. However, what is different about the high-temperature structures that lie close to the x-ray value and those that are far away from it on the PCA plot for stagger? Are there conformations/structural features of the DNA that cause stagger to deviate in the structures that are far away from the x-ray value? To answer these questions, structures that lay farthest from the x-ray value on the PCA stagger plot were examined by RMSD analysis, and the largest contribution to RMSD in structures was the DNA bases (not including backbone or sugars). An example of one of the simulated structures is shown below in Figure 3.22.

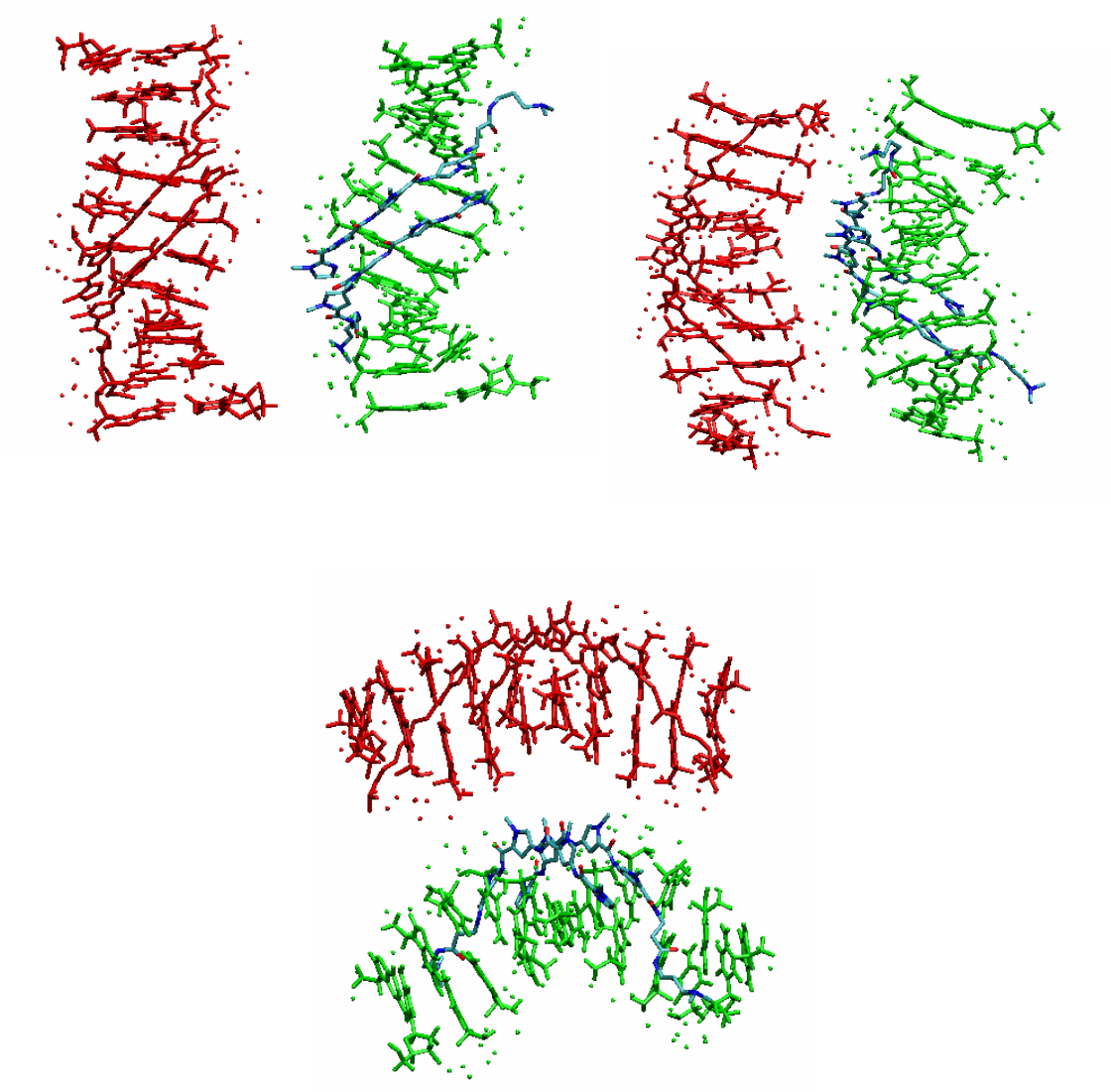


Figure 3.22 Comparison of x-ray structure and high temperature simulated structure. The x-ray structure is in red; polyamides are bound in the minor groove (also in red). One of the simulated structures that was far away from the x-ray point on the stagger PCA plot is shown with the DNA in green with the polyamides bound in the minor groove in colors. Gaps appear in the structures because only the bases of the DNA are shown (backbone of the DNA is not shown for clarity). The simulated structure has a wider minor groove and a more bent helix than the x-ray structure does.

The simulated structure is obviously distorted and has a wider minor groove (Figure 3.22, top left) and a more bent helix (Figure 3.22, bottom) than the x-ray structure does. The polyamides appear to have been "pushed" out the groove; the polyamides in the x-ray structure are deeper inside the groove than those in the simulated structures. The top

right picture shows the view from the major groove instead of the minor groove. This view is just to show that the polyamides really are not coming out of the groove (they look like like they are in the top left picture). The hydrogen bonding contacts are still where they should be; the polyamide tail parts that are flipped out of the groove appear to make hydrogen bonds with nearby waters. The distortion of the helix that is observed is why structures that look like this are farthest from the x-ray point on the stagger PCA plot.

The tilted/bent structures that are observed exhibit characteristics of A-DNA.[47-49] In A-DNA, the bases are more tilted with respect to the helical axis, which is seen in the structures. Also, the average inclination and x-displacement of the bases in those structures are closer to those of A-DNA, which are around 20° and -4 Å, respectively. A-DNA has a more shallow and a wider minor groove than B-DNA. The structures where the polyamide is getting "pushed out" of the minor groove are in line with A-form; the DNA adopts features of A-form, and the polyamide gets pushed out, because the groove has become narrower, and it also appears wider from the snapshots. However, the structures have not completely converged to the A-form; the rise (~2.6 Å in A-DNA and 3.4 Å in B-DNA) and twist helical parameters (31° in A-DNA and 36° in B-DNA), the distance between the phosphates in each strand (~5.0 Å for A-DNA and ~7.0 Å for B-DNA, and the width of the helix (26 Å in A-DNA and 20 Å in B-DNA) in these structures have B-form (or closer to B-form) values. An overstabilization of A-DNA was previously observed the CHARMM22 force field for nucleic acids;[50] this has since been corrected with revised parameters,[16, 26] which were used in this study, so the force field parameters should not be an issue.

Structures from the low temperature simulations were also examined to see if they exhibited the same distortion of the DNA helix. The low temperature structures do not exhibit the widened minor groove or the distorted helix, and the polyamides stay where they are; they are not pushed out of the groove as in the high temperature structures.

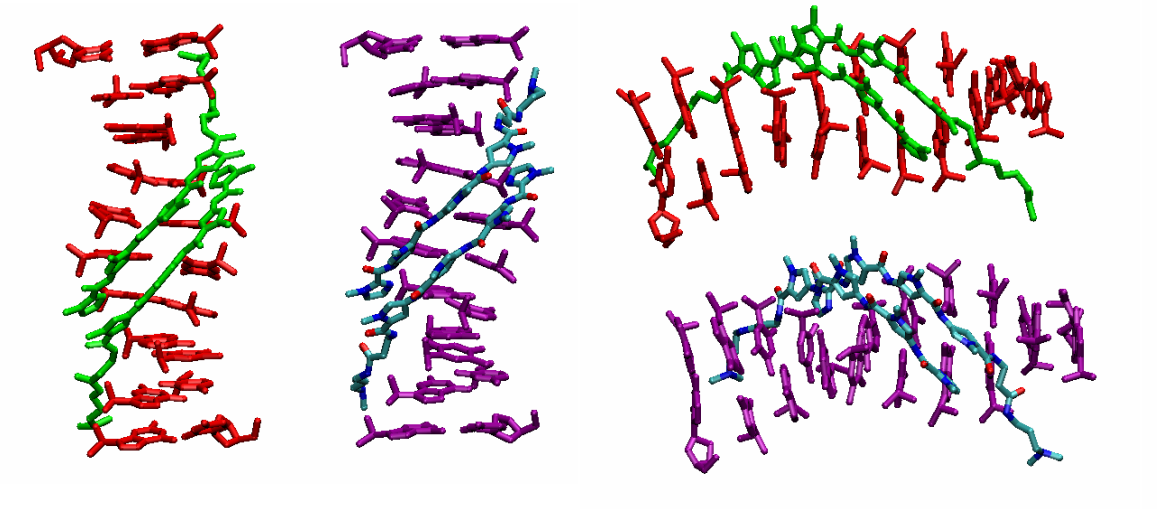


Figure 3.23 Comparison of x-ray structure and low temperature simulated structure. The x-ray structure is in red; polyamides are bound in the minor groove (green). A representative example of a simulated low temperature structure is shown with the DNA in purple with the polyamides bound in the minor groove in colors. Gaps appear in the structures because only the bases of the DNA are shown (backbone of the DNA is not shown for clarity).

To further examine the high and low temperature structures, histograms were made based on the minor groove width of the structures. The histogram from the eight low temperature simulations (Figure 3.24) show that a few structures have a wider groove, but most have a narrower minor groove (~13.0-13.6 Å), which is consistent with the representative structure shown in Figure 3.23 and the other structures examined from the low temperature simulations. The histogram from the six high temperature simulations (Figure 3.25) shows approximately a normal distribution of minor groove

widths, with most of them being wider (~14.0-15.0 Å) than the low temperature simulations.

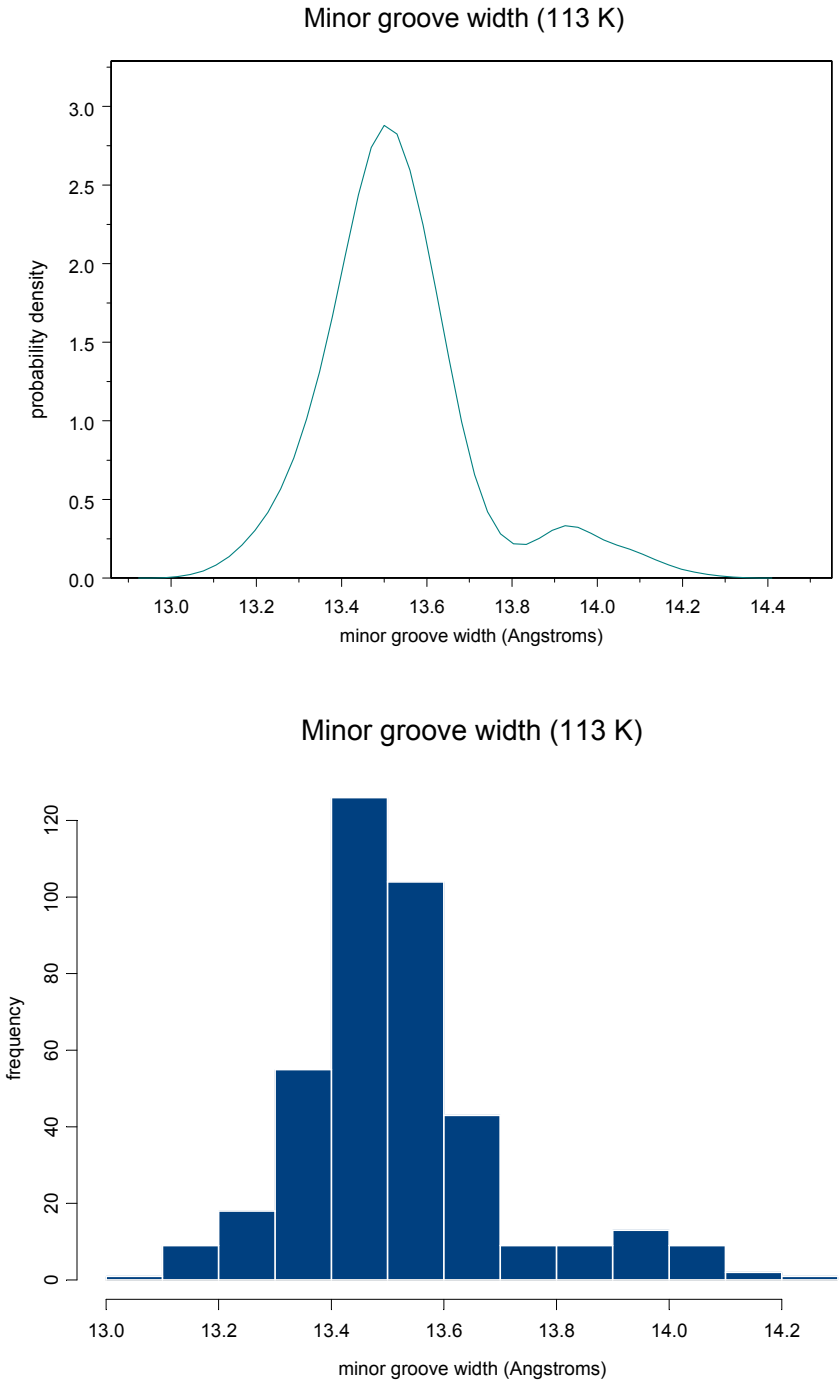


Figure 3.24 Probability density curve (top) and histogram bar plot (bottom) for minor groove width of low temperature simualtions.

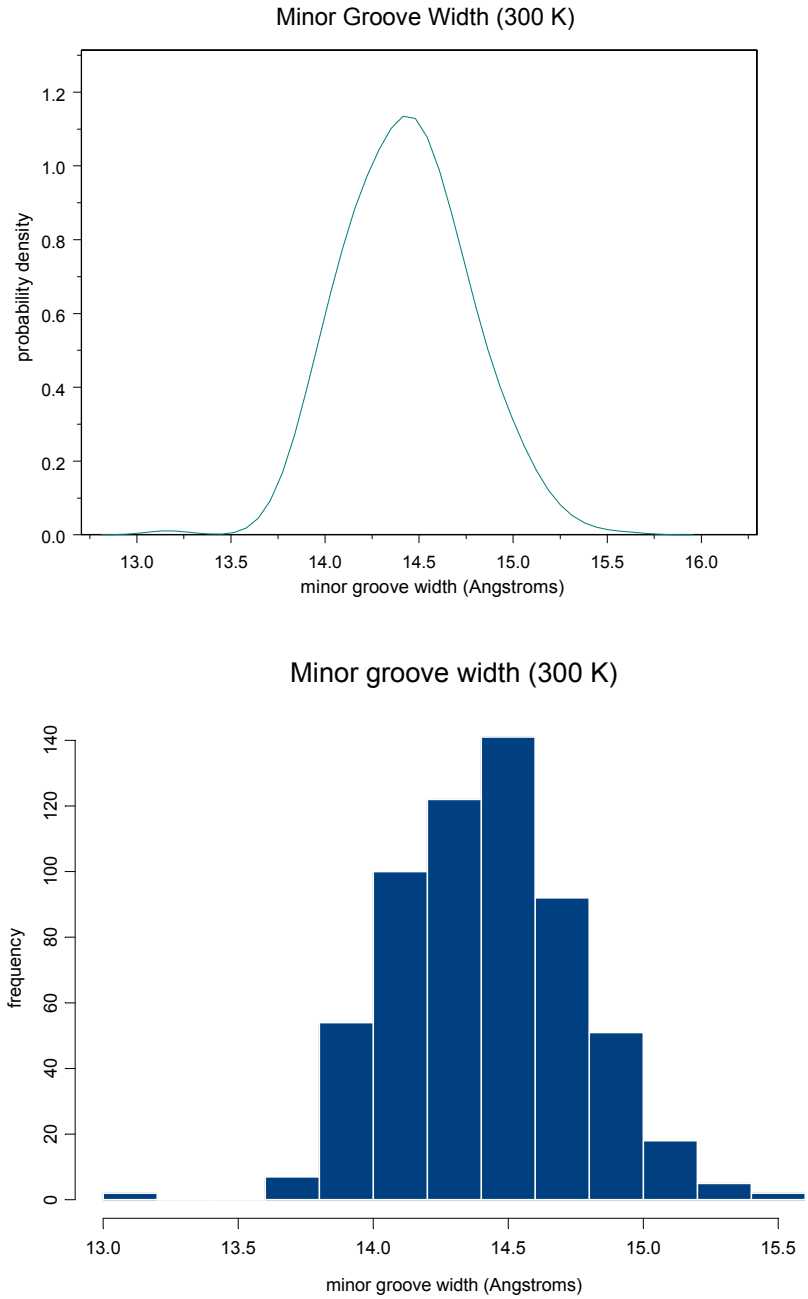


Figure 3.25 Probability density curve (top) and histogram bar plot (bottom) for minor groove width of high temperature simualtions.

Because the low temperature simulations do not show the distortion of the helix, and most of the minor groove widths are narrower, the B-form is retained. However, in the

high temperature simulations, almost all structures have the wider minor groove that is characteristic of A-form DNA. Why are the distortion and the A-like features observed only in the high temperature structures? Also, because more distortion is observed in the high temperature simulations, why is more statistical significance not observed with the helical parameters (the high temperature simulations actually show better results for stagger than the low temperature)?

Force fields often do not get melting temperatures correct.[51-56] This particular fragment of DNA has a theoretical melting temperature of 305 K; the melting temperature of the DNA/polyamide complex is probably a few degrees higher. Polyamides have been found to increase the melting temperature[57, 58] of the DNA alone; the magnitude of the increase depends on the sequence of the polyamide, the DNA fragment, and how tightly the polyamide binds to the fragment.[59] The simulation temperature for the high temperature simulations is 300 K. The simulation temperature is too close to that of the melting temperature, and because the force field does not account perfectly for the melting temperature, the DNA/polyamide complex is starting to melt in the high temperature simulations. The melting causes the distortion and thus the A-like features, which makes the minor groove wider and narrower, which in turn pushes the polyamides out. As far as the sampling is concerned, the higher temperature simulations do sample better and thus less statistical significance is observed for the stagger parameter. The low temperature simulations yield solid and undistorted structures, but they just do not sample well.

Polyamide/DNA interaction. Polyamides bind to the minor groove of DNA by hydrogen bonding, so these contacts were examined to ensure that they stayed in place

over time during the simulations and that the polyamide was not sliding around in the minor groove. For direct comparison back to the crystal structure, only heavy atom distances were measured. The distances that were measured are shown in the Figure 3.26 and denoted by the "x" with a number above the distance lines. As with the helical parameters, a statistical comparison back to the starting x-ray crystal structure was also performed with the DNA/polyamide distances (95% confidence level) as a test of the parameters.

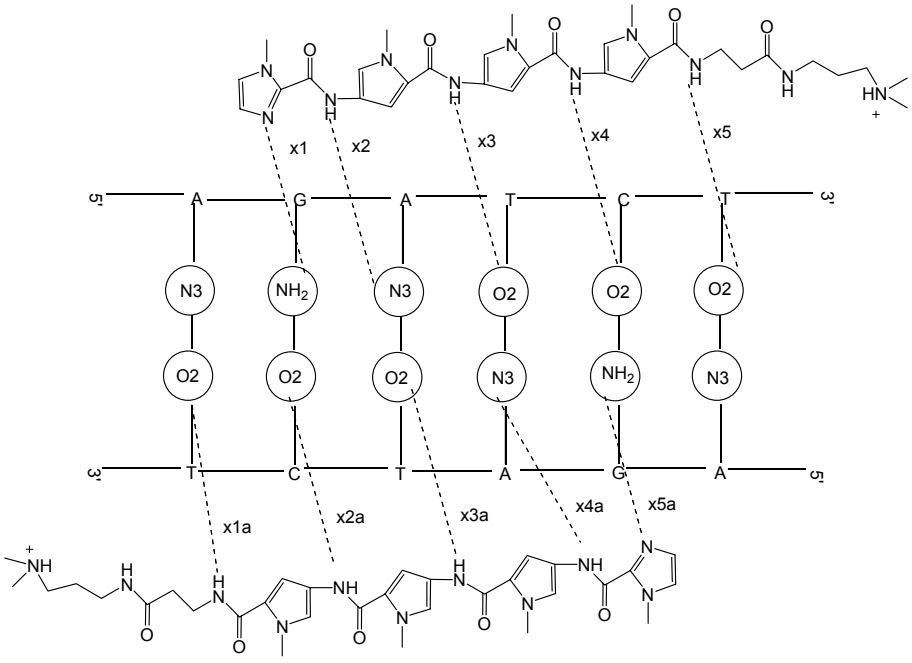


Figure 3.26 Polyamide distances measured (dashed lines) in the x-ray structure and from the low and high temperature crystal simulations. The "ladder" in the middle represents the DNA binding site of the polyamides. The circles in the middle represent the hydrogen bond donors or acceptors in the minor groove of the DNA.

The distances measured from the simulations are in good agreement with the x-ray structure (tables of measured distances and t-values can be found in Appendix A). Over the seven initial individual simulations, two distances were statistically different from the x-ray structure. The distance (x1a) is much larger than that in the crystal structure (x1a =

4.99 Å for the calculated average, and x1a = 3.3 Å for the x-ray structure) in one low temperature simulation. Upon closer examination of the atoms of the polyamide and DNA that show this large distance, a water molecule was discovered above the minor groove. The x1a distance is from an NH group in the polyamide backbone to the O2 of thymine. However, the NH group appears to be forming a strong hydrogen bond (~ 1.8 Å) with the oxygen of the water molecule above the groove rather than forming a strong hydrogen bond with the O2 of thymine. The other simulations do not have a water molecule in this position or anywhere near this position in order to form a hydrogen bond.

The distance from the H of the NH group to the water oxygen over the whole 10 ns of simulation 2 was plotted (shown in Figure 3.27, top). At the start of the simulation, the distance was ~3.5 Å; also, waters were deleted within 2.5 Å of the DNA/polyamide complex, but the water molecule moves right in to the polyamide, and the distance then stays small (~1.9 Å). Upon closer examination of the starting structure, the water forming the hydrogen bond to the polyamide was actually in the minor groove between the two polyamides, but it was still at least 3 Å from either of them and the DNA. It was trapped and moved closer to the NH of the polyamide backbone. The water and the DNA are therefore competing for hydrogen bonding to the polyamide.

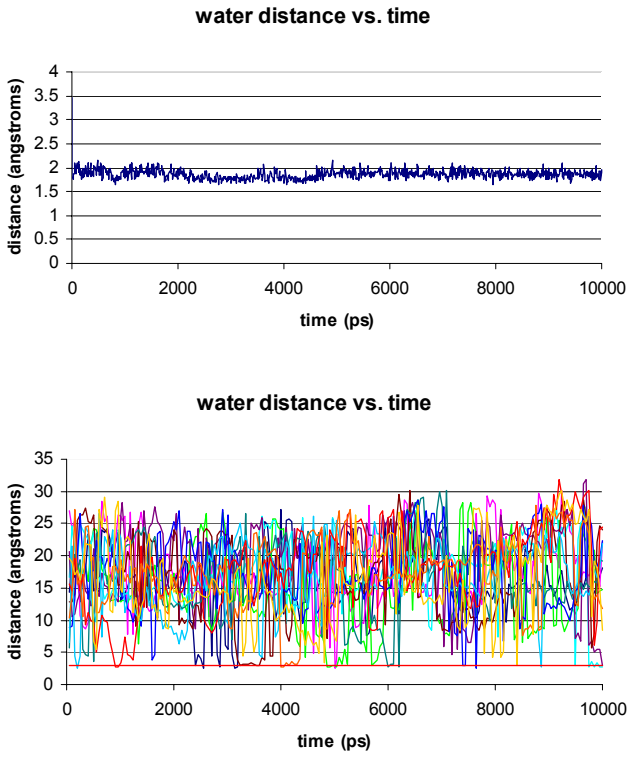


Figure 3.27 Plot of distance from water to thymine O2 over the 10 ns time span of the simulation for low temperature simulation (top). The bottom plot shows the distance from polyamide to nearby water molecules. The water molecules exchange, but they still hydrogen bond to the polyamide. The straight red line on the bottom plot is at 3 Å, which is the average length of the hydrogen bond formed from the polyamide to the exchanging waters. There are water molecules hydrogen bonding to the polyamide at most times during the simulation.

One of the distances (x4) in one high temperature simulation is much larger (~1 Å) than that of the crystal structure (calculated x4 = 3.95 Å; x-ray x4 = 2.95 Å). Upon closer examination of the simulation snapshots, a water molecule hydrogen was found to be hydrogen bonding to a carbonyl oxygen in the polyamide backbone. However, it was not the same water molecule hydrogen bonding to the polyamide for the duration of the simulation; the water molecules appear to be exchanging with each other, but nonetheless forming a strong hydrogen bond to a carbonyl oxygen in the polyamide backbone (Figure 3.27, bottom plot). The waters exchanging to form the hydrogen bond is probably a

consequence of the 300K simulations having more kinetic energy than the 113K simulations; the atoms move more and can thus exchange to form the hydrogen bond to the polyamide. Because good agreement between the calculated and x-ray distances was achieved with the initial seven simulations, distances were not measured for additional simulations (those were considered only for improvement of sampling).

Even though individual simulations exhibited statistical significance for DNA/polyamide distances due to intruding water molecules, when the distances were averaged over all simulations (low temperature and high temperature distances averaged separately), none of the distances showed statistical significance. The t-tests for the combined simulation distances were performed similarly to the combined t-tests for the helical parameters. For example, for distance x1 for the low temperature simulations, 1000 values were taken from each of the four simulations, the values at each time were averaged, and a t-test was performed between the simulation values and x-ray value at each time (1000 t-tests total for each distance). A table of the average t-values obtained for each distance can be found in Appendix A.

3.2.4. DNA/Dervan hydroxypyrrole polyamide complex. A 10 ns crystal simulation of a different DNA/Dervan polyamide complex was also performed. In this simulation, one pyrrole heterocycle in each of the polyamides is replaced by the hydroxypyrrole heterocycle. The polyamides in the simulations discussed in Section 3.2.3 and in this section recognize the same bases in the minor groove of DNA (5'-AGATCT-3') and the DNA sequence is the same; however, the hydroxypyrrole moiety reduces the binding affinity due to elongation of the Watson-Crick hydrogen bonds of the

target T-A base pairs.3 The starting structure for the simulation was PDB46 id 1CVX.3 The RMSD plot of the DNA/hydroxypyrrole polyamide is shown in Figure 3.28.

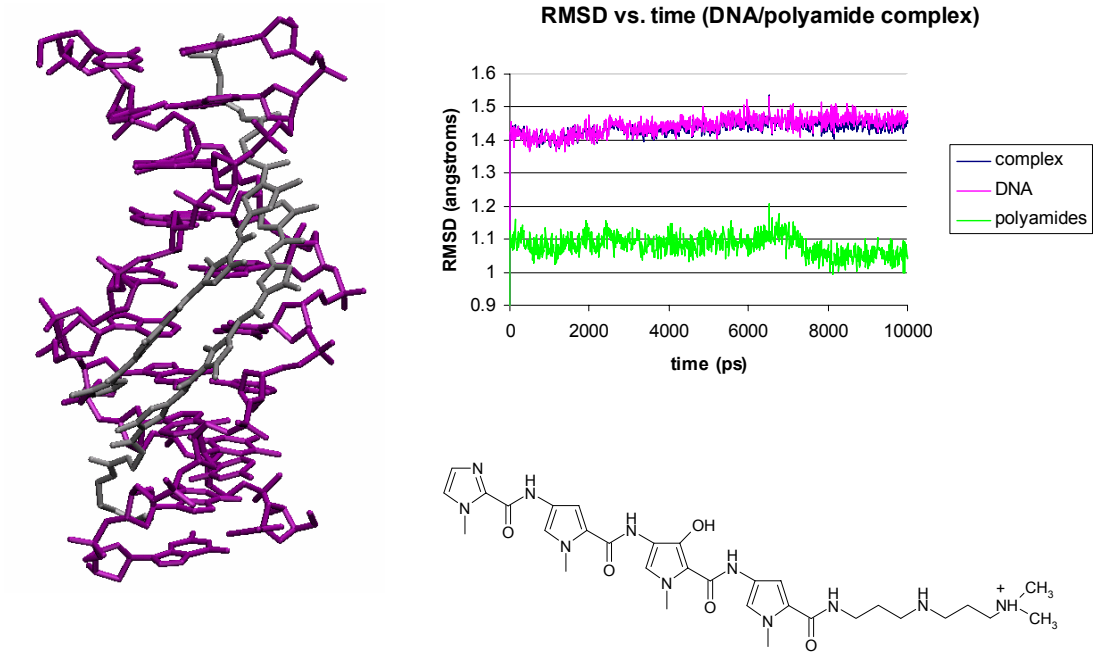


Figure 3.28 Starting structure for DNA/hydroxypyrrole polyamide simulation (PDB[46] id 1CVX[3]; left). The DNA is in purple, and the polyamides are in silver in the minor groove. The structure of the polyamides bound in the minor grove is shown on the bottom right (two are bound antiparallel in the minor groove). The plot of RMSD vs. time of DNA/polyamide complex is on the top right.

The average RMSDs of the complex, DNA, and polyamides are 1.44 ± 0.05 Å, 1.44 ± 0.05 Å, and 1.08 ± 0.05 Å, respectively. The RMSD of the complex and the DNA stabilizes at ~6 ns to 1.44 Å, while the RMSD of the polyamides is steady until about 7 ns. At 7 ns, the RMSD of the polyamides decreases to just above 1 Å. The DNA contributes more the RMSD than the polyamides, which is probably due to some fraying of the end bases. The contribution to the polyamide RMSD is most likely due to the polyamide tails, which contain C-C single bonds about which free rotation is allowed. This conformational freedom is reflected by the disorder in the starting x-ray structure,

because there were two possible conformations identified for the positions of the polyamide tails.[3]

The same analysis for the DNA helical parameters and DNA/polyamide distances was carried out for the hydroxypyrrole simulation. None of the 18 helical parameters examined were statistically significantly different from the x-ray structure values (a table of helical parameter and calculated t-values can be found in Appendix A). Figure 3.29 shows the DNA/polyamide distances measured (a table of measured distances is in Appendix A). Seven distances were measured for each polyamide to its respective DNA strand; only two distances out of the 14 measured were statistically significant according to the t-test. The distances showing statistical significance were the x5 and x1a distances. The x-ray values for these distances are 3.16 Å and 2.54 Å, respectively; whereas the average simulation distances are 4.93 Å and 5.93 Å (both are over an angstrom longer than the x-ray values).

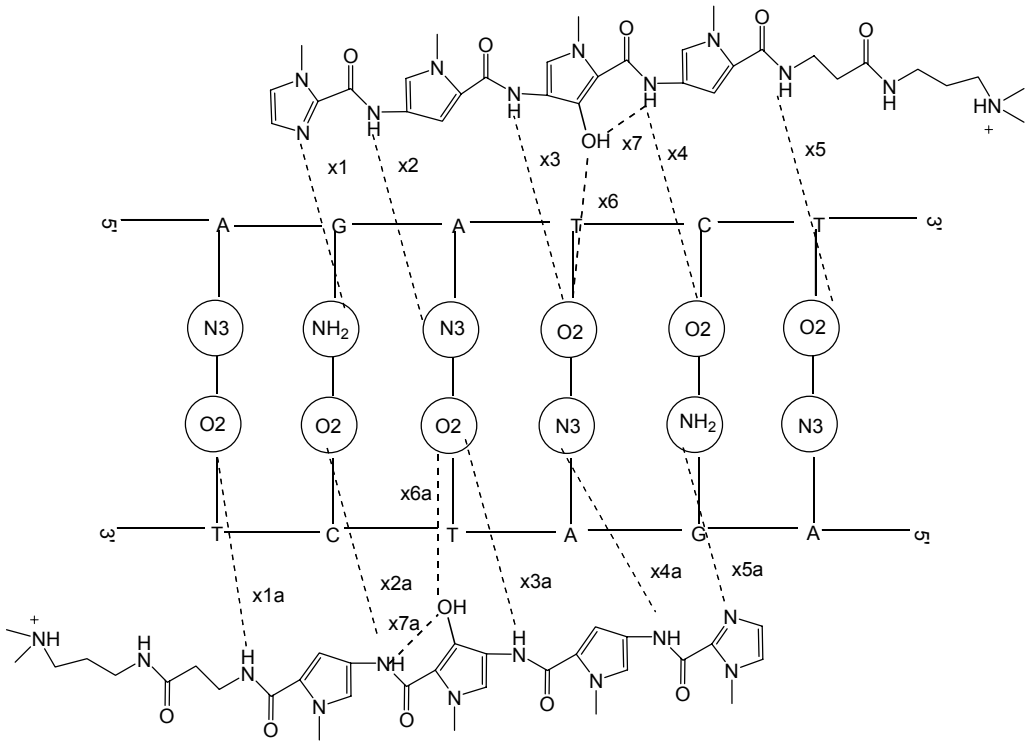


Figure 3.29 Polyamide distances measured (dashed lines) in the x-ray structure and from the hydroxypyrrole polyamide crystal simulations. The "ladder" in the middle represents the DNA binding site of the polyamides. The circles in the middle represent the hydrogen bond donors or acceptors in the minor groove of the DNA.

These distances are at the ends of the polyamide and the DNA strand, where water molecules were identified in the minor groove in two of the other DNA/polyamide simulations. In this simulation, water molecules were also found to be hydrogen bonding to the DNA or polyamide in this region, causing the distances to be large. Not only does water get into the groove at the ends, but the identification of two possible orientations of the polyamide tails in the x-ray structure indicates that there is a lot of conformational freedom in the tails. In one orientation, the tail is inside the minor groove where it can hydrogen bond with the DNA, and in the other orientation, the tail is projected outside of the minor groove. The simulation does sample the latter conformation, which contributes to the large distances observed.

3.2.5. DNA/Netropsin and DNA/Distamycin complexes. The fourth test of the new parameters was a 10 ns crystal simulation of a DNA/netropsin complex and a 10 ns crystal simulation of a DNA/distamycin complex. Netropsin and distamycin are both antibiotics that bind in the minor groove of DNA, preferentially to A-T rich sequences. Netropsin covers ~4 base pairs; distamycin covers ~5 base pairs. Like the Dervan polyamides, they contain pyrrole heterocycles linked by an amide backbone, and they recognize the minor groove by the hydrogen bonding pattern. Figure 3.30 shows the molecular structures of netropsin and distamycin and how each one binds in the minor groove of DNA.

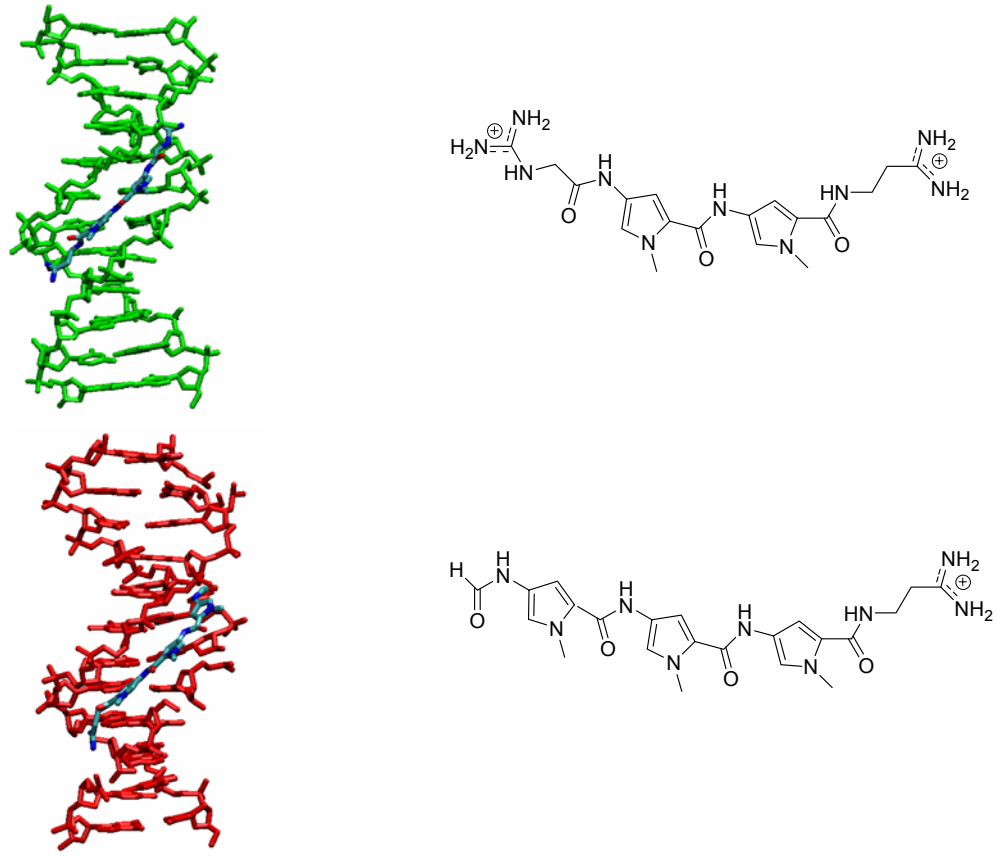


Figure 3.30 Starting structure for DNA/netropsin simulation (PDB[46] id 121D[60]; top left) and the molecular structure of netropsin (top right). The bottom panel shows the starting structure for the DNA/distamycin simulation (PDB[46] id 2DND[61]; bottom left) and the molecular structure of distamycin (bottom right).

DNA/netropsin complex. The RMSD plot of the DNA and netropsin complex is shown below (Figure 3.31). The average RMSD of the complex was 1.85 ± 0.29 Å. The DNA contributes the most to the RMSD (average RMSD was 1.83 ± 0.31 Å), while the netropsin has a lower RMSD (average RMSD was 0.91 ± 0.12 Å). The DNA/netropsin simulation was performed at 300K, and the RMSD for the DNA/netropsin complex is in the same range as the DNA/polyamide high temperature simulations. The geometric parameters of the netropsin molecule were examined and compared back to the x-ray structure, and the simulated structure is in good agreement with experiment. The average bond error was 0.03 ± 0.02 Å, the average angle error was 4 ± 3°, and the average dihedral error was 8 ± 12°. The large dihedral error stems from the rotation about the carbon-carbon bonds at the ends of the netropsin molecule. An overlay of the x-ray structure and the simulation average structure is shown in Figure 3.31. The tails of the molecule, especially the amidino group tail which contains two carbon-carbon single bonds, do not align with each other as the pyrrole rings in the middle. Free rotation is allowed about these carbon-carbon bonds, and this is responsible for the large dihedral angle error. The B-factors reported for the x-ray structure are slightly higher for the atoms of the tails of netropsin,[60] indicating that there is some disorder in the crystal for the tails. This is most likely observed because the tails contain C-C single bonds, and free rotation is allowed about these bonds; thus, the tail could adopt different orientations, leading to uncertainty in the positions of the atoms.

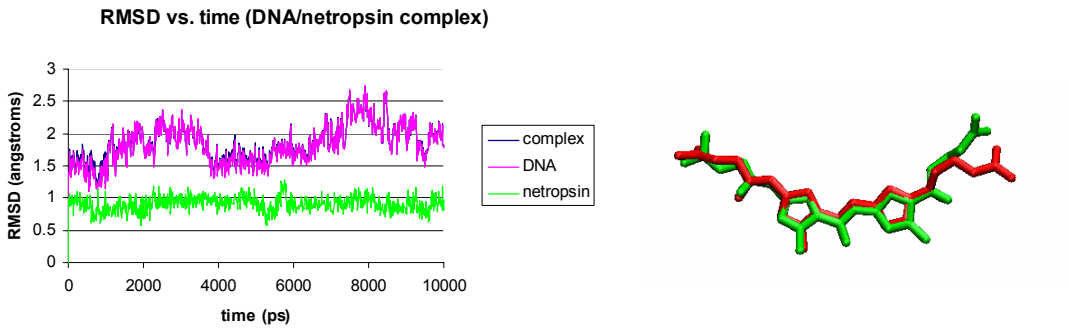


Figure 3.31 RMSD vs. time of DNA/netropsin complex (left). Overlay of x-ray structure (red) of netropsin and simulation average structure (green) (right).

The helical parameters of the DNA were examined from the DNA/netropsin simulation. The analysis was performed in the same way as for the DNA/polyamide complex simulations, using the t-test at the 95% confidence level. None of the helical parameters were found to be significantly different from those of the x-ray structure (a table of the calculated helical parameters and t-values can be found in Appendix A). The hydrogen bond distances from netropsin to the DNA were also examined (Figure 3.32). As for the DNA/polyamide complex, a t-test was used to analyze the hydrogen bonding distances (table of calculated distance is in Appendix A). For the DNA/netropsin complex, three of the 11 distances were found to be significantly different from the values determined from the x-ray structure.

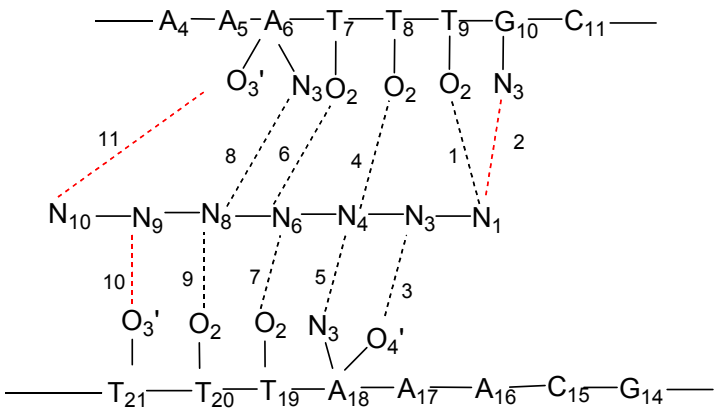


Figure 3.32 Hydrogen bond distances (numbered) from netropsin to DNA. The figure shows the middle eight base pairs of the DNA fragment and the hydrogen bonds formed from the netropsin N atoms to the atoms of the DNA bases (h-bonds represented by dashed lines). The netropsin molecule is represented by the numbered Ns in the middle of the figure. Red dashed lines show significant differences between experiment and simulation.

The three distances that were found to be significantly different from the x-ray structure are the G10 (N3) to netropsin (N1) distance, the T21 (O3') to netropsin (N9) distance, and the A6 (O3') to netropsin (N10) distance. These three distances involve nitrogens on the end of the netropsin molecule. It is observed that the netropsin moves most at the ends because free rotation is allowed at the single bonds. Therefore, it is not surprising that the length of these hydrogen bonds change more over time. Also, as in the simulations with DNA/polyamide complexes, there were water molecules found near the ends of the netropsin, which form hydrogen bonds with N1, N9, and N10 of netropsin. In the x-ray structure, no water molecules were identified as being hydrogen bonded to these three nitrogens of netropsin (they are hydrogen bonded to the DNA bases), but as the simulation progresses, those three nitrogens form hydrogen bonds with water molecules rather than the DNA. The water molecules appear to exchange during the simulation, but the netropsin is still hydrogen bonding to them (distances in the range of ~2.6-3.3 Å).

DNA/distamycin complex. The RMSD plot of the complex, DNA, and distamycin is shown below (Figure 3.33). The average RMSD of the complex was 2.33 ± 0.27 Å. The DNA contributes the most to the RMSD (average RMSD was 2.41 ± 0.27 Å), while the distamycin has a lower RMSD (average RMSD was 0.75 ± 0.10 Å). The cause of the high RMSD does not appear to be the distamycin; one base from the helix that does not interact with the distamycin flipped out slightly from the helix (THY7). The movement of this base did not appear to affect the DNA/distamycin distances or helical parameters greatly. The geometric parameters of the distamycin molecule were examined and compared back to the x-ray structure, and the simulated structure is in good agreement with the x-ray structure. The average bond error was 0.04 ± 0.04 Å, the average angle error was 4 ± 3°, and the average dihedral error was 12 ± 24°. The dihedral error stems from the rotation about the carbon-carbon bonds at the ends of the distamycin molecule. An overlay of the x-ray structure and the simulation average structure is shown in Figure 3.33. The tails of the molecule, especially the formamide group tail where the carbonyl oxygens are turned in opposite directions, do not align with each other as the pyrrole rings in the middle.

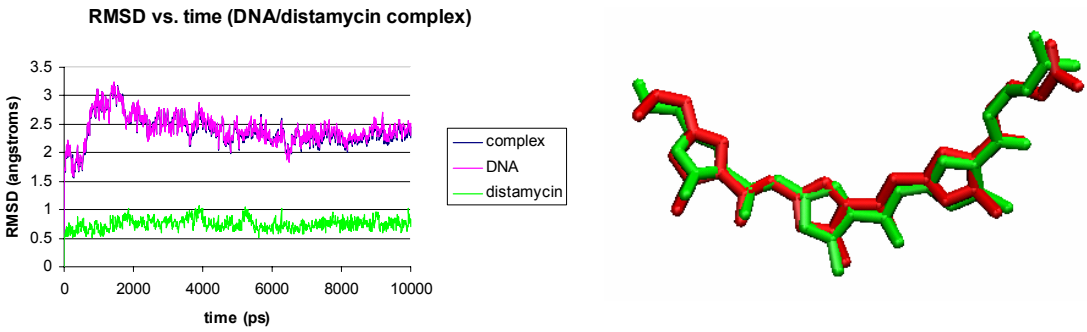


Figure 3.33 RMSD vs. time of DNA/distamycin complex (left). Overlay of x-ray structure (red) of distamycin and simulation average structure (green) (right).

The helical parameters of the DNA were examined from the DNA/distamycin simulation. The analysis was performed in the same way as for the other simulations. One helical parameter was found to be significantly different than the x-ray value, which was the opening parameter. This statistical significance can be attributed to the flipping out of Thy7. The hydrogen bond distances from distamycin to the DNA were also examined (Figure 3.34). For the DNA/distamycin complex, one of the 7 distances were found to be significantly different from the values determined from the x-ray structure.

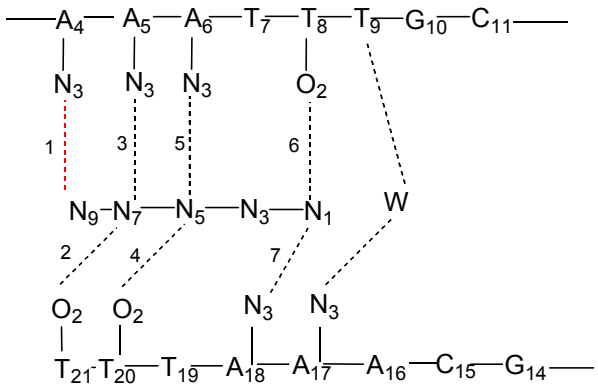


Figure 3.34 Hydrogen bond distances from distamycin to DNA (numbered). The figure shows the middle eight base pairs of the DNA fragment and the hydrogen bonds formed from the distamycin N atoms to the atoms of the DNA bases (h-bonds represented by dashed lines). The distamycin molecule is represented by the numbered Ns in the middle of the figure. The W indicates a water molecule found in the minor groove. Figure adapted from Coll, et al.[61]

The distance found to be significantly different from the x-ray structure was the A4 (N3) to distamycin (N9) distance. This distance involves a nitrogen on the end of distamycin. It is observed that the distamycin moves most at the ends because free rotation is allowed at the single bonds. Therefore, it is not surprising that this nitrogen may not maintain the hydrogen bond with DNA. Instead, there are water molecules in the groove near distamycin forming a hydrogen bond with N9. The water molecules appear to be exchanging, but they are still forming hydrogen bonds to N9 (distances were in the range

of 2.7-3.5 Å). The water molecule identified in the minor groove that forms a hydrogen bond to Thy9 and Ade17 also exchanges with other waters during the simulation. Initially, the water molecule is ~2.6 Å from Thy9 (O2) and 3.1 Å from Ade17 (N3). During the simulation, the water molecule exchanges with other waters but the water molecule always forms a hydrogen bond to the DNA bases. The distance between the water and the DNA bases changes slightly (distances range from 2.8 to 4.1 Å) and in some snapshots there are two water molecules (one forming a hydrogen bond to Thy9 and one forming the hydrogen bond to Ade17); however, the same trend is usually followed: the distance from Thy9 to the water is usually shorter than the distance from Ade17 to the water.

3.2.6. DNA/Netropsin Crystal Simulation with AMBER Parameters. A 10 ns crystal simulation of the DNA/netropsin complex was also performed using the previously reported parameters for netropsin[62] for use with the AMBER[63] force field. The netropsin molecule was not explicitly parameterized for the force field; instead, analogous parameters in the force field were adopted for netropsin. The same starting conditions (same structure, same number of water molecules, same system preparation) were used as with the CHARMM simulation. The RMSD plot for the DNA/netropsin complex is in Figure 3.35. The RMSD of the complex was 2.17 ± 0.21 Å, and the DNA contributes more to the RMSD (2.16 ± 0.20 Å) than the netropsin (1.16 ± 0.40 Å). The RMSDs of all parts of the system are higher than those obtained from the CHARMM parameters. There is also a large jump in RMSD for the netropsin at ~ 8ns. This is due to the large structural distortion of the netropsin relative to the starting x-ray structure (Figure 3.31).

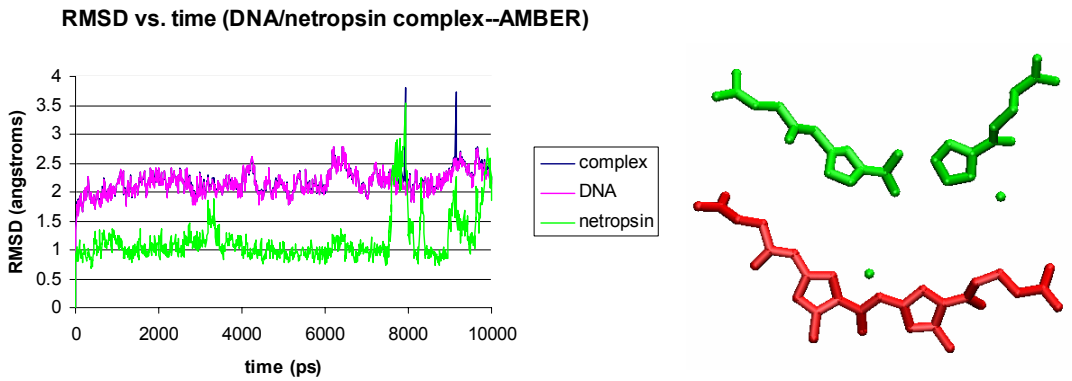


Figure 3.35 RMSD vs. time of DNA/netropsin complex using AMBER parameters (left). Comparison (right) of x-ray structure (red) of netropsin and simulation average structure (green).

The average simulation structure shows the failure of agreement of the AMBER parameters with the x-ray structure. The pyrrole ring on the left of the molecule is distorted, and the force constants for the connection between the pyrrole rings is obviously not large enough to keep the bond length correct (the VMD[64] viewer did not even draw a bond). The bond length between the amide nitrogen and the carbon of the pyrrole ring should be 1.44 Å (x-ray), but the average simulation structure yields 2.15 Å. The force constants are also flawed for the connection of the methyl groups to the nitrogen of the pyrrole rings. In the picture on the right, the green dots are the methyl groups that are supposed to be attached to the pyrrole rings (again, the viewer did not even draw a bond). The length in the x-ray structure for those bonds is ~1.50 Å, but the average simulation structure yields an average value of 3.76 Å.

The helical parameters and DNA-netropsin distances were also examined using the t-test. Of the 18 helical parameters measured, nine of them are statistically significant from the x-ray value. The significant parameters are x-displacement, inclination, stretch, stagger, opening, slide, twist, helical rise, and helical twist. Despite the fact that the

netropsin molecule does not stay bonded together, it does stay in the minor groove of the DNA. For the DNA/netropsin distances, 3 out of the 11 distances measured were statistically significant. These three distances are also found to be statistically significant using the CHARMM parameters (tables of calculated helical parameters and distances can be found in Appendix A). Overall, the AMBER parameters do not allow for a good comparison back to experiment; therefore, an explicit parameterization of the compounds of interest must be performed rather than simply taking analogous parameters from existing molecules in the force field.

3.3 Conclusions

New force field parameters for the CHARMM force field for DNA minor groove binding pyrrole-imidazole polyamides have been developed. To test the parameters, crystal simulations of DNA with different polyamides bound in the minor groove have been performed.

The computed helical parameters were in excellent agreement with experiment (17 out of the 18 examined were not statistically significant). The stagger helical parameter was the only parameter that was statistically different from the x-ray structure in both sets of DNA/polyamide simulations (113K and 300K). However, upon running additional trajectories, the sampling of stagger was improved. Overall, the low temperature simulations still yield statistical significance with respect to stagger, and stagger is still significant for the high temperature simulations at the 95% confidence level but not at the 97% confidence level. When only the three additional high temperature simulations are considered, stagger is not statistically significant. Temperature obviously influences the sampling of stagger; the higher temperature allows

for better sampling and thus non-statistically significant results according to the t-test. The fact that the high and low simulations do not show non-significance overall for all helical parameters underscores the importance of running multiple trajectories.

The high temperature simulations yield structures that are distorted and exhibit A-DNA features, and as a result, the polyamides are pushed out of the minor groove in these structures. The temperature of the simulations is too close to that of the melting temperature of the DNA fragment, causing the DNA to start melting, which gives rise to the distorted structures. However, the high temperature structures are still in good agreement with experiment; most helical parameters are non-significant, and we do not observe the rigid structure of the crystal. The authors of the x-ray structure most likely could not obtain crystals at 300 K either. The low temperature simulations yield solid and undistorted structures, much like the starting x-ray structure. Again, we get good agreement with experiment because we see the crystal at 113 K. Although the distortion is observed, the high temperature simulations sample better than the low temperature simulations, thus, stagger is sampled better at the high temperature and yields better t-test results.

The DNA/netropsin, DNA/distamycin, and DNA/hydroxypyrrole polyamide simulations show excellent agreement with the x-ray structure. The DNA/netropsin simulation and the DNA/hydroxypyrrole did not exhibit any statistically significant helical parameters. Opening was the only significantly different parameter in the DNA/distamycin simulation. This can be attributed to one base (Thy7) that flips out slightly out from the rest of the helix during the simulation. This slight flipping out did not affect the DNA/distamycin distances measured or the other helical parameters.

The distances measured from the atoms in the polyamides to their hydrogen bonding partners in the DNA show good agreement with the x-ray structure with the exception of two distances over all simulations. In one low temperature simulation, a water molecule was discovered to be hydrogen bonding to the H of an NH group in the polyamide backbone; as a result, the NH group was not hydrogen bonding to the DNA. In one high temperature simulation, a water molecule was again found to be hydrogen bonding to the polyamide; however, the water molecules appear to be exchanging with each other, nonetheless forming a hydrogen bond with the polyamide. The water molecules are therefore competing with the DNA for hydrogen bonding with the polyamide (which could happen with any molecule that relies on hydrogen bonding for recognition and binding to its target site), which is a possible factor for why some polyamides show reduced binding affinity for their target sequences. Even though individual simulations exhibited statistical significance for two of the DNA/polyamide distances, when the distances are averaged together over all low temperature and all high temperature simulations, none of the distances are significantly different. The distance measurements, like the helical parameters, show the utility of multiple trajectories; the behavior of the intruding water molecules may have gone unnoticed with only one simulation.

The DNA/distamycin, DNA/netropsin, and DNA/hydroxypyrrole polyamide simulations exhibited similar behavior with respect to water molecule hydrogen bonding to the ligand. The DNA/hydroxypyrrole simulation had two distances that were significantly different from the x-ray distances. These distances were at the ends of the DNA groove, which is accessible to water molecules. As in the other DNA/polyamide

simulations, water molecules were observed to be forming hydrogen bonds to the DNA or polyamide. The DNA/netropsin simulation had three distances that were significantly different from the x-ray structure and the DNA/distamycin simulation had one. These distances were from nitrogens on the ends of the ligands to DNA bases. Water molecules were observed to form hydrogen bonds to these nitrogens and although the water molecules appear to exchange during the simulation, they still form hydrogen bonds with the ligand. This does not seem unusual because the minor groove of DNA normally contains a spine of hydration, so water is present in the minor groove. The netropsin and distamycin ligands replace this spine of hydration in the middle of the groove, but water molecules were identified in the x-ray structures near the ends of the grooves closer to the ends of the ligands.[60, 61]

The simulation performed using the AMBER[63] parameters emphasizes that explicit parameterization of the force field for new molecules is necessary. The netropsin molecule was not explicity parameterized for the AMBER force field[63]; instead, the parameters were guessed at by simply using analogous parameters for existing molecules in the force field. The simulation exhibited large fluctuations in the RMSD of the DNA/netropsin complex, and this was due to large structural distortions of the pyrrole rings of netropsin. The guessed bond force constants were incorrect because the bond lengths of the methyl groups to the pyrrole rings are much too long for a nitrogen-carbon bond. The AMBER parameters did not yield a satisfactory comparison back to experiment; thus, unique and explicit parameterization must be performed.

The simulations performed have validated the new force field parameters for the polyamides, and new simulation studies can now be performed to further understand the

interactions and dynamics between polyamides and DNA. From this knowledge, new ligands can be designed that have a higher affinity for or bind better to their target site (new molecules or different heterocycles or linkages may improve the current design of polyamides). Also, new ligands can be designed that use polyamides as a sequence-specific recognition element that incorporate another type of molecule that will distort the DNA in a way that transcriptional proteins cannot bind to their target site.

References

(1) Kielkopf, C. L.; White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B.; Rees, D. C., A Structural Basis for Recognition of AT and TA Base Pairs in the Minor Groove of B-DNA, *Science*, 1998, 282, 111-115.

(2) Kielkopf, C. L.; Baird, E. E.; Dervan, P. B.; Rees, D. C., Structural Basis for GC Recognition in the DNA Minor Groove, *Nat. Struct. Biol.*, 1998, 5, 104.

(3) Kielkopf, C. L.; Bremer, R. E.; White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B.; Rees, D. C., Structural Effects of DNA Sequence on TA Recognition by Hydroxypyrrole/Pyrrole Pairs in the Minor Groove, *J. Mol. Biol.*, 2000, 295, 557-567.

(4) White, S.; Baird, E. E.; Dervan, P. B., Effects of the AT/TA Degeneracy of Pyrrole-Imidazole polyamide recognition in the Minor Groove of DNA, *Biochemistry*, 1996, 35, 12532-12537.

(5) White, S.; Baird, E. E.; Dervan, P. B., Orientation Preferences of Pyrrole-Imidazole Polyamides in the Minor Groove of DNA, *J. Am. Chem. Soc.*, 1997, 119, 8756-8765.

(6) White, S.; Turner, J. M.; Szewczyk, J. W.; Baird, E. E.; Dervan, P. B., Affinity and Specificity of Multiple Hydroxypyrrole/Pyrrole Ring Pairings for Coded Recognition of DNA, *J. Am. Chem. Soc.*, 1999, 121.

(7) Swalley, S. E.; Baird, E. E.; Dervan, P. B., Discrimination of 5'-GGGG-3', 5'-GCGC-3', and 5'-GGCC-3' Sequences in the Minor Groove of DNA by Eight-Ring Hairpin Polyamides, *J. Am. Chem. Soc.*, 1997, 119, 6953-6961.

(8) Swalley, S. E.; Baird, E. E.; Dervan, P. B., Effects of γ-Turn and β-Tail Amino Acids on Sequence-Specific Recognition of DNA by Hairpin Polyamides, *J. Am. Chem. Soc.*, 1999, 121, 1113-1120.

(9) Wurtz, N. R.; Pomerantz, J. L.; Baltimore, D.; Dervan, P. B., Inhibition of DNA Binding by NF-κB with Pyrrole-Imidazole Polyamides, *Biochemistry*, 2002, 41, 7604-7609.

(10) Trauger, J. W.; Baird, E. E.; Dervan, P. B., Recognition of DNA by Designed Ligands at Subnanomolar Concentrations, *Nature*, 1996, 382, 559-561.

(11) Trauger, J. W.; Baird, E. E.; Dervan, P. B., Recognition of 16 Base Pairs in the Minor Groove of DNA by a Pyrrole-Imidazole Polyamide Dimer, *J. Am. Chem. Soc.*, 1998, 120, 3534-3535.

(12) Turner, J. M.; Baird, E. E.; Dervan, P. B., Recognition of Seven Base Pair Sequences in the Minor Groove of DNA by Ten-Ring Pyrrole-Imidazole Polyamide Hairpins, *J. Am. Chem. Soc.*, 1997, 119, 7636-7644.

(13) Rucker, V. C.; Melander, C.; Dervan, P. B., Influence of β-alanine on Hairpin Polyamide Orientation in the DNA Minor Groove, *Helv. Chim. Acta*, 2003, 86, 1839-1851.

(14) Chiang, S.-Y.; Burli, R. W.; Benz, C. C.; Gawron, L.; Scott, G. K.; Dervan, P. B.; Beerman, T. A., Targeting the Ets Binding Site of the HER2/neu Promoter with Pyrrole-Imidazole Polyamides, *J.Biol. Chem.*, 2000, 275, 24246-24254.

(15) Dickinson, L. A.; Trauger, J. W.; Baird, E. E.; Ghazal, P.; Dervan, P. B.; Gottesfeld, J. M., Anti-Repression of RNA Polymerase II Transcription by Pyrrole-Imidazole Polyamides, *Biochemistry*, 1999, 38, 10801-10807.

(16) Foloppe, N.; MacKerell, A. D., Jr., All-Atom Empirical Force Field for Nucleic Acids: I. Parameterization Optimization Based on Small Molecule and Condensed Phase Macromolecular Target Data, *J. Comp. Chem.*, 2000, 21, 86-104.

(17) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Development and Testing of a General AMBER Force Field, *J. Comp. Chem.*, 2004, 25, 1157-1174.

(18) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M., All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins, *J. Phys. Chem. B*, 1998, 102, (18), 3586-3616.

(19) Kielkopf, C. L.; Ding, S.; Kuhn, P.; Rees, D. C., Conformational Flexibility of B-DNA at 0.74 Angstrom Resolution, *J. Mol. Biol.*, 2000, 296, 787-801.

(20) Allen, F. H., The Cambridge Structural Database: A Quarter of a Million Crystal Structures and Rising, *Acta Crystallographica Section B: Structural Science*, 2002, 58, 380.

(21) Berman, H. M.; Neidle, S.; Zimmer, C.; Thrum, H., Netropsin, A DNA-Binding Oligopeptide: Structural and Binding Studies, *Biochimica et Biophysica Acta--Nucleic Acids and Protein Synthesis*, 1979, 561, 124-131.

(22) Lavery, R.; Sklenar, H., The Definition of Generalized Helicoidal Parameters and of Axis Curvature for Irregular Nucleic Acids, *J. Biomolecular Structure and Dynamics*, 1998, 6, 63-91.

(23) Olson, W. K.; Bansal, M.; Burley, S. K.; Dickerson, R. E.; Gerstein, M.; Harvey, S. C.; Heinemann, U.; Lu, X. J.; Neidle, S.; Shakked, Z.; Sklenar, H.; Suzuki, M.; Tung, C. S.; Westhof, E.; Wolberger, C.; Berman, H. M., A Standard Reference Frame for the Description of Nucleic Acid-Base Pair Geometry, *J. Mol. Biol.*, 2001, 313, 229-237.

(24) Lu, X. J.; Olson, W. K., Resolving the Discrepancies Among Nucleic Acid Conformational Analyses, *J. Mol. Biol.*, 1999, 285, 1563-1575.

(25) Dickerson, R. E., Definitions and Nomenclature of Nucleic Acid Structure Components, *Nucleic Acids Research*, 1989, 17, 1797.

(26) MacKerell, A. D., Jr.; Banavali, N., All-Atom Empirical Force Field for Nucleic Acids: II. Application to Molecular Dynamics Simulations of DNA and RNA in Solution, *J. Comp. Chem.*, 2000, 21, 105-120.

(27) Berman, H. M.; Olson, W. K.; Beveridge, D. L.; Westbrook, J.; Gelbin, A.; Demeny, T.; Hsieh, S.-H.; Srinivasan, A. R.; Schneider, B., The Nucleic Acid Database: A Comprehensive Relational Database of Three-Dimensional Structures of Nucleic Acids, *Biophys. J.*, 1992, 63, 751-759.

(28) Lu, X. J.; Olson, W. K., 3DNA: A Software Package for the Analysis, Rebuilding, and Visualization of Three-Dimensional Nucleic Acid Structures, *Nucleic Acids Research*, 2003, 31, 5108.

(29) Dickerson, R. E., DNA Bending: the prevalence of kinkiness and the virtues of normality, *Nucleic Acids Research*, 1998, 26, 1906-1926.

(30) El Hassan, M. A.; Calladine, C. R., The assessment of the geometry of dinucleotide steps in double-helical DNA: a new local calculation scheme, *J. Mol. Biol.*, 1995, 251, 648-664.

(31) Lu, X. J.; El Hassan, M. A.; Hunter, C. A., Structure and conformation of helical nucleic acids analysis program (SCHNAap), *J. Mol. Biol.*, 1997, 273, 668-680.

(32) Gorin, A. A.; Zhurkin, V. B.; Olson, W. K., B-DNA twisting correlates with base-pair morphology, *J. Mol. Biol.*, 1995, 247, 34-48.

(33) Kosikov, K. M.; Gorin, A. A.; Zhurkin, V. B.; Olson, W. K., DNA Stretching and Compression: Large-Scale Simulations of Double Helical Structures, *J. Mol. Biol.*, 1999, 289, 1301-1326.

(34) Bansal, M.; Bhattacharyya, D.; Ravi, B., NUPARM and NUCGEN: Software for analysis and generation of sequence dependent nucleic acid structures, *CABIOS*, 1995, 11, 281-287.

(35) Tung, C. S.; Soumpasis, D. M.; Hummer, G., An Extension of the Rigorous Base-Unit Oriented Description of Nucleic Acid Structures, *J. Biol. Struct. Dynam.*, 1994, 11, 1327-1344.

(36) Babcock, M. S.; Olson, W. K., The effect of mathematics and coordinate systems on comparability and "dependencies" of nucleic acid structure parameters, *J. Mol. Biol.*, 1994, 237, 98-124.

(37) Lavery, R.; Zakrzewska, K., Base and Base Pair Morphologies, Helical Parameters, and Definitions. In *Oxford Handbook of Nucleic Acid Structure*, Oxford University Press: Oxford, 1999.

(38) Schlick, T., *Molecular Modeling and Simulation: An Interdisciplinary Guide*. Springer-Verlag: New York, 2000; 21.

(39) Clarage, J. B.; Romo, T.; Andrews, B. K.; Pettit, B. M.; Phillips, G. N., Jr., A Sampling Problem in Molecular Dynamics Simulations of Macromolecules, *Proc. Natl. Acad. Sci. USA*, 1995, 92, 3288-3292.

(40) Frauenfelder, H.; Sligar, S. G.; Wolynes, P., The Energy Landscapes and Motions of Proteins, *Science*, 1991, 254, 1598.

(41) Karplus, M.; McCammon, A., Molecular Dynamics Simulations of Biomolecules, *Nature Structural Biology*, 2002, 9, (9), 646.

(42) Auffinger, P.; Louise-May, S.; Westhof, E., Multiple molecular dynamics simulations of the anticodon loop of tRNAAsp in aqueous solution with counterions, *J. Am. Chem. Soc.*, 1995, 117, 6720-6726.

(43) Caves, L. S. D.; Evanseck, J. D.; Karplus, M., Locally Accessible Conformations of Proteins: Multiple Molecular Dynamics Simulations of Crambin, *Protein Science*, 1998, 7, 649-666.

(44) Schulze, B. G.; Evanseck, J. D., Cooperative Role of Arg45 and His64 in the Spectroscopic A3 State of Carbonmonoxy Myoglobin: Molecular Dynamics Simulations, Multivariate Analysis, and Quantum Mechanical Computations, *J. Am. Chem. Soc.*, 1999, 121, 6444-6454.

(45) Grossfield, A.; Feller, S. E.; Pitman, M. C., Convergence of Molecular Dynamics Simulations of Membrane Proteins, *Proteins: Structure, Functions, and BioInformatics*, 2007, 67, 31-40.

(46) 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 Research*, 2000, 28, 235-242.

(47) Sinden, R. R., *DNA Structure and Function*. Academic Press: San Diego, CA, 1994.

(48) Bloomfield, V. A.; Crothers, D. M.; Tinoco, I., *Nucleic Acids: Structures, Properties, and Functions*. University Science Books: Sausalito, CA, 2000.

(49) Neidle, S., *Nucleic Acid Structure and Recognition*. Oxford University Press: New York, NY, 2002.

(50) Feig, M.; Pettitt, M., Structural Equilibrium of DNA Represented with Different Force Fields, *Biophysical Journal*, 1998, 75, 134-149.

(51) Zheng, L.; Thompson, D. L., On the Accuracy of Force Fields for Predicting the Physical Properties of Dimethylnitramine, *J. Phys. Chem. B.*, 2006, 110, 16082-16088.

(52) Jas, G. S.; Kuczera, K., Equilibrium Structure and Folding of a Helix-Forming Peptide: Circular Dichroism Measurements and Replica-Exchange Molecular Dynamics Simulations, *Biophys. J.*, 2004, 87, 3786-3798.

(53) Pitera, J. W.; Swope, W., Understanding Folding and Design: Replica-Exchange Simulations of "Trp-Cage" Miniproteins, *Proc. Natl. Acad. Sci. USA*, 2003, 100, 7587-7592.

(54) Damm, W.; van Gunsteren, W. F., Reversible Peptide Folding: Dependence on Molecular Force Field Used, *J. Comp. Chem.*, 2000, 21, 774-787.

(55) Garcia, A. E.; Sanbonmatsu, K. Y., a-Helical Stabilization by Side Chain Shielding of Backbone Hydrogen Bonds, *Proc. Natl. Acad. Sci. USA*, 2002, 99, 2782-2787.

(56) Zhou, R., Trp-Cage: Folding Free Energy Landscape in Explicit Water, *Proc. Natl. Acad. Sci. USA*, 2003, 100, 13280-13285.

(57) Pilch, D. S.; Poklar, N.; Baird, E. E.; Dervan, P. B.; Breslauer, K. J., The Thermodynamics of Polyamide-DNA Recognition: Hairpin Polyamide Binding in the Minor Groove of Duplex DNA, *Biochemistry*, 1999, 38, 2143-2151.

(58) Pilch, D. S.; Poklar, N.; Gelfand, C. A.; Law, S. M.; Breslauer, K. J.; Baird, E. E.; Dervan, P. B., Binding of a Hairpin Polyamide in the Minor Groove of DNA: Sequence Specific Enthalpic Discrimination, *Proc. Natl. Acad. Sci. USA*, 1996, 93, 8306-8311.

(59) James, P. L.; Le Strat, L.; Ellervik, U.; Bratwall, C.; Norden, B.; Brown, T.; Fox, K. R., Effects of a Hairpin Polyamide on DNA Melting: Comparison with Distamycin and Hoechst 33258, *Biophysical Chem.*, 2004, 111, (3), 205-212.

(60) Tabernero, L.; Verdaguer, N.; Coll, M.; Fita, I.; van der Marel, G. A.; van Boom, J. H.; Rich, A.; Aymami, J., Molecular Structure of the A-tract DNA Dodecamer d(CGCAAATTTGCG) Complexed with the Minor Groove Binding Drug Netropsin, *Biochemistry*, 1993, 32, 8403-8410.

(61) Coll, M.; Frederick, C. A.; Wang, A. H.-J.; Rich, A., A Bifurcated Hydrogen-Bonded Conformation in the Base Pairs of the DNA Dodecamer d(GCGAAATTTGCG) and Its Complex with Distamycin, *Proc. Natl. Acad. Sci. USA*, 1987, 84, 8385-8389.

(62) Wellenzohn, B.; Winger, R. H.; Hallbrucker, A.; Mayer, E.; Liedl, K. R., Simulation of EcoRI Dodecamer Netropsin Complex Confirms Class I Complexation Mode, *J. Am. Chem. Soc.*, 2000, 122, 3927-3931.

(63) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeye, 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.

(64) Humphrey, W.; Dalke, A.; Schulten, K., VMD: Visual Molecular Dynamics, *J. Mol. Graphics*, 1996, 14, 33-38.

Chapter 4

Force Field Parameters for the Simulation of Conjugated Dienes with a Focus on Retinoids

This chapter describes the parameterization procedure for the retinoid molecules, the results of the crystal simulations performed to test the parameters, and conclusions derived from the simulations.

4.1 Parameterization of Retinoids

CHARMM force field topology and parameters for π-conjugated systems were developed and tested against four retinoid compounds (retinoic acid, retinol, fenretinide, and retinal). The parameterization strategy involved the subdivision of key retinoid functionality into 17 small organic (model) compounds. In force field development, one possible approach is to divide a larger molecular system into smaller representative, or model compounds that contain the important functionalities.[1] MacKerell and coworkers have used quantum mechanical data to parameterize various classes of molecules for the CHARMM force field.[2-12] Quantum mechanical data has also been used to parameterize the Merck Molecular Force Field,[13] the AMBER force field,[14] the OPLS-AA force field,[15, 16] the GROMOS force field,[17] and the CVFF force field.[18] Typically, experimental data is critical for force field development. However, geometries derived from quantum mechanical data are particularly useful for force field parameterization, since the time or length scale of experimental studies could introduce significant error or prohibit the measurement process.[13, 19]

Model compounds must be selected so that they capture the important bond and nonbond attributions of the molecular system, and be small enough to be computationally tractable. The model compounds are shown in Figure 4.1. Models **1-5** were selected because all retinoids contain a chain of alternating single and double bonds with methyl groups attached to various carbons on the chain. Structures **6** and **7** are for use in the fenretinide molecule, which contains amide and phenol groups at the tail end. The cyclohexene ring portion that is in all retinoid compounds was modeled by **8** and **9**. The tail ends were parameterized using **10-12**. Model **10** is for retinoic acid, **11** is for retinol, and **12** is for retinal. Models **13** and **14** were parameterized to examine rotation/energetics about the double bonds in retinoids. Models **15-17** were parameterized for the Schiff base linkage for retinal (when retinal binds to rhodopsin or bacteriorhodopsin, it binds covalently to a lysine residue through a Schiff base linkage). Models **18** and **19** are used to create patches for making the covalent linkage from retinal to the lysine of the protein; the patches simply delete the hydrogens on the nitrogen of the side chain of lysine and the oxygen of the carbonyl group on the tail of retinal and make a double bond between the lysine nitrogen and the carbonyl carbon of retinal.

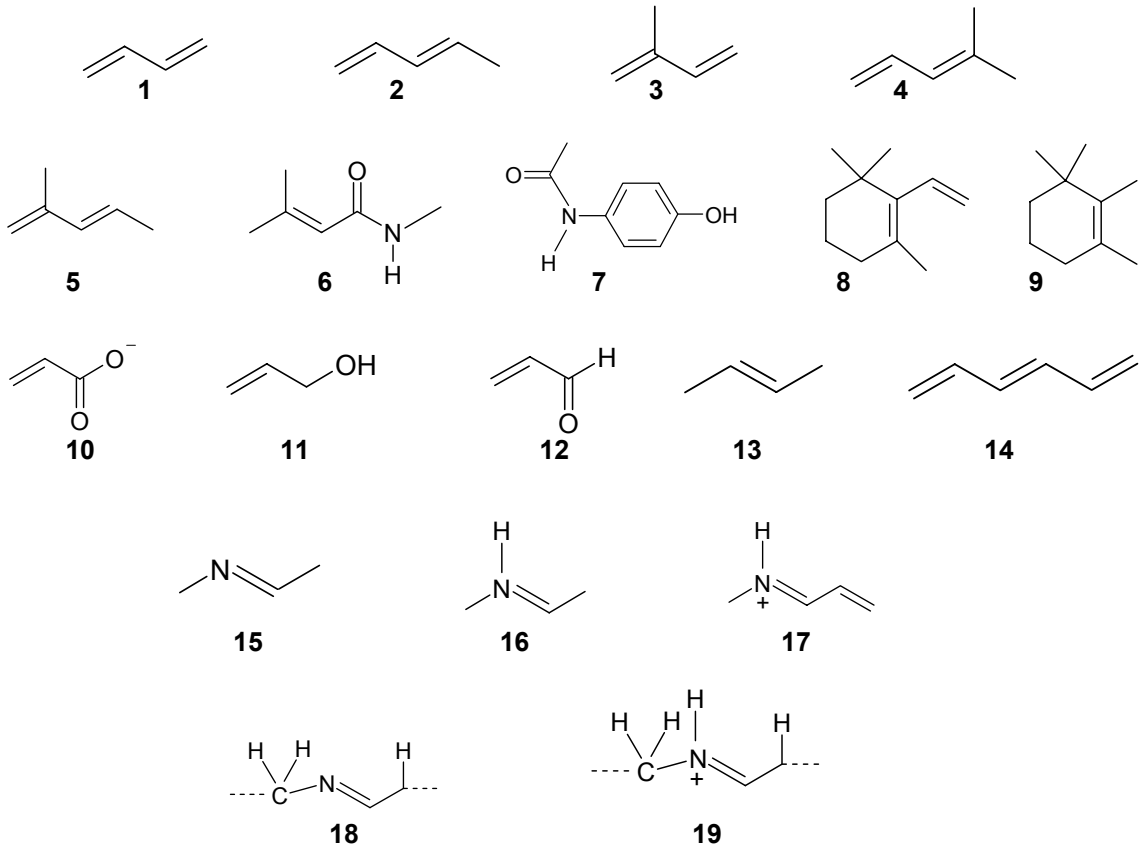


Figure 4.1 Model compounds and patch residues for retinoids (some hydrogens not shown for clarity). Models **1-17** are the model compounds, and **18** and **19** are used to create patches for forming Schiff bases.

New atom types were created to allow for proper treatment of alternating single and double bonds. The previous CHARMM parameters used only two atom types for conjugated systems, CE1 for internal carbons and CE2 for terminal carbons. Previously, alternation between single and double bonds was not possible with only one atom type for the internal carbons (Figure 4.2). The previous parameters (when the structure was energy minimized) showed only a 0.002 Å difference between the carbon-carbon single and double bonds, whereas microwave[20] and infrared data[21] resulted in a difference of 0.13 Å, Raman data gave a difference 0.14 Å,[22] and electron diffraction data showed a difference between 0.12 to 0.15 Å.[23-26]

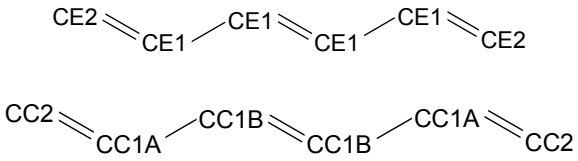


Figure 4.2 Previous 1,3,5-hexatriene carbon atom types (top). New 1,3,5-hexatriene atom types (bottom).

New atom types for π-conjugated systems have been created, which are CC2 for terminal carbons and CC1A and CC1B for internal carbons (Figure 4.2). Each model compound was energy minimized at the MP2/6-31G(d) level of theory.[13, 27, 28] The average MP2/6-31G(d) carbon-carbon bond lengths over all model compounds was 1.50 ± 0.03 Å for single bonds, and 1.35 ± 0.02 Å for double bonds. The average bond lengths calculated from MP2 were used as the equilibrium bond lengths for the new CHARMM parameters for the associated bond type. The force constants were based on MP2/6-31G(d) frequency computations for **1**, **13**, **15**, and **16** in Figure 4.1 as an initial guess. Each model compound was energy minimized with the CHARMM force field with steepest descent and Newton-Raphson (tolerance gradient was 10^{-6} kcal/mol/Å for each minimization). The parameters where then adjusted to minimize error with the MP2/6-31G(d) structure. The average value for the CHARMM single bonds is 1.51 ± 0.02 Å, and the average value for double bonds is 1.36 ± 0.02 Å. The average error between MP2 and CHARMM for single bonds is 0.01 ± 0.01 Å, and 0.005 ± 0.006 Å for double bonds. The average error between MP2 and CHARMM over all bonds (not just carbon-carbon bonds) in the model compounds is 0.012 ± 0.009 Å.

New angle terms and angle modifications were also made. The minimized angle values obtained from the MP2 optimized structures were used as the equilibrium angle

values for the new CHARMM parameters. The force constants were based on MP2/6-31G(d) frequency computations for **1**, **13**, **15**, and **16** in Figure 4.1 as an initial guess. Each model compound was energy minimized with the CHARMM force field with steepest descent and Newton-Raphson (tolerance gradient was 10^{-6} kcal/mol/Å). The parameters were adjusted to minimize the error across all model compounds. The average error between MP2 and CHARMM was calculated for angles as 2 ± 2º.

Torsion angle terms were also reparameterized. The torsion angles chosen were selected because they involve low energy barriers that are important in describing the dynamics allowing the retinoids to interact with various proteins; the single bonds and any bonds linking the conjugated hydrocarbon chain of the retinoids to a functional group must be able to rotate in order for the retinoid molecule to maximize favorable interactions with amino acids in the binding site of the protein. For each model compound, the potential energy surfaces (PES) for selected torsion angles (see Figures 4.3-4.9 for selected torsion angles; the bond with the arrows were the bonds which were rotated about) were calculated by holding the selected torsion angle fixed at different increments (20º increments from –180º to 180º) and geometry optimizing all other degrees of freedom. The torsion angles selected were for rotations about the single bond between two double bonds, rotation of a particular functional group that is necessary for optimizing ligand-protein contacts, or rotations necessary for ring puckering. The CHARMM dihedral angle parameters force constant (K_χ), multiplicity (n), and phase shift (δ) were then modified to match the quantum mechanical results from the potential energy surfaces. Each model compound was energy minimized in CHARMM (steepest descent and Newton-Raphson; tolerance gradient was 10^{-6} kcal/mol/Å) with the selected

torsion angle held fixed, and the potential energy surface of the dihedral was calculated. If the surface did not match the MP2 results, then the multiplicity, phase shift, or force constant were adjusted as necessary.

For models **1**-**5** in Figure 4.1, the single bond between the two double bonds of the molecules was rotated from –180º to 180º in 20º increments. Figure 4.3 shows the torsion angle results for 1,3-butadiene and 1,3-pentadiene.

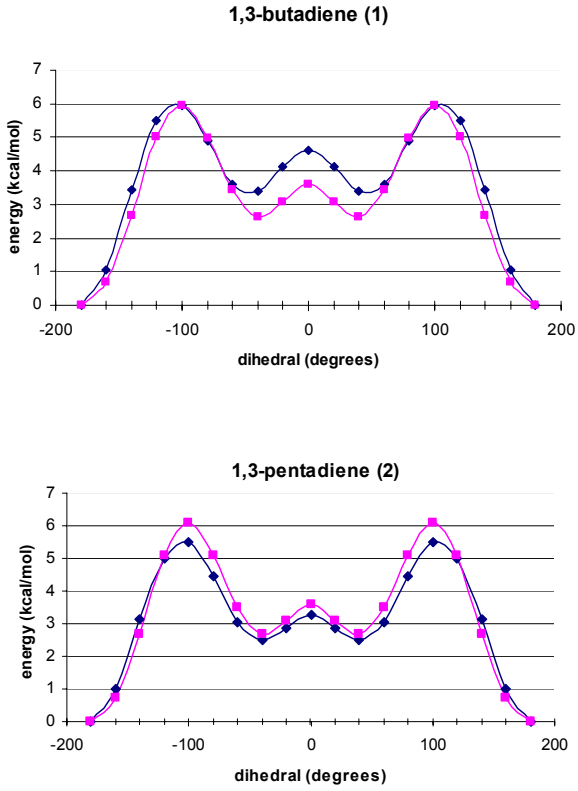


Figure 4.3 Torsion energy plots for 1,3-butadiene (top) and 1,3-pentadiene (bottom). Diamonds, final CHARMM; square, MP2/6-31G(d).

The initial CHARMM surface for 1,3-butadiene had barriers at ~64 kcal/mol (Appendix B), which is ~58 kcal/mol higher than the MP2 barriers; the initial CHARMM also did not show the correct multiplicity. This was corrected by changing the force constant (K_χ)

to lower the barriers and the multiplicity (*n*) to get the correct number of barriers. The final CHARMM results show the correct phase and multiplicity as the MP2 results. The CHARMM energy barrier at 0º, which corresponds to *cis*-1,3-butadiene, is 1.03 kcal/mol higher than the MP2 results, but overall, the results agree well.

1,3-Butadiene has been examined by several experiments, including electron diffraction,[23-26] UV,[29] IR,[21, 30, 31] microwave,[32] Raman,[22, 33] and computational studies.[34-38] The *trans* form is the dominant conformation, and the *cis* (0º) or *gauche* (~40º) conformation is the minor component. However, there is still debate on whether the *cis* or *gauche* conformer is the more stable minor conformer.[39-44] Computational studies show that the *cis* form is a "transition structure" between the two *gauche* conformers, but generally, experimental studies say that the *cis* form is the minor conformer.[31, 45] This difference has been attributed to conditions under which the torsional potential was calculated or measured. Theoretical and vapor phase spectroscopy results favor the *gauche* form as the second stable conformer, while spectroscopic measurements performed in an argon matrix favor the *cis* form as the second stable conformer of 1,3-butadiene (the interaction with the surrounding Ar matrix stabilizes the *cis* form and thus changes the torsional potential).[46] Also, it is difficult to distinguish the *cis* and *gauche* forms from one another by an experimental direct measurement because only ~1% of the minor conformer is present at room temperature, and 1,3-butadiene is reactive at higher temperatures.[42, 44]

For 1,3-pentadiene (bottom of Figure 4.3), the initial CHARMM had energy barriers that are too high (~92 kcal/mol; see Appendix B), and the multiplicity was incorrect. The energy barriers were lowered by decreasing the value of the force constant

(K_χ) to 0.5600 kcal/mol/Å^2, and the correct number of barriers was obtained by changing the multiplicity (n) to 2. The final CHARMM surface shows the correct phase and multiplicity, and the energy barriers are in good agreement with the MP2 results. The CHARMM energy barriers at 100º and –100º are ~0.58 kcal/mol lower than the MP2, and the CHARMM energy barrier at 0º is ~0.34 kcal/mol lower than the MP2.

Figure 4.4 shows the torsion angle results for the conjugated methylated butanes and pentenes. As for 1,3-butadiene, all of the methylated dienes have two minima (*gauche* conformers) in addition to the global minimum (*trans* conformer) on their potential energy surfaces. Also, the initial CHARMM for the methylated dienes shows very high energy barriers and incorrect multiplicities (Appendix B). 2-Methyl-1,3-butadiene shows a maximum at 116 kcal/mol for the initial CHARMM, 2-methyl-1,3-pentadiene shows a maximum at 125 kcal/mol, and 4-methyl-1,3-pentadiene has a maximum at 93 kcal/mol. The final CHARMM parameters are in good agreement with the MP2 results. For 2-methyl-1,3-pentadiene, the CHARMM barrier height is 0.62 kcal/mol lower than MP2 at 0º and 0.57 kcal/mol lower than MP2 at –100º and 100º. 4-Methyl-1,3-pentadiene has a CHARMM surface that is 0.90 kcal/mol higher than MP2 at 0º and 0.76 kcal/mol higher at 100º. 2-Methyl-1,3-butadiene has good agreement between CHARMM and MP2 with CHARMM having barriers only 0.28 kcal/mol higher than MP2 at the barrier at 0º.

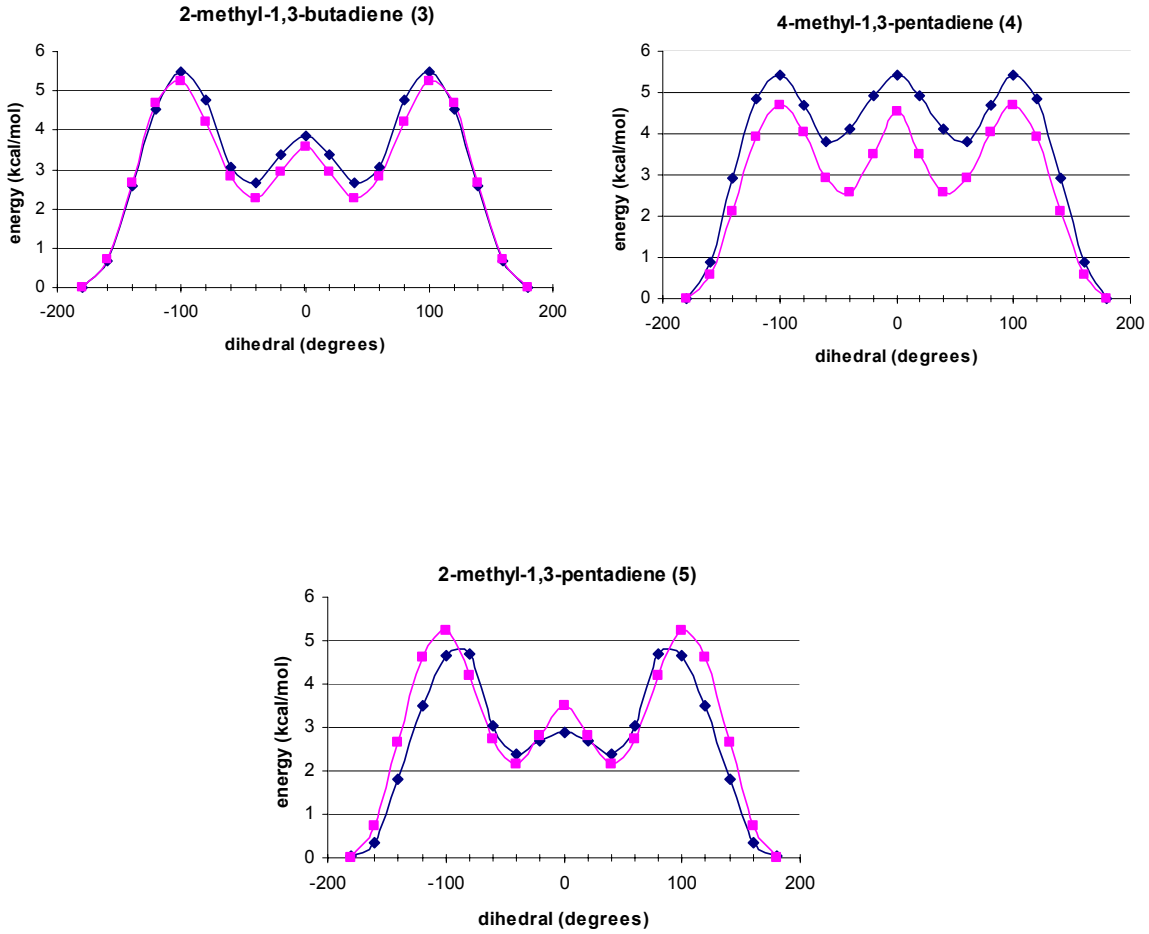


Figure 4.4 Torsion energy plots for 2-methyl-1,3-butadiene (top left), 4-methyl-1,3-pentadiene (top right), and 2-methyl-1,3-pentadiene (bottom). Diamonds—final CHARMM surfaces; square—MP2/6-31G(d) surface.

Model compounds **6** and **7** are for use in the fenretinide molecule, which contains and amide and phenol group at the tail end.

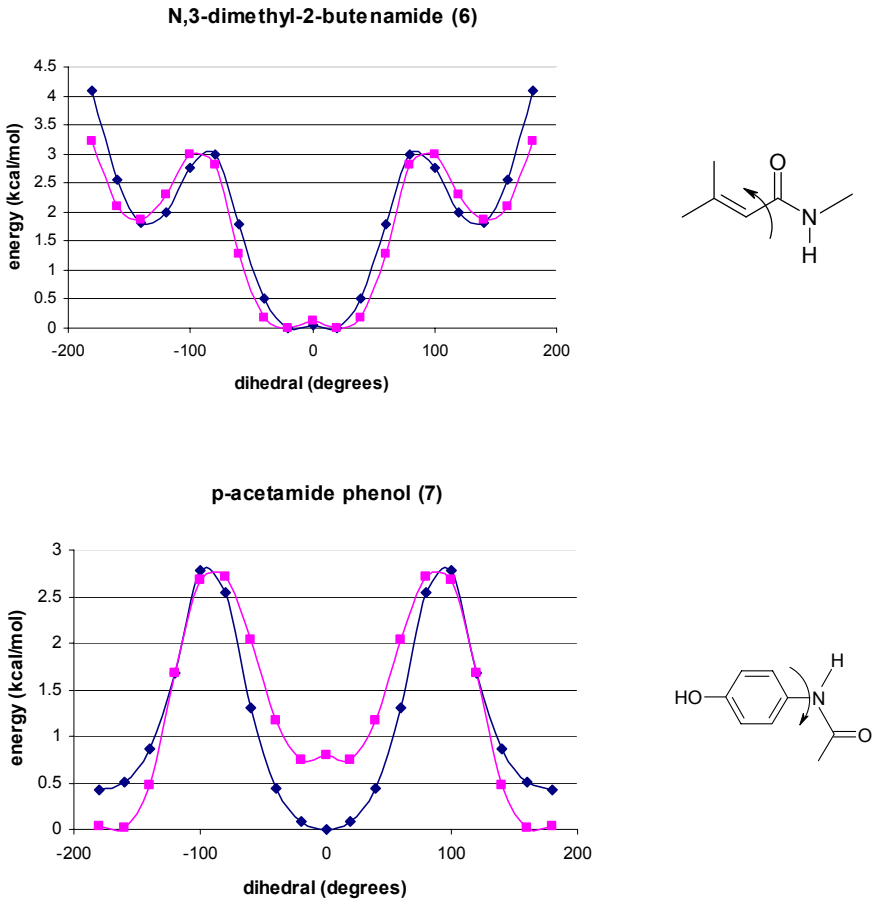


Figure 4.5 Torsion energy plot for N-3-dimethyl-2-butenamide (top) and *p*-acetamide phenol (bottom). Diamonds; MP2/6-31G(d); squares; final CHARMM.

N-3-dimethyl-2-buteneamide shows high energy barriers for the initial CHARMM (Appendix B) and also shows the wrong phase and multiplicity. The final CHARMM shows much better agreement with MP2, with the CHARMM being 0.86 kcal/mol lower than MP2 at the maxima (180° and -180°).

The initial CHARMM energy barriers for *p*-acetamide phenol, **7**, are too high (~13 kcal/mol; Appendix B), but the phase and multiplicity are correct. The final CHARMM maxima at 80º are in better agreement with MP2, and the minimum at 0º is ~0.80 kcal/mol higher than MP2.

Models **8** and **9** were parameterized for the cyclohexene ring portion that is in all retinoid compounds.

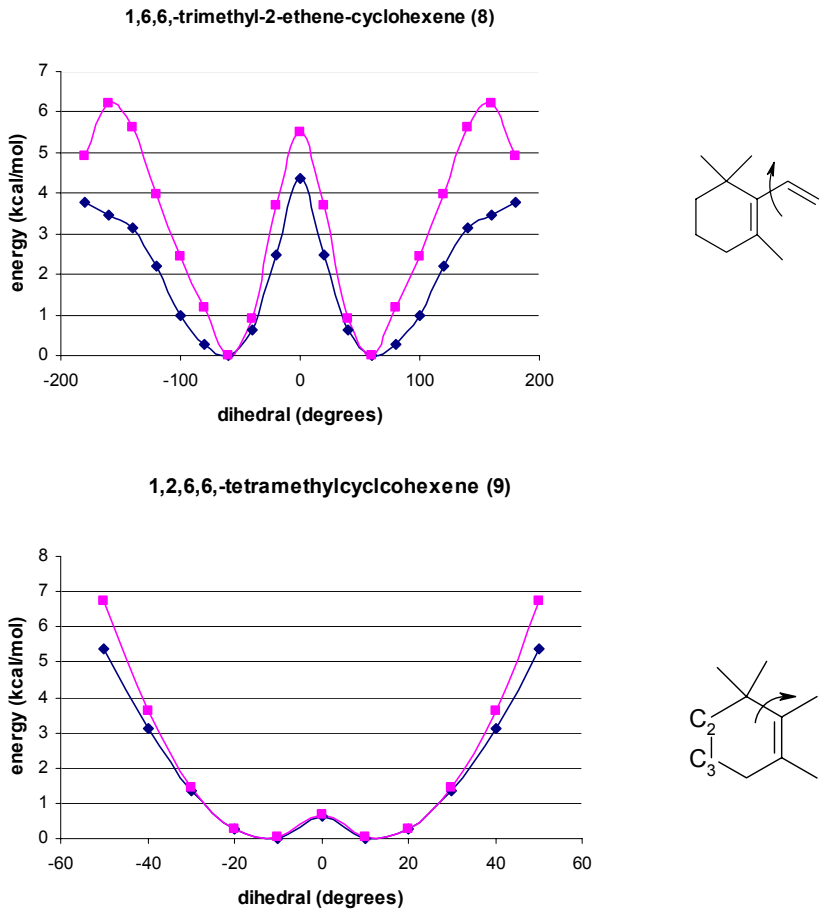


Figure 4.6 Torsion energy surfaces for 1,6,6-trimethyl-2-ethene-cyclohexene (top) and 1,2,6,6-tetramethylcyclohexene (bottom). Diamonds, MP2/6-31G(d); squares, final CHARMM.

The initial CHARMM parameters for the 1,6,6-trimethyl-2-ethene cyclohexene behave like the others; the energy barriers are too high (~120 kcal/mol; Appendix B). The final CHARMM parameters are in better agreement, but they show an increase in energy at 150º and -150º, whereas MP2 does not. Also, CHARMM is ~1.1 kcal/mol higher than MP2 at the maximum at 0º. The dihedral surface for 1,2,6,6-

tetramethylcyclcohexene was examined for the purpose of deciding which ring pucker was more favorable, C2-*endo* (C2 carbon above the plane of the ring) or C3-*endo* (C3 carbon above the plane of the ring). The side of the ring containing the double bond stays planar, but the other side of the ring can move and pucker with either the C2 or C3 above the ring plane. The dihedral examined was rotated from –50º to 50º in 10º increments (the dihedral is constrained being in the cyclohexene ring). When the dihedral is at –10º, the ring is in the C3-*endo* conformation, and when the dihedral is at 10º, the ring is in the C2-*endo* conformation. Since both conformations are at a minimum on the surface, each structure was geometry optimized to find which conformation was lowest in energy. The MP2 calculations showed that the both conformations were nearly the same in energy; the C2-*endo* structure was only 3.00×10^{-4} kcal/mol lower in energy than the C3-*endo* structure. The initial and final CHARMM parameters are almost the same for this model compound (Appendix B), and after revision, the final CHARMM parameters are only 0.5 kcal/mol higher than the initial parameters at 50º. MP2 is ~1.3 kcal/mol higher than CHARMM at 10º and -10º.

Models **10**, **11**, and **12** were parameterized for the tail ends of the retinoids. Model **10** is for retinoic acid, Model **11** is for retinol, and Model **12** is for retinal. For Models **10**, **11**, and **12**, rotation to obtain the surfaces was about the C-C single bond in the middle of the structure.

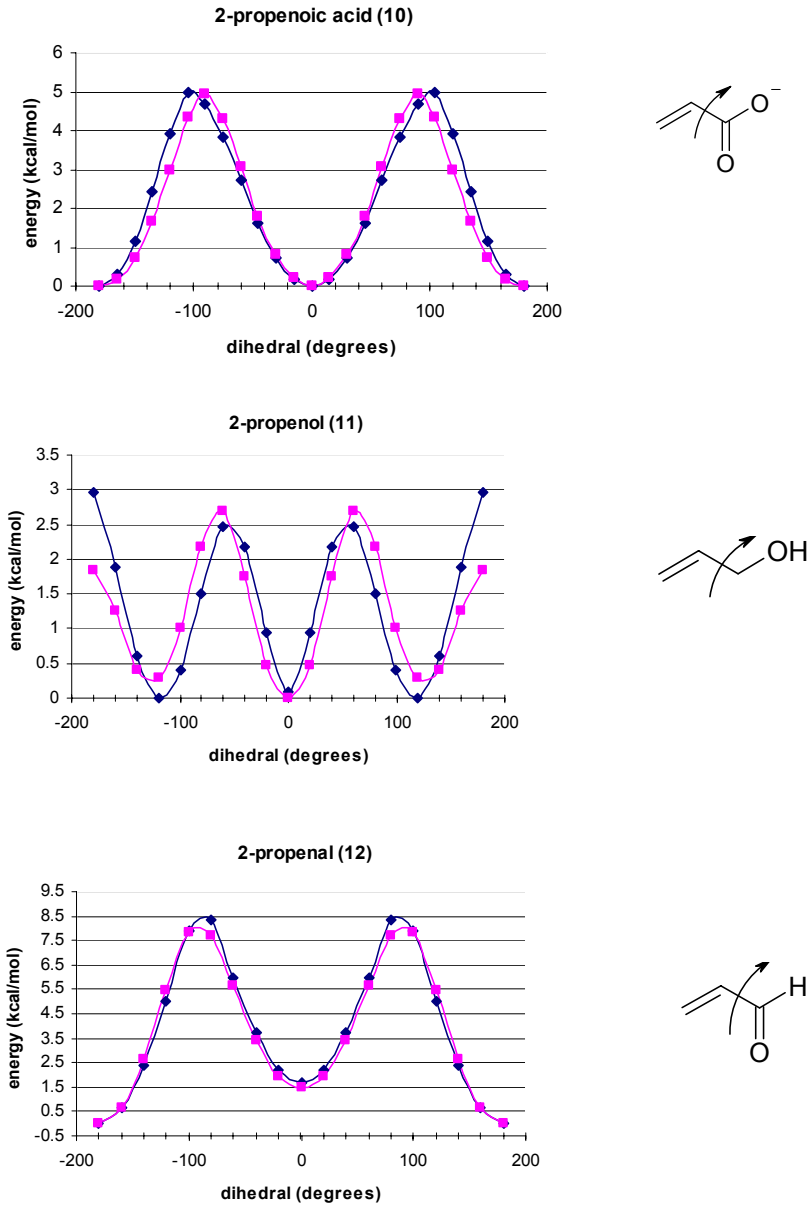


Figure 4.7 Torsion energy surfaces for 2-propenoic acid (top left), 2-propenol (top right), and 2-propenal (bottom). Diamonds, final CHARMM; squares, MP2/6-31G(d).

The initial CHARMM parameters for 2-propenoic acid, 2-propenol, and 2-propenal do not show the same magnitude of high energy barriers as the conjugated dienes. Nevertheless, the initial parameters do not agree with the MP2 results (see Appendix B).

The final CHARMM parameters for 2-propenoic acid are in good agreement; CHARMM shows the correct minima (at 0º, 180º, and -180º) and maxima; CHARMM is only 0.08 kcal/mol higher than MP2 at the maxima. For 2-propenol, the final CHARMM parameters are 1.1 kcal/mol higher than MP2 at 180º and –180º, and they give the correct phase and multiplicity. The final CHARMM parameters for 2-propenal are in also in good agreement with MP2, showing the correct barrier heights; CHARMM is only 0.5 kcal/mol higher than MP2 at the maxima.

In general, the initial CHARMM parameters delivered unrealistic torsional energetic barriers, which were parameterized by changing the force constants and multiplicities to match the MP2 data. For each case, the changes are given in Appendix B. The final CHARMM torsional surfaces agree well with those calculated from MP2. The largest difference observed between MP2 and CHARMM was 1.1 kcal/mol.

Models **13** and **14** were parameterized to examine rotation/energetics about the double bonds in the retinoids. CHARMM follows MP2 at the higher values for the dihedral (120º-180º), but as the compound approaches the lower values and gets closer to a *cis* conformation (105º down to 45º), CHARMM deviates from MP2, and at 45º, there is a 49 kcal/mol difference. As for 2-butene, the CHARMM parameters for 1,3,5-hexatriene follow MP2 when the molecule has a *trans* geometry (-180º to -120º), but as the structure approaches a more *cis* conformation about the middle double bond, CHARMM starts to deviate from MP2. At 60º, CHARMM even starts to show a decrease in energy. This deviation is due to a conformational change of the end groups that are on either side of the double bond being rotated. The CHARMM structures do not show as much movement of the end groups.

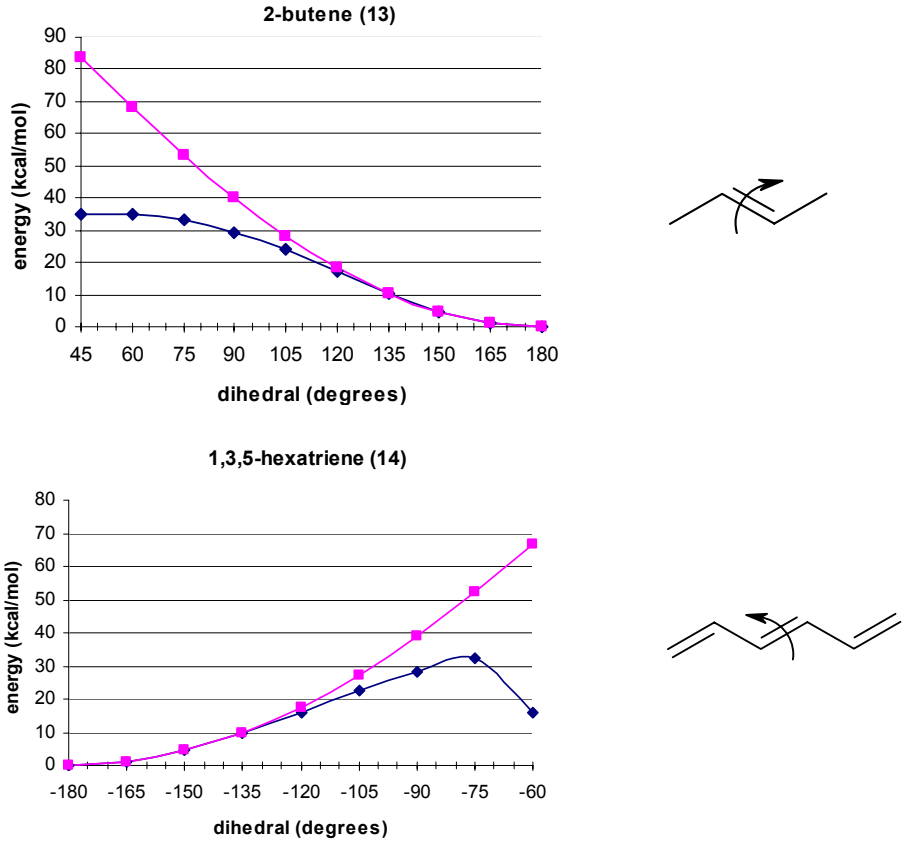


Figure 4.8 Torsion energy surfaces for 2-butene (left) and 1,3,5-hexatriene (right). Diamonds, final CHARMM; squares, MP2/6-31G(d).

Models **15**, **16**, and **17** in Figure 4.1 were parameterized for the Schiff base linkage for retinal. When retinal binds to rhodopsin or bacteriorhodopsin, it binds covalently to a lysine residue through a Schiff base linkage. Model **15** is for a deprotonated Schiff base linkage, and **16** and **17** are for protonated Schiff base linkages. The CHARMM for the first two Schiff base compounds (Figure 4.9) follows the MP2 up to the last two points tested, and then it starts to deviate by about 9 kcal/mol. The double bond was used to examine the rotation/energetics for the Schiff base linkage in retinal. The CHARMM parameters for the third Schiff base compound (Figure 4.9) are in excellent agreement with MP2; the multiplicity and phase are correct, and the barrier heights coincide almost exactly.

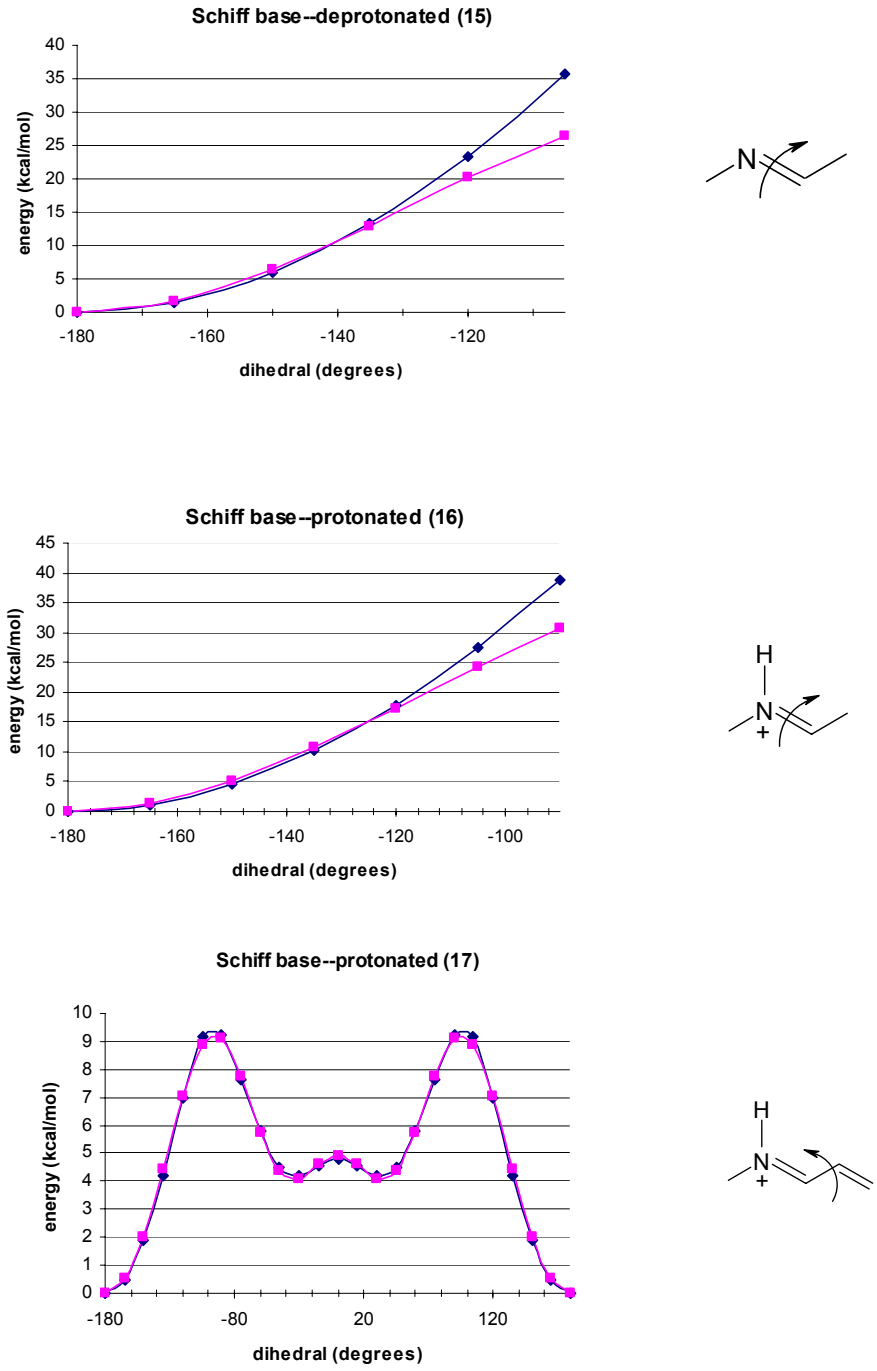


Figure 4.9 Torsion energy plot for deprotonated Schiff base (top left) and protonated Schiff base (top right and bottom). Diamonds, final CHARMM; squares, MP2/6-31G(d).

The last two models in Figure 4.1 (**18** and **19**) are for patches making the covalent linkage from retinal to the lysine of the protein. The Schiff base linkages they contain were parameterized (**15**, **16**, and **17**); the patches simply delete the hydrogens on the nitrogen of the side chain of lysine and the oxygen of the carbonyl group on the tail of retinal and make a double bond between the lysine nitrogen and the carbonyl carbon of retinal.

The final CHARMM parameters for the model compounds are in good agreement with the MP2/6-31G(d) results. The initial CHARMM parameters not only treated the alternating single and double bonds incorrectly, but for many of the model compounds, the energy barriers for the torsion angles were too high (over 60 kcal/mol), and some were out of phase. The final CHARMM parameters exhibit the correct barrier heights and phases, and largest difference between CHARMM and MP2 was 1.15 kcal/mol (CHARMM is higher than MP2 for 1,6,6-trimethyl-2-ethene-cyclohexene at the maxima at –180º and 180º). The average error between CHARMM and MP2 in bond distances is 0.012 ± 0.009 Å, the average error for angles is 1 ± 1º, and the average error for dihedrals is 2 ± 4º.

4.2 Comparison to Available Experimental Data

Some of model compounds and the retinoids built from them were compared to available experimental data. Experimental data for all model compounds is not available, nor is experimental data for all four retinoids. The most experimental data was for 1,3-butadiene, and this included mainly electron diffraction data;[23-26] however, there was infrared,[21] microwave,[32] and Raman[22] data as well. Electron diffraction data was also found for 2-methyl-1,3-butadiene[47] and 1,3,5-hexatriene,[25] and x-ray diffraction data was

found for *p*-acetamide phenol[48] and 2-propenoic acid.[49] For the final retinoids, x-ray structures of all-*trans* retinoic acid[50, 51] and all-*trans* retinal[52] were compared to the final retinoids in CHARMM. The experimental comparisons with the new CHARMM parameters and MP2 results (in the case of the smaller model compounds) for 1,3-butadiene are shown in Table 4.1 (CHARMM atom numbering referred to in Table 4.1 is shown above the table).

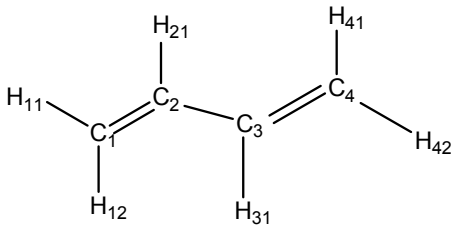


Table 0.1 Experimental, QM, and MM comparison for 1,3-butadiene.

Method	C1-C2	C2-C3	C3-C4	C-H (all)	C1-C2-C3	C2-C3-C4	C=C-C	C=C-H	C1-C2-C3-C4
ED[a]	1.35	1.46	1.35	1.06	124 ± 2	124 ± 2			180
ED[b]	1.337	1.483	1.337	1.082			122.4	119.8	180
ED[c]	1.344	1.467	1.344	1.094			122.8 ± 0.5	119.5 ± 1	180
ED[d]	1.341	1.463	1.341	1.086			123.3 ± 0.5	121.8 ± 1.2	180
IR[e]	1.338	1.464	1.338	1.086			123.2 ± 0.2	119.6 ± 0.2	180
Microwave[f]	1.337	1.467		1.087	123.5			120.9 ± 1.13	180
Raman[g]	1.337	1.476	1.337	1.085			122.9 ± 0.5	120	180
MP2/6-31G(d)	1.344	1.458	1.344	1.087	123.7	123.7	123.7 ± 0.2	120.9 ± 1.05	180
CHARMM	1.346	1.471	1.346	1.1	123.7	123.7	123.7 ± 0.3	120.1 ± 1.34	180

Bonds in Å; angles and dihedrals in degrees.
a see ref. 23
b see ref. 24; C=C bond error 0.005 Å, C-C bond error 0.01 Å, C-H bond error 0.01 Å.
c see ref. 25
d see ref. 26; C=C bond error 0.002 Å, C-C bond error 0.003 Å, C-H bond error 0.04 Å.
e see ref. 21; C-C bond error 0.003 Å.
f see ref. 32
g see ref. 22 C-C bond error 0.01 Å.

From Table 4.1, the MP2/6-31G(d) results and the final CHARMM results compare well to the experimental bonds, angles, and dihedral angle. MP2 and CHARMM both give the *trans* structure (180º) of 1,3-butadiene after energy

minimization; all of the experimental techniques find this as well. Also, the average error between MP2 and CHARMM compared with experimental bonds is 0.007 ± 0.005 Å and 0.009 ± 0.008 Å, respectively, and the average error between MP2 and CHARMM compared with experimental angles is 0.7 ± 0.5º and 0.6 ± 0.4º, respectively. When comparing the computed values to electron diffraction results, it should be kept in mind that the average carbon-carbon bond length can be determined with high accuracy when equivalent bonds are present, but if non-equivalent bonds are present, the individual values can be much less accurate.[53]

Table 4.2 shows the comparison of experimental data with the structures obtained from the final CHARMM parameters for isoprene (2-methyl-1,3-butadiene).

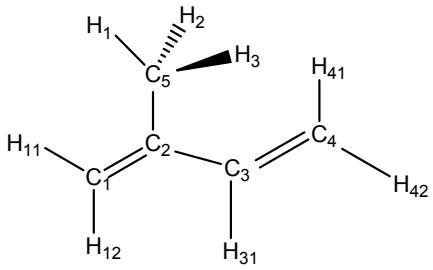


Table 0.2 Experimental, QM, and MM comparison for 2-methyl-1,3-butadiene.

	C=C	C2-C3	C2-C5	Csp^2-H	Csp^3-H	C1-C2-C3	C2-C3-C4	C1-C2-C5	H12-C1-C2	C2-C5-H1
ED[a]	1.34	1.46	1.51	1.08	1.11	121.4	127.3	121	124.3	109.1
MP2/6-31G(d)	1.35	1.47	1.51	1.09	1.09	122.6	122.6	119.8	121.7	111
CHARMM	1.35	1.49	1.51	1.1	1.1	121.8	128.4	119.4	121.3	114.3
MP2/ED error	0.01	0.01	0	0.01	0.02	1.2	4.7	1.2	2.6	1.9
CH/ED error	0.01	0.03	0	0.02	0.01	0.4	1.1	1.6	3	5.2
MP2/CH error	0	0.02	0	0.01	0.01	0.8	5.8	0.4	0.4	3.3

Bonds in Å; angles in degrees.
a see ref. 47; no error reported

For isoprene, the MP2 and CHARMM results compare well to the ED data. The average error for bonds between MP2 and the ED data and CHARMM and the ED data is 0.01 ± 0.01 Å and 0.01 ± 0.02 Å, respectively. The average error for angles compared to the ED

data is 2.5 ± 1.4º for MP2 and 3.1 ± 2.7º for CHARMM. The MP2 and CHARMM results are also in good agreement; the average error for bonds between MP2 and CHARMM is 0.008 ± 0.009 Å, and the average error for angles is 2.4 ± 2.2º.

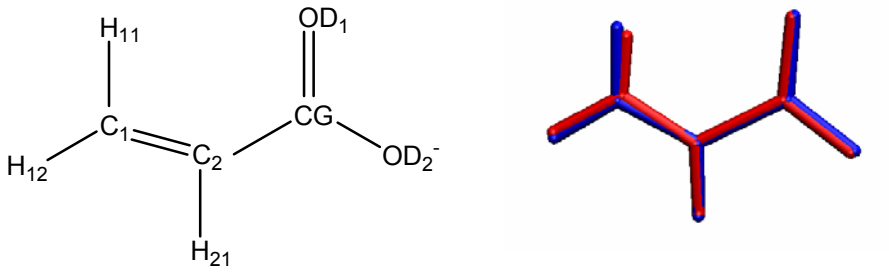


Table 0.3 Experimental, QM, and CHARMM comparison of 2-propenoic acid.

	C2-CG	C1-C2	CG-OD2	CG-OD1	OD1-OD2	OD1-CG-C2	OD1-CG-OD2	C1-C2-CG-OD2	C1-C2-CG-OD1
x-ray[a]	1.47	1.3	1.28	1.26	2.66	116	122	0	180
MP2/6-31G(d)	1.48	1.34	1.27	1.26	2.29	113.9	130	0	180
CHARMM	1.5	1.34	1.26	1.26	2.22	116.2	123.6	0	180
MP2/x-ray error	0.01	0.04	0.01	0	0.37	2.1	8	0	0
CH/x-ray error	0.03	0.04	0.02	0	0.44	0.2	1.6	0	0
MP2/CH error	0.02	0	0.01	0	0.07	2.3	6.4	0	0

Bonds in Å; angles and dihedrals in degrees.
h see ref. 49

An x-ray structure has been reported for 2-propenoic acid.[49] The average error for bonds between CHARMM and the x-ray data is 0.1 ± 0.2 Å (Table 4.3). This error is larger than the other comparisons when CHARMM is compared to experimental data; and this is probably due to the fact that CHARMM gives bond lengths that are slightly longer than the x-ray data for the carbon-carbon bonds, and one of the carbon-oxygen bonds is shorter than what was found in the x-ray structure. The error between MP2 and the x-ray data is smaller than that of CHARMM, with the error being 0.09 ± 0.1 Å. The error between MP2 and CHARMM is smallest at 0.02 ± 0.02 Å. Two of the MP2 angles are slightly less than the x-ray structure, and the angle between the two oxygens is larger

than the x-ray structure, giving an average error of 5.4 ± 3.0º. The error between CHARMM and the x-ray structure is smaller for the angles at 1.2 ± 0.8º. The average angle error between CHARMM and MP2 is 4.3 ± 2.1º. An overlay of the x-ray structure (red) and the CHARMM minimized structure (blue) is shown above Table 4.3; the RMSD between the two structures is 0.0001 Å.

An electron diffraction structure has been reported for 1,3,5-hexatriene.[25] All three structures give the *trans* conformation about the double bonds, and most of the bonds and angle from CHARMM and MP2 agree with those of the experimental structure (average bond errors are 0.01 ± 0.03 Å for MP2 and ED and 0.02 ± 0.02 Å for CHARMM and ED). However, CHARMM gives slightly longer carbon-carbon single bonds (~0.02 Å) than the x-ray and MP2 structures (Table 4.4).

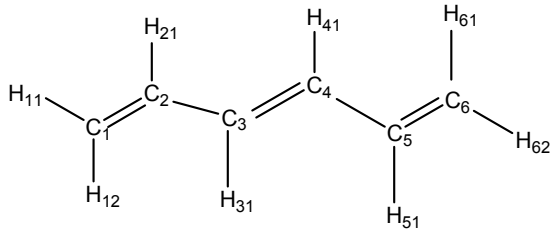


Table 0.4 Experimental, QM, and MM comparison of 1,3,5-hexatriene

	C1-C2	C2-C3	C3-C4	C4-C5	C5-C6	C-H (all)	C=C-C	C=C-H	C-C-C-C
ED[a]	1.345	1.45	1.345	1.45	1.345	1.103	124.3 ± 1.5	117.8	180
MP2/6-31G(d)	1.346	1.451	1.345	1.451	1.346	1.088	123.8 ± 0.04	120.4	180
CHARMM	1.346	1.475	1.346	1.476	1.345	1.1	123.8 ± 0.26	119.8	180
MP2/ED error	0.001	0.001	0	0.001	0.001				0
CH/ED error	0.001	0.025	0.001	0.026	0				0
CH/MP2 error	0	0.024	0.001	0.025	0.001				0

Bonds in Å; angles and dihedrals in degrees.
a see ref. 25

An x-ray structure was reported for *p*-acetamide phenol, and the bonds, angles, and dihedrals from MP2 and CHARMM agree with those of the x-ray structure (Table

4.5). The average errors between MP2 and the x-ray structure are 0.01 ± 0.01 Å for bonds, 1.1 ± 1.3º for angles, and 4.1 ± 5.8º for dihedrals. The average errors between CHARMM and the x-ray structure are 0.02 ± 0.01 Å for bonds, 1.0 ± 0.7º for angles, and 1.7 ± 1.1º for dihedrals. CHARMM is in good agreement with the x-ray structure for dihedrals; the largest deviations for MP2 and the x-ray structure are the dihedrals C23-C22-N21-C15 and C27-C22-N21-C15, which are the two dihedrals connecting the amide group to the phenol ring. An overlay of the x-ray structure (red) and the CHARMM minimized structure (blue) is shown above Table 4.5; the RMSD between the two structures is 0.0010 Å. For clarity, the ring hydrogens and methyl group hydrogens on C14 are not shown.

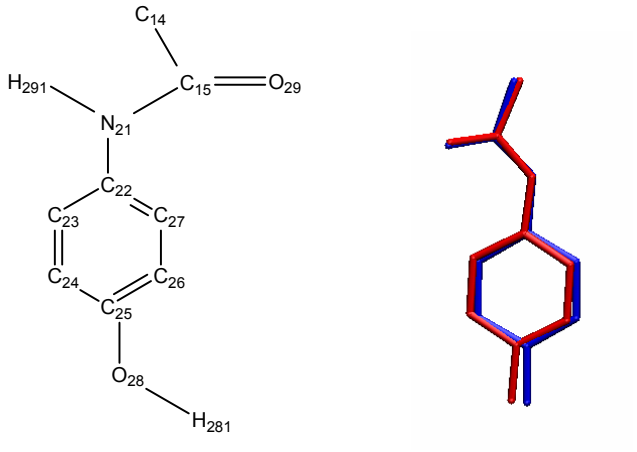


Table 0.5 Experimental, QM, and MM comparison of *p*-acetamide phenol[b]

	C22-C27	C22-N21	C15-O29	C27-C26	C22-C27-C26-C25	C25-C24-C23-C22	O28-C25-C26-C27
x-ray[a]	1.39	1.42	1.22	1.39	0.45	0.4	-178.2
MP2/6-31G(d)	1.4	1.41	1.22	1.39	0.09	0.07	-180
CHARMM	1.41	1.41	1.23	1.39	0.26	0.18	-179.9
MP2/x-ray error	0.01	0.01	0.01	0	0.36	0.33	1.8
CH/x-ray error	0.02	0.01	0	0.01	0.19	0.22	1.7
CH/MP2 error	0.01	0.01	0.01	0.01	0.17	0.11	0.1

Bonds in Å; dihedrals in degrees.
a see ref. 48
b Only some bonds and dihedrals are shown in the table. For a complete table, see Appendix B.

Two retinoids (retinal and retinoic acid) had available experimental data. The crystal structure of all-trans retinal was taken from the reference of Hamanaka, et al., [52](Cambridge Structural Database[54] id TRETAL01), and the average bond error between x-ray and CHARMM was 0.02 ± 0.02 Å, the average angle error was 1.93 ± 1.73º, and the average dihedral error was 15.6 ± 30.5º. The RMSD between CHARMM and x-ray is 1.92 Å. The cyclohexene ring of both structures has the same conformation (C2-*endo*), but several dihedrals in the chain show large differences between the structures; this is what causes the different orientations of the methyl groups (Figure 4.10). There are three dihedrals of the chain that show large differences between the CHARMM minimized and x-ray structures and two dihedrals of the ring that show large differences. These dihedrals are all rotations about single bonds of the chain, and because they are different in the two structures, the orientations of the methyl groups are different (see Appendix B for a table of bonds, angles, and dihedrals for CHARMM and the x-ray structure).

Figure 4.10 The left structure is that of retinal, and the right picture shows an overlay of the x-ray structure (red) and the CHARMM energy minimized structure (blue). The heavy atom RMSD between CHARMM and x-ray is 1.92 Å.

The triclinic crystal structure of all-*trans* retinoic acid was taken from the reference of Stam and MacGillavry[51] (Cambridge Structural Database[54] id VITAAC10). The average bond error between crystal and CHARMM is 0.04 ± 0.09 Å, the average angle error is 2.4 ± 1.7º, and the average dihedral error is 11.8 ± 20.2º. The RMSD

between x-ray and CHARMM is 2.18 Å. The methyl groups of the chain point in the same direction in both structures; however, some dihedrals along the chain show large differences between the two structures, and the cyclohexene rings are in different conformations (Figure 4.11). The x-ray structure shows a C2-*endo* conformation for the ring, while CHARMM gives a C3-*endo* conformation (see Appendix B for a comparison of CHARMM to the x-ray structure bonds, angles, and dihedrals).

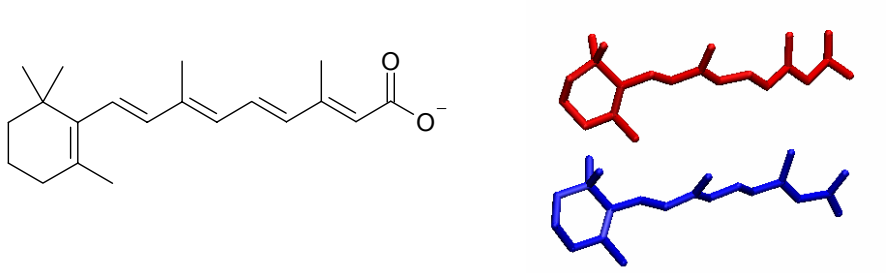


Figure 4.11 The structure on the right is that of retinoic acid, and the right picture is a comparison of the x-ray structure (red) and the CHARMM energy minimized structure (blue) for the triclinic form of retinoic acid.

The monoclinic crystal structure of all-trans retinoic acid was taken from the reference of Stam[50] (Cambridge Structural Database[54] id VITAAC01; search date 2/17/07). The average bond error between crystal and CHARMM is 0.04 ± 0.09 Å, the average angle error is 1.9 ± 1.6º, and the average dihedral error is 13.6 ± 26.0º. RMSD between x-ray and CHARMM is 2.72 Å. The cyclohexene rings have the same conformation; both are C3-*endo*. This is different than the triclinic crystal, where the chain sticks out from the ring, but the cyclohexene ring is in a different conformation (C2-*endo* in the triclinic form and C3-*endo* in the monoclinic form). However, the chain off the ring is oriented differently and so are the methyl groups. This is the result of the chain dihedrals being different (Figure 4.12; for a table of bonds, angle, and dihedrals, see Appendix B).

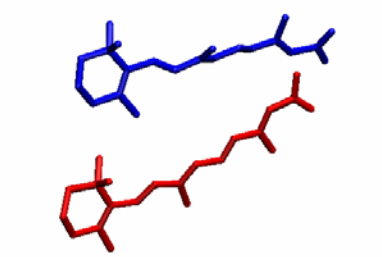


Figure 4.12 Comparison of monoclinic crystal and CHARMM forms (red, x-ray; blue, CHARMM). RMSD between x-ray and CHARMM is 2.72 Å.

Overall, the CHARMM minimized structures of retinal and retinoic acid agree well with the x-ray structures in terms of bonds and angles (largest average bond error is 0.04 ± 0.09 Å, and the largest angle error is 2.4 ± 1.7 º). However, there are some large differences in the dihedral angles, especially in the chain parts of the retinoids. In retinal, both the x-ray structure and CHARMM show the same C2-*endo* ring conformation for the cyclohexene ring, but three dihedrals of the chain cause the different orientations of the methyl groups. The largest dihedral difference in retinal arises because the last methyl group at the end of the chain (C20) is oriented differently in the CHARMM and x-ray structures. In retinoic acid, CHARMM shows differences with both the triclinic and monoclinic x-ray crystal forms. For the triclinic crystal, the cyclohexene rings are in different orientations (CHARMM gives C3-*endo* while the x-ray structure gives C2-*endo*). As for retinal, there are several dihedrals of the chain that cause different orientations of the methyl groups and the end carboxyl group. Unlike retinal, however, the differences in the methyl group orientations are not as pronounced for the CHARMM and triclinic forms of retinoic acid. For the monoclinic crystal, the cyclohexene ring of the x-ray structure and the CHARMM structure are in the same C3-*endo* conformation; however, the chain attached to the ring is oriented differently (the chain from the x-ray structure seems to point up, and the chain from the CHARMM structure points out),

which gives rise to large differences for the dihedrals that connect the ring to the chain. The triclinic and monoclinic crystals show differences with each other and CHARMM. The triclinic crystal and CHARMM have different ring conformations, but the chain off the ring seems to stick out from the ring more, and it doesn't stick up. The monoclinic crystal and CHARMM show the same ring conformations, but the chain off the ring sticks up more in the monoclinic crystal, whereas the CHARMM chain sticks out and not up. The differences between the ring conformations and how the chain is oriented off the cyclohexene ring in the triclinic and monoclinic structures are probably due to different crystal packing arrangements. In all structures, none of the chains off the cyclohexene ring are completely planar, both in the x-ray and CHARMM structures. The reason for the large differences in the dihedrals is that a solid (x-ray) structure with neighboring molecules is being compared to a gas-phase minimized structure (CHARMM).

4.3 Retinoid Crystal Simulations

Crystal simulations were performed for retinal and both crystalline forms of retinoic acid. Five 1 ns simulations were performed for each ligand in the crystal environment, and geometric parameters (bonds, angles, and dihedrals) of the ligands were examined. For simulation details, see Chapter 2.

4.3.1 Retinoic acid (Triclinic Form). Two molecules of retinoic acid were reported for the unit cell of the triclinic form of retinoic acid, CSD id[54] VITAAC10.[51] The bonds, angles, and dihedrals for each retinoic acid molecule were measured from the simulations and compared back to the x-ray structure. The RMSDs over time (compared to starting x-ray structure) for both retinoic acid molecules in the crystal for the five simulations are in Table 4.6. The RMSD is under 0.5 Å for all simulations, so the

simulated structure does not deviate much from the x-ray structure. The bond, angle, and dihedral error between the simulated and x-ray structure was also low (see Table 4.6 for average errors over all five 1 ns simulations). Unlike in the protein/retinoid simulations, which will be discussed later, the retinoic acid in the crystal simulation does not show large dihedral angle deviations about bonds of the chain. This may be because the simulation was done at a slightly lower temperature (123 K) than the protein/retinoic acid complex simulation (277 K), so the system here does not have as much energy to rotate as much around the single bonds of the chain. In the crystal, the chain remains relatively planar throughout the simulations, and the chain does not move much relative to the cyclohexene ring. The cyclohexene rings in both ligands retain the C2'-*endo* conformation found in the x-ray structure. Figure 4.13 shows overlays of the simulation average structures (blue) and the x-ray structure (red).

Table 0.6 RMSD of ligands compared to x-ray structure for triclinic form of retinoic acid

	RTAC1	RTAC2
Simulation 1 (1 ns)	0.33 ± 0.08 Å	0.34 ± 0.09 Å
Simulation 2 (1 ns)	0.35 ± 0.08 Å	0.44 ± 0.09 Å
Simulation 3 (1 ns)	0.38 ± 0.12 Å	0.32 ± 0.09 Å
Simulation 4 (1 ns)	0.32 ± 0.09 Å	0.40 ± 0.13 Å
Simulation 5 (1 ns)	0.33 ± 0.09 Å	0.30 ± 0.08 Å
Average RMSD (5 ns)	0.34 ± 0.09 Å	0.36 ± 0.11 Å
Average bond error	0.03 ± 0.02 Å	0.03 ± 0.03 Å
Average angle error	3.0 ± 3.1°	3.0 ± 3.0°
Average dihedral error	6.6 ± 8.1°	6.3 ± 6.6°

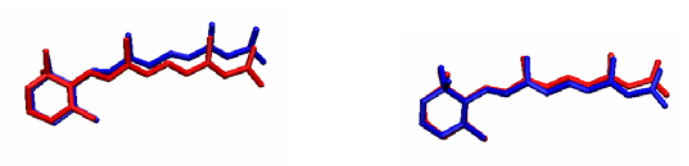


Figure 4.13 Examples of overlayed x-ray (red) and simulation average structures (blue) from Simulation 1.

4.3.2 Retinal (Monoclinic). Four molecules of retinal were reported for the unit cell of the monoclinic form of retinal, CSD id[54] TRETAL01.[52] The bonds, angles, and dihedrals for each retinal molecule were measured from the simulations and compared back to the x-ray structure. The RMSDs over time (compared to starting x-ray structure) for the four retinal molecules in the crystal for the five simulations are in Table 4.7. The RMSD is under 0.5 Å for all simulations, so the simulated structure does not deviate much from the x-ray structure. The bond, angle, and dihedral error between the simulated and x-ray structure was also low (See Table 4.7 for average errors over the five 1 ns simulations). As for the retinoic acid, the retinal in the crystal simulation does not show large dihedral angle deviations about bonds of the chain. In the crystal, the chain remains relatively planar throughout the simulations, and the chain does not move much relative to the cyclohexene ring (the chain is slightly twisted relative to the ring in the x-ray structure, and the retinal molecules retain this twisted configuration during the simulation). The cyclohexene rings in all four molecules in the crystal retain the C2'-*endo* conformation found in the x-ray structure. Figure 4.14 shows the overlays of simulated structures (blue) with the x-ray structure (red).

Table 0.7 RMSD of ligands compared to x-ray structure for retinal

	RTAL1	RTAL2	RTAL3	RTAL4
Simulation 1 (1 ns)	0.30 ± 0.06 Å	0.34 ± 0.06 Å	0.38 ± 0.09 Å	0.35 ± 0.06 Å
Simulation 2 (1 ns)	0.33 ± 0.09 Å	0.37 ± 0.07 Å	0.42 ± 0.10 Å	0.39 ± 0.09 Å
Simulation 3 (1 ns)	0.31 ± 0.08 Å	0.37 ± 0.08 Å	0.39 ± 0.09 Å	0.41 ± 0.09 Å
Simulation 4 (1ns)	0.37 ± 0.11 Å	0.39 ± 0.08 Å	0.41 ± 0.09 Å	0.42 ± 0.10 Å
Simulation 5 (1ns)	0.43± 0.11 Å	0.41 ± 0.08 Å	0.46 ± 0.10 Å	0.46 ± 0.09 Å
Average RMSD (5 ns)	0.35± 0.09 Å	0.38± 0.07 Å	0.41± 0.09 Å	0.41± 0.09 Å
Average bond error	0.02 ± 0.01 Å	0.02 ± 0.01 Å	0.02 ± 0.02 Å	0.02 ± 0.02 Å
Average angle error	3.0 ± 4.5°	3.4 ± 5.0°	4.3 ± 5.9°	4.0 ± 5.3°
Average dihedral error	6.6 ± 9.6°	5.9 ± 5.6°	7.4 ± 6.4°	5.7 ± 5.6°

Figure 4.14 Examples of overlayed x-ray (red) and simulation average structures (blue) (Simulation 2). Only ligands 1 (left) and 3 (right) of the x-ray structure are shown.

4.3.3 Retinoic acid (Monoclinic Form). Four molecules of retinoic acid were reported for the unit cell of the monoclinic form of retinoic acid, CSD id[54] VITAAC01.[50] The bonds, angles, and dihedrals for each retinoic acid molecule were measured from the simulations and compared back to the x-ray structure. The RMSDs over time (compared to starting x-ray structure) for the four retinoic acid molecules in the crystal for the five simulations are in Table 4.8.

Table 0.8 RMSD of ligands compared to x-ray structure for monoclinic form of retinoic acid

	RTAC	RTAC2	RTAC3	RTAC4
Simulation 1 (1 ns)	0.52 ± 0.10 Å	0.51 ± 0.11 Å	0.52 ± 0.10 Å	0.52 ± 0.11 Å
Simulation 2 (1 ns)	0.56 ± 0.10 Å	0.50 ± 0.10 Å	0.55 ± 0.10 Å	0.55 ± 0.09 Å
Simulation 3 (1 ns)	0.46 ± 0.11 Å	0.49 ± 0.11 Å	0.50 ± 0.11 Å	0.45 ± 0.11 Å
Simulation 4 (1 ns)	0.38 ± 0.11 Å	0.40 ± 0.09 Å	0.51 ± 0.09 Å	0.50 ± 0.10 Å
Simulation 5 (1 ns)	0.48± 0.11 Å	0.49 ± 0.09 Å	0.52 ± 0.10 Å	0.51 ± 0.08 Å
Average RMSD (5 ns)	0.48± 0.11 Å	0.48± 0.10 Å	0.52± 0.10 Å	0.51± 0.10 Å
Average bond error	0.02 ± 0.02 Å	0.02 ± 0.02 Å	0.03 ± 0.04 Å	0.03 ± 0.04 Å
Average angle error	3.6 ± 3.3°	3.6 ± 3.3°	3.4 ± 3.3°	3.8 ± 3.3°
Average dihedral error	21 ± 33°	21 ± 32°	19 ± 32°	16 ± 24°

Unlike the triclinic form of retinoic acid and retinal, the RMSD does go above 0.5 Å, but the RMSD does not go above 1 Å. The bond and angle error between the simulated and x-ray structure was also low (See Table 4.8 for average errors over the five 1 ns simulations). Unlike the triclinic form of retinoic acid and retinal, the monoclinic form of retinoic acid in the crystal simulation shows some large dihedral angle deviations about bonds of the chain, and the chain moves relative to the x-ray structure. The chain of

retinoic acid in the x-ray structure points upward from the ring, but during the simulation, the chain starts to point downward, and there is rotation about the single bonds of the chain. However, the cyclohexene rings in all four molecules in the crystal retain the C3'-*endo* conformation found in the x-ray structure. The reason for the higher RMSDs and more movement about the single bonds of the chain could be that this set of simulations was performed at a high temperature (393 K). Retinoic acid crystallizes in two forms (monoclinic and triclinic); however, the monoclinic form is metastable, and at ~ 80° C and above, the monoclinic form converts irreversibly to the triclinic form.[50, 51] This is what appears to be happening in the high temperature simulations. The chain starts to point more outward than upward, which is observed in the triclinic form, and rotation occurs about the single bonds, probably because of the higher temperature. Figure 4.15 shows examples of simulation average structures (blue) overlayed with the starting x-ray structure (red).

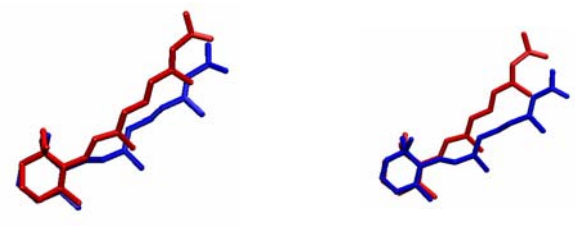


Figure 4.15 Example of overlayed x-ray (red) and simulation average structures (blue) (Simulation 1). The chain on the cyclohexene ring appears to point more outward than upward.

4.4 Retinoid/Protein Complex Simulations

Five simulations of retinoid protein complexes in the crystal environment were performed in order to validate the new force field parameters for the retinoids. In each simulation, amino acid-ligand contacts, ligand-water contacts, amino-acid water contacts,

and any other distances or parameters of interest were examined and compared back to the starting x-ray crystal structure. Each simulation was run at the temperature at which the crystal structure was determined, and distances, angles, and torsions measured from the simulations are for heavy atoms only in order to allow for direct comparison back to the x-ray structure. The structures of the four retinoids, all-*trans* retinoic acid, all-*trans* retinol, all-*trans* retinal, and fenretinide, which were built from the model compounds, are shown in Figure 4.16. All retinoids considered differ only in the functional group on the end of the chain. For simulation details, see Chapter 2.

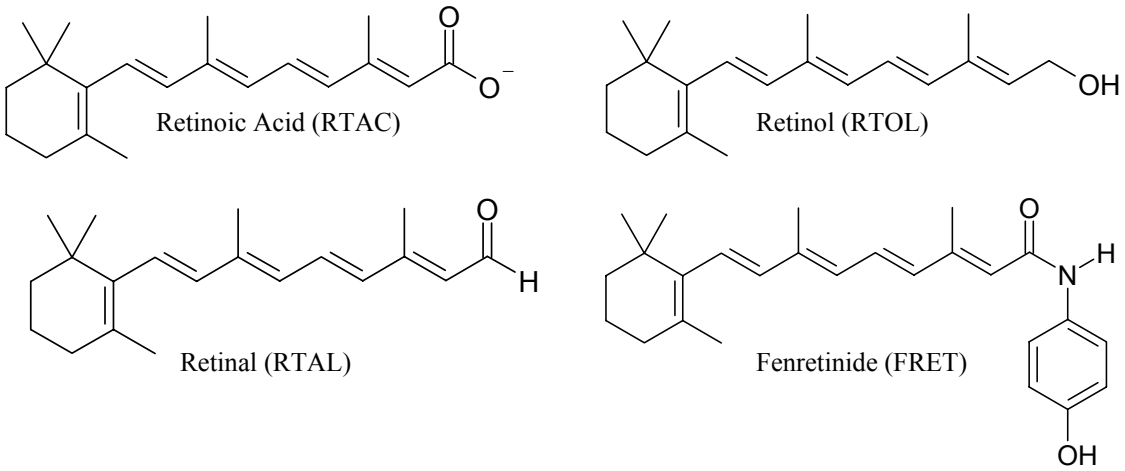


Figure 4.16 The four retinoids built from the model compounds for use in simulations. Retinoic acid, top left; retinol, top right; retinal, bottom left; fenretinide, bottom right.

4.4.1 Retinoic Acid Complexed with Celluar Retinoic Acid Binding Protein II.

Retinoic acid binds non-covalently to cellular retinoic acid binding protein II (CRABP II). The simulation was started from the crystal structure of Kleywegt, et al.,[55] which is shown in Figure 4.17.

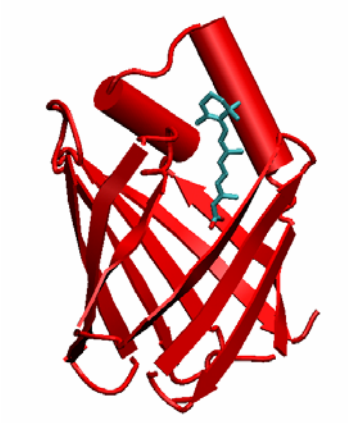
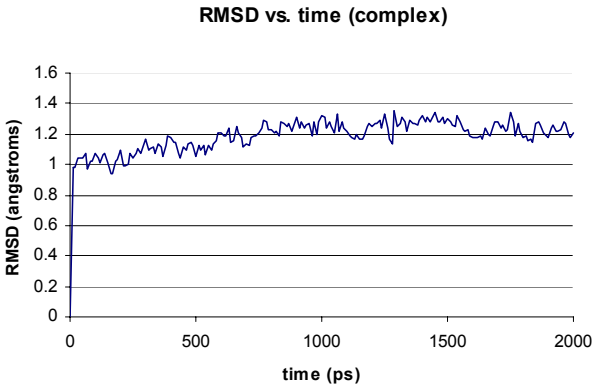


Figure 4.17 Retinoic acid complexed with CRABPII (PDB[56] id 1CBS[55]). Protein show in red; ligand shown in cyan with carboxyl group oxygens in red. Figure rendered using the VMD program.[57]

CRABPs (types I and II) are thought to modulate the amount of retinoic acid that is available to nuclear retinoic acid receptors.[55] From the x-ray structure, it was observed that CRABPII has a one domain with two orthogonal β-sheets that form a barrel, which in turn are made up of 10 β-strands.[55] The middle of the first β-strand contains a bulge at residue 10, which results in a directional change that allows this strand to form part of both β-sheets of the protein. This forms a cavity, which contains the retinoic acid ligand. The carboxyl group of the ligand points into the cavity, while the cyclohexene ring points out and is accessible to solvent. Shown below are the RMSD plots (compared to the starting x-ray crystal structure) for different parts of the complex.



Average RMSDs	
Protein/ligand complex	1.18 ± 0.12 Å
Protein only	1.15 ± 0.12 Å
Ligand only	1.44 ± 0.25 Å
Alpha C only	0.74 ± 0.01 Å

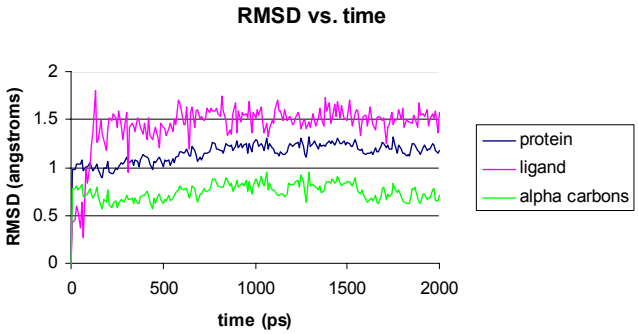


Figure 4.18 RMSD plots of the retinoid acid/CRABPII complex (top left) and different parts of the system individually (bottom).

The RMSD of the whole complex shows initial fluctuation, but appears to stabilize at ~700-800 ps. Upon examination of the individual RMSDs of the protein and ligand, the ligand appears to contribute the most to the RMSD of the complex (average of rmsd of ligand is ~0.29 Å higher than the average RMSD of the protein alone). The ligand RMSD is higher than that of the protein, but it follows a similar pattern of initial fluctuation and then stabilizes around 800 ps.

The geometric parameters (bonds, angles, and dihedrals) of the ligand from the simulation were examined, and the error with respect to the x-ray structure was calculated. The error for bonds of the ligand was 0.02 ± 0.02 Å, the error for angles of

the ligand was 5.5 ± 3.0°, and the error for dihedrals of the ligand was 13 ± 22º. The source of this large error is that the chain of the ligand turns with respect to the ring in the binding site of the protein. Also, there are three dihedrals of the chain that deviate from the orientation found in the x-ray structure that contribute to this large error. The chain rotates with respect to the cyclohexene ring during the simulation, which causes large deviations in the torsion angles that are involved around the connection of the ring to the chain. The other sources of large RMSD differences involve parts of the chain. These differences are seen during the simulation because these are single bonds, and the ligand will probably try to orient itself to optimize contacts with the protein side chains, and this does not appear to cause any major distortion of the surrounding protein residues (see Appendix B (Figure C.5) for figures of retinoic acid and large dihedral differences). The cyclohexene ring retains the C2'-*endo* conformation during the simulation. Because of the large deviations of these ligand dihedrals, different parts of the ligand were examined more closely. Figure 4.19 shows the RMSD for different parts of the retinoic acid (denoted as RTAC) ligand.

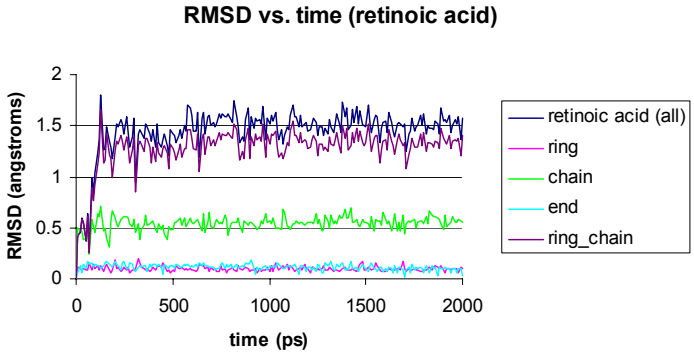


Figure 4.19 RMSD over time of different parts of the retinoic acid ligand.

The ring (including methyl groups) by itself does not give a high RMSD, and the chain (excluding the carboxyl group on the end) gives an RMSD intermediate between the ring by itself and ring and chain together. However, the ring and the chain together (including methyl groups but not including the carboxyl group on the end) do give higher RMSDs. This appears to account for most of the RMSD of the ligand. Upon closer examination of the average structure from the simulation and several snapshots, it appears that the cyclohexene ring rotates around, while the ligand is in the binding site and that the chain rotates around two of the single bonds in the chain. Figure 4.20 shows some snapshots from the simulation.

From the top pictures in Figure 4.20, cyclohexene ring of the ligand appears to rotate in the binding site because the methyl groups are not aligned when the structures are superimposed. The bottom pictures show the starting structure from the simulation (bottom left) and the ending structure (bottom right). The ring starts out with the two methyl groups of the ring pointed to the right; however, during the simulation, the ring turns so that the two methyl groups are pointed in the opposite direction. Also, it can be seen from the overlays of the ligand structures that the chain appears to reorient itself in several places, probably to optimize contacts with protein sidechains. The rotation of the cyclohexene ring in the binding site and the reorientation of some of the chain dihedrals is the cause for the large RMSD of the ligand and for the large average error in the torsion angles between the simulated structure and the x-ray structure.

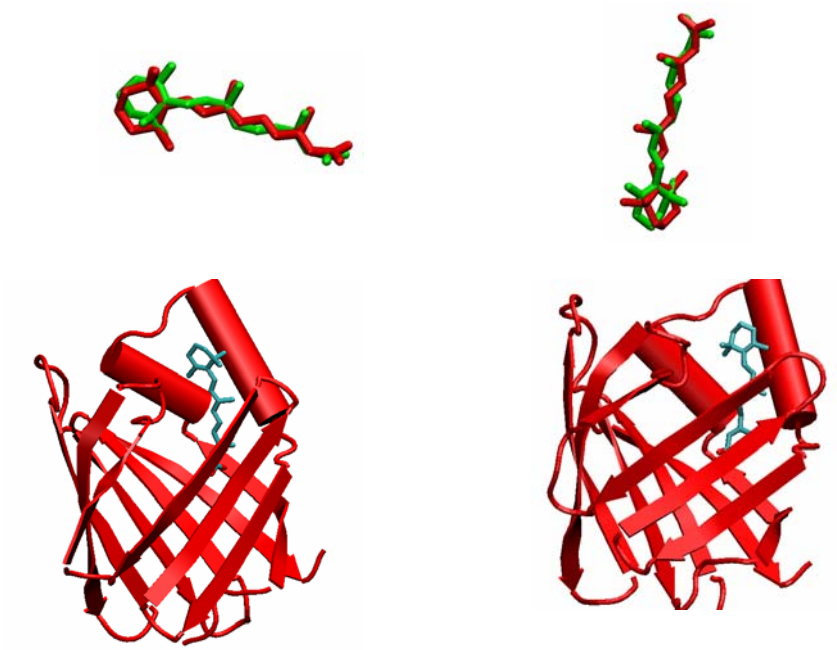


Figure 4.20 Overlay of the average structure of the ligand from the simulation (green) and the x-ray structure (red); top pictures. Starting structure (bottom left) and ending structure (bottom right) from the simulation.

Several ligand-protein and protein-protein contacts were mentioned by the authors of the x-ray structure. These distances were measured in the simulation, and they are shown in the table below and some of the plots are shown (for CHARMM atom names and remaining time series plots of these distances, see Appendix B).

Table 0.9 Protein-ligand and protein-protein distances (CRABPII/retinoic acid complex)

Protein-ligand contact	**x-ray (Å)**	**Computed (Å)**
Arg132(NH)—RTAC(O29)	2.73	2.69 ± 0.12
Tyr134(OH)--RTAC(O29)	2.57	2.91 ± 0.51
Ala32(CB)--RTAC(C8)	3.97	3.78 ± 0.30
Ala36(CB)--RTAC(C12)	4.49	4.04 ± 0.37
Leu28(CD1)--RTAC(C4)	5.14	4.87 ± 0.59
Ala35(CB)--RTAC(C19)	4.45	3.90 ± 0.39
Leu19(CD2)--RTAC(C16)	8.63	4.69 ± 1.32
Ile9(CD)--RTAC(C20)	7.01	7.67 ± 0.63
Phe15(CZ)--RTAC(C10)	3.91	4.22 ± 0.40
Pro39(CB)--RTAC(C20)	3.80	3.90 ± 0.38
Protein-protein		
Arg59(NH)--Gln74(OE1)	2.97	3.00 ± 0.50
Glu73(OE2)--Gln97(NE2)	2.97	3.09 ± 0.29
Glu73(OE2)--Trp109(NE1)	2.95	2.86 ± 0.14
Ser4(O)--Arg136(NE)	2.97	2.83 ± 0.12
Asn2(O)--Arg136(NH2)	2.97	3.18 ± 0.36
Glu73(OE2)--Ser83(OG)	2.81	2.71 ± 0.13
Arg111(NH1)—Trp109(CZ)	3.19	3.41 ± 0.20

The carboxyl oxygens make contact with three residues of the protein, which are Arg132, Tyr134, and Arg111, which are monitored in Figure 4.21. The retinoic acid (RTAC) ligand starts out and stays close to Arg132, and the x-ray value falls within the average and standard deviation computed from the simulation. The RTAC ligand starts out near Tyr134, moves away during the 100-550 ps time period, and then moves close to the Tyr residue again and stays there for the remainder of the simulation. The x-ray value falls into the overall average and standard deviation computed from the simulation.

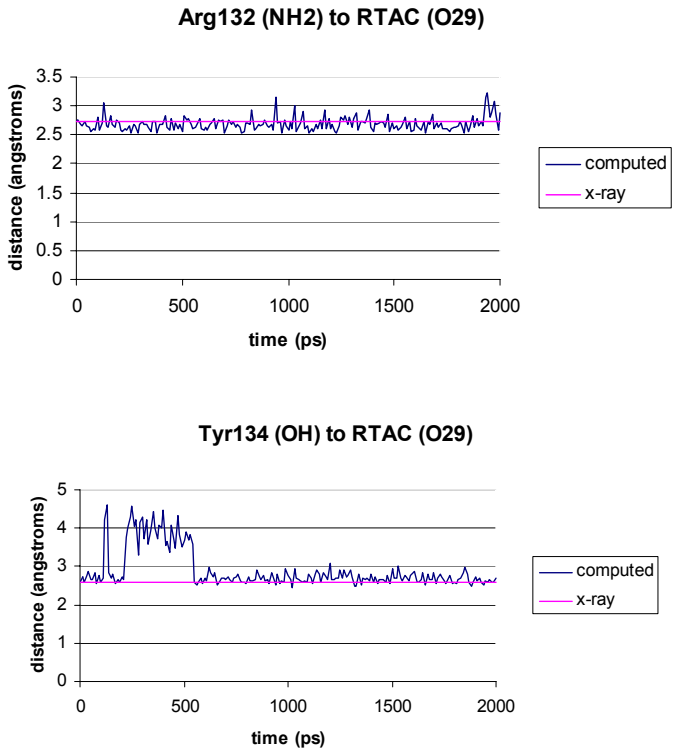


Figure 4.21 Distance vs. time plots of Arg132 and Tyr134 to retinoic acid.

The contact between Arg111 and one of the carboxylic oxygens of RTAC is mediated by a water molecule that was found in the crystal structure (Figure 4.22). From the top plot in Figure 4.22, the water molecules exchange but they still hydrogen-bond to Arg111. The closest water in the x-ray structure was 3.02 Å away (labeled as H_2O15). However, there are several waters that exchange positions, because at different time periods of the simulation, different water molecules are closest to Arg111. In the second plot in Figure 4.22, it can be seen again that H_2O15 and H_2O10 exchange positions. At the beginning of the simulation, H_2O15 is closest to the RTAC ligand, but as the simulation progresses, this water is replaced by H_2O10. The third plot shows the distance over time from Arg111 to the RTAC ligand. The distance is greater than that observed in the x-ray structure (average computed distance is 1.76 Å greater than x-ray distance); this

is probably because more water molecules are present in the simulation than in the x-ray structure, and thus more waters can get between the Arg111 and ligand, pushing them farther apart. The bottom plot in Figure 4.22 shows all the different water molecules that are close to the carboxyl oxygens of the ligand. The Arg111 distance is also shown for comparison, and water molecules appear to be closer to the ligand over most of the simulation time rather than Arg111.

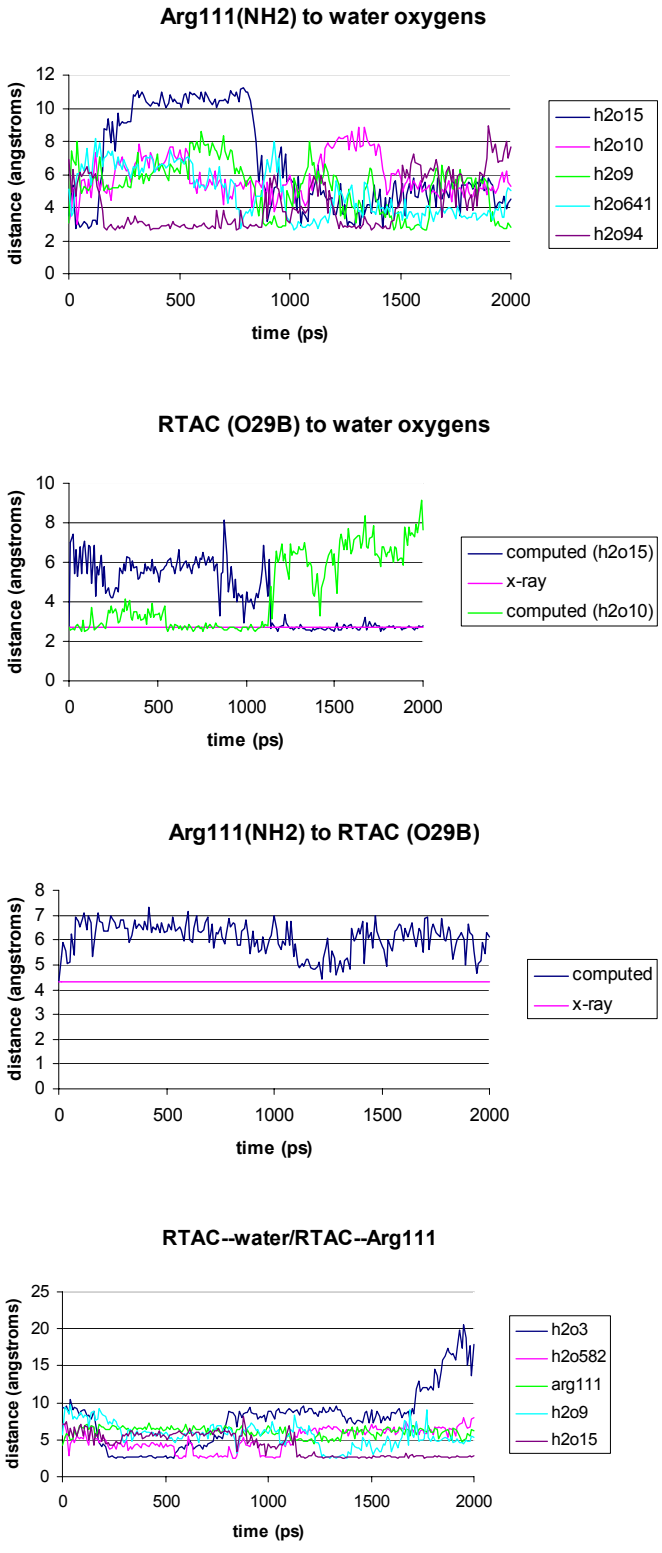


Figure 4.22 Distance from Arg111 to different water oxygens (top); distance from retinoic acid to different water oxygens (second from top), distance from Arg111 to RTAC (second from bottom), and ligand water distances (bottom).

Several other protein-ligand contacts from the simulation were also tracked. Most of these interactions were nonpolar residues with nonpolar parts of the ligand (most of the ligand is nonpolar, except the tail with the carboxyl group). Two of the contacts are shown in Figure 4.23.

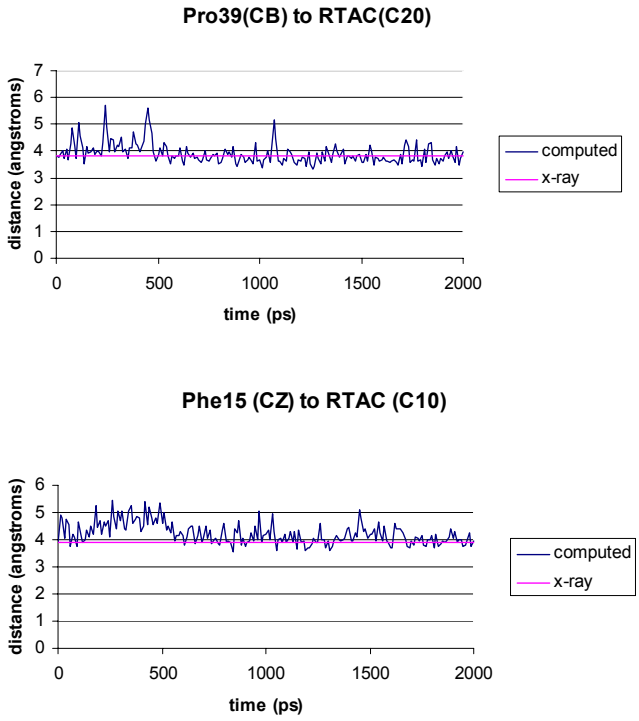


Figure 4.23 Two of the nonpolar ligand-protein contacts (retinoic acid/CRABPII).

Pro39 and Phe15 are two of the nonpolar residues that were in contact with the ligand. For both residues, there is some initial fluctuation at the beginning of the simulation, then the movement of them stabilizes and fluctuates around the x-ray value for the remainder of the simulation. The other nonpolar ligand-protein contacts showed similar behavior.

Several protein-protein contacts were examined to make sure that the new parameters for the ligand were not distorting the protein. Two of these plots are shown in Figure 4.24.

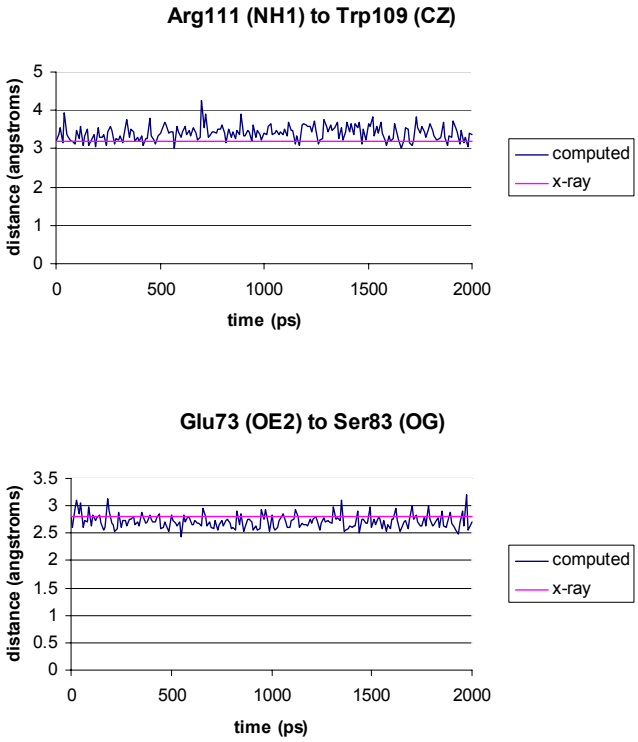


Figure 4.24 Two of the protein-protein distances (retinoic acid/CRABPII).

The Arg111-Trp109 contact is mediated through an "amido/aromatic ring hydrogen bond".[55] Burley and Petsko explain that side-chain amino groups interact with aromatic side chains, and that the amino groups of amino acids such as Lys, Arg, etc. are preferentially located within 6 Å of the ring centers of Phe, Tyr, and Trp.[58] The distance from the amino group of Arg111 to one of the carbons of the aromatic ring of Trp109 was measured (top plot in Figure 4.24). The computed average is 0.02 Å higher than the distance measured from the x-ray structure, which is in good agreement. The other protein-protein contact shown is for Glu73 and Ser83 (bottom plot in Figure 4.24), and this distance fluctuates around the x-ray value for the entire simulation.

Overall, the simulated retinoic acid/CRABPII complex agrees with the x-ray structure. The ligand gave a large RMSD over time and a large torsion angle error, which

can be attributed to the cyclohexene ring rotating in the binding site and the two chain dihedrals rotating. The ring is probably able to rotate freely because not many protein-ligand contacts were identified for the ring portion of the ligand; most are for the chain and carboxyl end group. Not many protein residues are contacting that ring strongly and forcing it to stay in any one position. One carbon of the ring was identified as a close contact with the protein (Leu28), but even that residue is 5.14 Å away from the ligand. The authors of the x-ray structure reported higher B-factors for the cyclohexene ring part of the ligand than for the chain part,[55] which means that there is more uncertainty for the positions of the ring atoms. It is possible the ring is free to rotate; however, a single x-ray structure will capture only one orientation, and through the simulations, the rotation is observed. The chain dihedrals probably rotate in order to optimize contacts with protein sidechains. The turning motion of the cyclohexene ring in the binding site could be important in ligand binding and dynamics or interaction of the complex with other proteins.

4.4.2 Retinol Complexed with Cellular Retinol Binding Protein. Retinol binds non-covalently to cellular retinol binding proteins (CRBPs). The simulation was started from the crystal structure determined by Calderone, et al.[59] The retinol ligand complexed with zebrafish CRBP is shown in Figure 4.25.

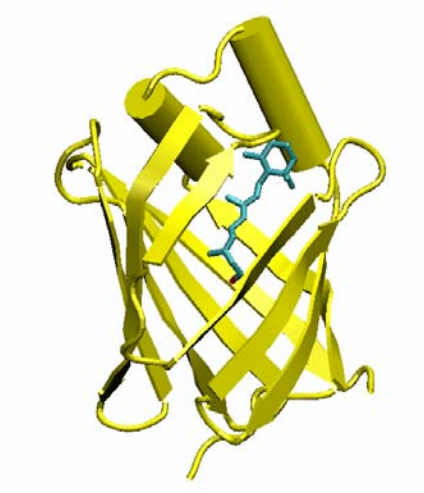
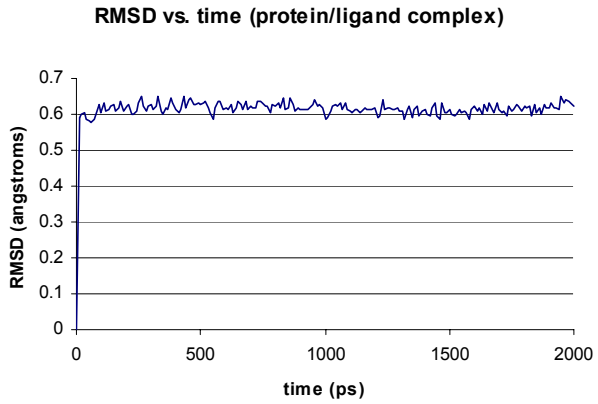


Figure 4.25 Retinol complexed with zebrafish CRBP (PDB[56] id 1KQW[59]). The ligand is in cyan with the OH oxygen in red, and the protein in yellow. Figure was rendered with VMD.[57]

CRBPs are found in a variety of tissues, are retinol-specific, and are suspected to aid in retinol storage, uptake, and metabolism.[60, 61] Four types of mammalian CRBPs have been identified and characterized, and zebrafish CRBP appears to have a similar amino acid sequence and structure to the mammalian types (sequence identity is highest with mammalian CRBPII—73%).[59] The CRBPs, intracellular lipid binding proteins (which include the CRABPs), and fatty acid binding proteins have been show to have low sequence identity, but they have a conserved structural fold,[62] so the CRBPs look similar to the CRABPs. They have the same β-barrel structure formed from 10 antiparallel β-strands, which forms the cavity where the ligand binds, and the open end of the β-barrel is capped by an α-helical region. In the x-ray structure of zebrafish CRBP complexed with retinol (as with rat CRBPI and CRBPII), the retinol is almost completely enclosed within the cavity of the protein. The hydroxyl group on the end is toward the inside of the cavity, while the cyclohexene ring is the near the α-helical region near the top of the

cavity.[63, 64] Shown below in Figure 4.26 are the RMSD plots for different parts of the CRBP/retinol complex.

The RMSD of the protein/ligand complex stabilizes quickly to 0.6 Å at the beginning of the simulation, and fluctuates around 0.6 Å for the remainder of the simulation. The alpha carbons follow the same trend as the protein alone but at a lower RMSD (alpha carbons alone fluctuate at around 0.4 Å. As for the retinoic acid/CRABP complex, the retinol of the retinol/CRBP complex contributes the most to the RMSD of the complex (protein alone is 0.6 Å, and the ligand alone is 0.74 Å). The ligand and protein RMSDs follow different patterns; the protein RMSD stabilizes quickly at the beginning of the simulation, but the ligand RMSD is initially at ~0.8 Å and then falls to ~0.7 Å at 1 ns and stays around 0.7 Å for the rest of the simulation.



Average RMSDs	
Protein/ligand complex	0.61 ± 0.05 Å
Ligand only	0.74 ± 0.09 Å
Protein only	0.60 ± 0.05 Å
Alpha carbons	0.44 ± 0.04 Å

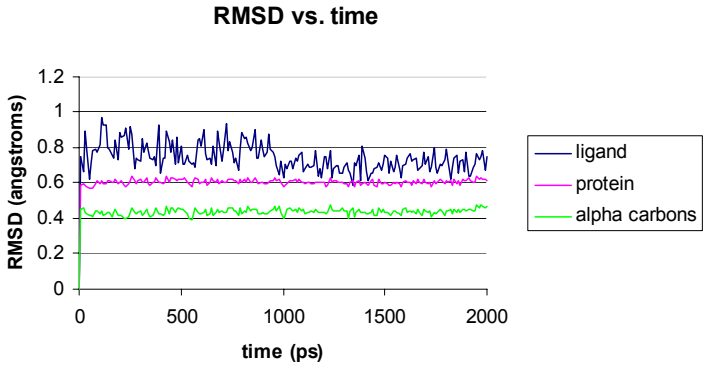


Figure 4.26 RMSD plots of the retinol/CRBP complex (top) and different parts of the system individually (bottom).

The geometric parameters (bonds, angles, and dihedrals) of the ligand from the simulation were examined, and the error with respect to the x-ray structure was calculated. The error for bonds was 0.02 ± 0.01 Å, the error for angles was 2.9 ± 3.0°, and the error for dihedrals was 21 ± 25º. The origin of this large dihedral error is rotation about the C12-C13 bond of the chain of retinol.

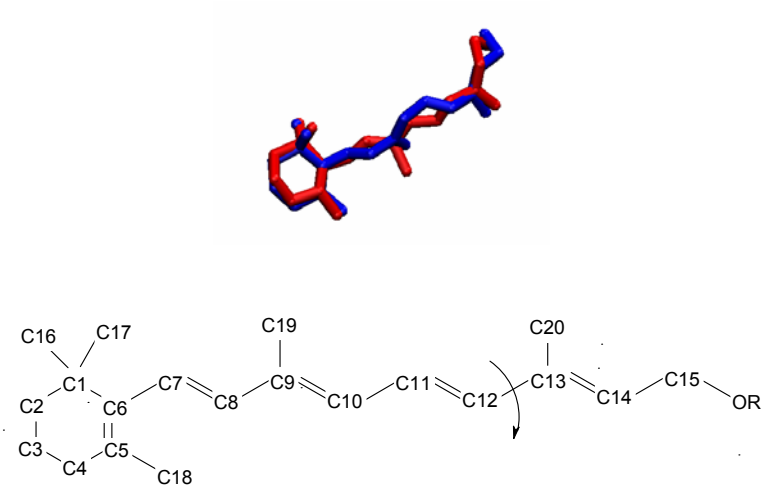


Figure 4.27 Overlay of x-ray (red) and simulation average structure (blue) (top) and CHARMM atom names in retinol (bottom). The arrow indicates the dihedrals where the large deviations from the starting x-ray crystal structure occur.

The two dihedrals that show the largest deviations from the x-ray structure are rotations about the C12-C13 bonds. The deviation can be seen in the overlayed structures; the C11-C12 bond points in opposite directions in the x-ray structure and in the simulation average structure. As with the retinoic acid ligand in the CRAPBII simulation, the ligand probably tries to optimize contacts with protein side chains; thus, there is rotation about single bonds. The retinol ligand has only two dihedrals showing a major deviation from the x-ray structure along the chain, while the retinoic acid has several dihedrals with major deviations from the x-ray structure. However, unlike the retinoic acid/CRAPBII complex, the retinol does not show any rotation of the cyclohexene ring with respect to the chain of the ligand. This large deviation from the x-ray dihedral does not appear to cause any major distortion of the surrounding protein residues. C11 and C13 move closer to surrounding protein residues, while C12, C14, and C20 remain relatively unaffected (see Appendix B for time series plots from these atoms to surrounding protein residues). The cyclohexene ring retains the C2'-*endo* conformation during the simulation.

The reference for the x-ray structure listed the protein residues with at least one atom within 4.5 Å of the ligand (some contacts were up to 6 Å away), which was used for the protein-ligand contacts to measure from the simulation. Table 4.10 lists the calculated and x-ray distances. Some plots are shown; the rest can be found in Appendix B.

Table 0.10 Protein-ligand distances (CRBP/retinol complex)

Protein-ligand contact	**x-ray (Å)**	**Computed (Å)**
Ile25(CD)--RTOL(C18)	4.00	3.83 ± 0.17
Thr29(CG2)--RTOL(C4)	5.10	4.72 ± 0.20
Gln38(CD)--RTOL(C17)	3.88	4.32 ± 0.33
Ile42(CD)--RTOL(C17)	6.31	6.54 ± 0.23
Thr53(OG1)--RTOL(C12)	4.16	4.30 ± 0.14
Phe57(CD2)--RTOL(C3)	3.79	4.17 ± 0.20
Val62(CG2)--RTOL(C14)	7.36	7.78 ± 0.31
Trp106(CZ2)--RTOL(C14)	4.59	4.28 ± 0.27
Leu117(CD2)--RTOL(C20)	3.86	4.05 ± 0.09
Ile119(CD)--RTOL(C19)	3.95	4.35 ± 0.26
Arg58(CB)--RTOL(C16)	4.24	3.97 ± 0.16
Thr53(CB)--RTOL(C11)	5.10	4.04 ± 0.16
Trp106(CZ2)--RTOL(C13)	4.96	4.56 ± 0.27
Met20(SD)--RTOL(C18)	3.90	3.81 ± 0.15
Ser55(OG)--RTOL(C17)	4.04	3.90 ± 0.17
Ala33(CB)--RTOL(C4)	4.01	3.92 ± 0.14
Leu36(CD1)--RTOL(CZ)	4.25	4.00 ± 0.26
Tyr60(CD2)--RTOL(C11)	5.00	4.46 ± 0.29
Thr51(OG1)—RTOL(OR)	5.15	5.86 ± 0.22
Gln108(NE2)—RTOL(OR)	4.51	4.13 ± 0.24
Gln108(NE2)—RTOL(OR)	3.00	2.67 ± 0.07
Val19(CG1)—RTOL(C19)	6.41	7.30 ± 0.41

There were only two polar contacts mentioned, which are shown below (these would be hydrogen bonds, but since no hydrogens were identified in the x-ray structure, they are measured as heavy atom distances). The retinol ligand is designated as RTOL.

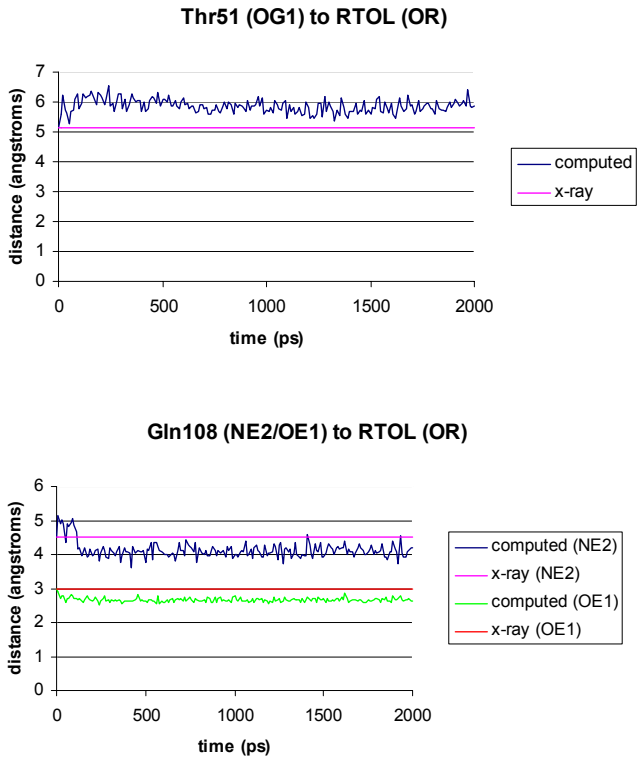


Figure 4.28 Polar protein/ligand contacts (retinol/CRBP).

The Thr51—RTOL distance from the simulation is larger than that measured from the x-ray structure (~0.71 Å larger). However, there is a residue in closer contact with the hydroxyl group than Thr51, which is Gln108. The nitrogen of the sidechain of Gln108 is not only closer to the ligand in the x-ray structure (4.51 Å vs. 5.15 Å), but the ligand stays near Gln108 during the simulation. The distance between the ligand and oxygen of the sidechain of Gln108 fluctuates around ~2.7 Å during the whole simulation, which is about 0.3 Å closer than that reported in the x-ray structure. The authors of the crystal structure refined two positions for the hydroxyl group of retinol so that in one position a hydrogen bond forms between the hydroxyl group and the oxygen of the sidechain of Gln108, and in the other position, the hydrogen bond from the ligand is formed to Thr51.[59] For the simulation, we chose the hydroxyl position in which the hydrogen bond

is formed with Gln108; this mode of retinol binding was also observed in structures of rat CRBP I and rat CRBP II complexed with retinol.[63, 64] Therefore, even though the retinol ligand is not forming a hydrogen bond with Thr51, it is forming (keeping during the simulation) a hydrogen bond with sidechain oxygen of Gln108, and the amino group of Gln108 presents another possible hydrogen bond donor/acceptor for the hydroxyl group.

All of the other ligand-protein contacts listed by the paper were contacts of protein residues with the chain or ring of the ligand (not the hydroxyl group). Four of them are shown below.

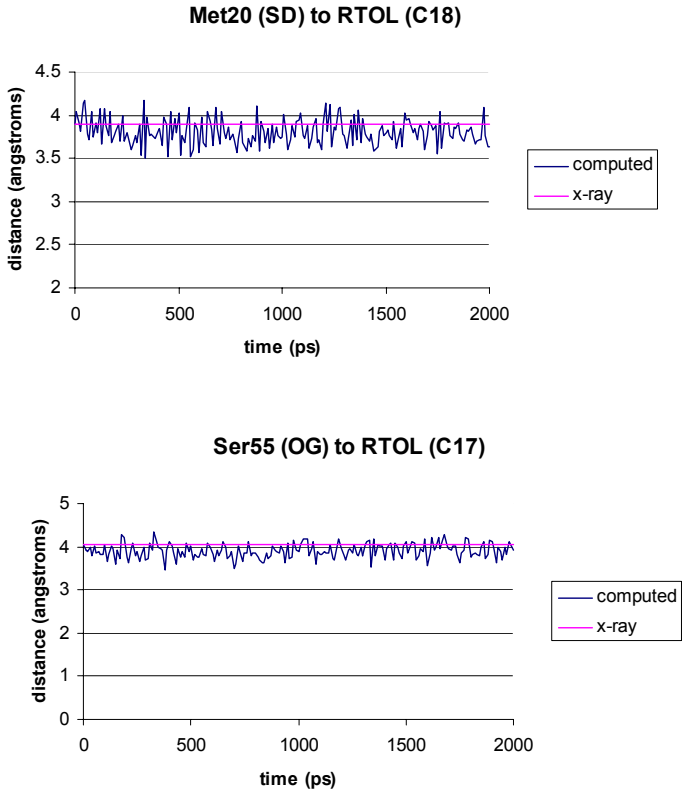


Figure 4.29 Two of the contacts from the retinol ligand chain to the protein (retinol/CRBP).

The amino acid residue contacts with the ligand chain measured from the simulation are in good agreement with the x-ray structure distances. For both distances shown, the

ligand chain and amino acid residues do not appear to move away from each other during the simulation. Met20 and the retinol ligand move ~0.5 Å closer to each other; Ser55 and retinol stay about the same distance apart during the whole simulation.

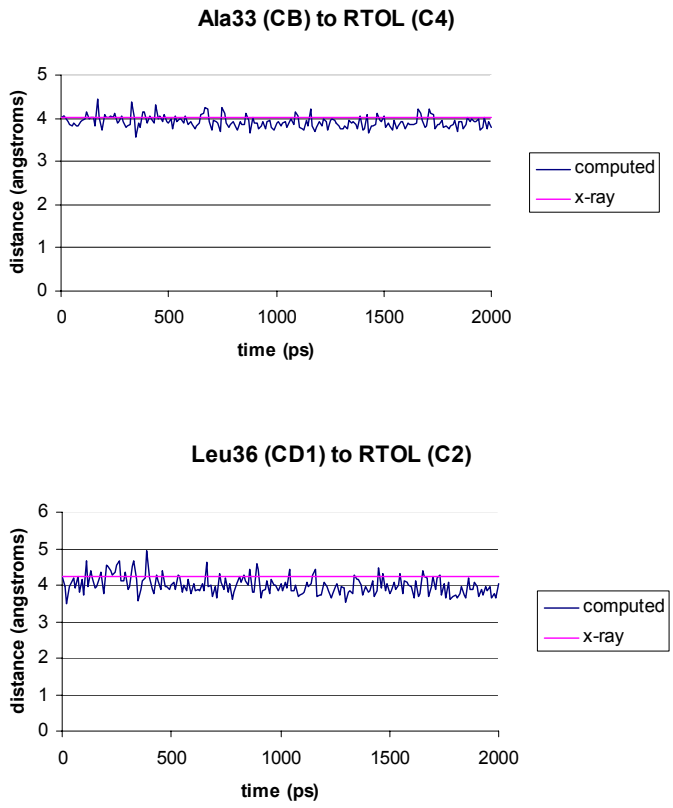


Figure 4.30 Two of the contacts of the retinol ligand ring to the protein (retinol/CRBP).

As with the amino acid residue contacts with the ligand chain, the amino acid contacts with the ligand ring measured from the simulation are in good agreement with the x-ray structure. The ring of the retinol ligand does appear to move away from either Ala33 or Leu36 during the simulation, and the distances fluctuate around the x-ray value. The other distances measured from the simulation are shown in Appendix B. Three of the average distances measured from the simulation became shorter over time than the distance measured from the x-ray structure; the Thr29 distance to the retinol ligand

decreased the most, becoming 0.38 Å shorter than that of the x-ray structure. Several distances measured from the simulation increased over time (ligand and amino acid residues drifted apart), and the most increased distance was for Val19 to the ligand (plot is in Appendix B). This distance increased on average 0.89 Å (initially increased to ~8.0 Å and then decreased to ~7 Å); however, initially, this distance was 6.41 Å, which was not extremely close to the ligand in the x-ray structure. The distance increase of 0.89 Å does not seem unusual, considering the starting distance and that water is present in the binding cavity, which could have pushed the residues farther apart.

The authors of the x-ray structure identified seven water molecules in the ligand binding cavity of zebrafish CRBP. The structure on the left of Figure 4.31 shows the seven water molecules in the binding cavity in the x-ray structure. The structure on the top right shows the average structure of the complex from the simulation. Six of the original crystallographic waters are still in the cavity; however, one of them appears to have moved out (that is the water molecule farthest to the right in the average structure of the simulation). Upon closer examination of the average structure, this water molecule appears to have moved outside of the binding cavity because it forms a strong hydrogen bond with Asn13.

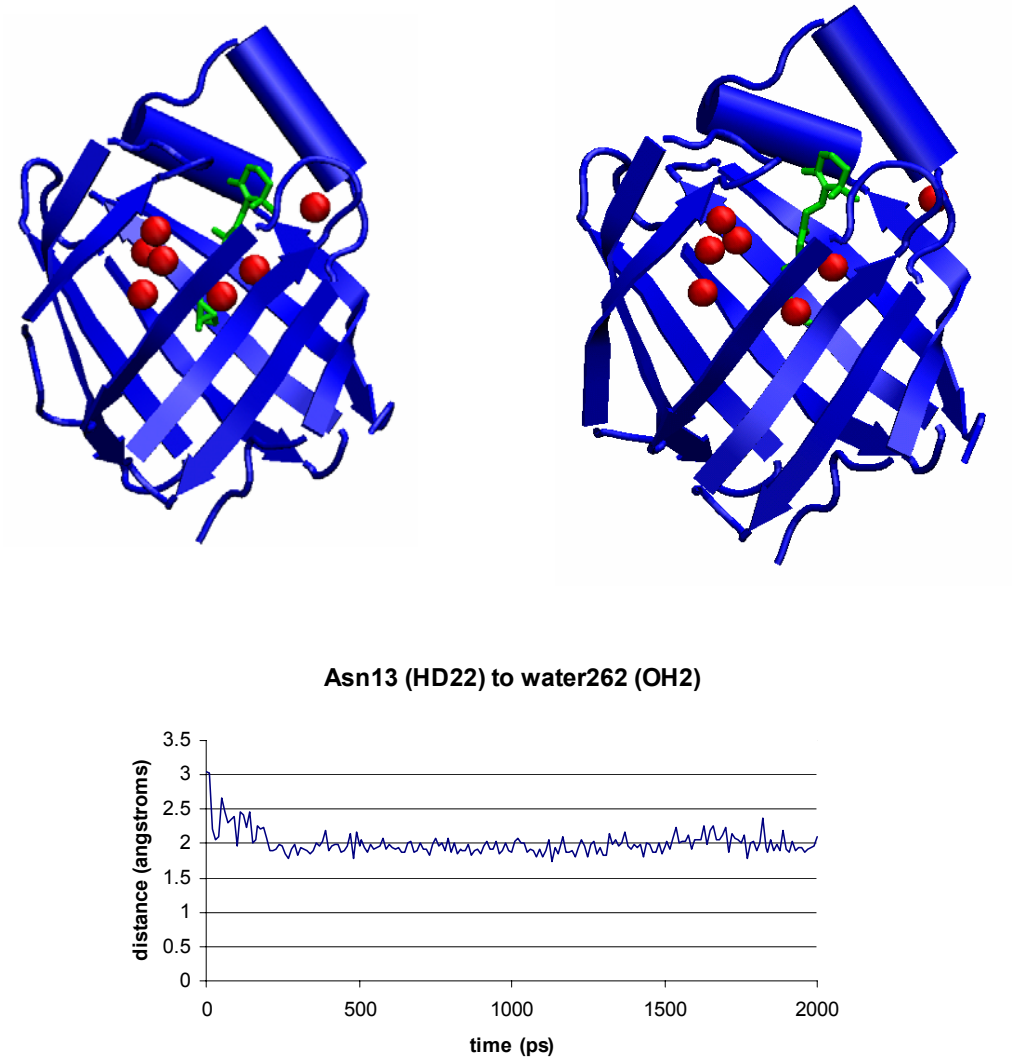


Figure 4.31 Water in binding cavity of zebrafish CRBP. The protein is shown in blue, the ligand is shown in green, and the water oxygens are shown in red. The starting structure is shown on the top left, and the average structure from the simulation is on the top right. The plot on the bottom is the distance from the water molecule to Asn13.

When the water molecule moves toward the edge of the cavity, it forms a strong hydrogen bond with Asn13 (2.00 ± 0.18 Å), and thus stays in that position for the remainder of the simulation.

Overall, the distances measured from the simulation agree well with the x-ray structure. Two of the distances from protein residues to the ligand decrease over time (by

0.38 Å at the most), and several of the distances increase. The largest increase was 0.89 Å, and the distance between the amino acid and ligand was over 6 Å in the starting structure, so movement of the protein and water in the binding site could have pushed the ligand and the residue apart. Six of the seven crystallographic waters stayed in the binding site during the simulation, and the one that did move to the edge was stabilized by a hydrogen bonding interaction with Asn13. Unlike the retinoic acid ligand complexed with CRABPII, the cyclohexene ring of the retinol ligand did not appear to rotate in the binding site of CRBP.

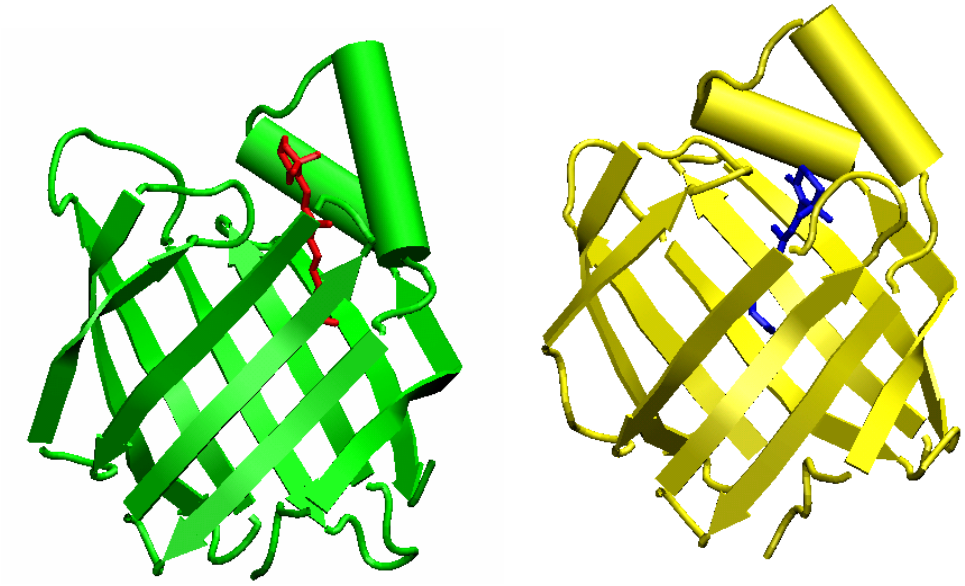


Figure 4.32 Starting structures for retinoic acid/CRABPII complex (left) and for retinol/CRBP complex (right).

Upon closer examination of the starting structures for both complexes, the retinol appears to be positioned deeper into the binding cavity of CRBP than retinoic acid is positioned into CRABPII (Figure 4.31). As a result, retinoic acid probably is not constrained as much by the protein residues around it (especially the ring at the top), so the ring has more space to rotate and move around. Because retinol is positioned farther down into

the binding site, it probably does not have as much room to move around due to the protein residues around it.

4.4.3 Fenretinide Complexed with Retinol Binding Protein (RBP). The starting x-ray crystal structure for the simulation was determined by Zanotti and coworkers.[65] The structure of the fenretinide/RBP complex is shown in Figure 4.33.

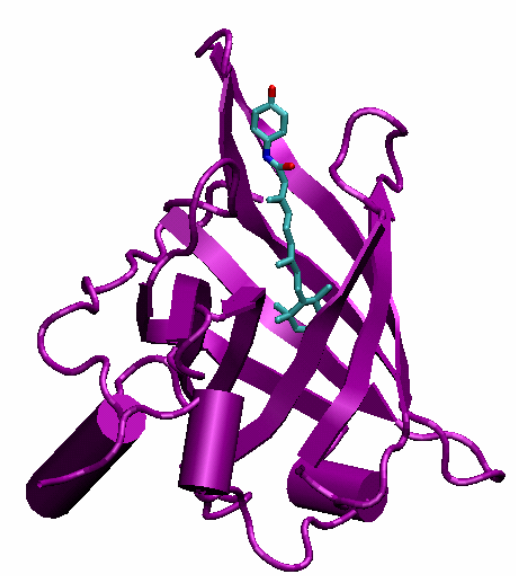


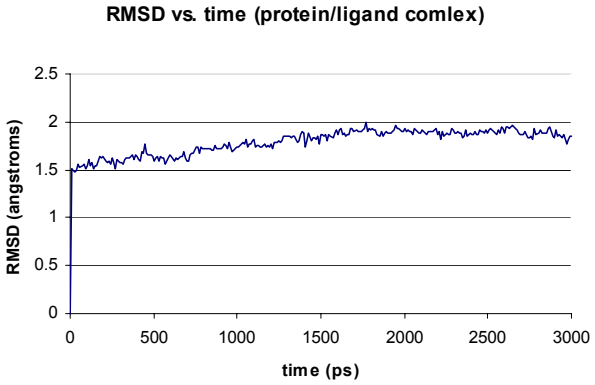
Figure 4.33 Fenretinide/RBP complex (PDB[56] id 1FEL[65]). The ligand is shown in cyan with the nitrogen atom in blue and oxygens in red. The protein is in purple. Figure rendered with VMD.[57]

Retinol binding protein carries retinol through the blood and delivers retinol from storage sites to target cells. RBP contains one binding site for retinol, and in the blood, RBP is found bound to transthyretin (TTR); the formation of the complex is believed to prevent filtration of the small RBP through the kidneys.[65] Like CRBPs and CRABPs, RBP binds the ligand noncovalently and contains a β-barrel which forms the binding cavity for the ligand. However, in the case of RBP, the cyclohexene ring binds inside the β-barrel, and the chain is extended out with the hydroxyl group reaching the surface (the ligand binds the opposite way in CRBPs and CRABPs).[65] Studies have shown that modifications to

ligands in the vicinity of the cyclohexene ring render them unable to bind to RBP, however, retinoids modified in the area of the hydroxyl group will bind well to RBP.[66-68]

With the increasing interest in synthetic retinoids that show pharmacological and antitumor activity, some questions about how they interact with retinoid-binding proteins, other proteins, and nuclear retinoid receptors have been raised. Synthetic retinoids may interact with retinoid-binding proteins involved in retinoid metabolism, and it has been shown that some synthetic retinoids may affect plasma transport of retinol, which could be a result of their interaction with RBP.[67] Fenretinide is a retinol analog in which the hydroxyl end group is replaced by a hydroxyphenyl amide group. This particular synthetic retinoid has been examined for use as an antitumor chemopreventative agent in humans.[65] Studies have shown the fenretinide binds well to apo RBP, but the fenretinide/RBP complex does not show any affinity for TTR.[69]

In the x-ray structure of the fenretinide/RBP complex, the fenretinide ligand fits into the β-barrel cavity where the retinol ligand would fit. The cyclohexene ring and chain take the place of the corresponding parts of retinol (cyclohexene ring is inside the cavity with the chain protruding up), while the hydroxyphenyl amide group comes out toward the solvent, replacing the position of the hydroxyl group of retinol. Shown below are the RMSD plots for the whole complex and different parts of the complex.



Average RMSDs	
Protein/ligand complex	1.78 ± 0.16 Å
Ligand only	1.11 ± 0.31 Å
Protein only	1.79 ± 0.16 Å
Alpha carbons	1.34 ± 0.13 Å

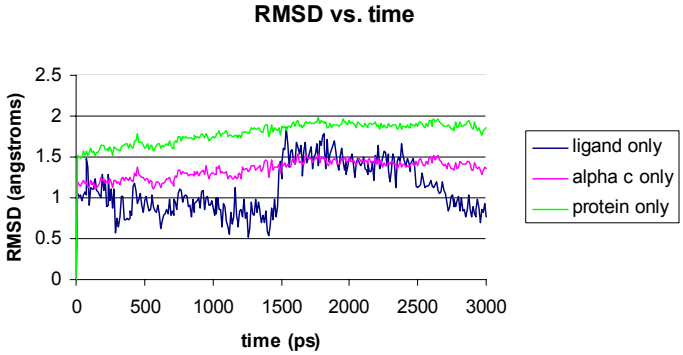


Figure 4.34 RMSD plots of the fenretinide/RBP complex (top) and different parts of the system individually (bottom).

The RMSD of the protein-ligand complex continues to increase until ~1.5 ns, after which the RMSD levels off and stays just under 2 Å. The RMSD for the alpha carbons of the protein follows the RMSD pattern of the whole protein, only the magnitude is ~0.5 Å lower for the alpha carbons. The ligand RMSD decreases from the beginning of the simulation to ~1.5 ns into the simulation, then it jumps up to over 1.5 Å and starts to decrease again. To examine what caused the large increase, the RMSDs for different parts of the ligand were plotted, which is shown at the top of Figure 4.35.

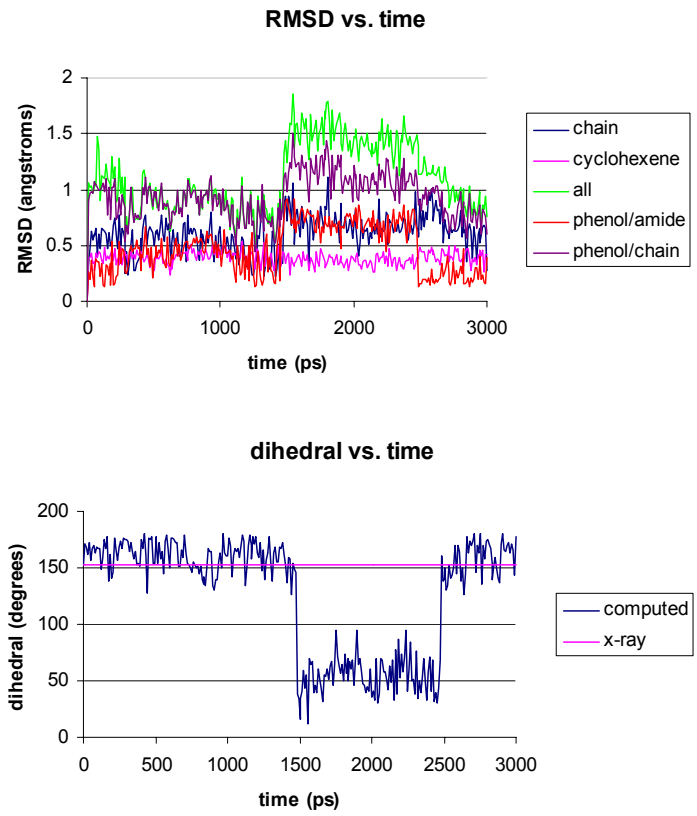


Figure 4.35 RMSDs for different parts of the ligand (top) and time series plot for rotation of the phenol ring of fenretinide (bottom).

The chain is the isoprene chain including the methyl groups on it, the cyclohexene ring includes the three methyl groups attached to it, the phenol/amide part is the phenol ring and the amide group, and the phenol/chain is the isoprene chain including methyl groups, the amide group, and the phenol ring. The cyclohexene ring contributes the least to the RMSD, while the phenol/amide and phenol/chain parts seem to contribute the most, so the phenol ring and amide are the parts that are moving around the most in the binding site. The time series of the dihedral for rotation of the phenol ring is shown in Figure 4.35 (bottom).

The time series shows a large increase in the dihedral value at ~1.5 ns, which is where the large increase is seen the RMSD plot for the ligand. From the plot, it appears that the phenol ring is turning in binding site during the simulation, and this will be supported further by distances measured from the ligand to nearby residues and from snapshots taken from the simulation.

The geometric parameters (bonds, angles, and dihedrals) of the ligand from the simulation were examined, and the error with respect to the x-ray structure was calculated. The error for bonds was 0.03 ± 0.03 Å, the error for angles was 4.9 ± 4.9°, and the error for dihedrals was 15 ± 12º. The origin of this large dihedral error is the rotation of the phenol ring, which was mentioned above. This dihedral shows an approximate 30° difference on average from the x-ray structure.

The authors of the x-ray structure mentioned few details about the ligand conformation and position; they explain that the resolution is not high enough and allows only a limited discussion of the bound ligand conformation.[65] In the x-ray structure, the cyclohexene ring is in a half-chair conformation (C3-endo), and from the picture below, the average structure from the simulation also shows the C3-endo conformation.

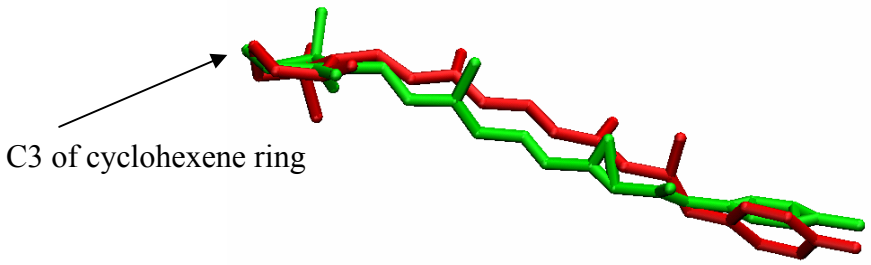


Figure 4.36 Overlay of x-ray structure of ligand and simulation average structure of ligand. The x-ray structure is shown in red, and the simulation average structure is shown in green.

From the overlay of the two structures, the rotation/shift of the phenol ring can also be seen. The authors also mention that two water molecules were found near the hydroxyl group of the phenol ring (each water was ~2.8 Å away). The fenretinide ligand is abbreviated as FRET.

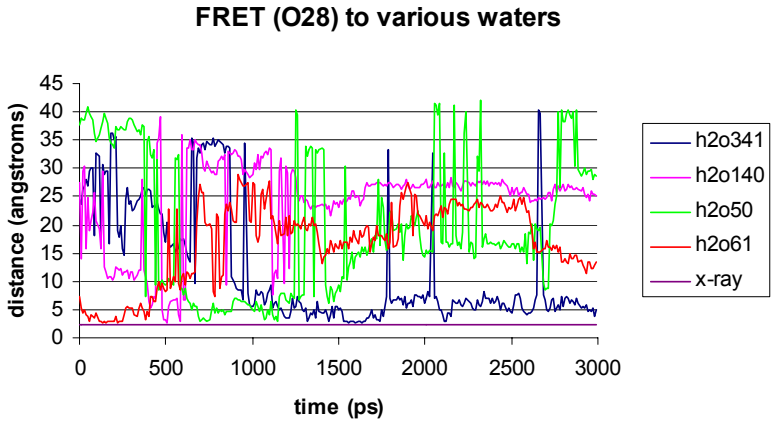


Figure 4.37 Distance from hydroxy group of the phenol ring of the ligand to various water molecules.

Over time, the water molecules near the hydroxy group of the phenol ring appear to be exchanging, but are still close to the phenol ring.

Due to limited resolution, the paper did not make mention of any ligand-protein contacts. Amino acid-ligand distances that were within ~6 Å in the x-ray structure were measured from the simulation. The table below shows the x-ray and simulation averages of these distances (for CHARMM atom names and remaining distance plots, see Appendix B).

Table 0.11 Protein-ligand distances (RBP/fenretinide complex)

Protein-ligand contact	**x-ray (Å)**	**Computed (Å)**
Ile41(CG2)--FRET(C17)	5.37	5.57 ± 0.58
Ala55(CB)--FRET(C4)	3.78	3.74 ± 0.21
Val61(CG1)--FRET(O29)	5.11	4.75 ± 0.48
Phe137(CZ)--FRET(C17)	4.61	4.07 ± 0.33
Gln117(CD)--FRET(C16)	4.64	4.28 ± 0.28
Ala43(CB)--FRET(C2)	4.55	3.81 ± 0.24
His104(CE1)--FRET(C16)	3.65	4.06 ± 0.56
Phe36(O)--FRET(N21)	3.00	3.58 ± 1.15
Tyr90(CE1)--FRET(C19)	5.46	4.57 ± 0.70
Trp91(O)--FRET(C18)	5.69	6.40 ± 0.64
Gly75(N)--FRET(C18)	4.27	3.90 ± 0.32
Ala57(CB)--FRET(C5)	3.69	3.93 ± 0.30
Ala71(CB)--FRET(C12)	5.64	6.13 ± 1.59
Leu97(CD1)--FRET(C14)	6.05	6.39 ± 0.54
Leu97(CD2)--FRET(C14)	4.23	4.88 ± 0.54
Gln98(N)--FRET(O29)	3.40	2.99 ± 0.25
Leu97(CA)--FRET(O29)	4.56	3.77 ± 0.40
Leu64(CG)--FRET(O28)	5.91	4.26 ± 0.77
Leu64(CB)--FRET(O28)	4.40	3.53 ± 0.51
Leu37(CD2)--FRET(C12)	4.27	4.05 ± 0.38
Leu64(CD1)—FRET(O28)	6.66	4.59 ± 1.01
Leu64(CD2)—FRET(O28)	6.70	4.88 ± 0.62

The distances listed in Table 4.11 show good agreement between the x-ray structure and the simulation; the x-ray value is within the averages and standard deviations from the simulation. However, some distances increase (ligand and protein residue drift apart) and some distances decrease (ligand and protein residue drift closer together); the largest deviation from the x-ray structure is 2.07 Å. This large change in distance occurs for Leu64 (sidechain carbons) to fenretinide.

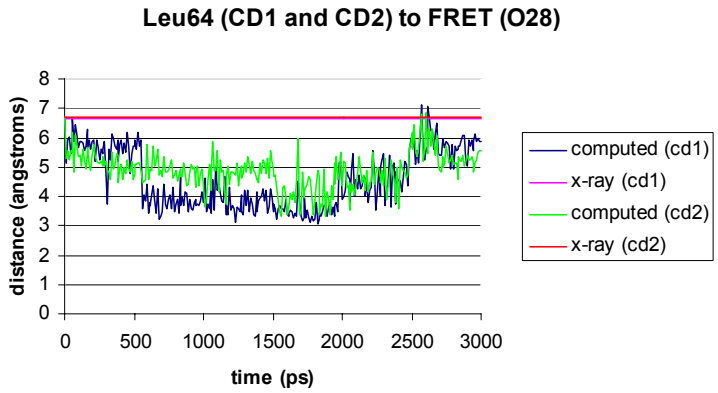
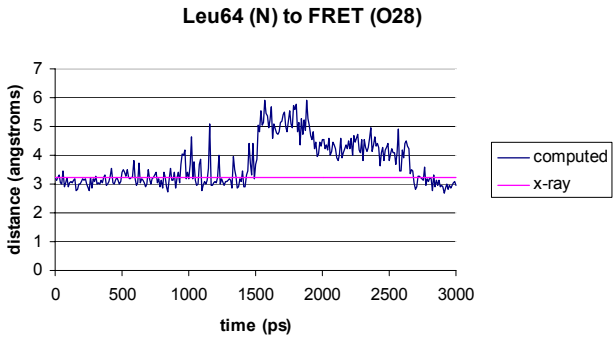
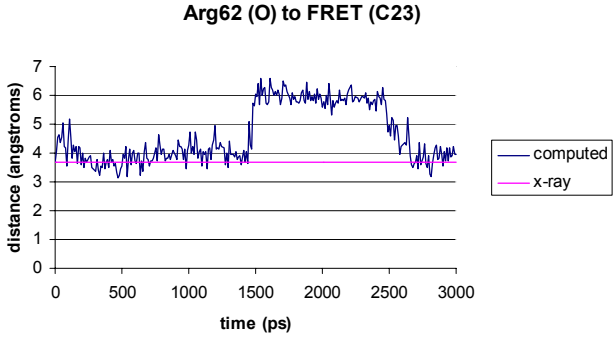
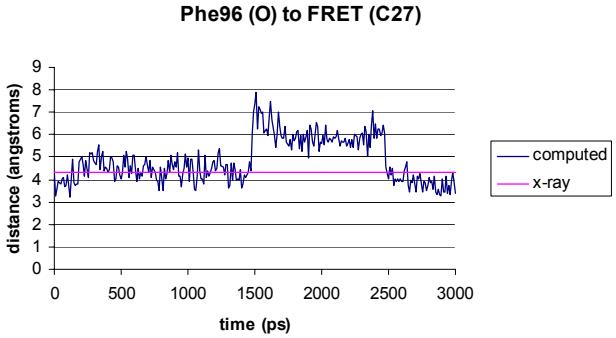


Figure 4.38 Distance over time for sidechain carbons of Leu64 to ligand.

The sidechain carbons of Leu64 and the oxygen of the phenol ring of the ligand move closer together during the simulation and then move farther apart near the end of the simulation. However, the backbone part of Leu64 moves away from the ligand during the 1.5-2.0 ns period of the simulation, which follows the pattern of the RMSD increase and the phenol ring rotation dihedral change. The backbone nitrogen of Leu64, the carbonyl oxygen of Arg62, and the carbonyl oxygen of Phe96 all show the same pattern of moving away from the ligand during 1.5-2.0 ns.







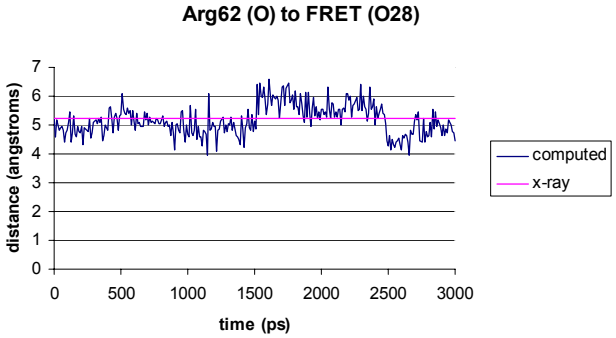


Figure 4.39 Distances over time for resides that move away from the ligand during the 1.5-2.0 ns time period.

The four distances shown in Figure 4.39 all exhibit the same pattern as the ligand RMSD and the dihedral for the phenol ring rotation (all show an increase during the 1.5-2.0 ns time period). Also, the ligand atoms in the plots of Figure 4.39 are all atoms of the phenol ring. An overlay of the ligand from different snapshots of the simulation shows that the fenretinide ligand does indeed move in the binding site.

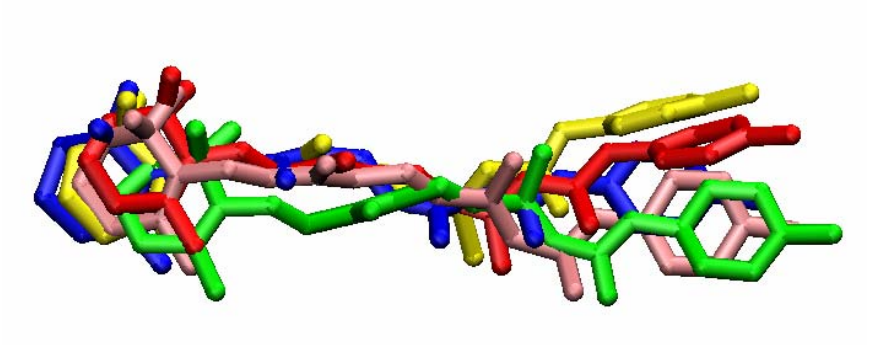


Figure 4.40 Overlay of snapshots of the fenretinide ligand.

Figure 4.40 shows an overlay of snapshots of the fenretinide ligand at different times during the simulation. The green (500 ps) and pink (1170 ps) structures are from the first part of the simulation before the RMSD and distances increased. The red (1600 ps) and yellow (1800 ps) structures are from the time period where the RMSD and distances increased. The blue (3000 ps) structure is from the last part of the simulation where the RMSD started to decrease again. The RMSD of the different parts of the ligand, the time series of dihedral for rotation of the phenol ring, and the increase in the protein-ligand distances for ligand atoms in or near the phenol ring are all evidence that the phenol ring at the end of the fenretinide ligand turns and shifts in the binding cavity, and that is what causes any large deviations in protein-ligand distances measured from the simulation.

A closer examination of some simulation snapshots reveals a possible explanation for the movement of the ligand. There are water molecules forming a "bridge" between the Leu35 oxygen to the hydroxyl group oxygen of the phenol ring of the ligand. The distance between Leu35 and the ligand decreases during the 1.5-2.0 ns; however, the distance is too large for a hydrogen bond--~4-6 Å (top plot in Figure 4.41). The water molecules start out in bulk solvent, but they move into the cavity with the ligand and form the "bridge" between Leu35 and the ligand.

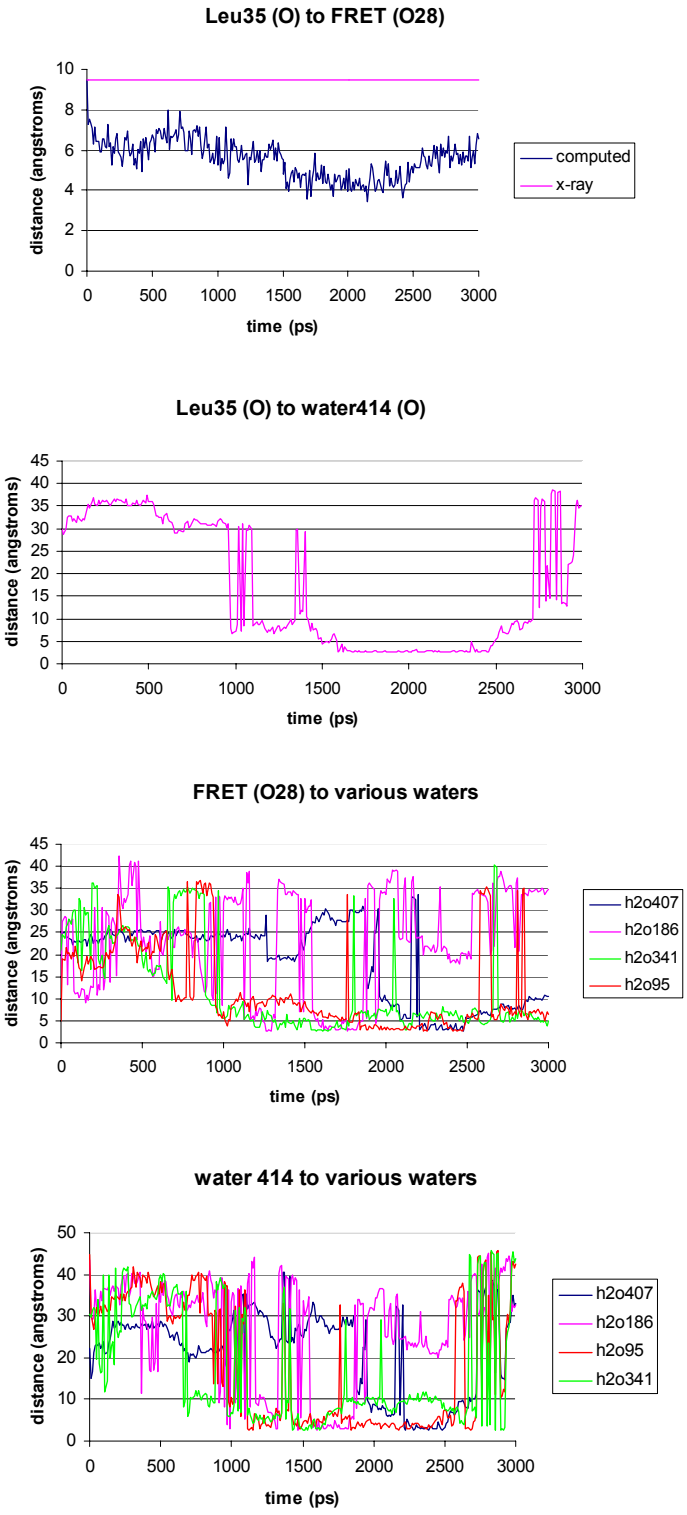


Figure 4.41 Distance from Leu35 to ligand (top), distance from Leu35 to water molecule 414 (second plot), and distance from ligand to various water molecules (third plot), and distance from water 414 to different nearby water molecules (bottom).

Figure 4.41 shows the distance over time from the water molecule to Leu35 and the distance from the ligand to various water molecules. The water molecule (numbered 414 in the simulation) is initially far away from the ligand and Leu35, but it moves in during the 1.5-2.0 ns time period and forms a hydrogen bond with the oxygen of Leu35 during most of that time period (second plot from the top in Figure 4.41). Different water molecules move near the ligand during this period to form a hydrogen bond with the hydroxyl group of the ligand and with water molecule number 414 (second from the bottom and bottom plots in Figure 4.41). The hydroxyl group of the ligand forms a hydrogen bond to different waters, which in turn form a hydrogen bond to water 414, which then forms a hydrogen bond to the carbonyl oxygen of Leu35 (some distances appear large because the O-O distance is shown in the plots. The O-H distances measured were all under 3.2 Å).

This water "bridge" may not be what is actually causing the ligand to shift and turn in the binding site, but this is probably what is stabilizing the conformation(s) seen during the 1.5-2.0 ns time period. The ligand appears to go back to its initial position in the binding site (see Figure 4.40) because the position of the phenol ring at 3 ns looks similar to that at 500 and 1170 ps (at 3 ns, the ligand conformation looks more similar to the structure at 500 ps than the conformation at 1600 ps does). The protein does not appear to undergo any major conformation changes (Figure 4.42) from the overlay of the structure at 500 ps (yellow ligand/blue protein) and the structure at 1600 ps (green ligand/red protein).

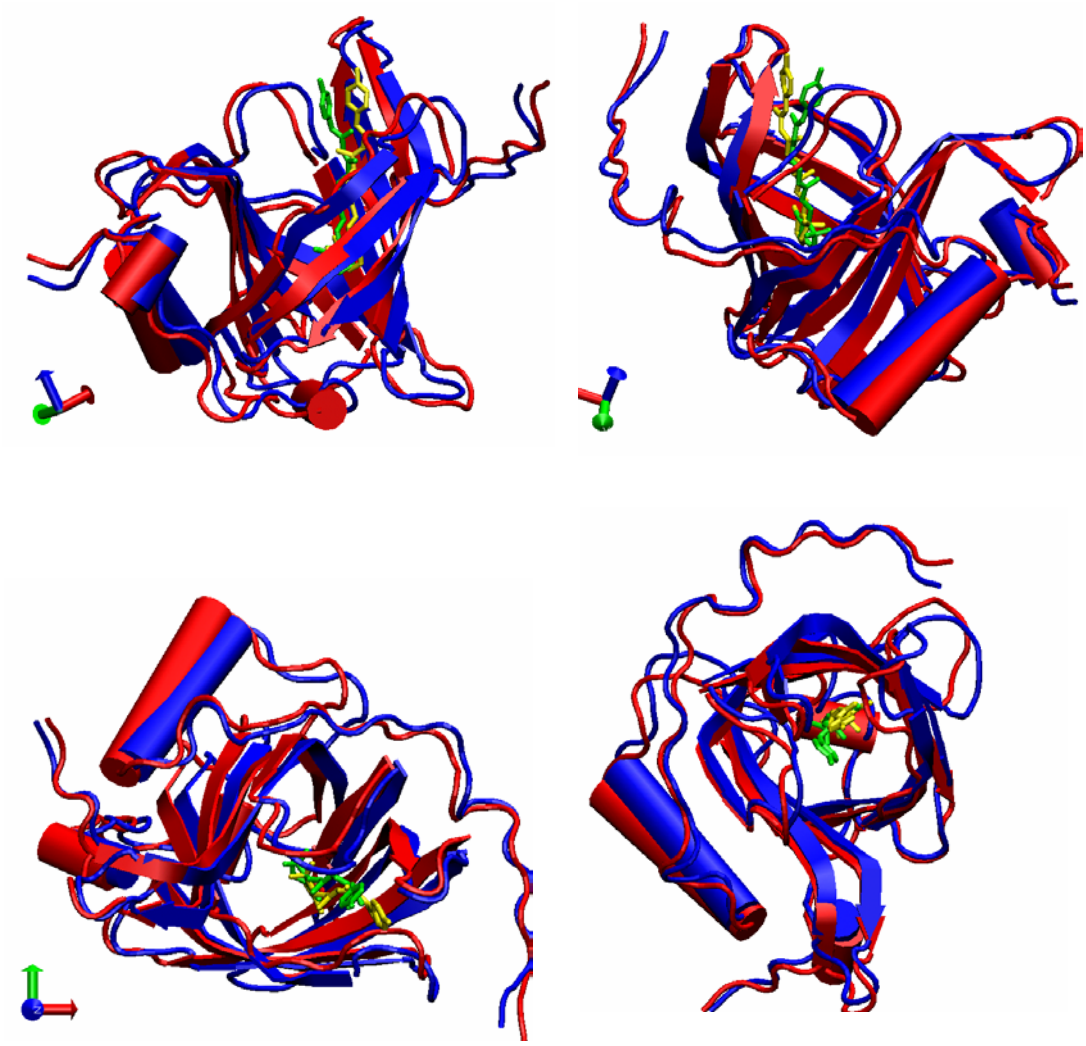


Figure 4.42 Overlays of the ligand/protein complex at 500 ps (yellow ligand/blue protein; before large RMSD change) and at 1600 ps (green ligand/red protein; during large RMSD change).

The end of the ligand with the phenol group looks like it shifts/turns in the binding cavity, but this does not appear to cause major structural changes to the protein (at least not on the timescale of the simulation). As with retinoic acid and CRABP, there is probably enough space in the binding site so the ligand can move around with little

consequence for the protein. Shown below are the residues around the ligand at 500 ps and at 1600 ps (Figures 4.43 and 4.44).

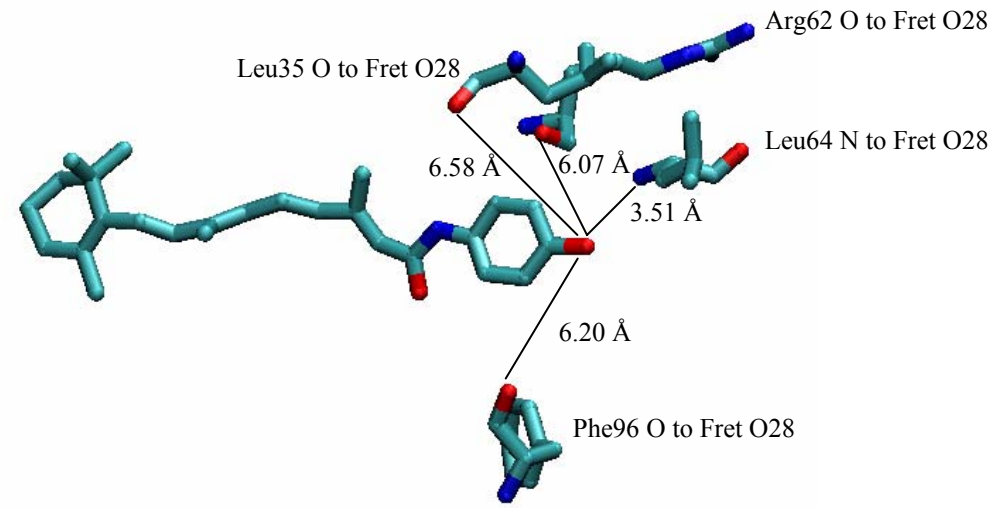


Figure 4.43 Snapshot from 500 ps (before distance increase). The labeled residues are the ones that showed a large increase in distance from the ligand during the 1.5-2.0 ns time period.

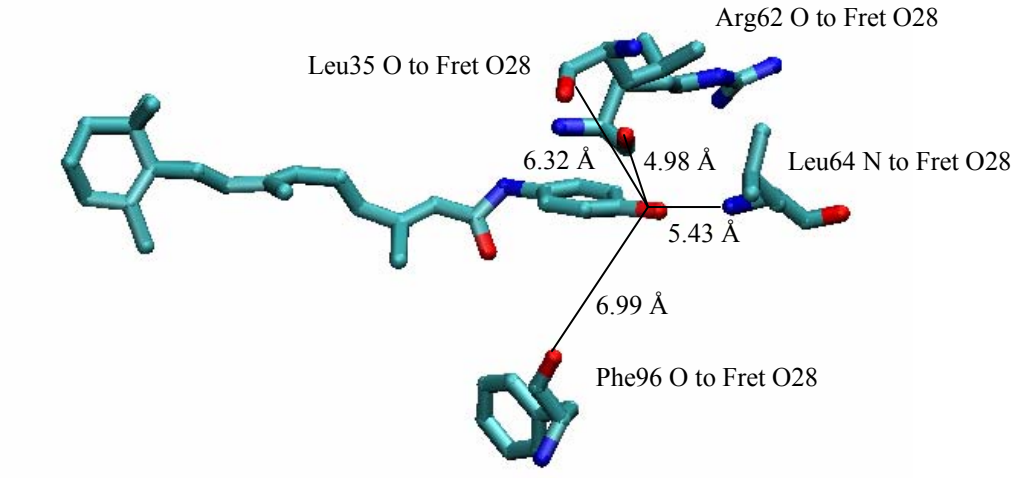


Figure 4.44 Snapshot from 1600 ps (after distance increase). The labeled residues are the ones that showed a large increase in distance from the ligand during the 1.5-2.0 ns time period.

Figures 4.43 and 4.44 show the orientation of the ligand with respect to the surrounding residues (those that showed a large distance change). The turn and shift of the ligand is visible from the snapshots. Overall, the simulated fenretinide/RBP complex agrees well with the x-ray structure. Most of the distances measured were in good agreement, the bonds and angle errors are small, and the fenretinide is not floating out of the binding cavity or distorting the protein much even with turning and shifting of the phenol ring. From the pictures in Figure 4.42, it appears that the fenretinide just has some space to move around in the binding cavity without disrupting too much of the protein, especially near the top of the β-barrel near the phenol ring. Because the turning/shifting of the phenol ring is not affecting the protein much (except for the amino acids right in the vicinity of the ligand), this does not appear to be a problem (this is the cause of the large dihedral error with respect to the x-ray structure). These motions of the ligand could be important in binding and dynamics of the fenretinide or other synthetic retinoid ligands.

4.4.4 Retinal Complexed with Sensory Rhodopsin II (SRII). The starting x-ray crystal structure for the simulation was determined by Royant and coworkers.[70] The structure of the SRII/retinal complex is shown in Figure 4.45.

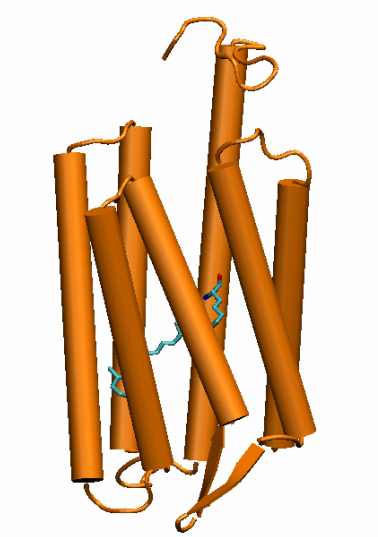
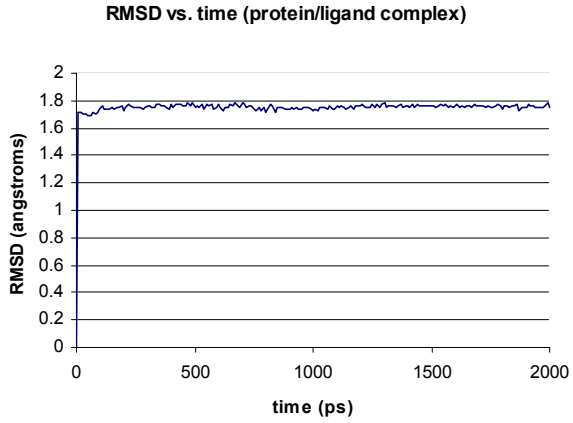


Figure 4.45 Sensory rhodopsin II complexed with retinal. The retinal ligand along with Lys205 are shown in cyan, and the protein is in orange.

Sensory rhodopsins I and II belong to the family of archaeal rhodopsins, which contain seven transmembrane helices (A through G) and a retinal chromophore covalently bound to a Lys residue.[70-72] SRI and SRII are coupled to the membrane-bound tranducer proteins, HtrI and HtrII. Excitation of SRI or SRII by light causes activation of these tranducer proteins, which in turn initiate phosphorylation cascades that result in attractant or repellent phototaxis, respectively (meaning the whole organism moves in response to the stimulus light).[72, 73] Analogous to bacteriorhodopsin (BR), SRI and SRII can work as light-driven proton pumps when they are not coupled to HtrI and HtrII, which indicates that the pump mechanism of BR is conserved in SRs.[73] Stimulation by light induces tilting of the helices on the cytoplasmic side of the SRs, and the how this structural

change in SRII translates into signal tranduction is unclear.[70, 72] The x-ray structure was determined in order to understand more about the mechanism of action of SRII.

The RMSD of the protein/ligand complex stabilizes quickly and stays at ~1.75 Å for the remainder of the simulation. The RMSD of the alpha carbons follows the pattern of the protein RMSD; the alpha carbon RMSD is ~0.5 Å lower that that of the protein RMSD. The retinal ligand appears to contribute most to the RMSD at ~1.77 Å; however, the ligand RMSD is stable and does not show any large increases or decreases.



Average RMSDs	
Protein/ligand complex	1.74 ± 0.13 Å
Ligand only	1.77 ± 0.14 Å
Protein only	1.70 ± 0.12 Å
Alpha carbons	1.54 ± 0.11 Å

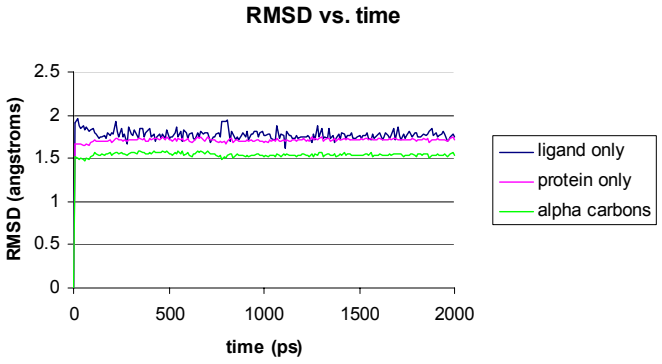


Figure 4.46 RMSD plots of the protein-ligand complex (top) and RMSDs of protein, ligand and alpha carbons (bottom).

The geometric parameters (bonds, angles, and dihedrals) of the ligand from the simulation were examined, and the error with respect to the x-ray structure was

calculated. The error for bonds was 0.03 ± 0.02 Å, the error for angles was 3.0 ± 2.3°, and the error for dihedrals was 25 ± 33º. The origin of this large dihedral error is the movement of the chain relative to the cyclohexene ring, and several rotations about bonds of the chain.

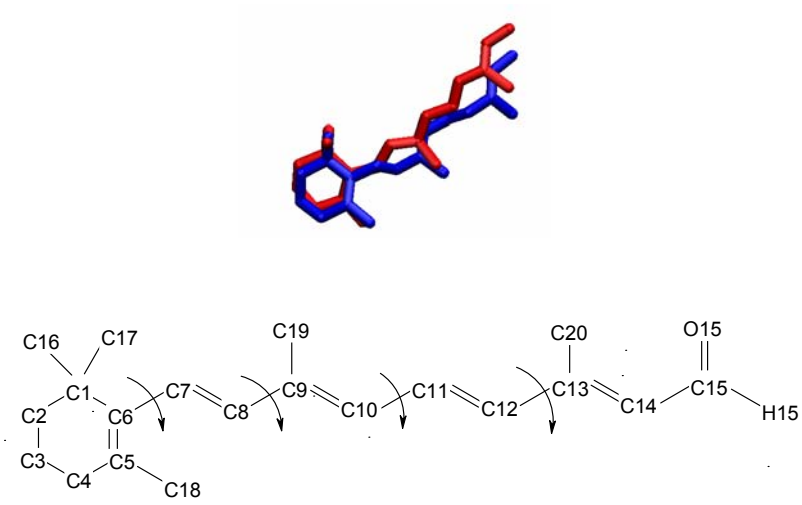


Figure 4.47 Overlay of the x-ray ligand (red) with the simulation average structure (blue) (top). The bottom picture shows the numbering of the carbon atoms of retinal and the dihedrals that contribute most to the dihedral error.

The chain of the ligand in the x-ray structure points upward, while the chain appears to move downwards during the simulation. The rotation about the C6-C7 bond contributes to the large dihedral error (C1-C6-C7-C8 and C5-C6-C7-C8). Three of the chain dihedrals also show large deviations from the x-ray dihedrals, and this can be seen in the overlayed structures in Figure 4.45 (C7-C8-C9-C19 and C11-C12-C13-C20). The rotation about the C10-C11 bond also shows deviation from the x-ray structure. This is similar to the retinoic acid ligand; rotations about the same single bonds of the chain show deviations from the x-ray structure. This movement probably occurs so that the ligand can optimize contacts with protein sidechains. The cyclohexene ring retains the

C2'-*endo* conformation. There does not appear to be much distortion of the surrounding protein residues, and most of the protein residue contacts stay in place over the time of the simulation. The largest deviation from the x-ray distances is ~0.5 Å (time series plots can be found in Appendix B).

Because the Schiff base linkage to the retinal ligand was parameterized, this distance was examined for the simulation to make sure it stayed in place. Figure 4.48 shows the Schiff base linkage over time, and the simulation shows excellent agreement with experiment (computed distance was 1.28 ± 0.015 Å; x-ray was 1.28 Å)

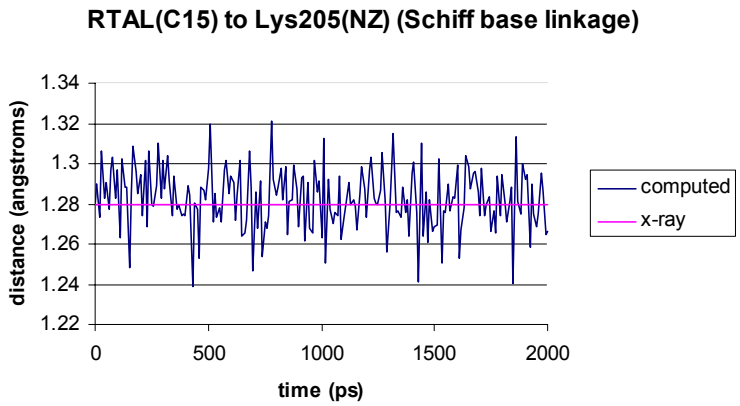


Figure 4.48 Schiff base linkage over time.

The authors described the binding site of SRII and some protein-protein interactions in detail, and these distances and features were all examined in the simulation. Some of these distances are shown in Table 4.12, some of the time series plots are shown, and the remaining time series plots can be found in Appendix B.

Table 0.12 Protein-ligand and protein-protein distances (SRII/retinal complex)

protein-ligand contact	**x-ray (Å)**	**Computed (Å)**
Trp76(CE3)--RTAL(C10)	4.18	4.10 ± 0.19
Trp171(CZ2)--RTAL(C20)	3.83	3.98 ± 0.12
Tyr174(CZ)--RTAL(C12)	3.60	3.85 ± 0.15
Ile83(CD)--RTAL(C20)	3.96	3.81 ± 0.17
Ala111(CB)--RTAL(C3)	5.69	5.62 ± 0.32
Leu126(O)--RTAL(C3)	5.38	4.98 ± 0.16
Gly130(O)--RTAL(C4)	4.24	4.01 ± 0.15
Val108(CG1)--RTAL(C7)	3.72	4.31 ± 0.15
Pro175(CB)--RTAL(C4)	3.87	4.17 ± 0.17
Ile43(CG2)--RTAL(C15)	5.75	6.10 ± 0.21
Ala47(CB)--RTAL(C15)	5.91	5.94 ± 0.15
Asp75(OD2)--RTAL(C15)	5.30	4.60 ± 0.32
Thr79(OG1)--RTAL(C15)	4.00	3.83 ± 0.19
Met109(CE)--RTAL(C10)	3.87	4.07 ± 0.22
Trp171(CB)--RTAL(C19)	3.72	4.15 ± 0.22
Asp201(OD2)--RTAL(C14)	3.44	3.36 ± 0.11
Thr204(OG1)--RTAL(C15)	3.74	3.99 ± 0.27
Arg72(NH1)--Cl- ion	3.56	4.28 ± 0.18
Tyr73(OH)--Cl- ion	3.51	3.47 ± 0.28
Phe69(O)--Cl- ion	3.35	3.51 ± 0.12
Water406(OH2)--Cl- ion	4.04, 2.83	3.28 ± 0.12
Water407(OH2)--Cl- ion	3.68	3.13 ± 0.08
Lys205(NZ)--Asp75(OD1)	4.07	3.58 ± 0.16
Lys205(NZ)--Thr79(OG1)	3.67	4.15 ± 0.34
Thr79(OG1)--Asp75(OD1)	2.81	2.81 ± 0.14

The authors identified one Cl^- ion in the structure, and they explain that this ion is coordinated to Arg72, Tyr73, Phe69, and two water molecules. Figure 4.49 shows these distances over time during the simulation.

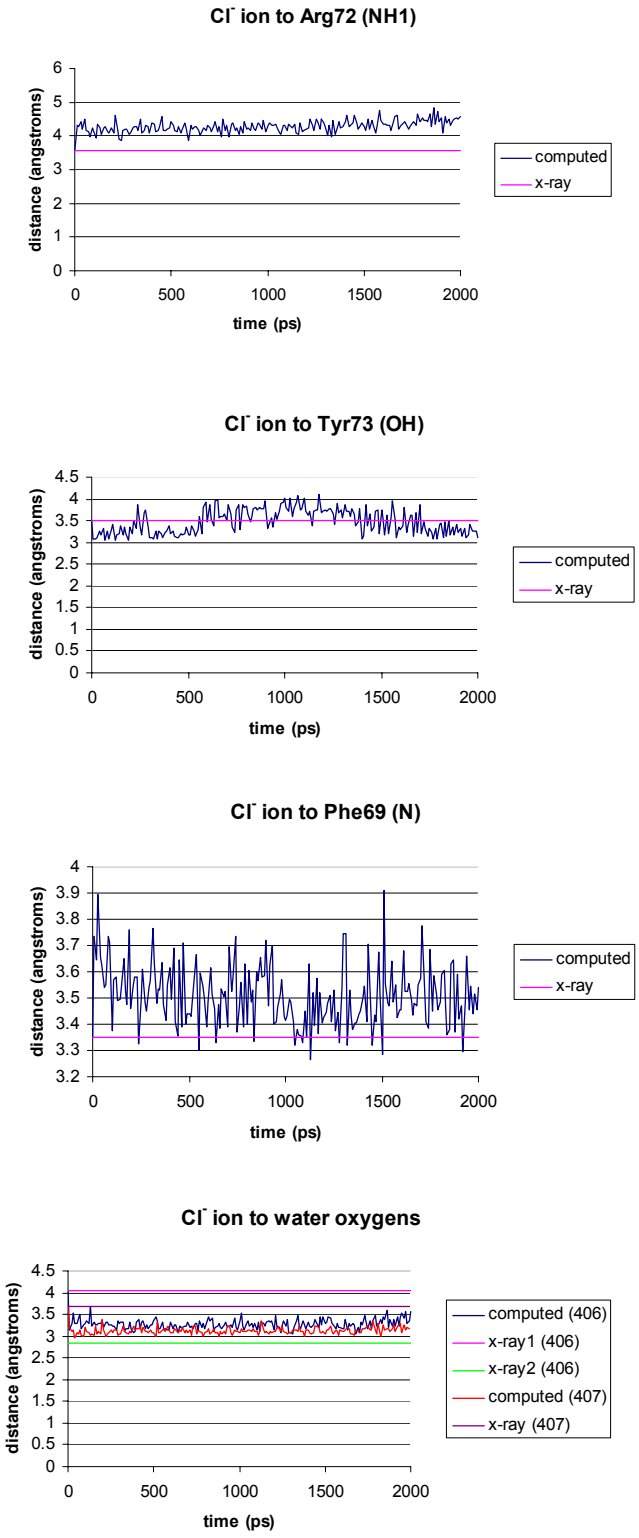
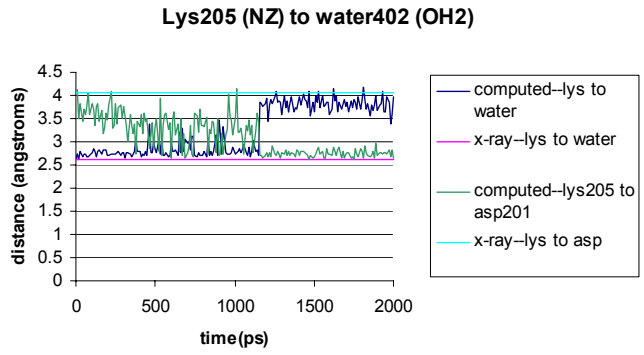
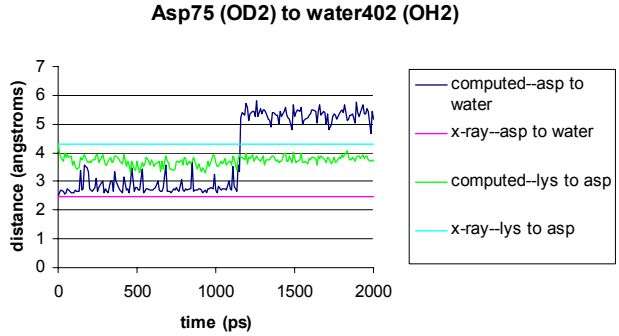


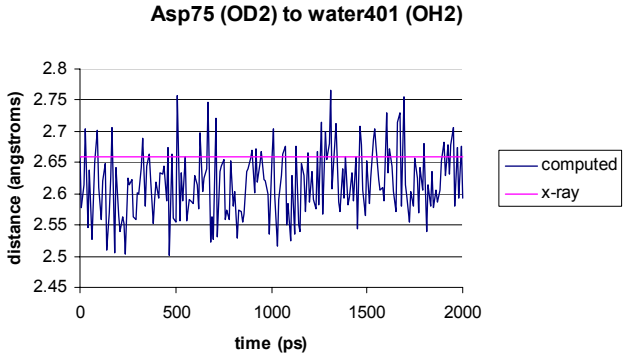
Figure 4.49 Cl^- ion distances to nearby residues (retinal/SRII).

The Cl^- ion moves away from Arg72 by ~0.72 Å, but the ion stays close to Tyr73 and Phe69. Two positions were identified for water 406, and the Cl^- ion stays in between the two positions. The ion and water 407 move ~0.5 Å closer together.

The protonated Schiff base is hydrogen bonded to a water molecule (numbered 402), which is also hydrogen bonded to two aspartic acid residues (Asp75 and Asp201). From the top plot of Figure 4.50, water 402 moves away from the Schiff base nitrogen (NZ) of Lys205, and an oxygen (OD1) of Asp201 appears to moves closer to the Schiff base nitrogen of Lys205, replacing the hydrogen bond that was formed between water 402 and Lys205. The second plot in Figure 4.50 shows the distance from Asp75 to water 402 and Lys205 over time. Water 402 moves away from the oxygen (OD2) of Asp75, but the Schiff base nitrogen of Lys75 stays close throughout the simulation. The contact between the Schiff base nitrogen and Asp75 (OD1) is part of the salt bridge that is hypothesized to keep SRII in its inactive conformation (the other part of the salt bridge is the Schiff base nitrogen to Asp75 (OD2).[73] Even though water 402 moves away from Asp75, water 401 stays close during the entire simulation (third plot in Figure 4.50) and forms a hydrogen bond. The bottom plot of Figure 4.50 shows the distance over time from Asp201 to water 402, which stays about the same throughout the simulation.







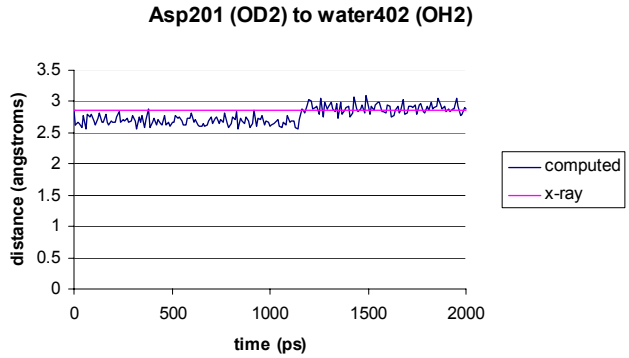


Figure 4.50 Distances for Lys205, Asp75, Asp201, water 401, and water 402.

The chain of the retinal ligand is constrained at the Schiff base end by aromatic residues (Trp76, Trp171, and Tyr174).

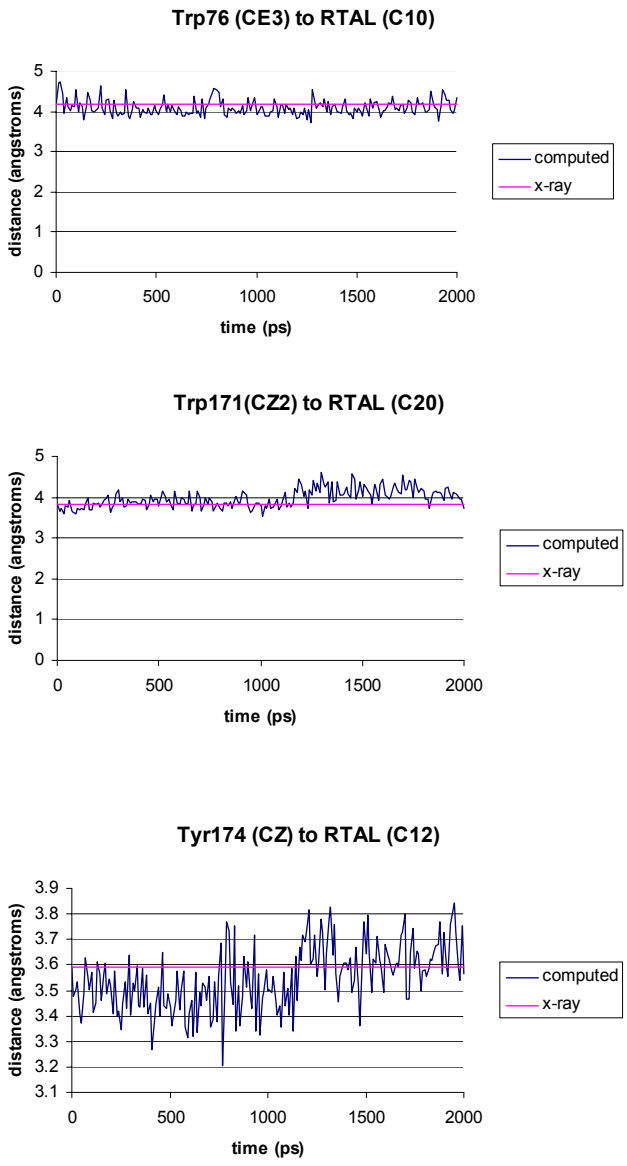


Figure 4.51 Distances from retinal ligand to aromatic residues.

Figure 4.51 shows the distances from the Schiff base end of the ligand to nearby aromatic residues. All of the measured simulation distances are in good agreement with the x-ray values; the retinal ligand appears to stay in place at the Schiff base end.

The authors made note of several protein-protein and protein-water contacts in and near the binding site of the ligand.

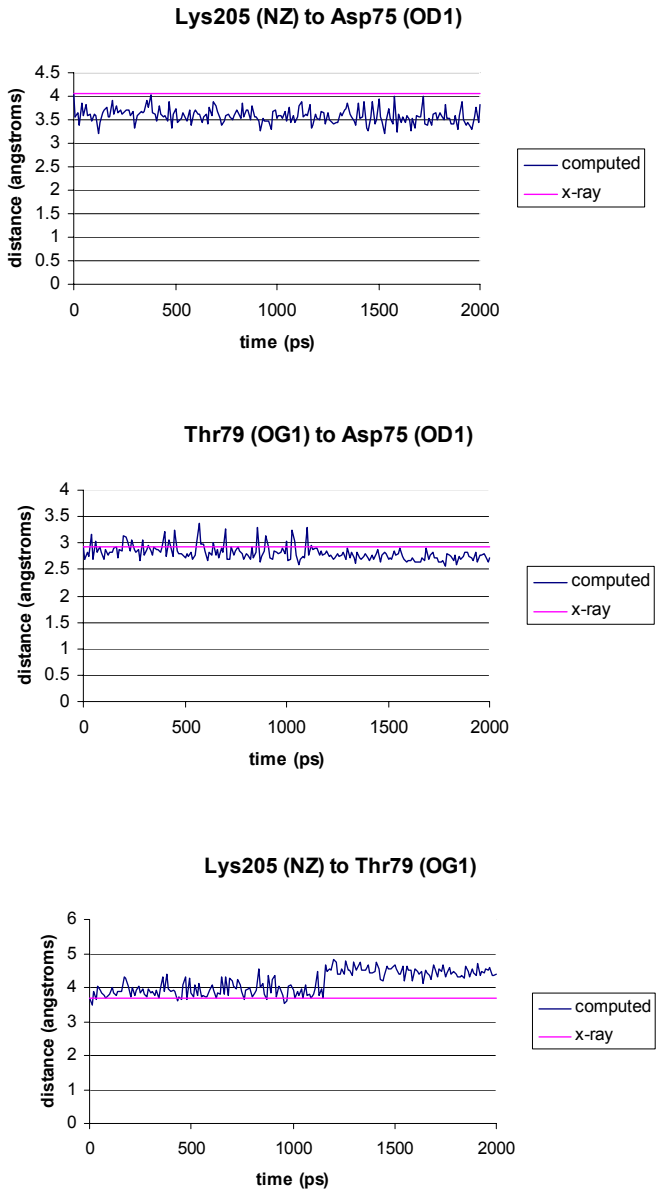


Figure 4.52 Protein-protein contacts near the ligand (retinal/SRII).

Lys205 stays near Asp75 throughout the simulation (top plot in Figure 4.52). This distance is the other part of the salt bridge that keeps the receptor in its inactive conformation. During the simulation, these two residues move about 0.5 Å closer

together. Lys205 moves away from Thr79 (~0.5 Å) about half way through the simulation (second plot in Figure 4.50); however, Asp75 stays close enough to Thr79 to form a hydrogen bond from one its sidechain oxygens to the hydroxyl group of Thr79. Thr79 and Lys205 move away from each other at about the same time that water 402 moves away from Lys205 (see Figure 4.50, top plot); however, Asp201 moves in and forms a hydrogen bond to the Schiff base nitrogen of Lys205 (Figure 4.50). Figure 4.53 shows the protein-water distances mentioned by the authors of the x-ray structure. The waters that are present in the x-ray structure in the active site stay there during the course of the simulation and form hydrogen bonds to Asp201 and Arg72.

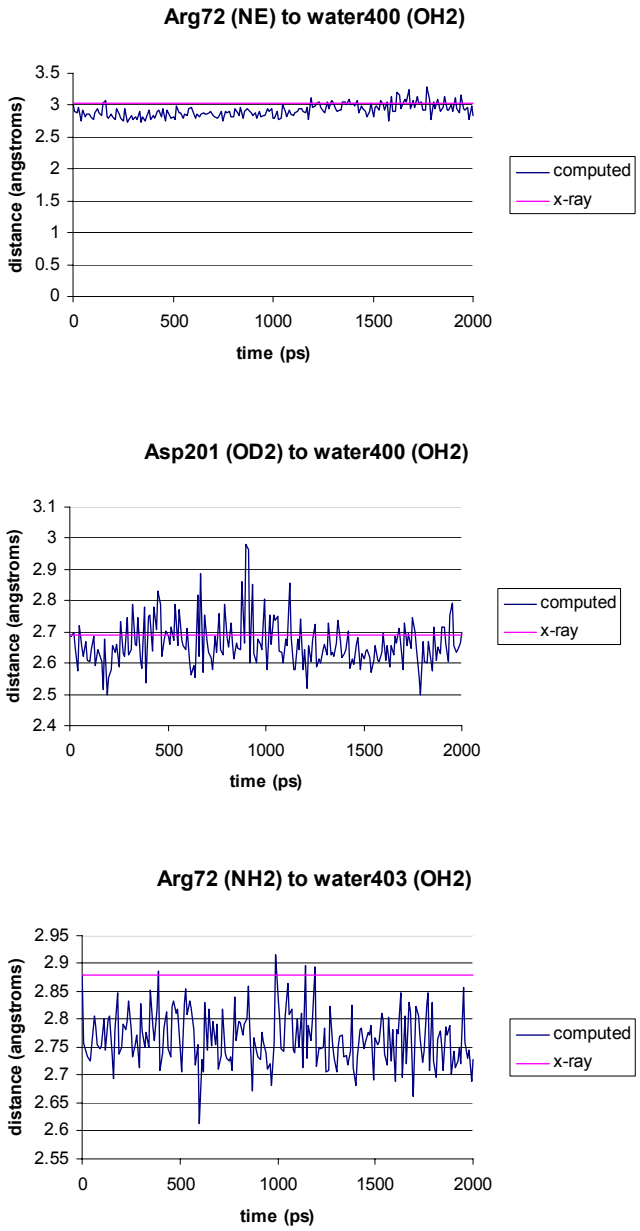


Figure 4.53 Protein water distances near the ligand (retinal/SRII).

The authors of the x-ray structure did not mention any other protein-ligand contacts, but several were measured from the simulation for comparison back to the x-ray structure. Three of them are shown here, and the rest can be found in Appendix B.

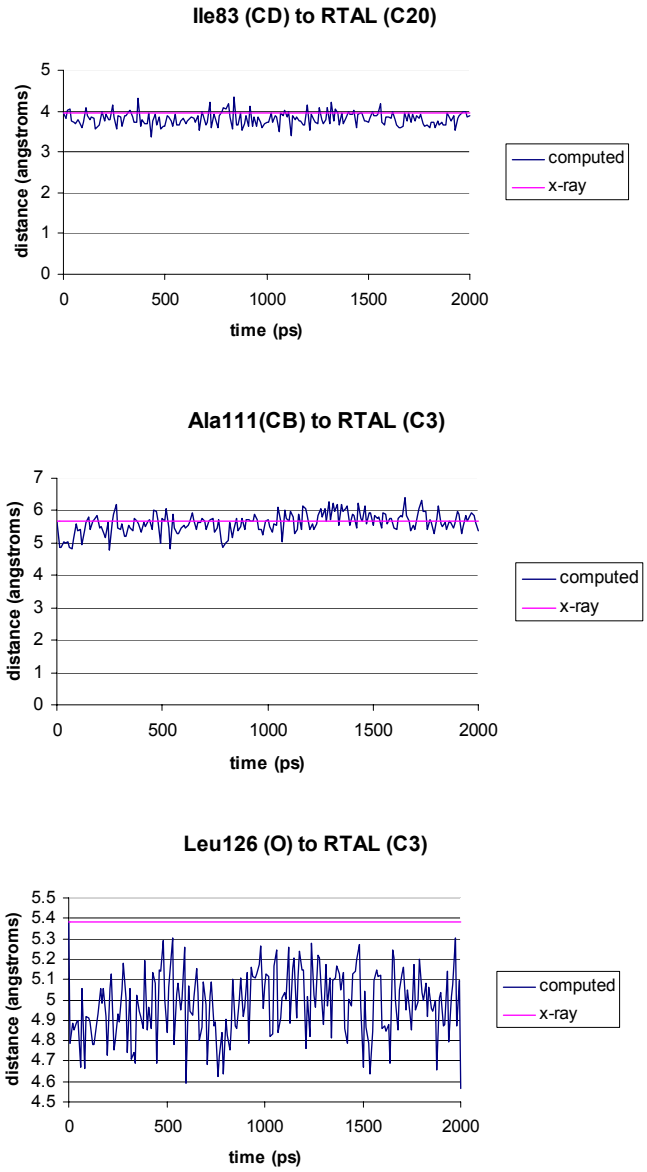


Figure 4.54 Three of the protein-ligand contacts measured from the simulation (retinal/SRII).

The protein-ligand contacts shown in Figure 4.54 show that there is good agreement between the crystal structure and the simulation. The top plot is from Ile83 to the isoprene chain of retinal, and the bottom two plots are from protein residues to the cyclohexene ring of retinal. In Figure 4.54, the protein residues and the ligand fluctuate around the value from the x-ray structure with the exception of Leu126, in which case the

residue and ligand move ~0.4 Å closer together. Figure 4.51 also contains time series plots for distances from the isoprene chain of retinal to nearby protein residues.

The rest of the protein-ligand contacts measured from the simulation can be found in Appendix B. Some of the distances for the protein-ligand contacts deviate from the starting x-ray value over time; the largest deviation (increase in distance) from the x-ray structure for protein-ligand distances observed from the simulation was ~0.5 Å (Val108 to retinal (C7); Appendix B). Solvating the protein in water instead of using a lipid bilayer could be the cause of some of these deviations. Most other distances fluctuated around the x-ray value or did not deviate by more than 0.5 Å. The simulated retinal/SRII complex agrees well with the x-ray structure. The Schiff base linkage from Lys205 to retinal stays intact, and the retinal ligand stays inside the protein. Most of the protein-ligand distances measured are in good agreement with the x-ray structure; in fact, this simulation had the most simulation distances that simply fluctuated around the x-ray structure values for the whole simulation time (did not show much deviation from x-ray). This is probably due to the fact that the retinal is covalently bound to SRII, rather than noncovalently, as in the other complexes. The average bond and angle error for the retinal is low; however, the average dihedral error is high. This is due to the chain of the ligand moving with respect to the cyclohexene ring and rotation around several of the single bonds of the chain. There does not seem to be any major distortion in the protein residue surrounding the ligand, so the ligand is probably moving simply to optimize contacts with the nearby sidechains.

4.4.5. Retinal Complexed with Bacteriorhodopsin from Halobacterium salinarum (BR). The starting x-ray crystal structure for the simulation was determined by Faham and coworkers.[74]

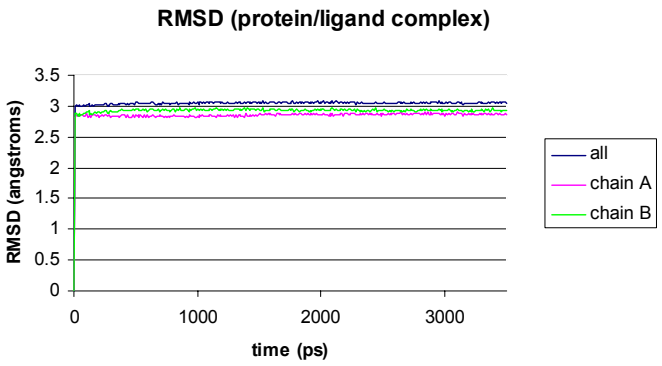
Figure 4.55 Retinal complexed with bacteriorhodopsin. The retinal ligand for each chain is shown in cyan, and the two protein chains of BR are shown in dark blue.

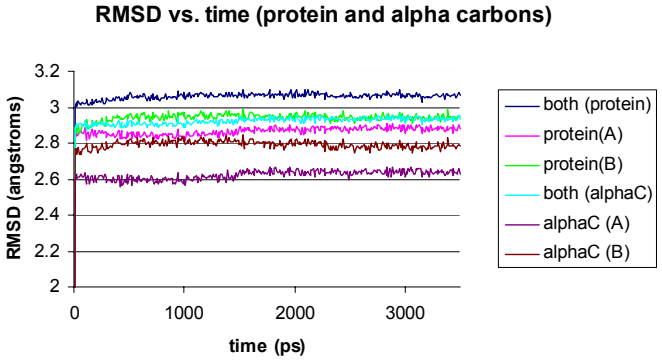
Bacteriorhodopsin (BR) is a light-driven proton pump (converts light energy into a proton gradient) present in the cell membrane of *Halobacterium salinarum*, an organism found in highly saline environments. BR contains seven transmembrane helices (A through G), which are linked by short loops on either side of the membrane. Each BR chain contains one molecule of retinal, which is covalently bound by a Schiff base linkage to Lys216 in helix G. Absorption of a photon by BR causes isomerization of all-trans retinal (protonated Schiff base linkage) to 13-cis retinal. The Schiff base is then depronated (which causes the retinal chromophore to straighten and push against helix F, causing it to tilt), and the proton is transferred to Asp85. Asp96 reprotonates the Schiff base, and a proton is taken in from the inner, cytoplasmic side to reprotonate Asp96 (helices F and G open a narrow channel in which a proton can come through). The proton on Asp85 is transferred through a network of hydrogen bonds and water molecules to the outside. The retinal then relaxes back to the all-trans configuration (ground state), and another

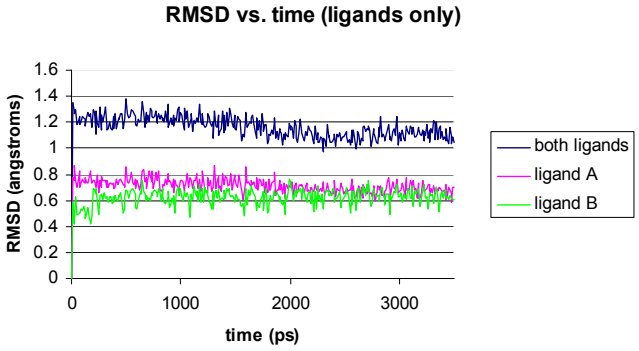
photocycle can then start. The proton pumping mechanism of BR has been studied extensively by a variety of experimental[75-84] and computational methods,[85-94] and the proton route through the membrane can now be followed in detail. The great interest in BR has continued for several reasons: BR is the most well-understood ion pump, it serves as a model for G-protein coupled receptors and other proteins which contain transmembrane helices, and it has become the paradigm of a membrane transporter.[95]

The simulation was performed on the asymmetric unit of the crystal structure, which was of a dimer of BR, rather than the native form, which is a trimer, or a monomer, which is commonly the result when the trimer form is solubilized by nonionic detergent.[75, 78] The RMSD and distance plots for protein-ligand or protein-protein distances were calculated for both chains, and only one x-ray distance is plotted because the x-ray distances were same for both chains of BR. Shown below are the RMSD plots for the whole complex, the protein only, the alpha carbons, and the ligands only.

The RMSD of the ligand/protein complex and the protein chains without the ligands both stabilize quickly to ~3 Å (both chains together and individually). The RMSD of the alpha carbons also stabilizes at the beginning of the simulation to just under 3 Å. The ligands together show an RMSD of ~1.2 Å until 2 ns into the simulation, where the RMSD then decreases to just above 1 Å. Ligand A follows the same pattern; a decrease in RMSD is observed at 2 ns. Ligand B shows an initial increase in RMSD, stabilizes at ~1 ns, and then converges to where Ligand A converges to (~0.7 Å).







Average RMSDs	all	chain A	chain B
Protein/ligand complex	3.04 ± 0.16 Å	2.85 ± 0.15 Å	2.92 ± 0.16 Å
Protein only	3.05 ± 0.16 Å	2.86 ± 0.15 Å	2.93 ± 0.16 Å
Alpha carbons only	2.92 ± 0.16 Å	2.62 ± 0.14 Å	2.79 ± 0.15 Å
Ligands only	1.16 ± 0.10 Å	0.71 ± 0.06 Å	0.62 ± 0.07 Å

Figure 4.56 RMSD plots for retinal/BR complex. Ligand/protein complex (top), protein chains and alpha carbons only (middle), and retinal ligands only (bottom).

The geometric parameters (bonds, angles, and dihedrals) of the ligands from the simulation were examined, and the error with respect to the x-ray structure was calculated. The error for bonds was 0.02 ± 0.02 Å for both ligands, the error for angles was 5 ± 4° for the first ligand and 5 ± 5° for the second ligand, and the error for dihedrals was ~33 ± 20º for both ligands. The origin of these large dihedral errors is the rotation about single bonds of the chain, and movement of the chain relative to the cyclohexene ring.

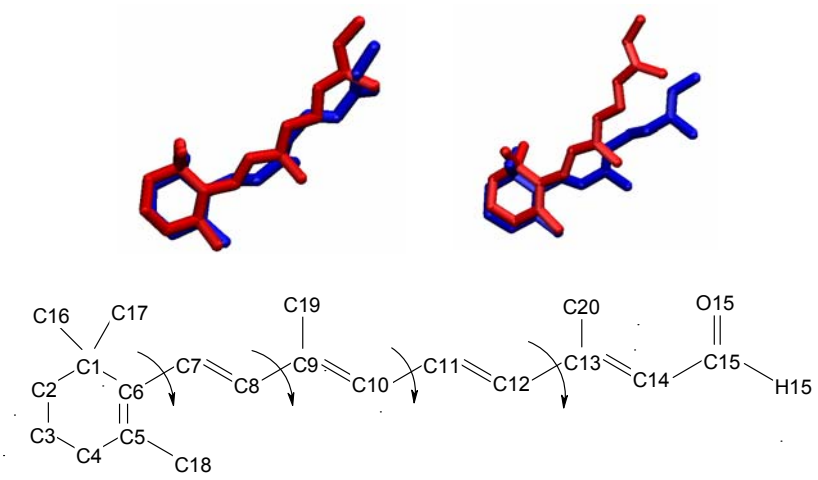


Figure 4.57 Overlays of x-ray structure (red) and simulation average structure (blue). The bottom picture shows the rotations about the single bonds that deviate from the x-ray structure.

The largest differences in the dihedrals are seen for the attachment of the ring to the chain (rotation about the C6-C7 bond), and for rotation about several bonds of the chain (C8-C9, C10-C11, and C12-C13). This is similar to the behavior seen the SRII/retinal and CRABPII/retinoic acid simulations; the deviations from the x-ray structure are rotations about the same bonds. The bonds that show rotation are single bonds, so the ligand

probably rotates about these to optimize contacts with protein sidechains. Both ligands retain the C3'-*endo* conformation for the cyclohexene ring found in the x-ray structure.

As for the retinal complexed with SRII, the Schiff base linkage was examined to make sure it stayed in place for both ligands. Figure 4.58 shows that the Schiff base linkage stays in place for both BR chains.

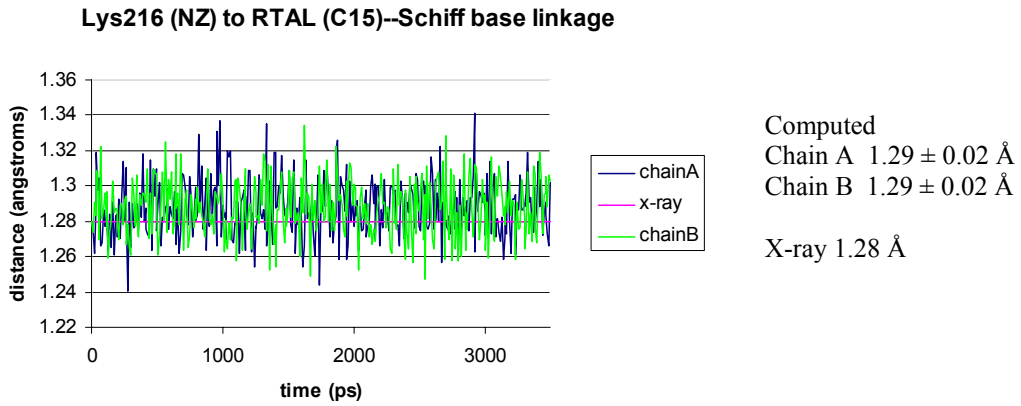


Figure 4.58 Schiff base linkages for retinal/BR complex.

Table 4.13 contains many of the protein-ligand distances measured from the simulation. Several time series plots are shown; additional properties measured can be found in Appendix B. The x-ray distances were measured for both chains, and the distance was found to be the same for both; thus, only one value is given for the x-ray distance in the table and on the time series plots.

Table 0.13 Protein-ligand and protein-protein distances (BR/retinal complex)

protein-ligand contact	**x-ray (Å)**	**computed chain A (Å)**	**computed chain B (Å)**
Ser141(OG)—RTAL(C2)	3.61	3.54 ± 0.12	3.65 ± 0.16
Tyr185(CD)—RTAL(C10)	3.95	3.42 ± 0.12	3.78 ± 0.19
Trp86(CD1)—RTAL(C14)	3.41	3.42 ± 0.10	3.49 ± 0.09
Met145(SD)—RTAL(C18)	3.89	3.62 ± 0.12	3.68 ± 0.15
Pro186(CB)—RTAL(C3)	4.97	4.00 ± 0.13	3.80 ± 0.14
Tyr185(CE1)—RTAL(C11)	3.86	3.73 ± 0.13	4.19 ± 0.22
Thr90(CG2)—RTAL(C12)	3.77	3.81 ± 0.16	4.36 ± 0.19
Asp85(CG)—RTAL(C15)	5.36	4.33 ± 0.20	4.42 ± 0.13
Met20(CE)—RTAL(C15)	5.79	6.64 ± 0.15	6.87 ± 0.42
Thr89(CB)—RTAL(C15)	4.32	4.31 ± 0.13	4.72 ± 0.16
Leu93(CD1)—RTAL(C20)	3.55	4.20 ± 0.37	3.63 ± 0.14
Lys41(NZ)—Asp38(OD1)	4.14	4.58 ± 0.10	3.87 ± 0.12
Lys41(NZ)—Asp38(OD2)	3.78	2.59 ± 0.07	2.67 ± 0.08
Thr46(OG1)—Asp96(OD2)	2.88	2.62 ± 0.05	2.80 ± 0.13

Several protein-protein distances were mentioned by the authors of x-ray structure, and three of those are shown below. Lys41 (NZ) forms an intrahelical salt bridge with Asp38 (top plot in Figure 4.59), and Thr46 forms a hydrogen bond to Asp96 (second plot in Figure 4.59). The hydrogen bond from Thr46 to Asp96 appears to stay intact over the entire simulation for both chains. The intrahelical salt bridge distance (Lys41 to Asp38) shortens for chain B but gets larger for chain A by ~0.44 Å. However, the distance decreases for both chains when measured from Lys41 to the other oxygen of Asp38 (distances decreases from 3.78 Å to ~2.60 Å), so the intrahelical salt bridge contact stays in place, just to the other oxygen of Asp38. Tyr30 makes another intrahelical salt bridge with Lys30 (bottom of Figure 4.59). The original hydrogen bond is broken in both chains; however, new ones are formed. In chain A, Tyr43 forms a hydrogen bond to Leu224, and Lys30 forms a hydrogen bond to Tyr26. In chain B, Tyr43 forms hydrogen bond to Met32, and Lys30 forms a hydrogen bond to Arg225.

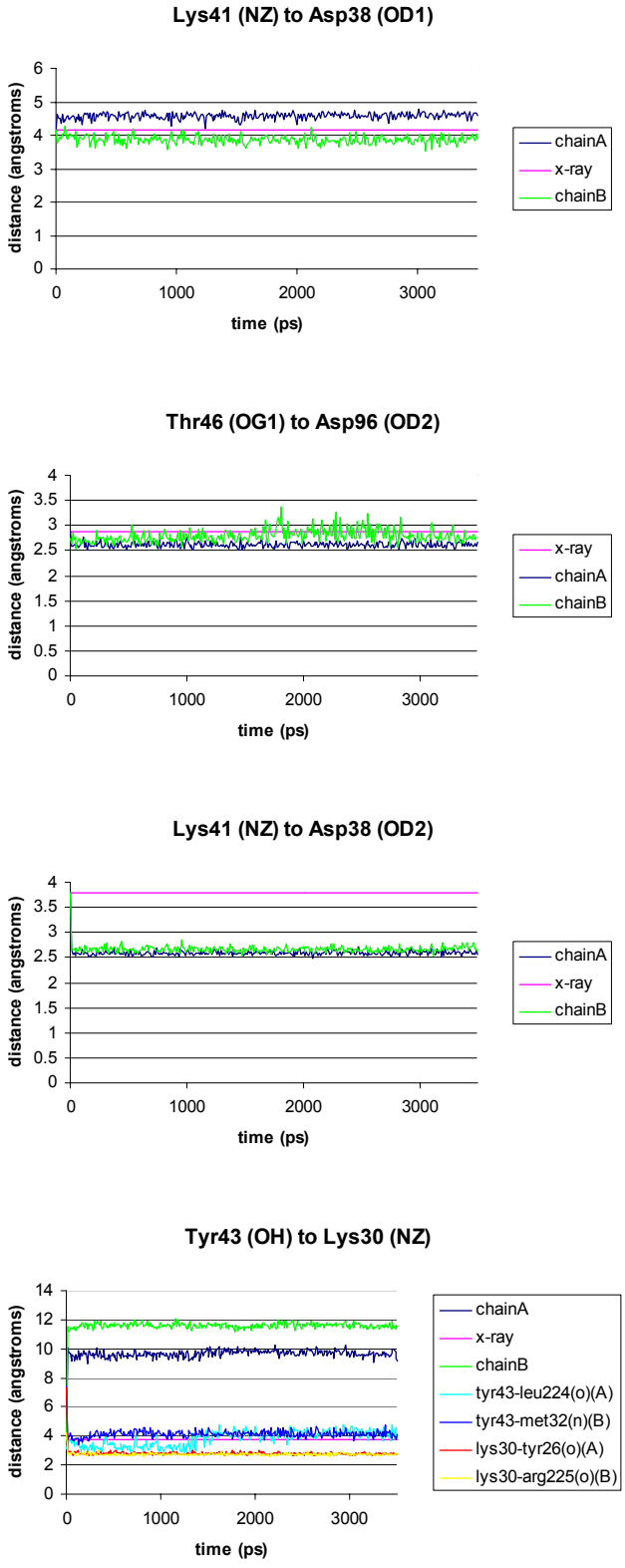


Figure 4.59 Protein-protein contacts measured from BR/retinal simulation.

Many contacts were measured from the retinal ligand to surrounding protein residues; however, only four of them will be shown here and the remaining ones are in Appendix B. The protein-ligand distances shown in Figure 4.60 show good agreement between the simulation and the x-ray structure. The distances measured from chain A and chain B appear to fluctuate around the x-ray value for the whole simulation time; however, the Tyr185 and Met145 distances to the ligand decrease ~0.5 Å from the x-ray value. The distances shown in Appendix B also show good agreement between the computed values and the x-ray structure. If any contacts appeared to be broken (i.e. ligand or protein distances increased), usually another protein residue near the ligand was found that replaced the original contact. This did not seem unusual; the ligand probably has some room to move around in the binding site even though it is covalently bound to the protein.

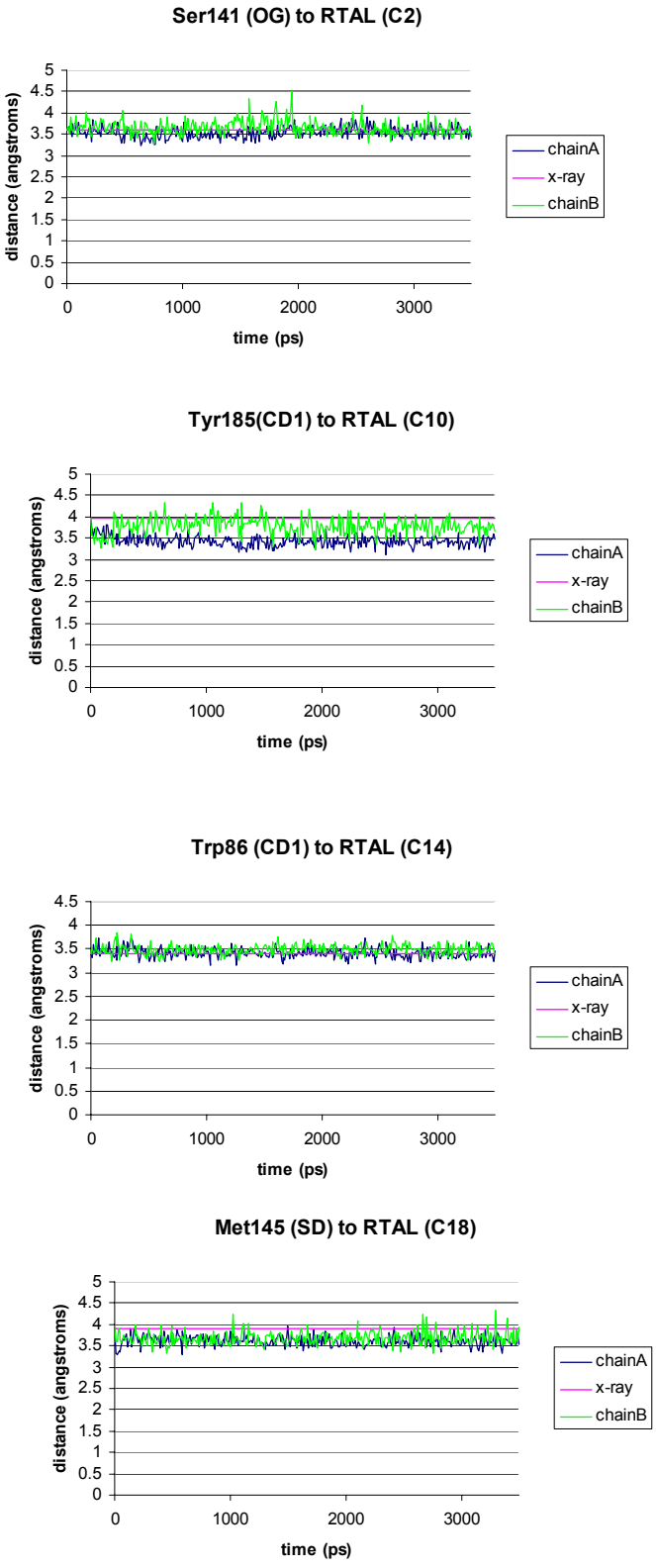


Figure 4.60 Four of the ligand-protein distances measured from the BR/retinal simulation.

Studies on bacteriorhodopsin have found that water molecules near the retinal ligand are extensively involved in both the structure and function of the protein.[79, 82, 83, 96] Water molecules were identified in the active site of the x-ray structure, and some of the distances of the water molecules to the ligand are shown in Figure 4.61. The top two plots in Figure 4.61 are for the ligand of chain A, and the bottom two plots are for the ligand of chain B. In chain A, a crystallographic water (denoted WATX14) stays close to the ligand for the entire simulation. WATX6 and WATX114 drift away from the ligand but are replaced by other crystallographic waters or waters added for the simulation. In chain B, WATX 23 and the ligand move closer together as the simulation progresses, and a bulk water replaces the replaces the WATX32-retinal contact. Even though other crystallographic or bulk waters replace the original water-retinal contacts, the interactions appear to stay intact over the simulation time period.

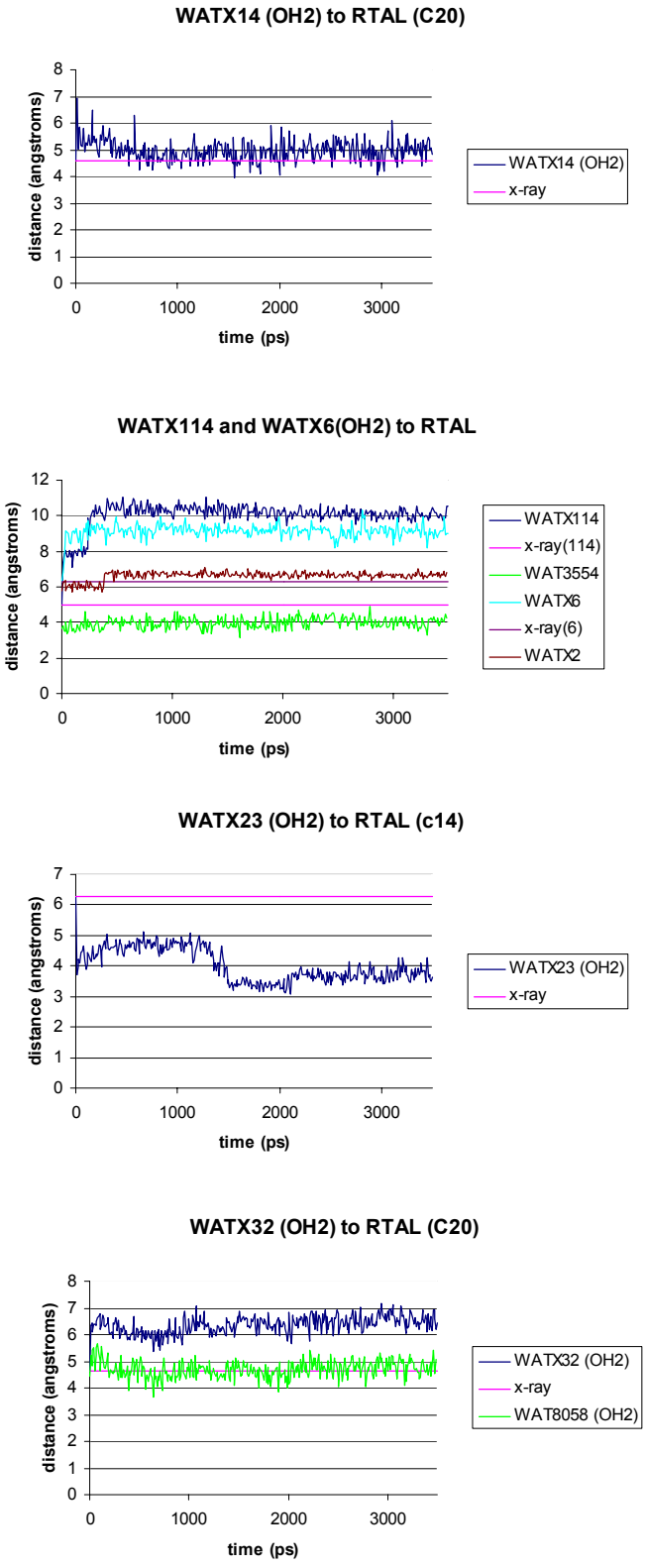


Figure 4.61 Water-ligand distances measured from the BR/retinal simulation.

Overall, the BR/retinal simulation compares well to the x-ray structure. The Schiff base linkage stays in place and if any ligand-protein or ligand-water contacts were broken, usually another protein residue or water was found near the ligand to make a new contact. The RMSD is highest for the BR/retinal complex, and this could be due to the fact that the protein is membrane-bound but the simulation was performed in water in order to save computational time. The average bond and angle errors are low; however, the average dihedral error is high. This is due to the chain moving relative to the cyclohexene ring and to the rotation about several single bonds of the chain. The surrounding protein residues do not appear to be disturbed much by the movement of the ligands.

4.6 Conclusions

Presented are parameters for the CHARMM force field that allow for molecular modeling and simulation studies of systems with alternating single and double bonds, specifically, retinoids. The bonds, angles, and dihedral angles have been parameterized against MP2/6-31G(d) target data, and the error between CHARMM and the quantum mechanical data is in excellent agreement. The CHARMM model compounds and full retinoids obtained from the new parameters were compared against available experimental data (spectroscopic and x-ray crystal data), and the comparison yields good agreement with respect to geometric parameters. The retinoids made from new parameters were also compared against five reported x-ray structures of protein/retinoid complexes. The structural and geometric analysis of these simulations also compares well to the experimental structures.

The CHARMM torsion energy surfaces show excellent agreement with those obtained from MP2/6-31G(d). The largest difference over all model compounds between barrier heights is ~1.15 kcal/mol (CHARMM is higher than MP2 for model **8**). The bond, angle, and dihedral error for between CHARMM and MP2 is low, with bond error being 0.012 ± 0.009 Å, angle error is 1 ± 1°, and dihedral error is 2 ± 4°.

Comparison of the new CHARMM parameters with available experimental data shows that CHARMM is in good agreement with experiment. The model compounds were compared to available experimental structures; the average bond error is 0.03 ± 0.06 Å, the average angle error is 2 ± 2°, and the average dihedral error is 3 ± 3°. Crystal simulations of the new retinoids show that the new parameters yield the proper bond lengths for compounds with alternating single and double bonds. The simulated structures of retinal and retinoic acid do not deviate much from the starting x-ray crystal structures; the cyclohexene ring retains the proper conformation, the chain does move much relative to the cyclohexene ring, and the single bonds of the chain show minimal rotation (chain stays relatively planar). Another important observation from the crystal simulations is that the new CHARMM parameters show the proper crystal transformations at high temperature. Retinoic acid crystallizes in the monoclinic and triclinic forms; however, at high temperature, the monoclinic form transforms irreversibly to the triclinic form. The crystal simulations at high temperature of retinoic acid in the monoclinic form show the beginnings of this transformation. The chain starts to move outward from the cyclohexene ring during the simulation, which is what is seen in the triclinic form (the monoclinic structure has the chain pointing upward from the ring).

In the crystal simulations of the protein/retinoid complexes, some retinoids show rotation about the single bonds of the chain and rotation of the chain relative to the cyclohexene ring. This behavior is observed because the retinoid ligand will try to optimize contacts with nearby protein side chains. These movements of the retinoids do not appear to cause any major conformational changes or disruptions of the protein, and the retinoid ligands stay inside the binding cavities of the proteins during the simulations and retain contacts found in the x-ray structures. Some ligand motions were also discovered that could be important in ligand binding and dynamics or interaction of the protein/ligand complex with other molecules. In the case of the retinoic acid/CRABPII complex, the cyclohexene ring turns in the binding site of the protein. In the fenretinide/RBP complex, the phenol ring turns and shifts in the binding site of RBP.

The new parameters yield good agreement with experiment and now exhibit the proper bond lengths, angles, and torsion angles for compounds with alternating single and double bonds. The new parameters can now be used in simulations of proteins that bind retinoids to better understand the structure and dynamics of those systems and in drug design to make new retinoids that have therapeutic and anticancer potential.

References

(1) MacKerell, A. D., Jr., Atomistic Models and Force Fields. In *Computational Biochemistry and Biophysics*, Marcel Dekker, Inc.: New York, 2001; 7-38.

(2) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B. R., W.E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M., All-atom empirical potential for molecular modeling and dynamics studies of proteins, *Journal of Physical Chemistry*, 1998, 102, 3586.

(3) Foloppe, N.; MacKerell, A. D., All-Atom Empirical Force Field for Nucleic Acids: I. Parameter Optimization Based on Small Molecule and Condensed Phase Macromolecular Target Data, *J. Comp. Chem.*, 2000, 21, (2), 86.

(4) Klauda, J. B.; Brooks, B. R.; MacKerell, A. D., Jr.; Venable, R. M.; Pastor, R. W., An ab initio Study on the Torsional Surface of Alkanes and Its Effect on Molecular Simulations of Alkanes and and DPPC Bilayer, *J. Phys. Chem. B*, 2005, 109, 5300-5311.

(5) MacKerell, A. D.; Feig, M.; Brooks, C. L., Improved Treatment of the Protein Backbone in Empirical Force Fields, *J. Am. Chem. Soc.*, 2004, 126, 698.

(6) Foloppe, N.; MacKerell, A. D., Jr., Contribution of the Phosphodiester Backbone and Glycosyl Linkage Intrinsic Torsional Energetics to DNA Structure and Dynamics, *J. Phys. Chem. B*, 1999, 103, 10955-10964.

(7) Feller, S. E.; MacKerell, A. D., Jr., An Improved Empirical Potential Energy Function for Molecular Simulations of Phospholipids, *J. Phys. Chem. B*, 2000, 104, 7510-7515.

(8) Boresch, S.; Leitgeb, M.; Beselman, A.; MacKerell, A. D., Jr., Unexpected Relative Aqueous Solubilities of a Phosphotyrosine Analogue and Two Phosphonate Derivatives, *J. Am. Chem. Soc.*, 2005, 127, 4640-4648.

(9) Feng, M.-H.; Philippopoulos, M.; MacKerell, A. D., Jr.; Lim, C., Structural Characterization of the Phosphotyrosine Binding Region of a High-Affinity SH2 Domain--Phosphopeptide Complex by Molecular Dynamics Simulation and Chemical Shift Calculations, *J. Am. Chem. Soc.*, 1996, 118, 11265-11277.

(10) Reiling, S.; Schenkrich, M.; Brickmann, J., Force Field Parameters for Carbohydrates, *J. Comp. Chem.*, 1996, 17, 450-468.

(11) Pavelites, J. J.; Gao, J.; Bash, P.; MacKerell, A. D., Jr., A Molecular Mechanics Force Field for NAD+, NADH, and the Pyrophosphate Groups of Nucleotides, *J. Comp. Chem.*, 1997, 18, 221-239.

(12) Bindal, R. D.; Golab, J. T.; Katzenellenbogen, J. A., Ab initio calculations on N-methylmethanesulfonamide and methyl methanesulfonate for the development of force field torsional parameters and their use in the conformational analysis of some novel estrogens, *J. Am. Chem. Soc.*, 1990, 112, 7861-7868.

(13) Halgren, T. A., Merck Molecular Force Field. I. Basis, Form, Scope, Parameterization, and Performance of MMFF94, *J. Comp. Chem.*, 1996, 17, 490-519.

(14) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, 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.

(15) 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.

(16) Jorgensen, W. L.; Tirado-Rives, J., The OPLS Potential Functions for Proteins. Energy Minimizations for Crystals of Cyclic Peptides and Crambin, *J. Am. Chem. Soc.*, 1988, 110, 1657-1666.

(17) Soares, T. A.; Hunenberger, P. H.; Kastenholz, M. A.; Krautler, V.; Lenz, T.; Lins, R. D.; Oostenbrink, C.; van Gunsteren, W. F., An Improved Nucleic Acid Parameter Set for the GROMOS Force Field, *J. Comp. Chem.*, 2004, 26, 725-737.

(18) Ewig, C. S.; Berry, R.; Dinur, U.; Hill, J.-R.; Hwang, M.-J.; Li, H.; Liang, C.; Maple, J.; Peng, Z.; Stockfisch, T. P.; Thacher, T.; Yan, L.; Ni, X.; Hagler, A. T., Derivation of Class II Force Fields. VIII. Derivation of a General Quantum Mechanical Force Field for Organic Compounds, *J. Comp. Chem.*, 2001, 22, 1782-1800.

(19) Smith, B. J.; Radom, L., Structure of Vinyl Alcohol: A Resolution of the Discrepancy between Theory and Experiment, *J. Am. Chem. Soc.*, 1990, 112.

(20) Kukolich, S. G.; Reohrig, M. A.; Wallace, D. W.; Henderson, G. L., Microwave measurements of the Rotational Spectrum and Structure of (Butadiene) iron Tricarbonyl, *J. Am. Chem. Soc.*, 1993, 115, 2021.

(21) Cole, A. R. H.; Mohay, G. M.; Osborne, G. A., Rotational Fine Structure of an infra-red band of butadiene and the length of the central C-C bond, *Spectrochimica Acta*, 1967, 23A, 909-920.

(22) Marais, D. J.; Sheppard, N.; Stoicheff, B. P., An investigation of the structures of Butadiene by high resolution infra-red and raman spectroscopy, *Tetrahedron*, 1962, 17, 163.

(23) Schomaker, V.; Pauling, L., The Electron Diffraction Investigation of the Structure of Benzene, Pyridine, Pyrazine, Butadiene-1,3, Cyclopentadiene, Furan, Pyrrole, and Thiophene, 1939.

(24) Almenningen, A.; Bastiansen, O.; Traetteberg, M., An electron diffraction Reinvestigation of the Molecular Structure of 1,3-Butadiene, *Acta. Chem. Scand.*, 1958, 12, 1221.

(25) Haugen, W.; Traetteberg, M., Molecular Structure of 1,3-Butadiene and 1,3,5-trans-hexatriene, *Acta Chemica Scandinavica (1947-1973)*, 1966, 20, 1726.

(26) Kuchitsu, K.; Fukuyama, T.; Morino, Y., Average Structures of Butadiene, Acrolein, and Glyoxal Determined by Gas Electron Diffraction and Spectroscopy, *Journal of Molecular Structure*, 1968, 1, 463-479.

(27) Defrees, D. J.; Levi, B. A.; Pollack, S. K.; Hehre, W. J.; Binkley, J. S.; Pople, J. A., Effect of Electron Correlation on Theoretical Equilibrium Geometries, *J. Am. Chem. Soc.*, 1979, 101, 4085-4089.

(28) Foloppe, N.; MacKerell, A. D., Jr., Conformational Properties of the Deoxyribose and Ribose Moieties of Nucleic Acids: A Quantum Mechanical Study, *J. Phys. Chem. B*, 1998, 102, 6669-6678.

(29) Mui, P. W.; Grunwald, E., Thermodynamics of Conformational Changes in 1,3-Butadiene Studied by High-Temperature Ultraviolet Absorption Spectroscopy, *J. Am. Chem. Soc.*, 1982, 104, 6562-6566.

(30) Fisher, J. J.; Michl, J., S-Cis vs. Gauche 1,3-Butadiene: Evidence of Planarity from Polarized Matrix-Isolation IR Spectroscopy, *J. Am. Chem. Soc.*, 1987, 109, 1056-1059.

(31) Halonen, M.; Halonen, L.; Nesbitt, D. J., Structural Issues in Conjugated Hydrocarbons: High-Resolution Infrared Slit-Jet Spectroscopy of trans-1,3-Butadiene, *J. Phys. Chem. A*, 2004, 108, 3367-3372.

(32) Caminati, W.; Grassi, G.; Bauder, A., Microwave Fourier Transform Spectrum of s-trans-1,3-butadiene-1,1-d_2, *Chem. Phys. Lett.*, 1988, 148, (1), 13-16.

(33) Carreira, L. A., Determination of the torsional potential function of 1,3-butadiene, *J. Chem. Phys.*, 1975, 62, (10), 3851-3854.

(34) Breulet, J.; Lee, T. J.; Schaefer, H. F., III, Comparison between the s-Cis and Gauche Conformers of 1,3-Butadiene, *J. Am. Chem. Soc.*, 1984, 106, 6250-6253.

(35) Rice, J. E.; Bowen, L.; Lee, T. J.; Rohlfing, C. M., The Structure of Cis-Butadiene, *Chem. Phys. Lett.*, 1989, 161, 277-284.

(36) Alberts, I. L.; Schaefer, H. F., III, The Second Stable Conformer of 1,3-Butadiene. Geometry Optimizations with Configuration Interaction and Coupled Cluster Methods, *Chem. Phys. Lett.*, 1989, 161, 375-382.

(37) Bock, C. W.; Panchenko, Y. N., An ab initio structural investigation of 1,3-Butadiene, isoprene, and 2,3-dimethyl-1,3-butadiene rotamers, *J. Mol. Structure (Theochem)*, 1989, 187, 69-82.

(38) Skaarup, S.; Boggs, J. E.; Skancke, P. N., Contributions of Resonance, Hybridization, and Nonbonded Interactions to the Structure of Butadiene, *Tetrahedron*, 1976, 32, 1179-1181.

(39) Aston, J. G.; Szasz, G.; Woolley, H. W.; Brickwedde, F. G., Thermodynamic Properties of Gaseous 1,3-Butadiene and the Normal Butenes about 25oC, *J. Chem. Phys.*, 1946, 14, (2), 67.

(40) Wiberg, K. B.; Rosenberg, R., Butadiene. I. A Normal Coordinate Analysis and Infrared Intensities. Structure of the Second Rotamer, *J. Am. Chem. Soc.*, 1990, 112, 1509.

(41) De Mare, G. R.; Neisius, D., Ab initio study of rotational isomerism in 1,3-butadiene. Effect of Geometry Optimization and Basis Set Size on the Barriers to Rotation and on the Stable Rotamers, *Journal of Molecular Structure*, 1984, 109, 103-126.

(42) De Mare, G. R.; Panchenko, Y. N.; Auwera, J. V., Structure of the High-Energy Conformer of 1,3-Butadiene, *J. Phys. Chem. A*, 1997, 101, 3998-4004.

(43) Karpfen, A.; Choi, C. H.; Kertesz, M., Single-Bond Torsional Potentials in Conjugated Systems: A Comparison of ab Initio and Density Functional Results, *J. Phys. Chem. A*, 1997, 101, 7426-7433.

(44) Engeln, R.; Consalvo, D.; Reuss, J., Evidence for a gauche minor conformer of 1,3-butadiene, *Chem. Phys.*, 1992, 160, 427-433.

(45) Murcko, M. A.; Castejon, H.; Wiberg, K. B., Carbon-Carbon Rotational Barriers in Butane, 1-Butene, and 1,3-Butadiene, *J. Phys. Chem.*, 1996, 100, 16162-16168.

(46) Kofranek, M.; Karpfen, A.; Lischka, H., Gauche-verus s-cis-butadiene revisited: a molecular dynamics simulation of the Ar matrix effect, *Chem. Phys. Lett.*, 1992, 189, 281-286.

(47) Traetteberg, M.; Paulen, G.; Cyvin, S. J.; Panchenko, Y. N.; Mochalov, V. I., Structure and Conformations of Isoprene by Vibrational Spectroscopy and Gas Electron Diffraction, *J. Mol. Structure (Theochem)*, 1984, 116, 141-151.

(48) Haisa, M.; Kashino, S.; Maeda, H., The Orthorhombic Form of p-Hydroxyacetanilide, *Acta Crystallogr.*, 1974, B30, 2510-2512.

(49) Higgs, M. A.; Sass, R. L., The Crystal Structure of Acrylic Acid, *Acta Crystallogr.*, 1963, 16, 657-661.

(50) Stam, C. H., The Crystal Structure of a Monoclinic Modification and the Refinement of a Triclinic Modification of Vitamin A Acid (Retinoic Acid), $C_{20}H_{28}O_2$, *Acta Crystallographica*, 1972, B28, 2936-2945.

(51) Stam, C. H.; MacGillavry, C. H., The Crystal Structure of the Triclinic Modification of Vitamin-A Acid, *Acta Crystallographica*, 1963, 16, 62-68.

(52) Hamanaka, T.; Mitsui, T.; Ashida, T.; Kakudo, M., The Crystal Structure of All-trans Retinal, *Acta Crystallographica*, 1972, B28, 214-222.

(53) Allinger, N. L.; Li, F.; Yan, L.; Tai, J. C., Molecular Mechanics (MM3) Calculations on Conjugated Hydrocarbons, *J. Comp. Chem.*, 1990, 11, 868-895.

(54) Allen, F. H., The Cambridge Structural Database: A Quarter of a Million Crystal Structures and Rising, *Acta Crystallographica Section B: Structural Science*, 2002, 58, 380.

(55) Kleywegt, G. J.; Bergfors, T.; Senn, H.; Le Motte, P.; Gsell, B.; Shudo, K.; Jones, T. A., Crystal Structures of Cellular Retinoic Acid Binding Proteins I and II in Complex with all-trans retinoic acid and a synthetic retinoid, *Structure*, 1994, 2, 1241.

(56) 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 Research*, 2000, 28, 235.

(57) Humphrey, W.; Dalke, A.; Schulten, K., VMD--Visual Molecular Dynamics, *J. Mol. Graphics*, 1996, 14, 33-38.

(58) Burley, S. K.; Petsko, G. A., Amino-aromatic interactions in proteins, *FEBS*, 1986, 203, 139-143.

(59) Calderone, V.; Folli, C.; Marchesani, A.; Berni, R.; Zanotti, G., Identification and Structural Analysis of a Zebrafish Apo and Holo Cellular Retinol-Binding Protein, *J. Mol. Biol.*, 2002, 321, 527-535.

(60) Li, E.; Norris, A. W., Structure/Function of Cytoplasmic vitamin A-binding proteins, *Annu. Rev. Nutr.*, 1996, 16, 205-234.

(61) Ghyselinck, N. B.; Bavik, C.; Sapin, V.; Mark, M.; Bonnier, D.; Hindelang, C.; Dierich, A.; Nilsson, C. B.; Hakansson, H.; Sauvant, P.; Azais-Braesco, V.; Frasson, M.; Picaud, S.; Chambon, P., Cellular retinol-binding protein I is essential for vitamin A homeostasis, *EMBO Journal*, 1999, 18, 4903-4914.

(62) Bernlohr, D. A.; Simpson, M. A.; Hertzel, A. V.; Banaszak, L. J., Intracellular lipid-binding proteins and their genes, *Annu. Rev. Nutr.*, 1997, 17, 277-303.

(63) Cowan, S. W.; Newcomer, M. E.; Jones, T. A., Crystallographic studies on a family of cellular lipophilic transport proteins. Refinement of P2 myelin protein and the structure determination and refinement of cellular retinol-binding protein in complex with all-trans-retinol., *J. Mol. Biol.*, 1993, 230, 1225.

(64) Winter, N. S.; Bratt, J. M.; Banaszak, L. J., Crystal Structures of Holo and Apo-cellular Retinol-binding Protein II, *J. Mol. Biol.*, 1993, 230, 1247-1259.

(65) Zanotti, G.; Marcello, M.; Malpeli, G.; Folli, C.; Sartori, G.; Berni, R., Crystallographic Studies on Complexes Between Retinoids and Plasma Retinol-Binding Protein, *J. Biol. Chem.*, 1994, 269, 29613-29620.

(66) Cogan, U.; Kopelman, M.; Mokady, S.; Shinitzky, M., Binding Affinities of Retinol and Related Compounds to Retinol Binding Proteins, *Eur. J. Biochem.*, 1976, 65, 71-78.

(67) Berni, R.; Clerici, M.; Malpeli, G.; Cleris, L.; Formelli, F., Retinoids: in vitro interaction with retinol-binding protein and influence on plasma retinol, *FASEB J.*, 1993, 7, 1179-1184.

(68) Zanotti, G.; Malpeli, G.; Berni, R., The Interaction of N-Ethyl Retinamide with Plasma Retinol-binding Protein (RBP) and the Crystal Structure of the Retinoid-RBP Complex at 1.9-A Resolution, *J. Biol. Chem.*, 1993, 268, 24873-24879.

(69) Berni, R.; Formelli, F., In vitro interaction of fenretinide with plasma retinol binding protein and its functional consequences, *Febs*, 1992, 308, 43.

(70) Royant, A.; Nollert, P.; Edman, K.; Neutze, R.; Landau, E. M.; Pebay-Peyroula, E.; Navarro, J., X-ray Structure of Sensory Rhodopsin II at 2.1-A resolution, *Proc. Natl. Acad. Sci. USA*, 2001, 98, 10131-10136.

(71) Spudich, J. L., Variations on a molecular switch: transport and sensory signalling by archael rhodopsins, *Molecular Microbiology*, 1998, 28, 1051-1058.

(72) Wegener, A.-A.; Chizhov, I.; Engelhard, M.; Steinhoff, H.-J., Time-Resolved Detection of Transient Movement of Helix F in Spin-labelled Pharaonis Sensory Rhodopsin II, *J. Mol. Biol.*, 2000, 301, 881-891.

(73) Sasaki, J.; Spudich, J. L., Proton Transport by sensory rhodopsins and its modulaiton by transducer-binding, *Biochimica et Biophysica Acta*, 2000, 1460, 230-239.

(74) Faham, S.; Yang, D.; Bare, E.; Yohannan, S.; Whitelegge, J. P.; Bowie, J. U., Side-Chain Contributions to Membrane Protein Structure and Stability, *J. Mol. Biol.*, 2004, 335, 297-305.

(75) Wang, J. M.; Link, S.; Heyes, C. D.; El-Sayed, M., Comparison of the Dynamics of the Primary Events of Bacteriorhodopsin in Its Trimeric and Monomeric States, *Biophys. J.*, 2002, 83, 1557-1566.

(76) Isenbarger, T. A.; Krebs, M. P., Role of Helix-Helix Interactions in Assembly of the Bacteriorhodopsin Lattice, *Biochemistry*, 1999, 38, 9023-9030.

(77) Scherrer, P.; Matthew, M. K.; Sperling, W.; Stoeckenius, W., Retinal Isomer Ratio in Dark-Adapted Purple Membrane and Bacteriorhodopsin Monomers, *Biochemistry*, 1989, 28, 829-834.

(78) Kovacs, I.; Hollos-Nagy, K.; Varo, G., Dark Adaptation and Spectral Changes in Triton-X-100-treated bacteriorhodopsin, *Journal of Photochemistry and Photobiology B: Biology*, 1995, 27, 21-25.

(79) Edman, K.; Nollert, P.; Royant, A.; Belrhali, H.; Pebay-Peyroula, E.; Hajdu, J.; Neutze, R.; Landau, E. M., High-Resolution X-ray structure of an early intermediate in the bacteriorhodopsin photocycle, *Nature*, 1999, 401, 822-826.

(80) Kamihira, M.; Vosegaard, T.; Mason, A. J.; Straus, S. K.; Nielsen, N. C.; Watts, A., Structural and orientational constraints of bacteriorhodopsin in purple membranes determined by oriented-sample solid-state NMR spectroscopy, *J. Structural Biology*, 2005, 149, 7-16.

(81) Pebay-Peyroula, E.; Rummel, G.; Rosenbusch, J. P.; Landau, E. M., X-ray Structure of Bacteriorhodopsin at 2.5 A from Microcrystals Grown in Lipidic Cubic Phases, *Science*, 1997, 277, 1676-1681.

(82) Luecke, H.; Schobert, B.; Richter, H.-T.; Cartailler, J.-P.; Lanyi, J. K., Structure of Bacteriorhodopsin at 1.55 A Resolution, *J. Mol. Biol.*, 1999, 291, 899-911.

(83) Shibata, M.; Kandori, H., FTIR Studies of Internal Water Molecules in the Schiff Base Region of Bacteriorhodopsin, *Biochemistry*, 2005, 44, 7406-7413.

(84) Kamihira, M.; Watts, A., Functionally Relevant Coupled Dynamic Profile of Bacteriorhodopsin and Lipids in Purple Membranes, *Biochemistry*, 2006, 45, 4304-4313.

(85) Humphrey, W.; Logunov, I.; Schulten, K.; Sheves, M., Molecular Dynamics Study of Bacteriorhodopsin and Artificial Pigments, *Biochemistry*, 1994, 33, 3668-3678.

(86) Rohrig, U. F.; Guidoni, L.; Rothlisberger, U., Early Steps of the Intramolecular Signal Transduction in Rhodopsin Explored by Molecular Dynamics Simulations, *Biochemistry*, 2002, 41, 10799-10809.

(87) Hayashi, S.; Tajkhorshid, E.; Schulten, K., Structural Changes during the Formation of Early Intermediates in the Bacteriorhodopsin Photocycle, *Biophys. J.*, 2002, 83, 1281-1297.

(88) Tajkhorshid, E.; Baudry, J.; Schulten, K.; Suhai, S., Molecular Dynamics Study of the Nature and Origin of Retinal's Twisted Structure in Bacteriorhodopsin, *Biophys. J.*, 2000, 78, 683-693.

(89) Baudry, J.; Tajkhorshid, E.; Molnar, F.; Phillips, J.; Schulten, K., Molecular Dynamics Study of Bacteriorhodopsin and the Purple Membrane, *J. Phys. Chem. B*, 2001, 105, 905-918.

(90) Rajamani, R.; Gao, J., Combined QM/MM study of the opsin shift in bacteriorhodopsin, *J. Comp. Chem.*, 2002, 23, 96-105.

(91) Ulmschneider, M. B.; Tieleman, D. P.; Sansom, M. S. P., Interactions of a Transmembrane Helix and a Membrane: Comparative Simulations of Bacteriorhodopsin Helix A, *J. Phys. Chem. B*, 2004, 108, 10149-10159.

(92) Janovjak, H.; Knaus, H.; Muller, Transmembrane Helices Have Rough Energy Surfaces, *J. Am. Chem. Soc.*, 2006.

(93) Seeber, M.; Fanelli, F.; Paci, E.; Caflisch, A., Sequential unfolding of individual helices of bacteriorhodopsin observed in molecular dynamics simulation of extraction from the purple membrane, *Biophys. J.*, 2006, 91, 3276-3284.

(94) Sato, Y.; Hata, M.; Neya, S.; Hoshino, T., Computational Analysis of the Proton Translocation from Asp96 to Schiff Base in Bacteriorhodopsin, *J. Phys. Chem. B*, 2006, 110, 22804-22812.

(95) Kuhlbrandt, W., Bacteriorhodopsin--the movie, *Nature*, 2000, 406, 569-570.

(96) Papadopoulos, G.; Dencher, N. A.; Zaccai, G.; Buldt, G., Water Molecules and exchangeable hydrogen ions at the active center of bacteriorhodopsin localized by neutron diffraction, *J. Mol. Biol.*, 1990, 214, 15-19.

Chapter 5

Convergence Testing in Biomolecular Simulations

This chapter discusses the statistical tests performed to test for convergence in long and short simulations of DNA and proteins. The purpose of this study was to try to assess convergence in biomolecular systems, and this was done by performing a few very long simulations and many very short simulations. A simple statistical test (potential scale reduction test) was then applied to observe which simulation procedure exhibited better convergence among simulations.

5.1 Potential Scale Reduction Test

The potential scale reduction test, also termed the variance ratio method, was developed by Gelman and Rubin in 1992 [1] and is a popular convergence diagnostic used by statisticians in order to monitor convergence in multiple Markov chain Monto Carlo simulations.[2-4] The most obvious way to compare the parameters measured from each simulation (and see if they converge) is to look at time series plots of a particular parameter for each simulation. The potential scale reduction test (PSRT) is a quantitative way (does not rely on visual inspection of the simulations) to compare a particular parameter and is based on analysis of variance. Approximate convergence is assumed when the variance between the different simulations for a particular parameter is no

larger than the variance seen in that particular parameter from an individual simulation.[5] This approach to monitoring convergence is based on whether or not the simulations have "forgotten" their starting points by comparing multiple simulations to see if they are indistinguishable. With these particular sets of simulations, different properties (18 helical parameters for DNA and 16 geometric parameters for crambin) were monitored for convergence. It is assumed that when the parameters measured have converged by the criteria of the PSRT, that they are sampling the same regions of conformational space at that point and the value of that parameter overlap. In other words, they have "forgotten" their starting points and have become indistinguishable because their values overlap.

For each parameter measured ψ, we have ψ_{ij}, where $j = 1,\ldots,n$ and $i = 1,\ldots,m$. The parameter m is the number of simulations and n is the number of samples used from each simulation. Two variances are calculated, the between-sequence variance (variance between different simulations), B, and the within-sequence variance (variance within an individual simulation), W:

$$B = \frac{n}{m-1}\sum_{i=1}^{m}(\overline{\psi_{i.}} - \overline{\psi})^2 \qquad \text{Eqn. 5.1}$$

$$W = \frac{1}{m}\sum_{i=1}^{m} s_i^2 \qquad \text{Eqn. 5.2}$$

where

$$\overline{\psi_i} = \frac{1}{n}\sum_{j=1}^{n}\psi_{ij} \qquad \text{Eqn. 5.3}$$

$$\overline{\psi} = \frac{1}{m}\sum_{i=1}^{m}\overline{\psi}_i \qquad \text{Eqn. 5.4}$$

$$s_i^2 = \frac{1}{n-1}\sum_{j=1}^{n}(\psi_{ij} - \overline{\psi_i})^2 \qquad \text{Eqn. 5.5}$$

B contains a factor of *n* because it is based on the variance of the within-sequence means, $\overline{\psi_i}$ (each of these is an average of *n* values ψ_{ij}).

From *B* and *W*, two estimates of variance of ψ are calculated. The first, $\hat{v}(\psi)$, is an overestimate of the variance under the assumption that the starting points (in our case the seed for the random number generator for the starting velocities) of the simulations are drawn from a wide range of values (this is termed overdispersion).

$$\hat{v}(\psi) = \frac{n-1}{n}W + \frac{1}{n}B \qquad \text{Eqn. 5.6}$$

For any finite *n*, *W* should underestimate the variance of ψ because the individual simulations have not had time to range over all of the possible values and will thus have less variability. As *n* approaches infinity, both $\hat{v}(\psi)$ and *W* will approach the actual variance, var(ψ), but they do so from opposite directions.

The convergence of the simulations is monitored by estimating the factor by which the overestimate of the variance, $\hat{v}(\psi)$, can be reduced. In other words, it is the

ratio between the upper ($\hat{v}(\psi)$) and lower bounds (W) for the standard deviation of ψ, which is termed the estimated potential scale reduction.

$$\hat{R} = \frac{\hat{v}(\psi)}{W} \qquad \text{Eqn. 5.7}$$

The potential scale reduction is designated as $\hat{R}$ instead R because both the numerator and denominator are estimates of the upper and lower bounds of the variance.

As the simulations converge (and as n goes to infinity), the within-sequence variance exceeds the between-sequence variance (there is more variability within an individual simulation than between the different simulations), and the potential scale reduction will approach 1. The different simulations will now have overlapping values for the parameter being tested. If the potential scale reduction is high, then the value of the parameter being measured is not converging for all simulations, and the values sampled for a particular parameter are not overlapping. In practice, if $\hat{R}$ values are less than or equal 1.2, then the particular property being measured is said to have converged.[5]

5.2 Protein Convergence

5.2.1. Long-Time Simulations. The PSRT was applied to 10 long (45 ns each) simulations of crambin in the crystal environment (see Chapter 2 for simulation methods). A total of 16 geometric parameters were examined, which were the radius of gyration, helix-helix distance (measured from the backbone nitrogen of Asn14 to the backbone nitrogen of Ala27), three beta turn distances (listed in Table 5.1; CA denotes the alpha carbon), and a total of 11 hydrogen bonds in the two α-helices (helix 1 has seven hydrogen bonds and helix 2 has 4 hydrogen bonds; these distances are listed in Table 5.1). The hydrogen bonds were measured from the amide hydrogen (denoted HN)

to the backbone carbonyl oxygen (denote O). Table 5.1 lists the $\hat{R}$ values for each parameter. All snapshots saved were used for the analysis (total of 4500 snapshots for each simulation).

Table 5.1 PSRT results (long-time crambin simulations)

parameter	$\hat{R}$ **long**
radius of gyration	3.3
helix-helix distance	1.4
beta turn 1 (Arg17 CA to Gly20 CA)	1.6
beta turn 2 (Pro41 CA to Tyr44 CA)	1.7
beta turn 3 (Gly42 CA to Ala45 CA)	2.6
helix 1	
hbond 1 (Ser11 HN to Ile7 O)	1.3
hbond 2 (Asn12 HN to Val8 O)	2.1
hbond 3 (Phe13 HN to Ala9 O)	1.2
hbond 4 (Asn14 HN to Arg10 O)	1.4
hbond 5 (Val15 HN to Ser11 O)	1.5
hbond 6 (Cys16 HN to Asn12 O)	1.2
hbond 7 (Arg17 HN to Phe13 O)	1.2
helix 2	
hbond 1 (Ala27 HN to Glu23 O)	1.2
hbond 2 (Thr 28 HN to Ala24 O)	1.1
hbond 3 (Tyr29 HN to Ile25 O)	1.2
hbond 4 (Thr30 HN to Cys26 O)	3.7

Only six of the measured parameters converge for all ten simulations, and all of these parameters are hydrogen bond distances. Two of the time series plots for the hydrogen bond distances that did converge are shown in Figure 5.1. All of the simulations overlap and thus sample approximately the same range of values.

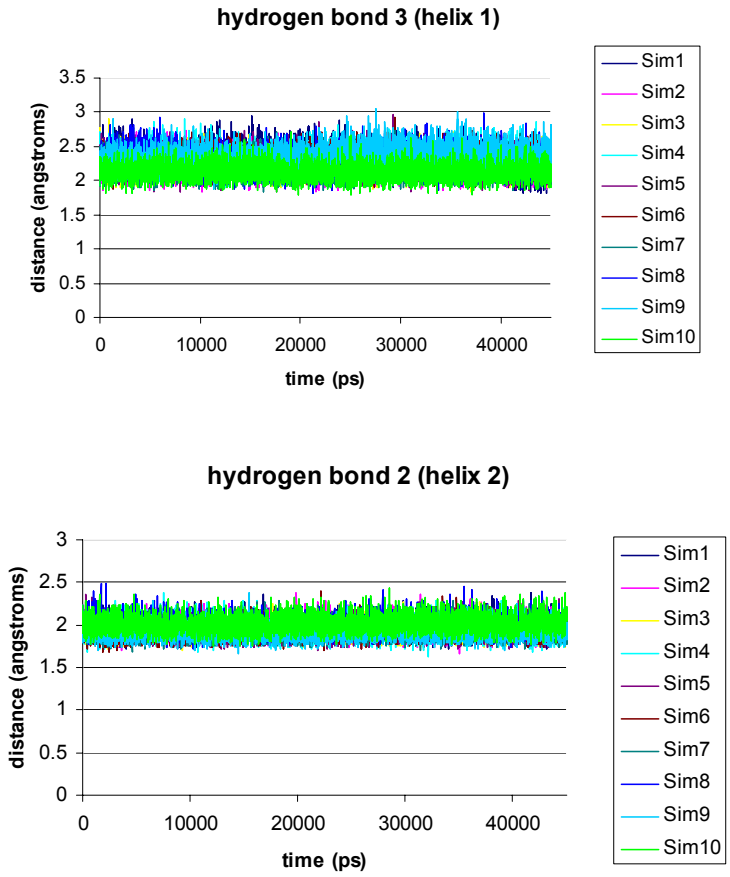


Figure 5.1 Examples of time series plots for hydrogen bonds. Both of these parameters converged according to the PSRT. The different colors represent the different simulations.

The parameters that did not converge were also examined with time series plots to see which simulations were not sampling the same values as the others. Figure 5.2 shows the time series plots for two of the parameters that did not converge. The top plot shows the time series for the radius of gyration, and it is very obvious that the simulations do not sample the same values. For example, the red and dark blue simulations sample lower values than the light blue and green simulations. The bottom plot is for the helix-helix distance, which shows more overlap between the different simulations than the radius of gyration; however, the helix-helix distance still does not converge according to the PSRT.

Even though the simulations exhibit overlap for the helix-helix distance, the PSRT yields non-convergence, which shows the importance of using a quantitative and non-visual test. The other time series plots looked similar.

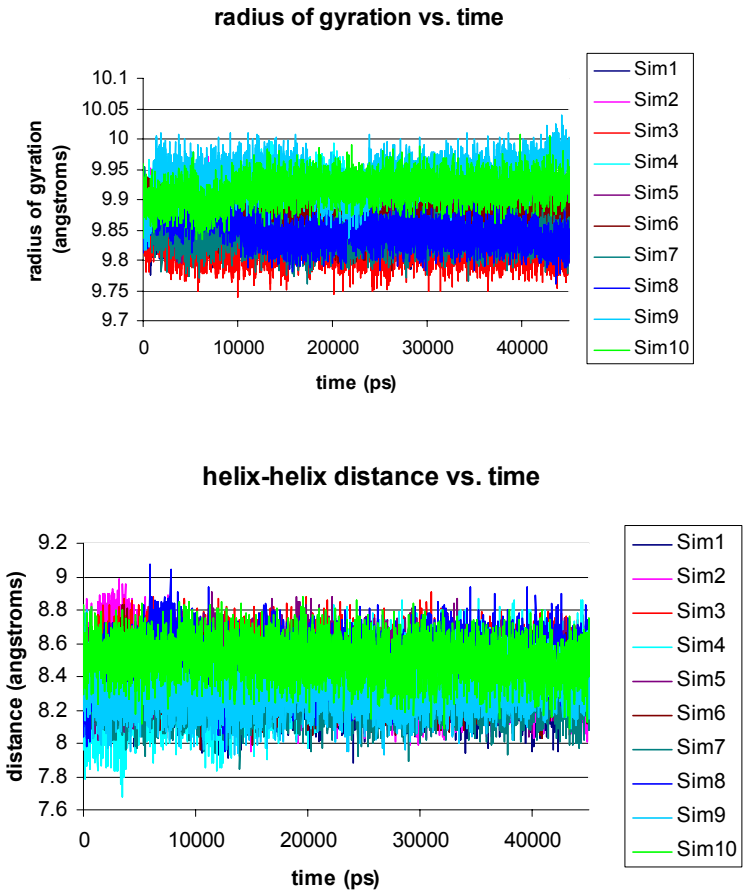


Figure 5.2 Time series plots for radius of gyration (top) and helix-helix distance (bottom). Neither of these parameters converged according to the PSRT. The different colors represent the different simulations.

The time series plots and Table 5.1 illustrate the importance of running multiple simulations. Some simulations sample different ranges of values than other simulations and thus sample different regions of conformational space. If only one simulation had been performed, none of these differences in behavior would have been discovered. The conclusions made from only one simulation would not apply to the rest of the simulations

and thus would not have been representative of the system. One simulation is not going to sample all possible values that a particular parameter can take on, and this would have gone unnoticed if only a single simulation had been run.

Because convergence could not be achieved for the radius of gyration by removing simulations (and because PCA has been used in other studies to assess convergence), a PCA plot was made to observe which simulations did not sample the same regions of conformational space as the others (Figure 5.3, top). None of the simulations appear to be outliers; they all appear to sample in the same regions. The Scree plot (Figure 5.3, bottom) does not show any large contributions to variance. The highest principal component only contributes ~18% of the variance. The PCA plot clearly shows that the simulations sample in the same areas, but according the PSRT, the simulations do not converge. The ten simulations have averages for the radius of gyration in the range of 9.82 to 9.92, which is only a tenth of an angstrom difference. PCA does not reflect this small difference, but the convergence test is sensitive enough that even a small difference is magnified.

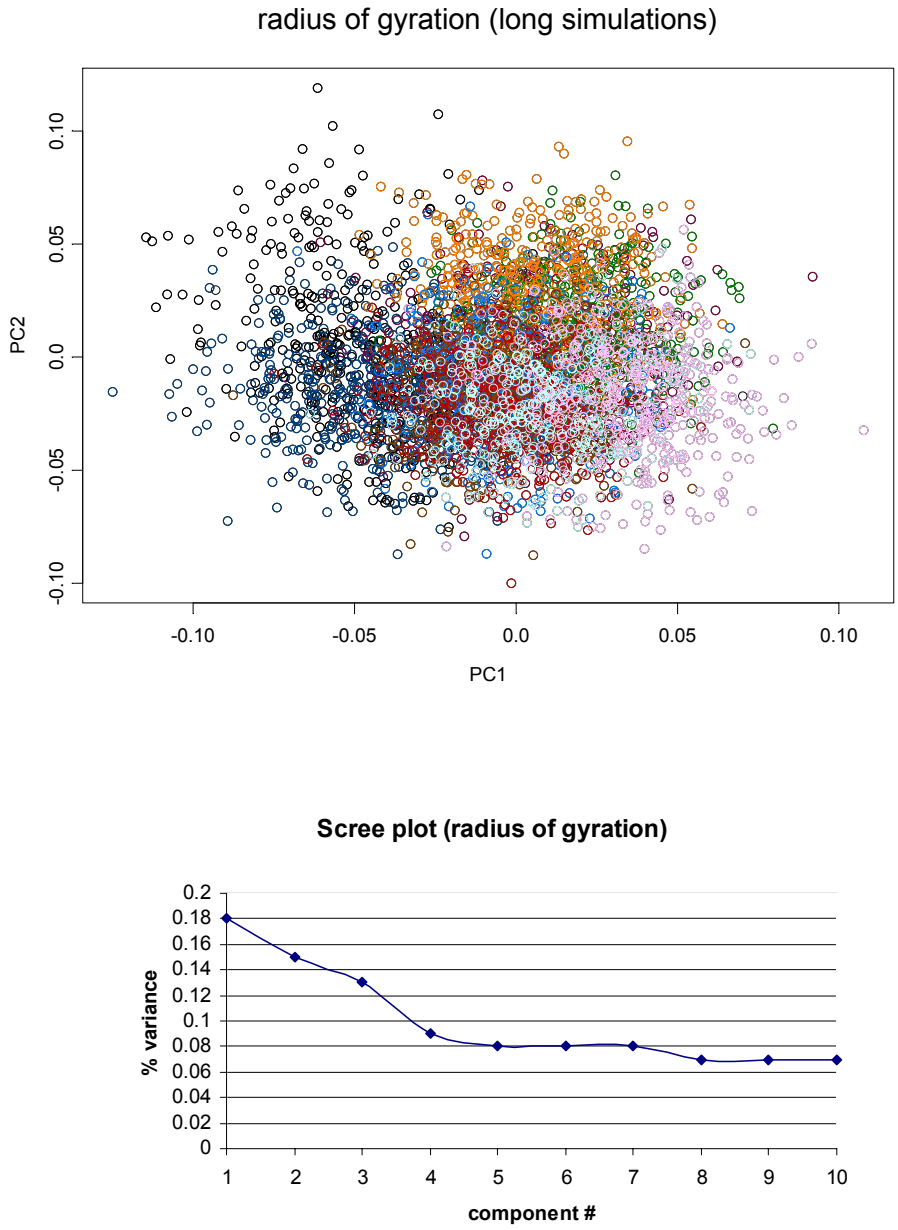


Figure 5.3 PCA plot for radius of gyration for the long-time simulations (top) and Scree plot corresponding to PCA plot (bottom).

5.2.2. Short-time Simulations. A total of 20 short-time (2 ns each) simulations were performed of crambin in the crystal environment (see Chapter 2 for simulation methods). The same geometric parameters were tested in the short simulations, and all saved snapshots were used for analysis (200 snapshots for each simulation). Table 5.2

lists the $\hat{R}$ values for each parameter for the short simulations. As for the long-time simulations, only six parameters converge according to the PSRT, and they are all hydrogen bond distances. The converging hydrogen bond distances are not the same in the long and short simulations, but they have four in common, which are the third and seventh distances in helix 1 and the first and second distances in helix 2. Two of the time series plots for the short-time simulations are in Figure 5.4. The top plot is for the second hydrogen in helix 2, and as for the long-time simulations, the short-time simulations overlap and sample the same range of values. The bottom plot is for the radius of gyration; some simulations sample higher values (the green and yellow simulations) while others (blue simulations) sample lower values.

Table 5.2 PSRT results (short-time crambin simulations)

parameter	$\hat{R}$ **short**
radius of gyration	2.6
helix-helix distance	2.1
beta turn 1 (Arg17 CA to Gly20 CA)	1.8
beta turn 2 (Pro41 CA to Tyr44 CA)	1.7
beta turn 3 (Gly42 CA to Ala45 CA)	2.1
helix 1	
hbond 1 (Ser11 HN to Ile7 O)	1.1
hbond 2 (Asn12 HN to Val8 O)	1.4
hbond 3 (Phe13 HN to Ala9 O)	1.2
hbond 4 (Asn14 HN to Arg10 O)	1.2
hbond 5 (Val15 HN to Ser11 O)	1.3
hbond 6 (Cys16 HN to Asn12 O)	1.3
hbond 7 (Arg17 HN to Phe13 O)	1.2
helix 2	
hbond 1 (Ala27 HN to Glu23 O)	1.1
hbond 2 (Thr 28 HN to Ala24 O)	1.1
hbond 3 (Tyr29 HN to Ile25 O)	1.3
hbond 4 (Thr30 HN to Cys26 O)	3.2

As for the long-time simulations, time series plots were examined for the short-time simulations to see which simulations were the outliers. Figure 5.4 shows the time series plots for hydrogen bond 2 of helix 2 (top) and the radius of gyration (bottom). For the hydrogen bond, the values of all simulations overlap and thus the parameter converges. However, for the radius of gyration, the opposite is observed; the simulations do not have overlapping values, and the variance within individual simulations is greater than that among the different simulations.

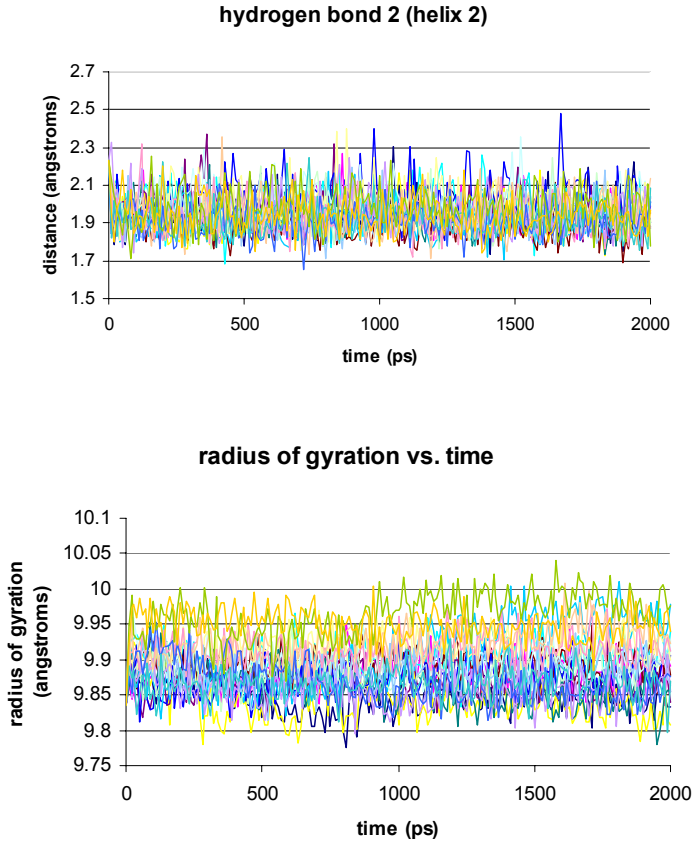


Figure 5.4 Examples of time series plots for short-time simulations. The top plot is for the second hydrogen bond of helix 2, which converged. The bottom plot is for the radius of gyration, which did not converge. The different colors represent the different simulations.

Table 5.3 shows the averages and standard deviations for the parameters for both the long and short sets of simulations.

Table 5.3 Parameter averages for long and short crambin simulations

parameter	**long average (Å)**	**short average (Å)**
radius of gyration	9.88 ± 0.14	9.89 ± 0.19
helix-helix distance	9.40 ± 0.51	8.37 ± 0.84
beta turn 1	5.31 ± 0.67	5.30 ± 1.35
beta turn 2	5.52 ± 0.45	5.51 ± 0.65
beta turn 3	5.08 ± 0.83	5.08 ± 1.06
helix 1		
hydrogen bond 1	1.87 ± 0.27	1.88 ± 0.37
hydrogen bond 2	1.98 ± 0.37	1.96 ± 0.43
hydrogen bond 3	2.25 ± 0.51	2.26 ± 0.71
hydrogen bond 4	1.98 ± 0.33	1.94 ± 0.40
hydrogen bond 5	2.25 ± 0.51	2.24 ± 0.68
hydrogen bond 6	1.93 ± 0.30	1.94 ± 0.44
hydrogen bond 7	2.05 ± 0.33	2.05 ± 0.48
helix 2		
hydrogen bond 1	1.86 ± 0.26	1.87 ± 0.36
hydrogen bond 2	1.96 ± 0.30	1.95 ± 0.44
hydrogen bond 3	2.01 ± 0.35	1.97 ± 0.48
hydrogen bond 4	2.21 ± 0.86	2.31 ± 1.39

Most of the average values calculated from the short and long simulations agree with one another or they are at least within the standard deviations of each other. Both sets of simulations appear to reach the same range of values, whether this happens in a long time or a short time. The long and short simulations appear to converge to the same values, even though the individual simulations within the sets do not converge according to the potential scale reduction test. Because they reach the approximately the same average values and show similar convergence behavior, the short simulations are probably the better choice for convergence when time and effort are considered. 5.3 DNA Convergence

5.3.1. Long-Time Simulations. The PSRT was applied to 4 long (150 ns each) simulations of DNA in the crystal environment (see Chapter 2 for simulation methods). The DNA fragment was a 12 base pair fragment of DNA (5'-CGTAGACTCAGC-3'; no ligands bound) and the same 18 helical parameters that were examined in DNA/polyamide study were analyzed with the PSRT. All 15,000 snapshots saved from each simulation were used for analysis. Table 5.4 lists the $\hat{R}$ values for each helical parameter for the four long simulations.

Table 5.4 PSRT results (long-time DNA simulations; 12 bases)

parameter	$\hat{R}$
shear	1.2
stretch	2.9
stagger	2.6
buckle	1.2
propeller	3.1
opening	2.5
shift	2.0
slide	1.4
rise	2.2
tilt	2.4
roll	4.6
twist	2.6
x-displacement	3
y-displacement	1.8
helical rise	2.6
inclination	5.1
tip	2.8
helical twist	1.1

Almost all of the helical parameters with the exception of three (shear, buckle, and helical twist) do not converge according to the PSRT. Upon examination of time series plots of the parameters, it was obvious that the simulations had a wide range of values for some

parameters and that the simulations appear to sample very different ranges of values. Shown below are the time series plots for propeller and stretch.

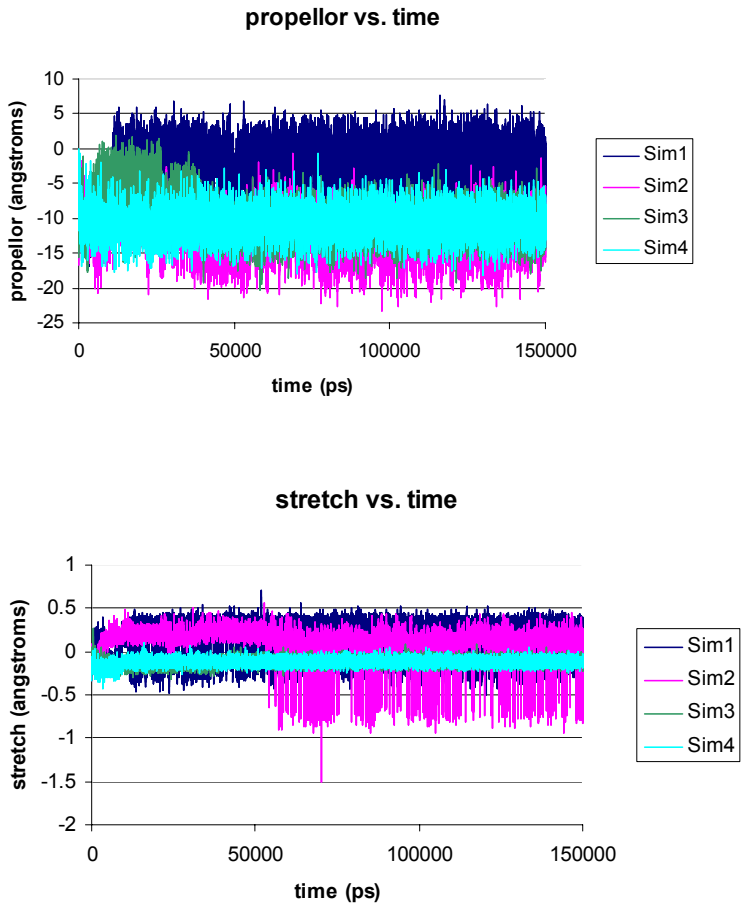
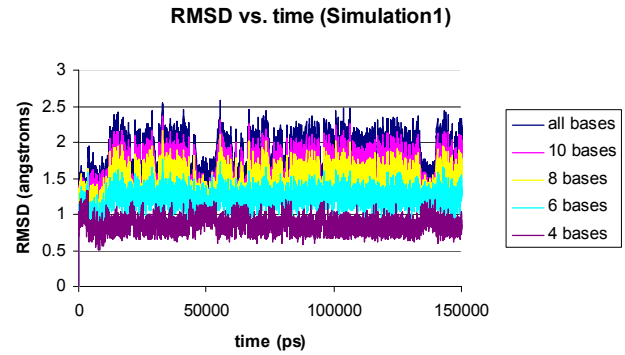
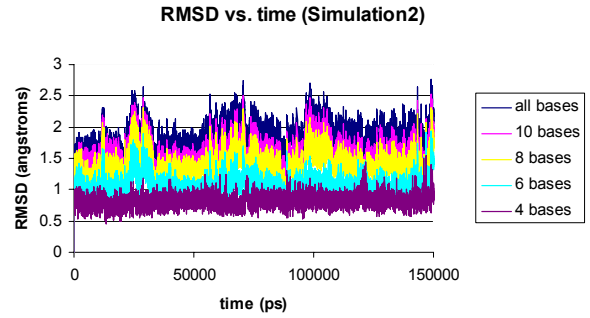


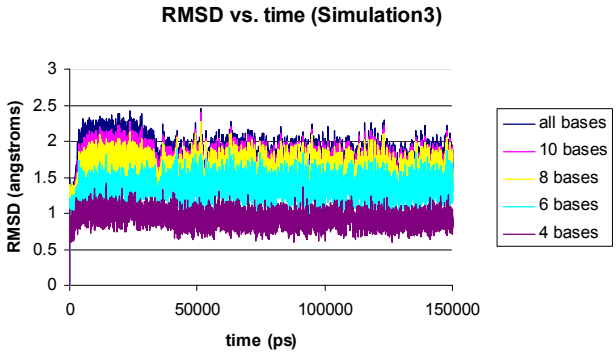
Figure 5.5 Time series plots for propeller and stretch for long DNA simulations. The values plotted are averages over all 12 base pairs of the DNA fragment.

For propeller, the values sampled range from ~5º to -25º, and Simulation 1 has higher values of propeller than the other simulations. Simulation 2 samples the widest range of values with a range of ~20º. Simulation 2 also exhibits the widest range of values for stretch (0.5 to –1.5 Å). Simulation 1 has a large range of value for stretch (0.5 to –0.5 Å), but not as large as Simulation 2. Simulations 3 and 4 appear to sample the same values for stretch and opening, and they do not appear to have as much variation.

Because fraying of the end base pairs is a commonly observed phenomenon in both experimental[6-15] and computational[16-23] studies of DNA, this is most likely occurring here. The terminal base pairs are not as constrained as internal bases, which are held in place by the stacking of the bases.[24] The conformational flexibility possessed by the end base pairs contributes to non-hydrogen bonded base pairs in the Watson-Crick double helix. Fraying of the end bases causes the helical parameter values to sample different values than the bases in the middle of the helix, and when the averages are taken over the 12 base pairs, the large variation between the ends and the middle bases is reflected in the average. In order to see which bases in the four simulations contributed most to fraying, the RMSD of the DNA with different bases was plotted. The RMSD was plotted for all 12 bases, 10 bases (one base on each end removed), 8 bases (two bases on each end removed), 6 bases (three bases on each end removed), and 4 bases (central four bases).







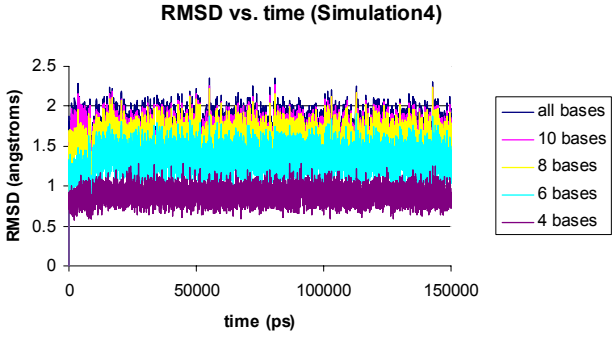


Figure 5.6 RMSD plots for 12, 10, 8, 6, and 4 DNA bases.

From the RMSD plots, it can be seen that there are a lot of fluctuations and bumps (the system is most likely moving between various substates) in the RMSD, in particular for Simulations 1 and 2. Upon removal of one, two, and three base pairs, the RMSD over time is still bumpy. When four base pairs are removed from each end of the helix, most of the fluctuation is eliminated, which means that fraying effects penetrate four base pairs in from the ends of the helix (at least for this DNA fragment). From the RMSD plots, only the middle bases are left for analysis. Because most simulation studies on DNA use only small fragments (10-20 base pairs), many authors eliminate one base from each end for analysis or add an extra base pair "cap" to each end, which is then left out of the analysis. However, the RMSD plots obtained from the four 150 ns simulations described here show that fraying affects not only the end bases, but four bases in from the ends. This RMSD analysis was helpful in deciding which bases should be included in the convergence analysis. When the helical parameters were averaged over the 12 base pairs, the effects of fraying obviously affect the convergence analysis. The abnormal helical parameter values obtained from the end bases are reflected in the averages and thus skew the convergence test. Even though the whole fragment shows large changes in RMSD, the middle 4 pairs do not, and this shows that not only are the middle four bases not affected as much by fraying, but because they are not affected as much by the fraying, their dynamics appear to converge more quickly.

To see if the helical parameters of the middle really did exhibit better convergence than the whole fragment, the PSRT was performed on only the middle four base pairs of the DNA fragment (5'-GATC-3'). The helical parameter values were averaged over the middle four bases. Table 5.5 gives the $\hat{R}$ values.

Table 5.5 PSRT results (long-time DNA simulations; 4 bases)

parameter	$\hat{R}$
shear	1.1
stretch	1.0
stagger	1.0
buckle	1.1
propeller	1.0
opening	1.0
shift	1.2
slide	1.1
rise	1.1
tilt	1.1
roll	1.0
twist	1.0
x-displacement	1.2
y-displacement	1.3
helical rise	1.5
inclination	1.0
tip	1.0
helical twist	1.0

When only the middle four bases are used, only two helical parameters do not converge according the PSRT, which are y-displacement and helical rise. The fraying of the end bases do influence the convergence of the simulations, and even in 150 ns, the dynamics of the end bases do allow the helical parameters to converge. A total of 16 out of 18 the helical parameters for the middle four bases converge; the dynamics of the middle four bases obviously converge more quickly than those of the ends. Thus, the middle bases appear to be the better choice for analysis rather than the whole DNA fragment because they are not influenced by the fraying of the ends. The time series plots below are for the middle four base pairs for twist (which converged) and y-displacement (did not converge).

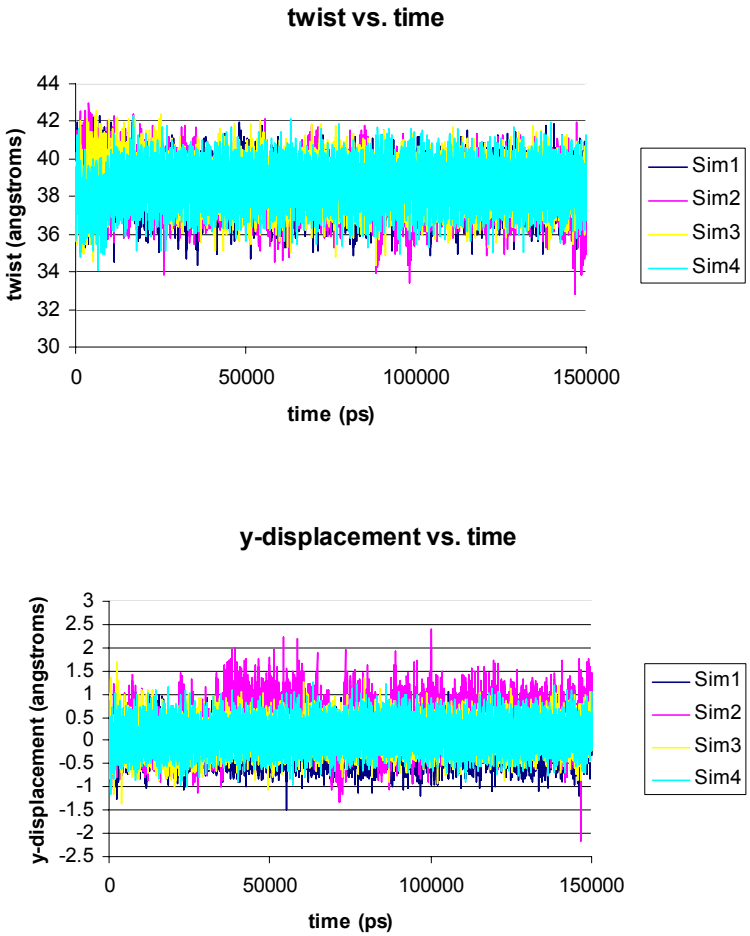


Figure 5.7 Time series plots for twist and y-displacement for the middle four bases of the DNA fragment.

Unlike the time series plots shown in Figure 5.5 for propeller and stretch for all 12 bases, twist does not show the large variation in values for the four simulations (other helical parameter plots were similar). All simulations appear to sample in the same small range of ~34º to 42º and thus exhibit convergence. However, even for twist, which did converge, the blue simulation shows an increase up until 35 ns into the simulation. This simulation showed similar behavior for other parameters, so even up to 35 ns, the system structural properties have not relaxed. Simulations 1,3,and 4 have about the same values

for y-displacement; however, Simulation 2 has more variation in values, and this is most likely why y-displacement does not converge even for the middle four bases.

5.3.2. Short-Time Simulations. From the long-time DNA simulations, the middle four base pairs of this DNA fragment obviously converged more quickly and are the better choice for helical parameter analysis. However, due to the amount of computational time needed to run 150 ns of DNA dynamics, many short simulations were run to see if they yielded similar results with the PSRT. A total of 20 short (2 ns each) simulations were run with same DNA fragment, and the PSRT was applied to all 12 bases and the four middle bases. Table 5.6 gives the results of the PSRT for the short simulations using all 12 bases.

Table 5.6 PSRT results (short-time DNA simulations; 12 bases)

parameter	$\hat{R}$
shear	3.3
stretch	3.1
stagger	1.8
buckle	1.7
propellor	1.7
opening	1.6
shift	2.9
slide	1.9
rise	2.0
tilt	1.4
roll	2.1
twist	1.5
x-displacement	1.5
y-displacement	1.4
helical rise	2.3
inclination	2.0
tip	1.6
helical twist	1.4

The convergence results for the short simulations show that none of the helical parameters converge when an average over all 12 base pairs is used. The time series plots below are for opening and shift for the 12 base pairs (other time series plots looked similar).

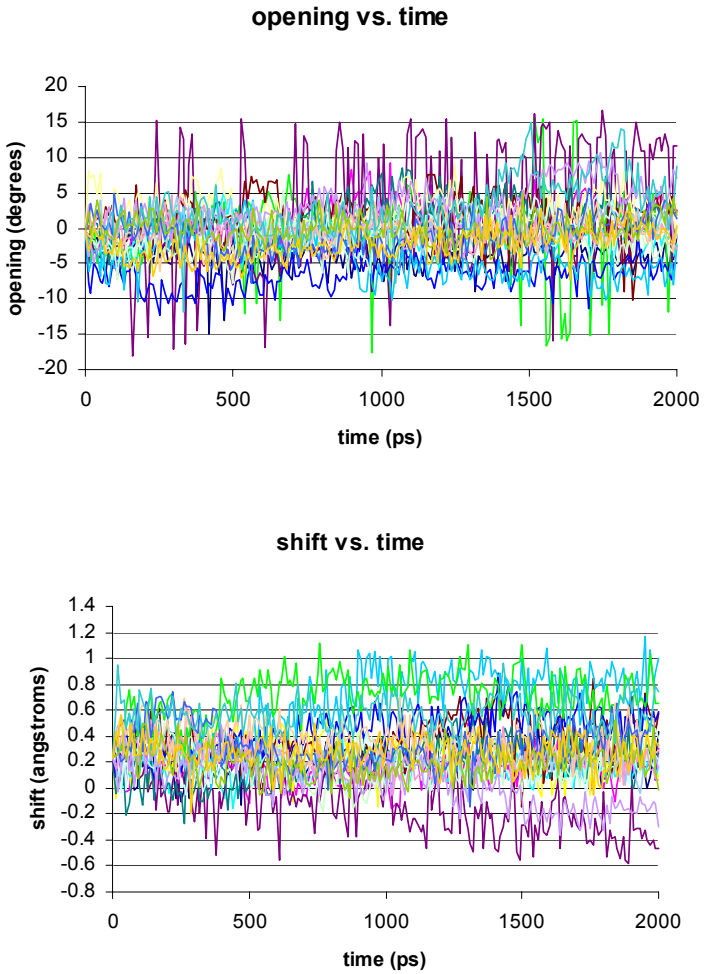


Figure 5.8 Time series plots for opening and shift for all 12 base pairs of the the 20 short DNA simulations.

As for the long simulations, the helical parameter values exhibit large variation (opening) and some simulations simply sample a different range of values than others (shift). Again, this is due to fraying, and since the majority of the helical parameters did not

converge for the 150 ns simulations for all 12 base pairs, it is not surprising that none of the helical parameters converge in two 2 ns when all bases are considered. Even though more simulations were run for 2 ns than 150 ns, each of the 20 simulations will diverge, and the dynamics of the end bases in each of those will need time plenty of time to relax (from the long simulations, they obviously need > 150 ns). From the plot of shift versus time in Figure 5.8, the system needs ~500 ps for the trajectories to exhibit divergence (for shift, the trajectories coincide up to 500 ps, and then start to diverge); therefore, in 2 ns, the systems will not have enough time for the helical parameters to converge for all bases.

As for the long simulations, the convergence of the helical parameters for the middle four bases was examined using the PSRT. Table 5.7 shows the results of the PSRT for the middle four bases of the short DNA simulations.

Table 5.7 PSRT results (short-time DNA simulations; 4 bases)

parameter	$\hat{R}$
shear	1.2
stretch	1.2
stagger	1.1
buckle	1.2
propellor	1.3
opening	1.4
shift	1.2
slide	1.7
rise	1.2
tilt	1.2
roll	1.5
twist	1.6
x-displacement	1.3
y-displacement	1.4
helical rise	1.2
inclination	1.7
tip	1.3
helical twist	1.4

When only the middle four bases are used, more of the helical parameters converge than when all 12 bases are used, so there is some improvement. The plots below are for shift and opening (compare to Figure 5.8 for all 12 bases). Shift converged when only 4 bases are used; opening did not.

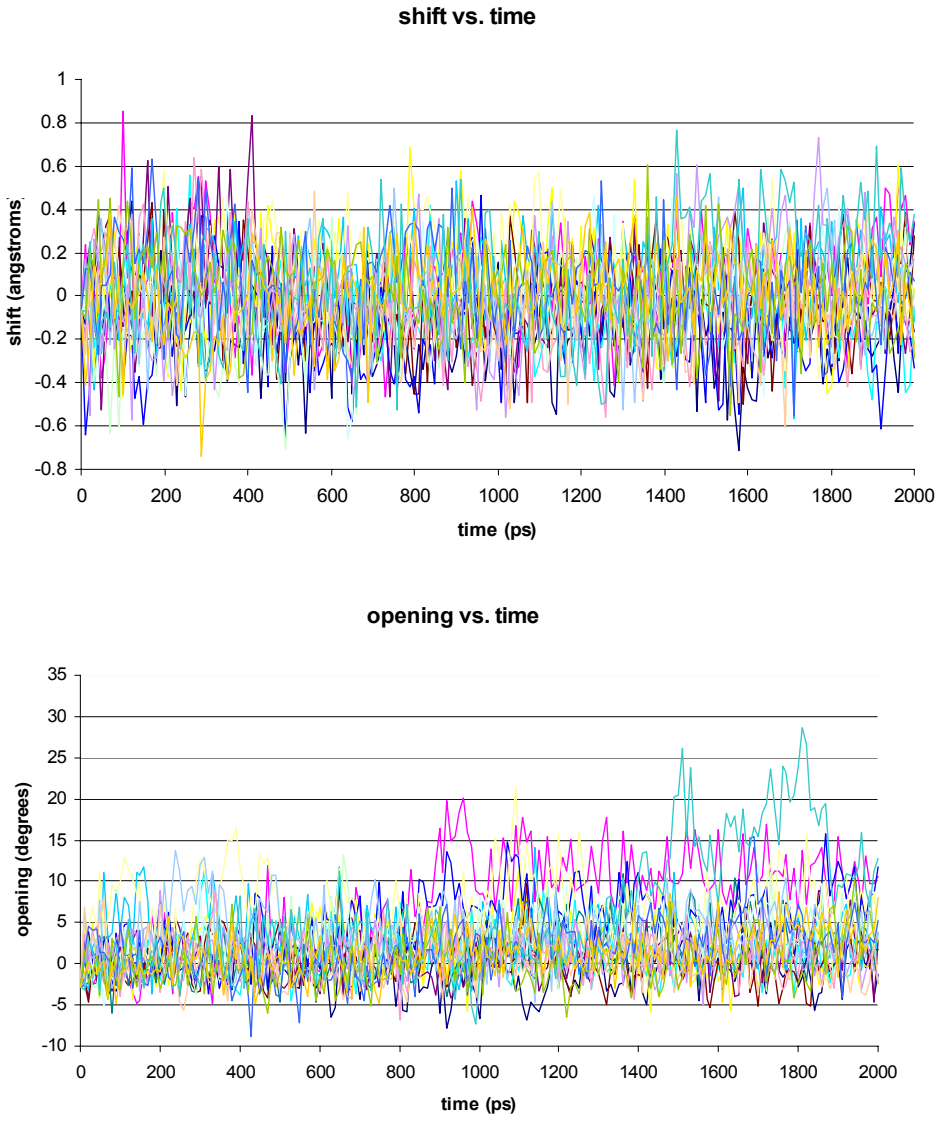


Figure 5.9 Time series plots for shift and opening for the middle four base pairs of the the 20 short DNA simulations.

Shift improved when the middle four bases were used; most of the simulations sample the same range of values for the middle four bases, while some trajectories clearly diverged when all 12 bases were used. For opening, some trajectories (most notably, the pink and blue simulations in Figure 5.9) sample a higher range of values than others even for the middle four bases and thus do not converge.

Even though half of the helical parameters converged when the middle four bases were used, nine parameters still did not converge. The long-time simulations exhibited convergence for all parameters with the exception of two for the middle four bases. The short-time simulations obviously do not show similar convergence behavior to the long-time simulations. A simulation of 2 ns is not enough for most of the helical parameters even for the middle four bases to converge. As with the 12 bases for 2 ns, some of the simulations diverge from the others, and it takes ~500 ps or longer for this to occur (for example, in Figure 5.9, it took 800 ps for opening to diverge). It takes > 500 ps for divergence to be observed, and in 2 ns, the simulations do not have enough time to both diverge and then start to sample the same values again, and therefore, converge. The long simulations, in the case of the DNA, are the better choice. However, since half of the helical parameters did converge for the short simulations, the simulations probably do not need to be run for 150 ns nor do 20 of them need to be run. A smaller number of simulations and a short time period (for example, 10 simulations run for 10 ns each) may yield similar convergence results to the long-time simulations for the middle four bases.

Table 5.8 shows the averages and standard deviations for the parameters for both the long and short sets of DNA simulations.

Table 5.8 Parameter averages for long and short DNA simulations (12 and 4 bases)

parameter	long simulations (12bp)	short simulations (12bp)	long simulations (4bp)	short simulations (4bp)
shear	-0.04 ± 0.36	0.05 ± 1.35	-0.012 ± 0.73	0.05 ± 0.88
stretch	0.03 ± 0.45	-0.10 ± 0.88	-0.11 ± 0.16	-0.11 ± 0.43
stagger	0.04 ± 1.13	0.06 ± 1.04	0.03 ± 0.40	0.05 ± 0.93
buckle	2.33 ± 8.83	-0.04 ± 23.5	5.49 ± 15.1	4.40 ± 25.4
propellor	-8.75 ± 11.32	-8.75 ± 16.6	-20.5 ± 13.7	-18.7 ± 18.8
opening	2.48 ± 6.27	-0.31 ± 18.4	1.89 ± 5.90	2.43 ± 18.6
shift	-0.01 ± 0.38	0.28 ± 1.14	-0.03 ± 0.88	0.01 ± 1.07
slide	0.04 ± 0.34	0.37 ± 1.10	-0.07 ± 0.48	-0.11 ± 1.27
rise	3.31 ± 0.31	3.39 ± 0.60	3.51 ± 0.32	3.44 ± 0.51
tilt	-0.53 ± 5.38	-0.21 ± 8.85	0.37 ± 4.59	0.39 ± 8.80
roll	0.04 ± 8.19	0.88 ± 10.6	0.28 ± 5.04	1.34 ± 10.1
twist	33.6 ± 1.91	34.0 ± 8.79	38.5 ± 7.56	38.1 ± 6.91
x-displacement	-0.25 ± 1.39	0.28 ± 2.68	-0.26 ± 1.01	-0.59 ± 2.12
y-displacement	-0.16 ± 1.01	-0.52 ± 2.67	0.10 ± 1.28	0.07 ± 2.17
helical rise	3.08 ± 0.45	3.18 ± 1.23	3.47 ± 0.30	3.38 ± 0.61
inclination	0.46 ± 15.0	1.54 ± 20.2	0.75 ± 7.62	2.86 ± 19.0
tip	0.46 ± 8.02	-0.15 ± 14.4	-0.67 ± 7.17	-0.69 ± 15.1
helical twist	36.1 ± 4.33	36.4 ± 12.2	39.2 ± 7.51	38.8 ± 6.41

As for the protein simulations, most of the average values calculated from the short and long DNA simulations (12 and 4 bases) agree with one another or they are at least within the standard deviations of each other. Both sets of simulations appear to reach a similar range of values, whether this happens in a long time or a short time. The long and short simulations appear to sample and thus converge to similar values, even though the individual simulations within the sets do not converge according to the potential scale reduction test. The long and short simulations sample the same range of values for both 12 and 4 base pairs; however, they show different convergence behavior with the potential scale reduction test, especially when the middle four bases are used.

5.4 Conclusions

The potential scale reduction test (PSRT) was applied to multiple simulations of two biomolecular systems in order to assess convergence of several geometric properties. The systems examined were a small protein (crambin) and a 12-base pair fragment of DNA, both in the crystal environment. For both systems, a set of a few long simulations and a set of many short simulations were run in order to assess which set would exhibit better convergence with respect to the properties measured.

For the crambin system, 16 geometric parameters were measured for both sets of simulations (see Table 5.1). Ten long (45 ns each) and 20 short (2 ns each) simulations were run for crambin, and the PSRT was applied to the parameters measured for both sets of trajectories. For the long simulations, only six parameters measured converged for the ten simulations, all of which were hydrogen bond distances. The same was observed for the 20 short simulations; six parameters converged, and all were hydrogen bond distances. The distances that converged in the long and short simulations were not all the same distances, but they were all hydrogen bond distances. The other parameters measured (beta turn distances, radius of gyration, and helix-helix distance) are "looser" distances, and the hydrogen bond distances are "tighter"—the former are distances that can vary from simulation to simulation more easily than the hydrogen bond distances. The hydrogen bond distances measured were all in the α-helices of the protein, and the hydrogen bonds in the helices do not deviate greatly unless the α-helices break apart. The α-helices stayed intact during the simulations, so most of the hydrogen bonds stayed around the same values for most simulations. A distance such as the helix-helix distance, however, can fluctuate more among simulations because the helices can drift apart during

the simulations (more or less in different simulations) without the protein unfolding, secondary structure elements breaking down, or any major conformational changes occurring. Because the hydrogen bond distances are "tighter", this is most likely why all the parameters that do converge are hydrogen bond distances; it is more likely that most simulations will sample the same values for hydrogen bonds rather than for a distance like the helix-helix distance. Even for the hydrogen bond distances that do not converge, the distances measured for all simulations are a normal value for a hydrogen bond (~1.8-3.0 Å). Both the long and short simulations exhibit similar convergence behavior when examined with the PSRT. Six parameters converge for both sets, all of which are hydrogen bond distances.

When the averages over all long simulations and all short simulations are examined (Table 5.3), most of the average values calculated from the short and long simulations agree with one another or they are at least within the standard deviations of each other. Both the long and short sets of simulations appear to reach the same range of values, even though the individual simulations within the sets do not converge according to the potential scale reduction test (even in 45 ns). Because they reach the approximately the same average values and show similar convergence behavior, the short simulations are probably the better choice for convergence when time and effort are considered.

For the DNA system, 18 helical parameters were examined. Four long (150 ns each) and 20 short (2 ns each) simulations were run for the DNA, and the PSRT was applied to the parameters measured for both sets of trajectories. The PSRT was applied to the helical parameter averages for all 12 bases of the DNA fragment and also to the

averages over the middle four bases of the fragment. When the PSRT was applied to the averages over all 12 bases of the long simulations, only three helical parameters converged. The lack of convergence is due to fraying of the end base pairs. Fraying causes the end bases to sample a different (and wider) range of values than the bases in the middle of the helix, and when the averages are taken over the 12 base pairs, the large variation between the ends and the middle bases is reflected in the average. The lack of convergence observed for the helical parameters over 12 bases shows agreement with experimentally observed results for DNA relaxation. Local structural relaxation of three different B-DNA 16-mers was measured, and complete relaxation and convergence was not observed even at 40 ns (the experiment did not go past 40 ns because reliability is lost beyond that).[25] However, some relaxation times were only tens of picoseconds; the authors suggest that this broad timescale indicates a complex convergence over a large number of conformational substates.[25] The simple convergence test used here shows that even at 150 ns, convergence of the helical parameters is not achieved for all 12 bases.

When the RMSD was plotted for different parts of the DNA fragment (eliminating 1, 2, 3, and 4 base pairs from the ends), the fluctuations in the RMSD were removed only when four bases on each end were chopped off. The effects of fraying penetrate four base pairs in from the ends, and this left only the middle four bases for analysis. To see whether the helical parameter averages over the middle four bases exhibited better convergence than the helical parameter averages over all 12 bases, the PSRT was applied to the helical parameters obtained from the middle four bases. When only the middle four bases are used, the convergence was improved (only two helical parameters do not converge according the PSRT). The fraying of the end bases did

influence the convergence of the simulations, and even in 150 ns, the dynamics of the end bases do allow the helical parameters to converge. For a total of 16 out of 18 the helical parameters for the middle four bases converge; the dynamics of the middle four bases obviously converge more quickly than those of the ends. Thus, the middle bases are the better choice for analysis rather than the whole DNA fragment (at least for this particular piece of DNA). The RMSD analysis (chopping off successive bases from each end) should be performed on simulated DNA fragments to decide which bases should be used in further analysis of the simulations. The middle four bases are not affected as much by fraying, the values sampled for the helical parameters of these bases do not show as much variation, and thus, their dynamics converge more quickly for the simulations.

The same analysis was carried out on the short-time DNA simulations. When all 12 bases were used for analysis with the PSRT, none of the helical parameters converged. This is again due to fraying, and for some parameters, the system needs ~500 ps for the trajectories to even diverge. In the long simulations, the helical parameters did not even converge over 150 ns; therefore, 2 ns is obviously not enough time for the trajectories to diverge and then start to sample the same values in order to converge. When only the middle four bases were used in the analysis of the short simulations, improvement was observed. Nine of the helical parameters converged, but nine still did not converge. The convergence behavior for the short simulations is obviously different than that of the long simulations, in which only two helical parameters did not converge.

Simulations of 2 ns is not enough for most of the helical parameters even for the middle four bases to converge. As with the 12 bases for 2 ns, some of the simulations diverge from the others, and it takes over 500 ps for divergence to be observed, and in 2

ns, the simulations do not have enough time to both diverge and then converge. Because the long and short simulations do not exhibit similar convergence behavior, the long simulations are the better choice in the case of the DNA.

Other computational studies have reported different results for convergence times for DNA helical parameters. From the results of 10 ns DNA (10 base pairs) simulations, Feig and Pettitt report that structural parameters of the DNA appear to relax around 4 ns into the simulation (for the middle eight bases) and suggest that a timescale of several nanoseconds is probably necessary to obtain convergence of structural parameters, although they did not specify the actual amount of time necessary.[16] From the results of a 60 ns DNA simulation (12 base pairs), Ponomarev, et al. report that helical parameters relax at ~500 ps into the simulation; however, this was for only one base pair step in the middle of the helix and only one simulation was performed.[26] From our multiple short simulations, 500 ps is not enough time for the helical parameters to relax for even the middle four base pairs. In some of the simulations for some parameters, 500 ps appears to be long enough for relaxation, but this does not occur in all simulations. Our results appear to be in better agreement with those of Feig and Pettitt; the structural parameters need a nanosecond timescale for relaxation (> 150 ns for all 12 bases and at least 35 ns for relaxation for the middle four bases). Overall, for each system of interest, there is a balance between the number of simulations and the length of time that they need to be run. Detailed studies for each unique system are required in order to determine what that balance is.

References

(1) Gelman, A.; Rubin, D. B., Inference from Iterative Simulation Using Multiple Sequences, *Statistical Science*, 1992, 7, 457-472.

(2) Cowles, M. K.; Carlin, B. P., Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review, *J. Am. Stat. Assoc.*, 1996, 91, 883-904.

(3) Brooks, S. P.; Roberts, G. O., Convergence Assessment Techniques for Markov Chain Monte Carlo, *Statistics and Computing*, 1998, 8, 319-335.

(4) Chen, M.-H.; Shao, Q.-M.; Ibrahim, J. G., *Monte Carlo Methods in Bayesian Computation*. Springer-Verlag: 2000.

(5) Gelman, A., Inference and monitoring convergence. In *Markov Chain Monte Carlo in Practice*, Chapman and Hall/CRC Press: 1996; 131-141.

(6) Cheung, S.; Arndt, K.; Lu, P., Correlation of lac operator DNA imino proton exchange kinetics with its function, *Proc. Natl. Acad. Sci.*, 1984, 81, 3665-3669.

(7) Chou, S.-H.; Wemmer, D. E.; Hare, D. R.; Reid, B. R., Sequence-Specific Recognition of DNA: NMR studies of the Imino Protons of a Synthetic RNA Polymerase Promoter, *Biochemistry*, 1984, 23, 2257-2262.

(8) Dornberger, U.; Leijon, M.; Fritzsche, H., High Base Pair Opening Rates in Tracts of GC Base Pairs, *J. Biol. Chem.*, 1999, 274, 6957-6962.

(9) Louis, J. M.; Martin, R. G.; Clore, G. M.; Groneborn, A. M., Preparation of Uniformly Isotope-labeled DNA oligonucleotides for NMR Spectroscopy, *J. Biol. Chem.*, 1998, 273, 2374-2378.

(10) Orbons, L. P. M.; van der Marel, G.; van Boom, J. H.; Altona, C., Hairpin and duplex formation of the DNA octamer d(m^5C-G-m^5C-G-T-G-m^5C-G) in solution: An NMR Study, *Nucleic Acids Research*, 1986, 14, 4187.

(11) Pardi, A.; Tinoco, I., Kinetics for Exchange of Imino Protons in Deoxyribonucleic Acid, Ribonucleic Acid, and Hybrid Oligonucleotide Helices, *Biochemistry*, 1982, 21, 4686-4693.

(12) Pasternack, L. B.; Bramham, J.; Mayol, L.; Galeone, A.; Jia, X.; Kearns, D. R., 1H NMR studies of the 5-(hydroxymethyl)-2'-deoxyuridine containing TF1 binding site, *Nucleic Acids Research*, 1996, 24, 2740-2745.

(13) Singh, S.; Patel, P. K.; Hosur, R. V., Structural Polymorphism and Dynamism in the DNA Segment GATCTTCCCCCCGGAA: NMR investigations of Hairpin, Dumbbell, Nicked Duplex, Parallel Strands, and i-Motif, *Biochemistry*, 1997, 36, 13214-13222.

(14) Trotta, E.; Paci, M., Solution Structure of DAPI selectively bound in the minor groove of a DNA T-T mismatch-containing site: NMR and molecular dynamics studies, *Nucleic Acids Research*, 1998, 26, 4706-4713.

(15) Wemmer, D. E.; Benight, A. S., Preparation and Melting of single strand circular DNA loops, *Nucleic Acids Research*, 1985, 13, 8611-8621.

(16) Feig, M.; Pettitt, M., Structural Equilibrium of DNA Represented with Different Force Fields, *Biophysical Journal*, 1998, 75, 134-149.

(17) Beveridge, D. L.; Barreiro, G.; Byun, K. S.; Case, D. A.; Cheatham, T. E., III; Dixit, S. B.; Giudice, E.; Lankas, F.; Lavery, R.; Maddocks, J. H.; Osman, R.; Seibert, E.; Sklenar, H.; Stoll, G.; Thayer, K. M.; Varnai, P.; Young, M. A., Molecular Dynamics Simulations of the 136 Unique Tetranucleotide Sequences of DNA Oligonucleotides. I. Research Design and Results on d(CpG) Steps, *Biophysical Journal*, 2004, 87, 3799-3813.

(18) Chaoui, M.; Derreumaux, S.; Mauffret, O.; Fermandjian, S., An intrinsic curvature towards the minor groove in the cAMP-responsive element DNA found by combined NMR and molecular modelling studies, *Eur. J. Biochem.*, 1999, 259, 877-886.

(19) Cieplak, P.; Cheatham, T. E., III; Kollman, P. A., Molecular Dynamics Simulations Find That 3'-Phosphoramidate Modified DNA Duplexes Undergo a B to A transition and Normal DNA Duplexes Undergo an A to B transition, *J. Am. Chem. Soc.*, 1997, 119, 6722-6730.

(20) Cubero, E.; Sherer, E. C.; Luque, J.; Orozco, M.; Laughton, C. A., Observation of Spontaneous Base Pair Breathing Events in the Molecular Dynamics Simulation of a Difluorotoluene-Containing DNA Oligonucleotide, *J. Am. Chem. Soc.*, 1999, 121, 8653-8654.

(21) Drukker, K.; Wu, G.; Schatz, G. C., Model Simulations of DNA Denaturation Dynamics, *J. Chem. Phys.*, 2001, 114, 579-590.

(22) Guliaev, A. B.; Hang, B.; Singer, B., Structural insights by molecular dynamics simulations into differential repair efficiency for ethano-A versus etheno-A adducts by the human alkylpurine-DNA N-glycosylase, *Nucleic Acids Research*, 2002, 30, 3778-3787.

(23) Guliaev, A. B.; Sagi, J.; Singer, B., Sequence-dependent conformational perturbation in DNA duplexes containing an εA-T mismatch using molecular dynamics simulation, *Carcinogenesis*, 2000, 21, 1727-1736.

(24) Gorenstein, D. G., Conformation and Dynamics of DNA and Protein-DNA Complexes by ^{31}P NMR, *Chemical Reviews*, 1994, 94, 1315-1338.

(25) Brauns, E. B.; Madaras, M. L.; Coleman, R. S.; Murphy, C. J.; Berg, M. E., Complex Local Dynamics in DNA on the Picosecond and Nanosecond Time Scales, *Physical Review Letters*, 2002, 88, 158101-158101-158101-158104.

(26) Ponomarev, S. Y.; Thayer, K. M.; Beveridge, D. L., Ion Motions in Molecular Dynamics Simulations on DNA, *Proc. Natl. Acad. Sci.*, 2004, 101, 14771-14775.

Appendix A

This appendix contains tables of simulated helical parameter values and calculated t-values for the DNA/polyamide simulations. Tables A.1-A.4 contain the helical parameter values and t-values for the DNA/Dervan polyamide simulations (low and high temperature), Tables A.5-A.7 contain the DNA/Dervan polyamide distances, Tables A.9 and A.10 contain the DNA/hydroxpyrrole helical parameters and distances, Tables A.11 and A.12 contain the DNA/netropsin helical parameters and distances, Tables A.13 and A.14 contain the DNA/distamycin helical parameters and distances, Tables A.15 and A.16 contain the helical parameters and distances from the AMBER simulation of the DNA/netropsin complex, and Table A.17 contains the helical parameter t-values for the low and high temperature simulations combined. The polyamide topology and parameters are at the end of this appendix.

Table A.1. Computed and x-ray averages and standard deviations of helical parameters for initial four low temperature (113 K) simulations

Simulation 1				
helical parameter (Å)	**computed avg./sd**	**x-ray value/sd**	**t-test value**	**significant?**
x-displacement	0.29 ± 1.69	0.21 ± 2.21	0.10	no
y-displacment	0.017 ± 1.14	0.05 ± 0.96	0.10	no
shear	-0.038 ± 0.67	0.04 ± 0.40	0.58	no
stretch	-0.15 ± 0.11	-0.15 ± 0.18	0	no
stagger	-0.0095 ± 0.25	0.23 ± 0.13	5.47	yes
shift	0.12 ± 0.87	0.01 ± 0.95	0.36	no
slide	0.98 ± 0.66	0.82 ± 0.73	0.67	no
rise	3.33 ± 0.22	3.31 ± 0.17	0.42	no
helical rise	3.42 ± 0.34	3.29 ± 0.18	2.28	yes
helical parameter (°)				
inclination	12.27 ± 9.71	9.19 ± 16.06	0.57	no
tip	-0.09 ± 7.12	-0.57 ± 6.57	0.21	no
buckle	4.46 ± 6.38	-1.14 ± 5.39	3.11	no
propellor	-11.25 ± 7.51	-13.41 ± 9.86	0.66	no

Table A.1 (cont'd.)

opening	0.13 ± 5.78	-0.86 ± 6.05	0.49	no
tilt	0.10 ± 4.24	0.35 ± 4.17	0.18	no
roll	6.89 ± 4.99	4.63 ± 8.70	0.78	no
twist	34.4 ± 7.02	34.99 ± 6.41	0.26	no
helical twist	35.72 ± 7.53	36.62 ± 4.96	0.54	no
Simulation 2				
helical parameter (Å)	**computed avg./sd**	**x-ray value/sd**	**t-test value**	**significant?**
x-displacement	0.076 ± 2.04	0.21 ± 2.21	0.18	no
y-displacment	-0.053 ± 1.12	0.05 ± 0.96	0.32	no
shear	-0.029 ± 0.46	0.04 ± 0.40	0.51	no
stretch	-0.16 ± 0.12	-0.15 ± 0.18	0.13	no
stagger	-0.21 ± 0.29	0.23 ± 0.13	9.91	yes
shift	-0.51 ± 0.75	0.01 ± 0.95	0.19	no
slide	1.02 ± 0.69	0.82 ± 0.73	0.81	no
rise	3.29 ± 0.35	3.31 ± 0.17	0.37	no
helical rise	3.32 ± 0.48	3.29 ± 0.18	0.48	no
helical parameter (°)				
inclination	14.18 ± 12.87	9.19 ± 16.06	0.93	no
tip	0.91 ± 9.74	-0.57 ± 6.57	0.67	no
buckle	0.31 ± 9.23	-1.14 ± 5.39	0.80	no
propellor	-14.52 ± 4.96	-13.41 ± 9.86	0.34	no
opening	-1.76 ± 4.54	-0.86 ± 6.05	0.45	no
tilt	-0.46 ± 5.24	0.35 ± 4.17	0.58	no
roll	7.45 ± 6.29	4.63 ± 8.70	0.97	no
twist	33.65 ± 8.41	34.99 ± 6.41	0.62	no
helical twist	35.55 ± 7.09	36.62 ± 4.96	0.64	no
Simulation 3				
helical parameter (Å)	**computed avg./sd**	**x-ray value/sd**	**t-test value**	**significant?**
x-displacement	0.36 ± 2.31	0.21 ± 2.21	0.21	no
y-displacment	0.26 ± 1.13	0.05 ± 0.96	0.65	no
shear	-0.010 ± 0.45	0.04 ± 0.40	0.38	no
stretch	-0.16 ± 0.10	-0.15 ± 0.18	0.23	no
stagger	-0.02 ± 0.37	0.23 ± 0.13	5.55	yes
shift	-0.11 ± 0.92	0.01 ± 0.95	0.38	no
slide	1.22 ± 0.95	0.82 ± 0.73	1.63	no
rise	3.33 ± 0.36	3.31 ± 0.17	0.27	no
helical rise	3.39 ± 0.65	3.29 ± 0.18	1.60	no
helical parameter (°)				
inclination	13.17 ± 16.59	9.19 ± 16.06	0.74	no
tip	-0.17 ± 8.83	-0.57 ± 6.57	0.18	no
buckle	0.62 ± 11.95	-1.14 ± 5.39	0.96	no
propellor	-12.86 ± 5.67	-13.41 ± 9.86	0.17	no
opening	-1.27 ± 3.52	-0.86 ± 6.05	0.20	no
tilt	0.25 ± 5.02	0.35 ± 4.17	0.07	no

Table A.1 (cont'd.)

roll	6.98 ± 9.09	4.63 ± 8.70	0.81	no
twist	33.95 ± 7.80	34.99 ± 6.41	0.49	no
helical twist	36.23 ± 6.33	36.62 ± 4.96	0.24	no
Simulation 4				
helical parameter (Å)	**computed avg./sd**	**x-ray value/sd**	**t-test value**	**significant?**
x-displacement	0.017 ± 1.96	0.21 ± 2.21	0.26	no
y-displacment	0.066 ± 1.38	0.05 ± 0.96	0.05	no
shear	0.081 ± 0.51	0.04 ± 0.40	0.32	no
stretch	-0.15 ± 0.13	-0.15 ± 0.18	0.09	no
stagger	-0.08 ± 0.27	0.23 ± 0.13	7.36	yes
shift	-0.042 ± 0.93	0.01 ± 0.95	0.16	no
slide	0.708 ± 0.82	0.82 ± 0.73	0.46	no
rise	3.35 ± 0.30	3.31 ± 0.17	0.68	no
helical rise	3.35 ± 0.41	3.29 ± 0.18	0.92	no
helical parameter (°)				
inclination	10.12 ± 9.38	9.19 ± 16.06	0.17	no
tip	0.74 ± 8.30	-0.57 ± 6.57	0.59	no
buckle	1.39 ± 7.90	-1.14 ± 5.39	1.47	no
propellor	-11.99 ± 8.18	-13.41 ± 9.86	0.45	no
opening	-2.58 ± 4.86	-0.86 ± 6.05	0.90	no
tilt	-0.41 ± 4.52	0.35 ± 4.17	0.54	no
roll	5.36 ± 4.83	4.63 ± 8.70	0.25	no
twist	34.03 ± 7.33	34.99 ± 6.41	0.45	no
helical twist	35.12 ± 6.67	36.62 ± 4.96	0.90	no

*helical rise and helical twist are parameters which describe the regularity of the DNA helix, while rise and twist describe the geometry of a base pair step from a local perspective. For more details, see Lu, X.J. and Olson, W.K, *Nuc. Acids Res*. 2003, **31**, 5108.

Table A.2. Computed and x-ray averages and standard deviations of helical parameters for initial three high temperature (300 K) simulations

Simulation 1				
helical parameter (Å)	**computed avg./sd**	**x-ray value/sd**	**t-test value**	**significant?**
x-displacement	0.0081 ± 1.92	0.21 ± 2.21	0.27	no
y-displacment	0.0041 ± 1.46	0.05 ± 0.96	0.14	no
shear	-0.033 ± 0.60	0.04 ± 0.40	0.54	no
stretch	-0.081 ± 0.19	-0.15 ± 0.18	1.15	no
stagger	-0.198 ± 0.56	0.23 ± 0.13	9.15	yes
shift	-0.0072 ± 0.98	0.01 ± 0.95	0.05	no
slide	0.84 ± 0.79	0.82 ± 0.73	0.08	no
rise	3.28 ± 0.41	3.31 ± 0.17	0.46	no
helical rise	3.29 ± 0.46	3.29 ± 0.18	0.02	yes
Helical parameter (°)				
inclination	13.28 ± 13.4	9.19 ± 16.06	0.76	no

Table A.2 (cont'd).

tip	0.08 ± 12.2	-0.57 ± 6.57	0.22	no
buckle	0.77 ± 17.8	-1.14 ± 5.39	1.02	no
propellor	-12.17 ± 9.36	-13.41 ± 9.86	0.37	no
opening	-1.34 ± 8.39	-0.86 ± 6.05	0.24	no
tilt	0.29 ± 6.72	0.35 ± 4.17	0.04	no
roll	7.45 ± 7.68	4.63 ± 8.70	0.97	no
twist	33.8 ± 8.07	34.99 ± 6.41	0.54	no
helical twist	36.1± 7.28	36.62 ± 4.96	0.32	no
Simulation 2				
helical parameter (Å)	**computed avg./sd**	**x-ray value/sd**	**t-test value**	**significant?**
x-displacement	-0.571 ± 2.45	0.21 ± 2.21	1.05	no
y-displacment	0.034 ± 1.69	0.05 ± 0.96	0.05	no
shear	-0.0029 ± 0.64	0.04 ± 0.40	0.34	no
stretch	-0.099 ± 0.24	-0.15 ± 0.18	0.89	no
stagger	-0.123 ± 0.49	0.23 ± 0.13	8.04	yes
shift	-0.063 ± 1.18	0.01 ± 0.95	0.23	no
slide	0.69 ± 0.89	0.82 ± 0.73	0.55	no
rise	3.33 ± 0.37	3.31 ± 0.17	0.26	no
helical rise	3.27 ± 0.53	3.29 ± 0.18	0.31	no
Helical parameter (°)				
inclination	16.81 ± 15.9	9.19 ± 16.06	1.42	no
tip	0.71 ± 13.1	-0.57 ± 6.57	0.58	no
buckle	1.37 ± 12.3	-1.14 ± 5.39	1.44	no
propellor	-7.85 ± 10.5	-13.41 ± 9.86	1.77	no
opening	1.83 ± 9.76	-0.86 ± 6.05	1.39	no
tilt	-0.29 ± 7.33	0.35 ± 4.17	0.46	no
roll	9.33 ± 8.46	4.63 ± 8.70	1.61	no
twist	31.9 ± 7.46	34.99 ± 6.41	1.41	no
helical twist	35.4 ± 6.28	36.62 ± 4.96	0.75	no
Simulation 3				
helical parameter (Å)	**computed avg./sd**	**x-ray value/sd**	**t-test value**	**significant?**
x-displacement	-0.507 ± 2.04	0.21 ± 2.21	0.97	no
y-displacment	0.158 ± 1.43	0.05 ± 0.96	0.33	no
shear	0.0064 ± 0.59	0.04 ± 0.40	0.26	no
stretch	-0.107 ± 0.19	-0.15 ± 0.18	0.76	no
stagger	-0.164 ± 0.46	0.23 ± 0.13	9.04	yes
shift	-0.073 ± 1.14	0.01 ± 0.95	0.26	no
slide	0.68 ± 0.82	0.82 ± 0.73	0.58	no
rise	3.34 ± 0.35	3.31 ± 0.17	0.27	no
helical rise	3.34 ± 0.49	3.29 ± 0.18	0.77	no
Helical parameter (°)				
inclination	15.88 ± 13.3	9.19 ± 16.06	1.24	no
tip	-0.33 ± 11.9	-0.57 ± 6.57	0.11	no
buckle	2.79 ± 12.0	-1.14 ± 5.39	2.25	no

Table A.2 (cont'd).

propellor	-9.87 ± 9.11	-13.41 ± 9.86	1.13	no
opening	0.148 ± 6.79	-0.86 ± 6.05	0.52	no
tilt	0.259 ± 7.14	0.35 ± 4.17	0.06	no
roll	9.16 ± 7.42	4.63 ± 8.70	1.56	no
twist	33.5 ± 6.59	34.99 ± 6.41	0.67	no
helical twist	36.2 ± 5.58	36.62 ± 4.96	0.23	no

Table A.3. Computed and x-ray averages and standard deviations of helical parameters for additional three high temperature (300 K) simulations

Simulation 1				
helical parameter (Å)	**computed avg./sd**	**x-ray value/sd**	**t-test value**	**significant?**
x-displacement	-0.78 ± 2.54	0.21 ± 2.21	1.22	no
y-displacment	0.14 ± 1.66	0.05 ± 0.96	0.17	no
shear	0.02 ± 0.70	0.04 ± 0.40	0.08	no
stretch	-0.08 ± 0.25	-0.15 ± 0.18	0.91	no
stagger	-0.05 ± 0.52	0.23 ± 0.13	1.71	no
shift	-0.05 ± 1.19	0.01 ± 0.95	0.14	no
slide	0.63 ± 0.84	0.82 ± 0.73	0.70	no
rise	3.33 ± 0.39	3.31 ± 0.17	0.15	no
helical rise	3.24 ± 0.53	3.29 ± 0.18	0.27	no
Helical parameter (°)				
inclination	17.3 ± 17.2	9.19 ± 16.06	1.49	no
tip	-0.64 ± 12.3	-0.57 ± 6.57	0.02	no
buckle	3.61 ± 15.8	-1.14 ± 5.39	0.95	no
propellor	-8.43 ± 10.2	-13.41 ± 9.86	1.54	no
opening	2.64 ± 9.85	-0.86 ± 6.05	1.12	no
tilt	0.03 ± 6.97	0.35 ± 4.17	0.15	no
roll	9.21 ± 8.91	4.63 ± 8.70	1.62	no
twist	32.3 ± 8.59	34.99 ± 6.41	1.00	no
helical twist	35.5 ± 7.08	36.62 ± 4.96	0.51	no
Simulation 2				
helical parameter (Å)	**computed avg./sd**	**x-ray value/sd**	**t-test value**	**significant?**
x-displacement	-0.23 ± 2.29	0.21 ± 2.21	0.61	no
y-displacment	-0.11 ± 1.57	0.05 ± 0.96	0.33	no
shear	-0.003 ± 0.64	0.04 ± 0.40	0.21	no
stretch	-0.11 ± 0.15	-0.15 ± 0.18	0.80	no
stagger	-0.006 ± 0.54	0.23 ± 0.13	1.37	no
shift	0.08 ± 1.13	0.01 ± 0.95	0.21	no
slide	0.69 ± 0.72	0.82 ± 0.73	0.58	no
rise	3.32 ± 0.32	3.31 ± 0.17	0.12	no
helical rise	3.31 ± 0.44	3.29 ± 0.18	0.12	no
Helical parameter (°)				
inclination	11.7 ± 14.4	9.19 ± 16.06	0.54	no
tip	0.006 ± 12.4	-0.57 ± 6.57	0.15	no

Table A.3 (cont'd.)

buckle	1.34 ± 14.2	-1.14 ± 5.39	0.55	no
propellor	-10.4 ± 9.22	-13.41 ± 9.86	1.02	no
opening	0.03 ± 6.17	-0.86 ± 6.05	0.45	no
tilt	0.17 ± 7.02	0.35 ± 4.17	0.08	no
roll	5.98 ± 7.07	4.63 ± 8.70	0.60	no
twist	33.4 ± 8.60	34.99 ± 6.41	0.60	no
helical twist	35.4 ± 7.39	36.62 ± 4.96	0.52	no
Simulation 3				
helical parameter (Å)	**computed avg./sd**	**x-ray value/sd**	**t-test value**	**significant?**
x-displacement	-0.009 ± 2.52	0.21 ± 2.21	0.28	no
y-displacment	-0.07 ± 2.08	0.05 ± 0.96	0.18	no
shear	-0.007 ± 0.66	0.04 ± 0.40	0.23	no
stretch	-0.04 ± 0.52	-0.15 ± 0.18	0.70	no
stagger	-0.01 ± 0.53	0.23 ± 0.13	1.46	no
shift	-0.009 ± 1.39	0.01 ± 0.95	0.04	no
slide	1.11 ± 1.01	0.82 ± 0.73	0.91	no
rise	3.36 ± 0.38	3.31 ± 0.17	0.45	no
helical rise	3.44 ± 0.59	3.29 ± 0.18	0.82	no
Helical parameter (°)				
inclination	17.6 ± 17.9	9.19 ± 16.06	1.48	no
tip	1.25 ± 14.1	-0.57 ± 6.57	0.41	no
buckle	-0.62 ± 12.8	-1.14 ± 5.39	0.13	no
propellor	-6.42 ± 12.0	-13.41 ± 9.86	1.83	no
opening	4.09 ± 14.1	-0.86 ± 6.05	1.10	no
tilt	-0.27 ± 7.64	0.35 ± 4.17	0.26	no
roll	9.46 ± 9.52	4.63 ± 8.70	1.59	no
twist	31.5 ± 8.14	34.99 ± 6.41	1.36	no
helical twist	35.1 ± 6.43	36.62 ± 4.96	0.73	no

Table A.4. Computed and x-ray averages and standard deviations of helical parameters for additional four low temperature (113 K) simulations

Simulation 1				
helical parameter (Å)	**computed avg./sd**	**x-ray value/sd**	**t-test value**	**significant?**
x-displacement	0.09 ± 1.68	0.21 ± 2.21	0.21	no
y-displacment	0.03 ± 1.18	0.05 ± 0.96	0.05	no
shear	-0.02 ± 0.56	0.04 ± 0.40	0.35	no
stretch	-0.18 ± 0.11	-0.15 ± 0.18	0.84	no
stagger	-0.02 ± 0.26	0.23 ± 0.13	3.07	yes
shift	0.005 ± 0.90	0.01 ± 0.95	0.02	no
slide	0.87 ± 0.63	0.82 ± 0.73	0.23	no
rise	3.36 ± 0.23	3.31 ± 0.17	0.63	no
helical rise	3.42 ± 0.33	3.29 ± 0.18	1.27	no

Table A.4 (cont'd.)

Helical parameter (°)				
inclination	12.4 ± 9.9	9.19 ± 16.06	1.00	no
tip	-0.33 ± 6.72	-0.57 ± 6.57	0.11	no
buckle	3.21 ± 6.37	-1.14 ± 5.39	2.15	yes
propellor	-10.9 ± 9.01	-13.41 ± 9.86	0.88	no
opening	-0.95 ± 4.11	-0.86 ± 6.05	0.07	no
tilt	0.24 ± 3.99	0.35 ± 4.17	0.09	no
roll	7.08 ± 5.41	4.63 ± 8.70	1.42	no
twist	34.6 ± 6.74	34.99 ± 6.41	0.18	no
helical twist	35.9 ± 6.16	36.62 ± 4.96	0.34	no
Simulation 2				
helical parameter (Å)	**computed avg./sd**	**x-ray value/sd**	**t-test value**	**significant?**
x-displacement	0.23 ± 1.88	0.21 ± 2.21	0.04	no
y-displacment	-0.04 ± 1.12	0.05 ± 0.96	0.26	no
shear	-0.07 ± 0.71	0.04 ± 0.40	0.50	no
stretch	-0.15 ± 0.45	-0.15 ± 0.18	0.01	no
stagger	-0.002 ± 0.44	0.23 ± 0.13	1.68	no
shift	0.06 ± 0.90	0.01 ± 0.95	0.18	no
slide	0.94 ± 0.72	0.82 ± 0.73	0.51	no
rise	3.33 ± 0.26	3.31 ± 0.17	0.24	no
helical rise	3.39 ± 0.34	3.29 ± 0.18	0.96	no
Helical parameter (°)				
inclination	11.9 ± 10.3	9.19 ± 16.06	0.83	no
tip	-0.52 ± 7.94	-0.57 ± 6.57	0.02	no
buckle	6.53 ± 6.84	-1.14 ± 5.39	3.54	yes
propellor	-11.1 ± 7.98	-13.41 ± 9.86	0.93	no
opening	-1.31 ± 9.72	-0.86 ± 6.05	0.15	no
tilt	0.47 ± 4.49	0.35 ± 4.17	0.08	no
roll	6.56 ± 6.82	4.63 ± 8.70	0.89	no
twist	34.4 ± 9.92	34.99 ± 6.41	0.17	no
helical twist	35.7 ± 9.58	36.62 ± 4.96	0.30	no
Simulation 3				
helical parameter (Å)	**computed avg./sd**	**x-ray value/sd**	**t-test value**	**significant?**
x-displacement	0.33 ± 1.95	0.21 ± 2.21	0.19	no
y-displacment	0.13 ± 1.09	0.05 ± 0.96	0.23	no
shear	0.05 ± 0.78	0.04 ± 0.40	0.06	no
stretch	-0.14 ± 0.09	-0.15 ± 0.18	0.18	no
stagger	-0.03 ± 0.25	0.23 ± 0.13	3.35	yes
shift	-0.10 ± 0.88	0.01 ± 0.95	0.39	no
slide	1.03 ± 0.71	0.82 ± 0.73	0.93	no
rise	3.36 ± 0.23	3.31 ± 0.17	0.66	no
helical rise	3.44 ± 0.34	3.29 ± 0.18	1.36	no
Helical parameter (°)				
inclination	12.4 ± 11.3	9.19 ± 16.06	0.88	no

Table A.4 (cont'd.)

tip	0.46 ± 7.69	-0.57 ± 6.57	0.42	no
buckle	6.67 ± 6.97	-1.14 ± 5.39	3.53	yes
propellor	-11.4 ± 9.16	-13.41 ± 9.86	0.70	no
opening	-0.004 ± 7.21	-0.86 ± 6.05	0.37	no
tilt	-0.16 ± 4.86	0.35 ± 4.17	0.33	no
roll	6.96 ± 5.65	4.63 ± 8.70	1.29	no
twist	34.8 ± 7.35	34.99 ± 6.41	0.08	no
helical twist	36.3 ± 6.55	36.62 ± 4.96	0.16	no
Simulation 4				
helical parameter (Å)	**computed avg./sd**	**x-ray value/sd**	**t-test value**	**significant?**
x-displacement	0.21 ± 1.73	0.21 ± 2.21	0.002	no
y-displacment	-0.03 ± 1.06	0.05 ± 0.96	0.25	no
shear	-0.08 ± 0.56	0.04 ± 0.40	0.69	no
stretch	-0.17 ± 0.11	-0.15 ± 0.18	0.46	no
stagger	-0.03 ± 0.26	0.23 ± 0.13	3.06	yes
shift	0.03 ± 0.82	0.01 ± 0.95	0.09	no
slide	0.89 ± 0.71	0.82 ± 0.73	0.35	no
rise	3.36 ± 0.22	3.31 ± 0.17	0.76	no
helical rise	3.43 ± 0.35	3.29 ± 0.18	1.27	no
Helical parameter (°)				
inclination	11.7 ± 9.7	9.19 ± 16.06	0.80	no
tip	0.005 ± 6.91	-0.57 ± 6.57	0.26	no
buckle	4.64 ± 6.85	-1.14 ± 5.39	2.66	yes
propellor	-11.1 ± 7.54	-13.41 ± 9.86	0.98	no
opening	-1.23 ± 4.35	-0.86 ± 6.05	0.27	no
tilt	0.03 ± 3.90	0.35 ± 4.17	0.25	no
roll	6.55 ± 5.08	4.63 ± 8.70	1.18	no
twist	34.3 ± 6.82	34.99 ± 6.41	0.29	no
helical twist	35.6 ± 6.22	36.62 ± 4.96	0.53	no

Table A.5. Polyamide distances (distances in Å; standard deviation for the x-ray structure is the resolution, 2.15 Å) for three initial high temperature simulations (300K)

Simulation 1				
distance	**computed average**	**x-ray**	**t-value**	**significant?**
x1	3.58 ± 0.24	3.08	0.74	no
x2	3.55 ± 0.27	3.13	0.62	no
x3	3.58 ± 0.34	2.93	0.96	no
x4	3.95 ± 0.36	2.95	4.64	yes
x5	2.94 ± 0.22	3.13	0.28	no
x1a	2.96 ± 0.19	3.3	0.50	no
x2a	3.83 ± 0.36	3.39	0.65	no
x3a	3.44 ± 0.30	3.12	0.47	no
x4a	3.46 ± 0.24	2.77	1.01	no

Table A.5 (cont'd.)

x5a	3.59 ± 0.24	3.16	0.63	no
Simulation 2				
distance	**computed average**	**x-ray**	**t-value**	**significant?**
x1	3.30 ± 0.20	3.08	0.32	no
x2	3.27 ± 0.21	3.13	0.20	no
x3	3.13 ± 0.21	2.93	0.29	no
x4	3.12 ± 0.22	2.95	0.25	no
x5	2.81 ± 0.12	3.13	0.47	no
x1a	2.93 ± 0.20	3.3	0.54	no
x2a	3.14 ± 0.24	3.39	0.36	no
x3a	3.12 ± 0.22	3.12	0	no
x4a	3.30 ± 0.22	2.77	0.77	no
x5a	3.29 ± 0.20	3.16	0.19	no
Simulation 3				
distance	**computed average**	**x-ray**	**t-value**	**significant?**
x1	3.22 ± 0.18	3.08	0.20	no
x2	3.29 ± 0.21	3.13	0.23	no
x3	3.12 ± 0.21	2.93	0.28	no
x4	3.12 ± 0.22	2.95	0.25	no
x5	2.82 ± 0.12	3.13	0.45	no
x1a	2.91 ± 0.23	3.3	0.57	no
x2a	3.22 ± 0.27	3.39	0.25	no
x3a	3.17 ± 0.22	3.12	0.07	no
x4a	3.32 ± 0.22	2.77	0.80	no
x5a	3.30 ± 0.20	3.16	0.20	no

Table A.6. Polyamide distances (distances in Å; standard deviation for x-ray structure is 2.15 Å) for low temperature simulations (113K)

Simulation 1				
distance	**computed average**	**x-ray**	**t-value**	**significant?**
x1	3.15 ± 0.091	3.08	0.10	no
x2	3.20 ± 0.115	3.13	0.10	no
x3	3.34 ± 0.15	2.93	0.60	no
x4	2.99 ± 0.11	2.95	0.47	no
x5	3.27 ± 0.19	3.13	0.21	no
x1a	3.19 ± 0.11	3.3	0.66	no
x2a	3.12 ± 0.13	3.39	0.40	no
x3a	3.13 ± 0.12	3.12	0.01	no
x4a	2.85 ± 0.085	2.77	0.62	no
x5a	3.34 ± 0.12	3.16	0.26	no

Table A.6
(cont'd.)

Simulation 2				
distance	**computed average**	**x-ray**	**t-value**	**significant?**
x1	3.47 ± 0.12	3.08	0.57	no
x2	3.21 ± 0.10	3.13	0.12	no
x3	3.07 ± 0.11	2.93	0.21	no
x4	2.74 ± 0.06	2.95	0.21	no
x5	3.09 ± 0.12	3.13	0.57	no
x1a	4.99 ± 0.19	3.3	2.49	yes
x2a	4.09 ± 0.18	3.39	1.03	no
x3a	3.22 ± 0.14	3.12	0.15	no
x4a	3.19 ± 0.11	2.77	0.62	no
x5a	3.19 ± 0.10	3.16	0.04	no
Simulation 3	**computed average**	**x-ray**	**t-value**	**significant?**
x1	3.20 ± 0.10	3.08	0.17	no
x2	3.14 ± 0.10	3.13	0.01	no
x3	3.33 ± 0.14	2.93	0.58	no
x4	2.78 ± 0.07	2.95	0.13	no
x5	3.04 ± 0.11	3.13	0.51	no
x1a	3.19 ± 0.11	3.3	0.54	no
x2a	3.63 ± 0.21	3.39	0.35	no
x3a	3.13 ± 0.13	3.12	0.01	no
x4a	2.93 ± 0.11	2.77	0.61	no
x5a	3.27 ± 0.12	3.16	0.16	no
Simulation 4	**computed average**	**x-ray**	**t-value**	**significant?**
x1	3.23 ± 0.12	3.08	0.22	no
x2	3.19 ± 0.11	3.13	0.09	no
x3	3.24 ± 0.16	2.93	0.46	no
x4	2.90 ± 0.09	2.95	0.59	no
x5	3.35 ± 0.19	3.13	0.34	no
x1a	3.16 ± 0.10	3.3	0.47	no
x2a	3.49 ± 0.15	3.39	0.15	no
x3a	3.01 ± 0.10	3.12	0.16	no
x4a	2.98 ± 0.13	2.77	0.57	no
x5a	3.40 ± 0.12	3.16	0.35	no

Table A.7. T-test results for combined polyamide distances

Low temperature (113 K) simulations

distance	# significant tests	%	avg. t-value	significant?
x1	16	1.6	0.99	no
x2	7	0.7	0.72	no
x3	425	42.5	1.92	no
x4	5	0.5	0.42	no
x5	3	0.3	0.37	no
x1a	2	0.2	0.19	no
x2a	0	0	0.11	no
x3a	0	0	0.28	no
x4a	102	10.2	1.47	no
x5a	35	3.5	1.12	no

Table A.8. T-test results for combined polyamide distances

Low temperature (300 K) simulations

distance	# significant tests	%	avg. t-value	significant?
x1	49	4.9	1.10	no
x2	26	2.6	0.87	no
x3	22	2.2	0.94	no
x4	59	5.9	0.81	no
x5	0	0	0.38	no
x1a	81	8.1	0.92	no
x2a	5	0.5	0.33	no
x3a	1	0.1	0.34	no
x4a	318	31.8	1.82	no
x5a	22	2.2	0.78	no

Table A.9. Computed and x-ray averages and standard deviations of helical parameters for hydroxypyrrole polyamide simulation

helical parameter (Å)	computed avg./sd	x-ray value/sd	t-test value	significant?
x-displacement	0.59 ± 1.76	0.37 ± 1.32	0.41	no
y-displacment	-0.31 ± 3.00	-0.23 ± 1.25	0.08	no
shear	0.15 ± 1.72	-0.12 ± 0.56	0.49	no
stretch	-0.09 ± 0.31	-0.21 ± 0.12	1.22	no
stagger	0.01 ± 0.56	0.15 ± 0.30	0.79	no
shift	0.04 ± 1.51	0.01 ± 0.84	0.07	no
slide	0.76 ± 1.13	0.76 ± 0.72	0.006	no
rise	3.15 ± 0.43	3.29 ± 0.26	1.01	no
helical rise	3.20 ± 0.81	3.33 ± 0.33	0.50	no
Helical parameter (°)				
inclination	7.44 ± 14.1	7.36 ± 9.54	0.02	no
tip	1.66 ± 11.1	0.49 ± 8.79	0.33	no

Table A.9 (cont'd.)

buckle	2.72 ± 10.9	-0.55 ± 7.17	0.94	no
propellor	-9.12 ± 12.6	-15.13 ± 5.79	1.51	no
opening	-3.39 ± 6.04	-0.53 ± 5.60	1.49	no
tilt	-0.36 ± 5.70	0.11 ± 4.94	0.26	no
roll	3.88 ± 7.14	4.24 ± 4.75	0.16	no
twist	32.6 ± 9.00	35.04 ± 9.32	0.84	no
helical twist	34.1 ± 8.32	35.95 ± 8.88	0.69	no

Table A.10. Polyamide distances (distances in Å; standard deviation for x-ray structure is 2.27 Å) for hydroxypyrrole polyamide simulation

Simulation 1				
distance	**computed average**	**x-ray**	**t-value**	**significant?**
x1	3.43 ± 0.13	3.34	0.12	no
x2	3.31 ± 0.12	2.92	0.55	no
x3	3.13 ± 0.14	2.73	0.55	no
x4	4.17 ± 0.15	3.23	1.31	no
x5	4.93 ± 0.17	3.16	2.47	yes
x6	3.28 ± 0.15	2.51	1.07	no
x7	3.14 ± 0.07	2.99	0.21	no
x1a	5.53 ± 0.28	2.60	4.17	yes
x2a	4.36 ± 0.14	3.30	1.73	no
x3a	3.13 ± 0.12	2.67	0.64	no
x4a	3.06 ± 0.7	2.91	0.29	no
x5a	3.58 ± 0.14	3.25	0.47	no
x6a	2.98 ± 0.09	2.75	0.31	no
x7a	3.28 ± 0.08	3.00	0.94	no

Table A.11. Computed and x-ray averages and standard deviations of helical parameters for DNA/netropsin simulation

helical parameter (Å)	**computed avg./sd**	**x-ray value/sd**	**t-test value**	**significant?**
x-displacement	-0.55 ± 1.84	0.24 ± 1.31	1.49	no
y-displacment	-0.05 ± 0.20	0.04 ± 1.12	1.41	no
shear	0.004 ± 0.56	0.09 ± 0.55	0.53	no
stretch	-0.13 ± 0.17	-0.22 ± 0.24	1.86	no
stagger	-0.09 ± 0.47	0.13 ± 0.49	1.65	no
shift	0.06 ± 0.76	-0.14 ± 0.64	0.91	no
slide	0.05 ± 0.62	0.27 ± 0.61	1.24	no
rise	3.29 ± 0.33	3.31 ± 0.38	0.18	no
helical rise	3.18 ± 0.40	3.27 ± 0.40	0.80	no
Helical parameter (°)				
inclination	6.08 ± 13.1	0.82 ± 8.32	1.39	no

Table A.11 (cont'd.)

tip	-0.43 ± 9.34	0.87 ± 8.87	0.48	no
buckle	-1.62 ± 12.1	-3.23 ± 10.6	0.46	no
propellor	-10.7 ± 10.7	-11.81 ± 10.5	0.35	no
opening	1.30 ± 6.18	-0.09 ± 7.21	0.77	no
tilt	0.18 ± 5.48	-0.41 ± 5.24	0.37	no
roll	3.36 ± 7.46	0.07 ± 5.14	1.52	no
twist	34.8 ± 6.12	35.7 ± 5.44	0.54	no
helical twist	36.1 ± 5.52	36.4 ± 5.28	0.16	no

Table A.12. Polyamide distances (distances in Å; standard deviation for x-ray structure is 2.20 Å (resolution)) for DNA/netropsin simulation

Simulation 1				
distance	**computed average**	**x-ray**	**t-value**	**significant?**
1	3.59 ± 0.66	2.52	1.54	no
2	6.55 ± 0.60	3.24	4.76	yes
3	3.11 ± 0.20	3.40	0.42	no
4	3.25 ± 0.18	3.50	0.36	no
5	3.69 ± 0.35	2.70	1.42	no
6	3.38 ± 0.26	3.70	0.46	no
7	3.33 ± 0.29	2.70	0.91	no
8	3.35 ± 0.18	2.90	0.65	no
9	3.66 ± 0.31	3.30	0.52	no
10	5.43 ± 0.63	3.20	3.20	yes
11	5.87 ± 0.74	3.50	3.40	yes

Table A.13. Computed and x-ray averages and standard deviations of helical parameters for DNA/distamycin simulation

helical parameter (Å)	**computed avg./sd**	**x-ray value/sd**	**t-test value**	**significant?**
x-displacement	0.06 ± 2.08	0.25 ± 1.41	0.44	no
y-displacment	-0.28 ± 1.42	0.16 ± 0.79	1.81	no
shear	0.04 ± 0.85	-0.07 ± 0.28	1.26	no
stretch	-0.15 ± 0.38	-0.11 ± 0.08	1.38	no
stagger	0.36 ± 1.73	0.14 ± 0.36	1.83	no
shift	0.20 ± 1.01	-0.08 ± 0.47	1.92	no
slide	0.35 ± 1.05	0.14 ± 0.79	0.88	no
rise	3.36 ± 0.51	3.37 ± 0.38	0.07	no
helical rise	3.46 ± 1.05	3.33 ± 0.47	0.89	no
Helical parameter (°)				
inclination	3.01 ± 14.8	-0.49 ± 9.87	1.17	no
tip	1.29 ± 10.6	-0.02 ± 7.50	0.57	no
buckle	0.60 ± 19.9	1.38 ± 9.29	0.28	no

Table A.13 (cont'd).

propellor	-11.6 ± 20.5	-15.1 ± 5.50	1.91	no
opening	1.51 ± 9.58	-2.73 ± 4.17	3.29	yes
tilt	-0.89 ± 7.06	0.08 ± 4.84	0.66	no
roll	1.36 ± 11.1	-0.23 ± 5.95	0.87	no
twist	36.7 ± 11.3	36.5 ± 4.64	0.10	no
helical twist	38.5 ± 12.1	37.2 ± 4.47	0.88	no

Table A.14. Polyamide distances (distances in Å; standard deviation for x-ray structure is 2.20 Å) for DNA/distamycin simulation

Simulation 1				
distance	**computed average**	**x-ray**	**t-value**	**significant?**
1	4.65 ± 0.86	3.03	2.33	yes
2	3.26 ± 0.21	3.17	0.13	no
3	3.47 ± 0.42	2.93	0.77	no
4	3.54 ± 0.29	3.14	0.58	no
5	3.76 ± 0.41	3.25	0.74	no
6	4.18 ± 0.97	3.13	1.50	no
7	3.91 ± 0.69	3.24	0.96	no

Table A.15. Computed and x-ray averages and standard deviations of helical parameters for DNA/netropsin simulation with AMBER parameters

helical parameter (Å)	**computed avg./sd**	**x-ray value/sd**	**t-test value**	**significant?**
x-displacement	-0.78 ± 2.98	0.24 ± 1.31	2.53	yes
y-displacment	-0.25 ± 3.27	0.04 ± 1.12	0.82	no
shear	-0.22 ± 0.97	0.09 ± 0.55	1.86	no
stretch	0.12 ± 0.61	-0.22 ± 0.24	4.52	yes
stagger	-0.17 ± 0.62	0.13 ± 0.49	2.04	yes
shift	-0.38 ± 1.43	-0.14 ± 0.64	1.22	no
slide	-0.32 ± 0.61	0.27 ± 0.61	3.19	yes
rise	3.26 ± 0.39	3.31 ± 0.38	0.39	no
helical rise	2.95 ± 0.88	3.27 ± 0.40	2.57	yes
Helical parameter (°)				
inclination	10.5 ± 21.8	0.82 ± 8.32	3.76	yes
tip	-0.61 ± 11.2	0.87 ± 8.87	0.55	no
buckle	-1.30 ± 13.1	-3.23 ± 10.6	0.61	no
propellor	-16.5 ± 16.3	-11.81 ± 10.5	1.47	no
opening	-9.22 ± 42.8	-0.09 ± 7.21	3.65	yes
tilt	0.57 ± 6.83	-0.41 ± 5.24	0.62	no
roll	0.65 ± 11.7	0.07 ± 5.14	0.36	no
twist	30.1 ± 15.1	35.7 ± 5.44	3.30	yes
helical twist	29.5 ± 20.9	36.4 ± 5.28	4.03	yes

Table A.16. Polyamide distances (distances in Å; standard deviation for x-ray structure is 2.20 Å (resolution)) for DNA/netropsin simulation with AMBER parameters

Simulation 1				
distance	**computed average**	**x-ray**	**t-value**	**significant?**
1	3.17 ± 0.29	2.52	0.94	no
2	5.78 ± 0.41	3.24	3.65	yes
3	2.85 ± 0.12	3.40	0.79	no
4	3.33 ± 0.21	3.50	0.25	no
5	3.06 ± 0.22	2.70	0.52	no
6	3.17 ± 0.25	3.70	0.77	no
7	3.06 ± 0.20	2.70	0.52	no
8	3.53 ± 0.25	2.90	0.91	no
9	3.40 ± 0.27	3.30	0.15	no
10	5.29 ± 0.55	3.20	3.01	yes
11	6.62 ± 0.50	3.50	4.49	yes

Table A.17. T-values for combined high and low temperature DNA/polyamide simulations

Parameter (Å)	**300 K (6 simulations)**	**113 K (8 simulations)**	**all (14 simulations)**
shear	0.391	0.228	0.271
stretch	0.752	0.278	0.35
stagger	2.24	3.11	1.4
shift	0.214	0.089	0.127
slide	0.508	0.121	0.25
rise	0.231	0.267	0.171
x-displacement	0.727	0.078	0.366
y-displacement	0.214	0.102	0.119
helical rise	0.391	0.505	0.331
Parameter (°)			
buckle	0.642	1.92	1.276
propellor	0.963	0.306	0.663
opening	0.443	0.303	0.198
tilt	0.222	0.178	0.113
roll	1.33	0.533	0.98
twist	0.668	0.291	0.478
inclination	1.11	0.676	0.949
tip	0.253	0.307	0.244
helical twist	0.349	0.425	0.397

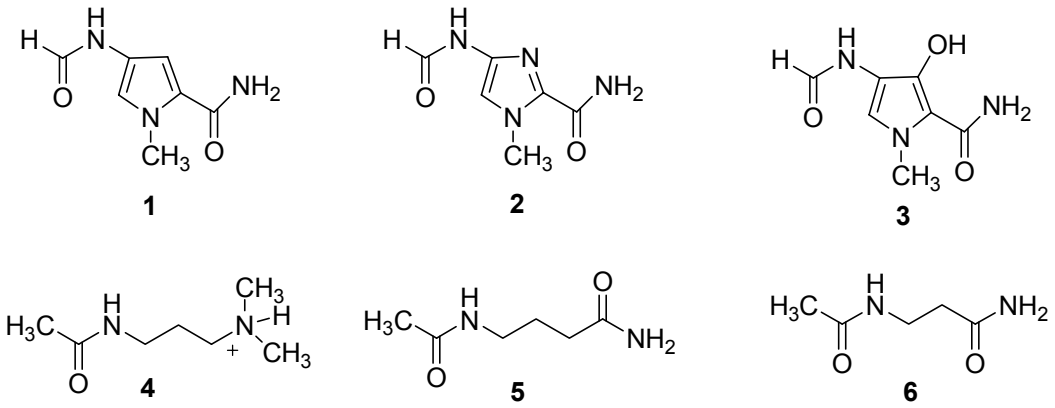


Figure A.1 Model compounds for polyamides (some hydrogens not shown for clarity).

Topology for **1** (topology was added to the topology for small molecules, top_all22_model.inp and topology for DNA/proteins, top_all27_prot_na.inp)

```
RESI PYRR      0.00    !N-methyl-pyrrole (with amide groups)
GROUP                  !                HM
ATOM HG   HP     0.115 !                |
ATOM CG   CY    -0.145 !          HZ    N2       HG
ATOM CD2  CPT   -0.020 !            \  /  \      /
ATOM CD1  CA     0.035 !             CB   CG--CD2       HA
ATOM HD1  HP     0.115 !             ||   ||   ||       /
ATOM NE1  NY    -0.230 !             O2   CD1  CE2   N1
ATOM CE2  CPT    0.130 !                 / \   /  \  / \
GROUP                  !               HD1  NE1    CA  HB
ATOM C1   CT3   -0.27  !                      |     ||
ATOM H1   HA     0.09  !                      C1    O1
ATOM H2   HA     0.09  !                     / | \
ATOM H3   HA     0.09  !                   H1 H2 H3
GROUP                  !
ATOM CA    C      0.51
ATOM O1    O     -0.51
ATOM N1    NH1   -0.62
ATOM HA     H     0.31
ATOM HB     H     0.31
GROUP
ATOM N2    NH1    -0.47
ATOM HM     H      0.235
ATOM CB     C      0.51
ATOM O2     O     -0.51
ATOM HZ     H      0.235

BOND CD2 HG  CD2 CG  NE1 CD1
BOND NE1 CE2
BOND CD1 HD1
DOUBLE CD1 CG  CE2 CD2
```

```
BOND C1 NE1
BOND C1 H1  C1 H2  C1 H3
BOND CA CE2  N1 CA  N1 HA  N1 HB
DOUBLE CA O1  CB O2
BOND N2 CG  N2 HM  N2 CB  CB HZ
!ic's for pyrrole ring
IC CG   CD1  NE1  CE2   0.0000  0.0000  0.0000  0.0000  0.0000
IC CD2  CB   *CG  CD1   0.0000  0.0000 180.0000  0.0000  0.0000
IC CD1  CG   CD2  CE2   0.0000  0.0000  0.0000  0.0000  0.0000
IC CD2  CG   CD1  NE1   0.0000  0.0000  0.0000  0.0000  0.0000
IC CD1  CE2  *NE1 C1   0.0000  0.0000 -180.0000  127.0  1.46
IC CG   NE1  *CD1 HD1   0.0000  0.0000 180.0000  0.0000  0.0000
IC CE2  CG   CD2  HG    0.0000  0.0000 -180.0000  126.9  1.08
!ic's for methyl group
IC  CD1 NE1 C1 H1          1.37    123.2   0.00    109.5   1.08
IC  CD1 NE1 C1 H2          1.37    123.2  -120.0   109.5   1.08
IC  CD1 ND1 C1 H3          1.37    123.2   120.    109.5   1.08
ic  CE2 NE1 C1 H3          1.37    123.2   -60.0   109.5   1.08
!ic's for amide group on CE2
ic  O1  CA  N1  HA     1.2233 122.84  180.  119.23  0.9933
ic  O1  CA  N1  HB     1.2233 122.84  0.00  119.23  0.9933
ic  HA  N1  CA  CE2    0.9933 119.23    0.  116.25  1.47
ic  HB  N1  CA  CE2    0.9933 119.23 -180.0 116.25  1.47
ic  N1  CA  CE2 NE1    1.3418 116.25  180.  122.1  1.38
ic  N1  CA  CE2 CD2    1.3418 116.25    0.  130.3  1.40
ic  NE1 CE2 CA  O1     1.38   122.0   0.00  123.5  1.2233
ic  CD2 CE2 CA  O1     1.40   130.3   180.0 123.5  1.2233
IC  CG  CD2 CE2 CA     1.40   107.1  -180.0 130.3  1.46
ic  CD1 NE1 CE2 CA     1.37   109.8   180.0 122.1  1.46
ic  C1  NE1 CE2 CA     1.46   127.0   0.00  122.1  1.46
!ic's for amide group on CG
ic  O2  CB  N2  HM      1.2233 122.84  -180.  119.23  0.9933
ic  HM  N2  CB  HZ      0.9933 119.23    0.   112.25  1.10
ic  O2  CB  N2  CG      1.2233 122.84    0.   125.57  1.40
ic  CB  N2  CG  CD2     1.3418 122.57  180.   124.9  1.41
ic  CB  N2  CG  CD1     1.3418 122.57    0.   127.2  1.40
IC  N2  CG  CD2 CE2     1.40   124.9   180.   107.1  1.40
ic  N2  CG  CD1 NE1     1.40   127.2  -180.0  107.6  1.37
```

Parameters for **1** (parameters were added to the parameters for DNA/proteins, par_all27_prot_na.inp)

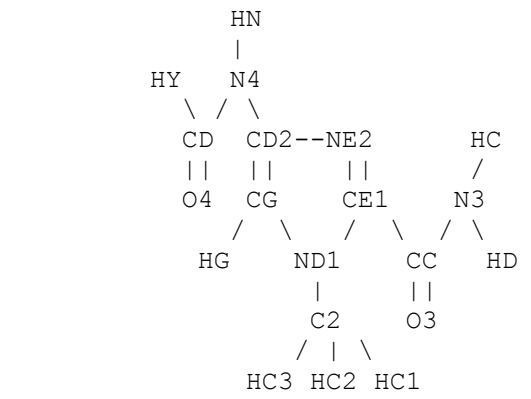
```
BONDS
!N-methyl-pyrrole
CPT  HP    500.00     1.08
CT3  NY    500.000    1.46
C    CPT   500.00     1.47
NH1  CY    500.00     1.40
C    H     500.00     1.10
ANGLES
!N-methyl-pyrrole
CPT  CY   NH1    40.000    124.9     CA   CA   NH1    40.000    127.2
HP   CPT  CY     40.000    126.9
HP   CPT  CPT    40.000    125.9
CA   NY   CT3    40.000    123.3
CPT  NY   CT3    40.000    127.
CPT  CPT  C      40.000    130.2
```

```
NY   CPT  C      40.000    122.1
NY   CT3  HA     40.000    109.5
CPT  C    O      40.000    123.5
CPT  C    NH1    40.000    114.4
H   NH1   H      40.000    115.7
CY  NH1   H      40.000    117.2
CY  NH1   C      40.00     125.2
NH1  C    H      40.00     112.0
CA   CY   NH1    40.000    127.2
CA   CY   C      40.00     100.3
CPT  CY   C      40.000    151.9
CPT  CY   CPT    40.000    36.3
DIHEDRALS
!N-methyl-pyrrole

!------------------------------------------------
CA  NY   CT3  HA     1.60     1       0.00   !methyl group
CA  NY   CT3  HA     0.000    3     -120.0
CA  NY   CT3  HA     0.000    2     -60.0
!------------------------------------------------
!----------------------------------------
CPT NY   CT3  HA     0.065   3        0.0    !methyl group
CPT NY   CT3  HA     0.00    1      -120.
CPT NY   CT3  HA     0.00    2       -60.0
!----------------------------------------
HP  CPT  CY   CA     2.0     1        180.0
HP  CPT  CY   NH1    2.8     2        180.0
CY  CA   NY   CT3    0.8     2        180.0
CY  NH1  C     O     1.5     2        180.0
CY  NH1  C     H     1.20    1        180.0
CPT CY   NH1   H     0.000   1         0.00
CPT CY   NH1   C     0.200   1        180.0
CPT  CPT  NY   CT3   0.8     2        180.0
CPT  CPT  C    O     2.5     2        180.0
CPT  CPT  C    NH1   1.5     2        180.0
CA   NY   CPT  C     2.00    1        180.0
HP   CA   CY   NH1   1.00    2        180.0
CA   CY   NH1  H     0.0     1         0.00
CA   CY   NH1  C     2.0     2        180.0
CA   CY   NH1  C     2.0     1        180.0

HP   CA   NY   CT3   0.4000  2      180.0
NY   CA   CY   NH1   2.0     2      180.0
NY   CPT  C    O     1.0     2      180.0
NY   CPT  C    O     3.0     1      180.0
NY   CPT  C    NH1   1.0     1        0.0
CPT  CPT  CY   NH1   3.0     2      180.0
CT3  NY   CPT  C     0.8     2      180.0
H    NH1  C    H     1.4     2      180.0
CPT  C    NH1   H    2.5     2      180.0
```

Topology for **2** (topology was added to topology for small molecules, top_all22_model.inp and to DNA/proteins, top_all27_prot_na.inp)

```
RESI IMID      0.00    !N-methyl-imidazole
```

```
GROUP                   !
ATOM CG   CPH1  -0.05 !
ATOM HG   HR3    0.09 !
ATOM CD2  CPH1   0.32 !
ATOM ND1  NR1   -0.36 !
ATOM CE1  CPH2   0.38 !
ATOM NE2  NR2   -0.38 !
                      !
GROUP                 !
ATOM C2   CT3    -0.27!
ATOM HC1   HA     0.09!
ATOM HC2   HA     0.09!
ATOM HC3   HA     0.09!
                      !
GROUP                 !
ATOM CC    C       0.51!
ATOM O3    O      -0.51!
ATOM N3    NH1    -0.62
ATOM HC     H      0.31
ATOM HD     H      0.31

GROUP
ATOM N4    NH1    -0.47
ATOM HN     H       0.235
ATOM CD     C       0.51
ATOM O4     O      -0.51
ATOM HY     H       0.235
 BOND NE2 CD2  ND1 CG  CE1 ND1
BOND CG HG
DOUBLE NE2 CE1  CD2 CG
BOND ND1 C2  C2 HC1  C2 HC2  C2 HC3
BOND CE1 CC  CC N3  N3 HC  N3 HD
DOUBLE CC O3  CD O4
BOND CD2 N4  N4 HN  N4 CD  CD HY
! KEEPS HYDROGENS IN RING PLANE
IMPH ND1  CG   CE1  C2       ND1  CE1  CG   C2
IMPH CD2  CG   NE2  N4       CD2  NE2  CG   N4
IMPH CE1  ND1  NE2  CC       CE1  NE2  ND1  CC
IMPH CG   CD2  ND1  HG        CG   ND1  CD2  HG
IC   HG    CG    ND1   CE1    1.5421  122.67 -173.67  109.79   1.2987
IC   CG    ND1   CE1   NE2    1.2854  109.79    0.21  110.31   1.3071
IC   ND1   CE1   NE2   CD2    1.2987  110.31    0.03  105.82   1.3165
IC   CE1   NE2   CD2   CG     1.3071  105.82   -0.23  108.68   1.3758
IC   NE2   CD2   CG    ND1    1.3165  108.68    0.35  105.39   1.2854
IC   NE2   CD2   CG    HG     1.3165  108.68  172.86  131.52   1.5421
IC   CD2   CG    ND1   CE1    1.3758  105.39   -0.34  109.79   1.2987
IC   CD2   NE2   CE1   CC     1.3165  105.82  -180.0  124.7    1.46
IC   NE2   CE1   ND1   C2      1.34  110.31    180.0  127.8    0.9770
IC   HG    CG    CD2   N4      1.07  131.52     0.00  129.3   1.40
!ic's for methyl group
IC  CG ND1 C2 HC1          1.37    124.7   0.00    109.5   1.08
IC  CG ND1 C2 HC2          1.37    124.7  -120.0   109.5   1.08
IC  CG ND1 C2 HC3          1.37    124.7   120.    109.5   1.08
ic  CE1 ND1 C2 HC3         1.37    124.7   -60.0   109.5   1.08
impr C2 CG ND1 CE1  !for now
!ic's for amide group on CE1
IMPR  CC CE1 N3 O3   !for now
```



```
ic  O3  CC  N3  HC     1.2233 122.84  -180.  119.23  0.9933
ic  O3  CC  N3  HD     1.2233 122.84  0.00  119.23  0.9933
ic  HC  N3  CC  CE1    0.9933 119.23    0.  116.25  1.47
ic  HD  N3  CC  CE1    0.9933 119.23  180.0 116.25  1.47
ic  N3  CC  CE1 ND1    1.3418 116.25  180.  122.1   1.38
ic  N3  CC  CE1 NE2    1.3418 116.25    0.  130.3   1.40
ic  ND1 CE1 CC  O3     1.38   122.0   0.00  123.5   1.2233
ic  NE2 CE1 CC  O3     1.40   130.3   180.0 123.5   1.2233
IC  CD2 NE2 CE1 CC     1.40   107.1  -180.0 130.3   1.46
ic  CG  ND1 CE1 CC     1.37   109.8   -180.0 122.1   1.46
ic  C2  ND1 CE1 CC     1.46   127.0   0.00  122.1   1.46
!ic's for amide group on CD2
IMPR  CD HY N4 O4
ic  O4  CD  N4  HN      1.2233 122.84   180.  119.23  0.9933
ic  HN  N4  CD  HY      0.9933 119.23    0.   112.25  1.10
ic  O4  CD  N4  CD2      1.2233 122.84    0.   125.57  1.40
ic  CD  N4  CD2  NE2     1.3418 122.57  -180.   124.9  1.41
ic  CD  N4  CD2  CG     1.3418 122.57    0.   127.2  1.40
IC  N4  CD2 NE2 CE1     1.40   124.9   -180.   107.1  1.40
ic  N4  CD2 CG  ND1     1.40   127.2  -180.0  107.6  1.37
impr HN CG CD2 N4
```

Parameters for **2** (parameters were added to parameters for DNA/proteins, par_all27_prot_na.inp)

```
BONDS
!N-methyl-imidazole
NR1 CT3   500.00      1.46
CPH2 C    500.000     1.47
CPH1 NH1  500.00      1.40
C    H    500.00      1.10
ANGLES
!N-methyl-imidazole
CPH1 CPH1 NH1    40.000   129.3
NR2  CPH1 NH1    40.000   120.0
CPH1 NR1  CT3    40.000   124.7
CPH2 NR1  CT3    40.000   127.8
NR1  CPH2 C      40.000   123.8
NR2  CPH2 C      40.000   124.8
NR1  CT3  HA     40.000   107.4
CPH2 C    O      40.000   122.6
CPH2 C    NH1    40.000   112.9
H   NH1   H      40.000   121.1
CPH1 NH1  H      40.000   115.7
CPH1 NH1  C      40.000   124.5
NH1  C    H      40.000   112.4
DIHEDRALS
!N-methyl-imidazole
!-----------------------------------------------------
CPH1 NR1  CT3  HA      1.60  1     0.00
CPH1 NR1  CT3  HA      0.000  3     -120.0     !methyl group
CPH1 NR1  CT3  HA      0.000  2    -60.0
!-----------------------------------------------------
!-----------------------------------------------------
CPH2 NR1 CT3 HA        0.0100   3    0.
CPH2 NR1 CT3 HA        0.0000   1   -120.0     !methyl group
CPH2 NR1 CT3 HA        0.0000   2   -60.0
```

```
!-----------------------------------------------

CPH1 CPH1 NH1 H        0.1      2   -180.0
CPH1 CPH1 NH1 H        0.00     1    0.00
CPH1 CPH1 NH1 C        1.5      1    180.0
CPH1 CPH1 NH1 C        1.0      2    00.0
HR3  CPH1 CPH1 NH1     0.5000   2    0.00
CPH1 NH1  C    O       0.2000   2    0.00
CPH1 NH1  C    O       3.000    1    180.0
CPH1 NH1  C   H        0.80000  2   -180.0
CPH1 NH1  C   H        0.5000   1     0.00
NR1  CPH1 CPH1 NH1     1.0000   1   -180.0
CPH2 NR2 CPH1 NH1      0.2000   1   -180.0
NR2 CPH1 NH1  H        1.3      2     0.0
NR2 CPH1 NH1  C        1.3      2   -180.0
H   NH1  C   H         0.2000   2    0.00
CPH1 CPH1 NR1  CT3     0.2000   1   -180.0
O  C  NH1  H           2.5000   2    180.0
!-----------------------------------------------------
CPH1 NR1  CPH2 C       0.2000  1    -180.0
HR3  CPH1 NR1  CT3     0.2000  1      0.00
CPH1 NR2  CPH2 C       0.2000  1    -180.0
NR1  CPH2 C   O        5.500   2     180.0
NR1  CPH2 C   O        3.000   1     180.0
NR1  CPH2 C  NH1       1.00    2      0.0
NR2 CPH2 C   O         1.0000  1      0.0
NR2 CPH2 C  NH1        0.0     1      0.00
CPH2 C   NH1  H        0.2000  1      0.00
CPH2 C   NH1  H        0.2000  2     180.0
NR2 CPH2 NR1 CT3       0.2000  1     180.0
CT3 NR1  CPH2 C        0.2000  1      0.00
IMPROPER
!N-methyl-imidazole
CT3 CPH1 NR1 CPH2    90.000    0      180.0
C   NR1  CPH2 NR2    90.000    0     -180.0
NR1 CPH1 CPH2 CT3    90.000    0       0.00
NR1 CPH2 CPH1 CT3    90.000    0       0.00
CPH1 CPH1 NR2 NH1    90.0000   0       0.00
CPH1 NR2 CPH1 NH1    90.000    0       0.00
CPH2 NR1 NR2  C      90.000    0       0.00
CPH2 NR2 NR1  C      90.000    0       0.00
```

Topology for **3** (added to topology for DNA/proteins, top_all27_prot_na.inp)

```
RESI OHPY       0.00   !Hydroxy pyrrole
GROUP                  !              HM
ATOM CG   CY    -0.030 !              |
ATOM CD2  CPT   -0.020 !        HZ    N2      OH--HH
ATOM CD1  CA     0.035 !          \  /  \     /
ATOM HD1  HP     0.115 !           CB   CG--CD2     HA
ATOM NE1  NY    -0.230 !           ||   ||  ||     /
ATOM CE2  CPT    0.130 !           O2  CD1  CE2   N1
                                      /  \   / \  / \
GROUP                  !            HD1  NE1   CA   HB
ATOM OH   OH1   -0.43  !                  |    ||
ATOM HH   H      0.43  !                 C1    O1
```

```
                                        / | \
GROUP                   !               H1 H2 H3
ATOM C1   CT3   -0.27  !
ATOM H1   HA     0.09  !
ATOM H2   HA     0.09  !
ATOM H3   HA     0.09  !

GROUP
ATOM CA    C     0.51
ATOM O1    O     -0.51
ATOM N1    NH1   -0.62
ATOM HA     H     0.31
ATOM HB     H     0.31

GROUP
ATOM N2    NH1    -0.47
ATOM HM     H      0.235
ATOM CB     C      0.51
ATOM O2     O     -0.51
ATOM HZ     H      0.235
BOND CD2 CG  NE1 CD1   !CD2 HG  CD2 CG  NE1 CD1
BOND CD2 OH  OH  HH
BOND NE1 CE2
BOND CD1 HD1
DOUBLE CD1 CG  CE2 CD2
BOND C1 NE1
BOND C1 H1  C1 H2  C1 H3
BOND CA CE2  N1 CA  N1 HA  N1 HB
DOUBLE CA O1  CB O2
BOND N2 CG  N2 HM  N2 CB  CB HZ
!ic's for pyrrole ring
IC CG   CD1  NE1  CE2   0.0000  0.0000  0.0000  0.0000  0.0000
IC CD2  CB   *CG  CD1   0.0000  0.0000 180.0000  0.0000  0.0000
IC CD1  CG   CD2  CE2   0.0000  0.0000  0.0000  0.0000  0.0000
IC CD2  CG   CD1  NE1   0.0000  0.0000  0.0000  0.0000  0.0000
IC CD1  CE2  *NE1 C1    0.0000  0.0000 -180.0000  127.0  1.46
IC CG   NE1  *CD1 HD1   0.0000  0.0000 180.0000  0.0000  0.0000
IC CE2  CG    CD2 OH    0.0000  0.0000 -180.0000  126.9  1.37
IC CG   CD2  OH   HH    1.39    128.9   180.0     106.4  0.98


!ic's for methyl group
IC  CD1 NE1 C1 H1          1.37    123.2   0.00    109.5   1.08
IC  CD1 NE1 C1 H2          1.37    123.2  -120.0   109.5   1.08
IC  CD1 ND1 C1 H3          1.37    123.2   120.    109.5   1.08
ic  CE2 NE1 C1 H3          1.37    123.2   -60.0   109.5   1.08
!ic's for amide group on CE2
ic  O1  CA  N1  HA     1.2233 122.84  180.  119.23  0.9933
ic  O1  CA  N1  HB     1.2233 122.84  0.00  119.23  0.9933
ic  HA  N1  CA  CE2    0.9933 119.23    0.  116.25  1.47
ic  HB  N1  CA  CE2    0.9933 119.23 -180.0 116.25  1.47
ic  N1  CA  CE2 NE1    1.3418 116.25  180.  122.1  1.38
ic  N1  CA  CE2 CD2    1.3418 116.25    0.  130.3  1.40
ic  NE1 CE2 CA  O1     1.38   122.0   0.00  123.5  1.2233
ic  CD2 CE2 CA  O1     1.40   130.3   180.0 123.5  1.2233
IC  CG  CD2 CE2 CA     1.40   107.1  -180.0 130.3  1.46
ic  CD1 NE1 CE2 CA     1.37   109.8   180.0 122.1  1.46
```

```
ic  C1  NE1 CE2 CA     1.46   127.0   0.00  122.1  1.46
!ic's for amide group on CG
ic  O2  CB  N2  HM      1.2233 122.84  -180.  119.23  0.9933
ic  HM  N2  CB  HZ      0.9933 119.23    0.   112.25  1.10
ic  O2  CB  N2  CG      1.2233 122.84    0.   125.57  1.40
ic  CB  N2  CG  CD2     1.3418 122.57  180.   124.9  1.41
ic  CB  N2  CG  CD1     1.3418 122.57    0.   127.2  1.40
IC  N2  CG  CD2 CE2     1.40   124.9   180.   107.1  1.40
ic  N2  CG  CD1 NE1     1.40   127.2  -180.0  107.6  1.37
```

Parameters for **3** (added to parameters for DNA/proteins, par_all27_prot_na.inp)

```
BONDS
!hydroxy pyrrole
CPT  OH1  500.000      1.37
CT3  NY   500.000      1.46
C   CPT   500.00       1.45
NH1 CY    500.000      1.40
C    H    500.000      1.10
ANGLES
!Hydroxy pyrrole
CPT  CY  NH1    40.00    123.1
CA   CY  NH1    40.000   129.7
CY  CPT  OH1    40.00    123.1
CPT CPT  OH1    40.00    129.0
CA  NY   CT3    40.000   124.3
CPT NY   CT3    40.000   125.9
CPT CPT  C      40.000   128.3
NY  CPT  C      40.000   124.3
CPT OH1  H      40.000   106.4
NY  CT3  HA     40.00    107.4
CPT  C   O      40.000   125.5
CPT  C   NH1    40.00    112.5
H   NH1  H      40.000   111.8
CY  NH1  H      40.00    116.2
CY  NH1  C      40.000   124.6
NH1  C   H      40.000   112.3
CPT CPT CC      40.00    128.25394
NY CPT CC       40.00    124.31412
CPT CC O        40.00    125.47897
CPT CC NH2      40.00    112.47125
DIHEDRALS
!Hydroxy pyrrole

!from pyrrole
!-----------------------------------------------
CA  NY   CT3  HA     1.60     1       0.00   !methyl group
CA  NY   CT3  HA     0.000    3     -120.0
CA  NY   CT3  HA     0.000    2     -60.0
!-----------------------------------------------
!-----------------------------------------
CPT NY   CT3  HA     0.065   3        0.0    !methyl group
CPT NY   CT3  HA     0.00    1      -120.
CPT NY   CT3  HA     0.00    2       -60.0
!-----------------------------------------

HP  CPT  CY   CA     2.0     1        180.0
```

```
HP  CPT  CY   NH1    2.8     2        180.0
CY  CA   NY   CT3    0.8     2        180.0
CY  NH1  C     O     2.0     2        180.0
CY  NH1  C     H     0.3     1        180.0
CPT CY   NH1   H     0.000   1         0.00
CPT CY   NH1   C     0.5     1        180.0
CPT  CPT  NY   CT3   0.8     2        180.0
CPT  CPT  C    O     2.5     2        180.0

CPT  CPT  C    NH1   2.5     2        180.0
CPT  CPT  C    NH1   2.0     1        0.00

CA   NY   CPT  C     2.30    2        180.0

HP   CA   CY   NH1   1.0     2        180.0

CA   CY   NH1  H     1.5     1         0.00

CA   CY   NH1  C     1.5     2        180.0
CA   CY   NH1  C     0.0     1        180.0

HP   CA   NY   CT3   0.4000  2        180.0
NY   CA   CY   NH1   1.0     2        180.0

NY   CPT  C    O     1.0     2        180.0
NY   CPT  C    O     5.5     1        180.0

NY   CPT  C    NH1   1.0     1        180.0

CPT  CPT  CY   NH1   4.0     2        180.0
CT3  NY   CPT  C     0.8     2        180.0

H    NH1  C    H     1.4     2        180.0
CPT  C    NH1   H    4.0     2        180.0
!-------------------------------------------------

CY  CPT  OH1 H       0.0000  2        90.0
CY  CPT  OH1 H       4.0     1         0.0


CPT CPT OH1 H         0.5      2    180.00
CPT CPT OH1 H         0.30     1    0.0


OH1 CPT CY NH1        0.200     2   180.0

CA  CY  CPT OH1       3.1       2   180.0

IMPROPER
!hydroxy-pyrrole
C   CPT  NH1  O     90.0         0    0.00
C   NH1  CPT  O     90.0         0    0.00
O    NY   CPT  C    90.00        0    0.00
CPT  CA   NY   CT3   90.000      0   -180.0
C  NH1   O   CPT    90.000       0    0.00
NH1  NY   CPT C     90.000       0    0.00
O1   CPT  CPT C     90.000       0    0.00
```

```
H   CY  CPT  OH1    45.000       0   -26.0
OH1  CY  CPT CPT    90.000       0   -180.0
H  CPT CPT OH1      90.000       0    14.0
NH1 CA  CY  CPT     90.00        0   -180.0
```

Topology for **4** (added to topology for DNA/proteins, top_all27_prot_na.inp)

```
RESI TAIL      1.00   !cationic tail group for polyamides
GROUP                !
ATOM CL   CT3   -0.27!
ATOM HL1  HA     0.09!                                  HX2 HX3
ATOM HL2  HA     0.09!                                   \ /
ATOM HL3  HA     0.09!   HL1       H    HR1  HS1  HT1  CX—HX1
ATOM C    C      0.51!    \        |    |    |    |  + /
ATOM O    O     -0.51!HL2-CL---C---N---CR---CS---CT---N1--HC
ATOM N    NH1   -0.47!    /    ||       |    |    |    \
ATOM H    H      0.47! HL3     O       HR2  HS2  HT2   CY—HY1
ATOM CR   CT2   -0.11!                                 /  |
ATOM HR1  HA     0.09!                                HY2 HY3
ATOM HR3  HA     0.09!
                     !
GROUP                 !
ATOM CS  CT2     -0.11!
ATOM HS1  HA      0.09!
AtOM HS2  HA      0.09!
ATOM CT   CT2    -0.11!
ATOM HT1  HA      0.09!
ATOM HT2  HA      0.09!

GROUP
ATOM N1   NH3     -0.30
ATOM H1   HC       0.33
ATOM CX   CT3      0.23
ATOM HX1  HA       0.05
ATOM HX2  HA       0.05
ATOM HX3  HA       0.05
ATOM CY   CT3      0.23
ATOM HY1  HA       0.05
ATOM HY2  HA       0.05
ATOM HY3  HA       0.05

BOND CL C  C N  N CR  N H
BOND CL HL1  CL HL2  CL HL3
BOND CR HR1   CR HR3  CR CS
DOUBLE C O
BOND CS HS1  CS HS2  CS CT
BOND CT HT1  CT HT2  CT N1
BOND N1 H1  N1 CX  N1 CY
BOND CX HX1  CX HX2  CX HX3
BOND CY HY1  CY HY2  CY HY3
!-------------------------------------------------------
!ic table
IMPR  N  C  CR H
IMPR  C  CL N  O
IC  O   C   N   H      1.2233 122.84  180. 119.23 0.9933
```

```
IC  H   N   C   CL     0.9933 119.23    0. 116.25 1.5118
IC  O   C   N   CR     1.2233 122.84    0. 122.57 1.4488
IC  N   C   CL  HL1    1.3418 116.25  154. 109.3  1.109
IC  N   C   CL  HL2    1.3418 116.25   33. 109.3  1.109
IC  N   C   CL  HL3    1.3418 116.25  -87. 109.3  1.109
IC  C   N   CR  HR1    1.3418 122.57   -180. 110.7  1.113
IC  C   N   CR  CS     1.3418 122.57   60. 110.7  1.50
IC  C   N   CR  HR3    1.3418 122.57  -60.  110.7  1.113

IC  N  CR  CS CT       1.47   109.5   180.0  109.5  1.50
IC  N  CR  CS HS1      1.47   109.5   60.0  109.5  1.10
IC  N  CR  CS HS2      1.47   109.5    -60.0  109.5  1.10
IC  CR CS  CT HT1      1.50   109.5    -60.0  109.5  1.10
IC  CR CS  CT HT2      1.50   109.5    60.0 109.5  1.10
IC  CR CS  CT N1       1.50   109.5    180.0 109.5  1.10

IC  CS CT N1 H1         1.50   109.5   60.00   109.5  1.10
IC  CS CT N1 CX        1.50   109.5  -60.0  109.5  1.47
IC  CS CT N1 CY        1.50   109.5   180.0  109.5  1.47

IC  CT N1 CX HX1       1.47   109.5   -60.0  109.5  1.10
IC  CT N1 CX HX2       1.47   109.5    60.0  109.5  1.10
IC  CT N1 CX HX3       1.47   109.5   -180.  109.5  1.10
IC  CT N1 CY HY1       1.47   109.5    60.0  109.5  1.10
IC  CY N1 CY HY2       1.47   109.5    180.0 109.5  1.10
IC  CY N1 CY HY3       1.47   109.5   -60.0  109.5  1.10

IC  HY2 CY N1 H1        1.10   109.5    -60.0    109.5  1.10
IC  HY3 CY N1 H1        1.10   109.5     60.0    109.5  1.10
```

Parameters for **4** (added to parameters for DNA/proteins, top_all27_prot_na.inp)

```
ANGLES
!cationic tail
CT2  NH3  CT3   40.00      111.4
CT3  NH3  CT3   40.0       110.8

DIHEDRALS
CT3  C   NH1  H          2.5      2     180.0
CT3  C   NH1  CT2        1.6      1     0.00
CT3  C   NH1  CT2        2.5      2     180.0

HA  CT3  C    O          0.0      3     180.0
HA  CT3  C   NH1         0.0      3     0.00

C   NH1  CT2  HA         0.00     3     0.00
C   NH1  CT2  CT2        1.0      2     0.00
C   NH1  CT2  CT2        0.5      1     0.00

O    C   NH1   H         2.5      2     180.0
O    C   NH1  CT2        2.5      2     180.0

NH1  CT2  CT2  HA        0.1950   3     0.00
NH1  CT2  CT2  CT2       0.1950   3     0.00

H  NH1  CT2  HA          0.000    3     0.00

H  NH1  CT2 CT2          0.000    1     0.00
```

```
CT2 CT2  CT2   HA        0.1950   3     0.00
CT2 CT2  CT2  NH3        0.1950   3     0.00

H  CT2   CT2  HA         0.1950   3     0.00
CT2  CT2  NH3  HC       0.100      3     0.00

CT2  CT2  NH3  CT3      0.4000     3     0.000

HA   CT2  CT2  NH3      0.1950     3     0.00
CT2  NH3  CT3  HA       0.09000    3     0.00

HA   CT2  CT2  HA       0.1950     3     0.00

HA   CT2  NH3  HC       0.1000     3     0.00
HA   CT2  NH3  CT3      0.100      3     0.00

HC   NH3  CT3  HA       0.09000    3     0.00
CT3  NH3  CT3  HA       0.0900     3     0.00


IMPROPER
NH1  C  CT2 H     20.00     0       0.00
C   CT3 NH1 O     120.0     0       0.00
```

Topology for **5** (added to topology for DNA/proteins, top_all27_prot_na.inp)

```
RESI GAMM      0.00   !gamma aminobutyric acid linker for polyamides
GROUP              !
ATOM CL   CT3   -0.27!
ATOM HL1  HA     0.09!     HL1       H   HR1  HS1  HT1  OM    H2
ATOM HL2  HA     0.09!      \        |    |    |    |   ||    /
ATOM HL3  HA     0.09! HL2--CL---C---N---CR---CS---CT---CM---N2
ATOM C    C      0.51!      /    ||       |    |    |         \
ATOM O    O     -0.51!     HL3   O       HR2  HS2  HT2         H3
ATOM N    NH1   -0.47!
ATOM H    H      0.47!
ATOM CR   CT2   -0.18!
ATOM HR1  HA     0.09!
ATOM HR3  HA     0.09!
                   !
GROUP              !
ATOM CS   CT2    -.18!
ATOM HS1  HA    0.09
ATOM HS2  HA    0.09
ATOM CT   CT2   -.18
ATOM HT1  HA    0.09
ATOM HT2  HA    0.09
GROUP
ATOM CM   CC     0.51
ATOM OM   O    -0.51
ATOM N2   NH2   -0.47
ATOM H2   H      0.235
ATOM H3   H      0.235
```

```
BOND CL HL1  CL HL2  CL HL3
BOND CL C  C N  N H  N CR
DOUBLE C O
BOND CR HR1  CR HR3  CR CS
BOND CS HS1  CS HS2  CS CT
BOND CT HT1  CT HT2  CT CM
DOUBLE CM OM
BOND CM N2  N2 H2  N2 H3
!ic table
IMPR  N  C  CR H
IMPR  C  CL N  O
IC  O   C   N   H      1.2233 122.84  180. 119.23 0.9933
IC  H   N   C   CL     0.9933 119.23    0. 116.25 1.5118
IC  O   C   N   CR     1.2233 122.84    0. 122.57 1.4488
IC  N   C   CL  HL1    1.3418 116.25  180. 109.3  1.109
IC  N   C   CL  HL2    1.3418 116.25   60. 109.3  1.109
IC  N   C   CL  HL3    1.3418 116.25  -60. 109.3  1.109
IC  C   N   CR  HR1    1.3418 122.57  -30. 110.7  1.113
IC  C   N   CR  CS    1.3418 122.57    90. 110.7   1.50
IC  C   N   CR  HR3    1.3418 122.57  -150. 110.7  1.113

IC  N CR CS CT         1.47   109.5  -180.0  109.5  1.50
IC  N CR CS HS1        1.47   109.5  -60.0  109.5  1.10
IC  N CR CS HS2        1.47   109.5   60.0  109.5  1.10
IC  CR CS CT HT1       1.50   109.5   60.0  109.5  1.10
IC  CR CS CT HT2       1.50   109.5  -60.0  109.5  1.10
IC  CR CS CT CM        1.50   109.5   180.0 109.5  1.50
IC  CS CT CM OM        1.50   109.5  -30.00  123.1  1.22233
IC  CS CT CM N2        1.50   109.5   150.0 115.0  1.37
IC  CT CM N2 H2        1.50   115.0  -180.0 120.6  1.10
IC  CT CM N2 H3        1.50   115.0   0.00  120.6  1.10

IMPR  CM CT N2 OM
```

Parameters for **5** (added to parameters for DNA/proteins, par_all27_prot_na.inp)

```
ANGLES
!gamma linker
H  NH1 H        25.00    117.0

DIHEDRALS

CT3  C   NH1  H     2.5     2    180.0

CT3  C   NH1  CT2   1.6     1     0.00
CT3  C   NH1  CT3   2.5     2    180.0

HA  CT3 C   O       0.000   3    180.0
HA  CT3 C   NH1     0.000   3    0.00

C   NH1  CT2  HA    0.00    3    0.00

C   NH1  CT2  CT2   0.5     2    0.00
O   C    NH1  H     2.5     2    180.0
O   C    NH1  CT2   2.5     2    180.0
```

```
NH1  CT2  CT2  HA   0.1950  3    0.00

NH1  CT2  CT2  CT2  0.1950  3    0.00
H  NH1  CT2 HA      0.00    3    0.00

H  NH1  CT2  CT2    0.00    1    0.00
CT2  CT2  CT2  HA   0.1950  3    0.00
CT2  CT2  CC  O     0.050   6    0.00

CT2  CT2  CC  NH2    0.020  6    180.00
!CT2  CT2  CC  NH2   0.00   3    180.0

HA   CT2  CT2  HA     0.1950  3    0.00
HA   CT2  CT2  CC     0.1950  3    0.00
CT2  CC    NH2  H     1.4     2    180.0
HA   CT2  CC   O      0.00    3    0.00
HA   CT2  CC   NH2    0.000   3    0.00
O    CC   NH2  H      1.4     2    180.0
CT2  CT2  CT2  CC     0.1950  3    0.00

IMPROPER
NH1  C   CT2  H     20.00    0    0.00
C  CT3  NH1  O      120.0    0    0.00
CC  CT2  NH2  O     45.0     0    0.00
```

Topology for **6** (added to topology for DNA/proteins, top_all27_prot_na.inp)

```
RESI BETA        0.00  !Beta alanine linker for polyamides
GROUP                  !
ATOM CL   CT3   -0.27!    HL2       H   HR1  HS1        H1
ATOM HL1  HA     0.09!     \        |    |    |         /
ATOM HL2  HA     0.09! HL1—CL---C---N---CR---CS---CT---NS
ATOM HL3  HA     0.09!     /    ||       |    |   ||    \
                       !   HL3    O       HR3  HS2  OL     H2
GROUP                  !
ATOM C    C      0.51!
ATOM O    O     -0.51!
ATOM N    NH1   -0.47!
ATOM H    H      0.47!
ATOM CR   CT2   -0.18!
ATOM HR1  HA     0.09!
ATOM HR3  HA     0.09!

GROUP
ATOM CS   CT2    -0.18
ATOM HS1  HA      0.09
ATOM HS2  HA      0.09
GROUP
ATOM CT   CC       0.51
ATOM OL   O      -0.51
ATOM NS   NH2     -0.47
ATOM H1   H       0.235
ATOM H2   H       0.235

BOND  HL1 CL  HL2 CL  HL3 CL
BOND  CL C  C N  N CR
```

```
BOND  N H
BOND  HR1 CR  HR3 CR
DOUBLE C O  CT OL
BOND CR CS  CS CT  CT NS
BOND NS H1  NS H2  CS HS1  CS HS2

IMPR  N  C  CR H
IMPR  C  CL N  O
IC  O C N H      1.2233 122.84  180. 119.23 0.9933
IC  H N C CL     0.9933 119.23    0. 116.25 1.5118
IC  O C N CR     1.2233 122.84    0. 122.57 1.4488
IC  N C CL HL1    1.3418 116.25  150. 109.3  1.109
IC  N C CL HL2    1.3418 116.25   25. 109.3  1.109
IC  N C CL HL3    1.3418 116.25  -95. 109.3  1.109
IC  C N CR  CS     1.3418 122.57  90. 110.7  1.50
IC  C N CR  HR1    1.3418 122.57  -150. 110.7  1.50
IC  C N CR  hR3    1.3418 122.57  -30.0 110.7  1.50

IC  N CR CS HS1    1.4488  110.7   60.0  109.5   1.10
IC  N CR CS HS2    1.4488  110.7   -60.0  109.5   1.10
IC  N CR CS CT     1.4488  110.7  -180.0 109.5   1.50
IC  CR CS CT NS    1.50    109.5   150.0 115.0   1.40
IC  CR CS CT OL    1.50    109.5   -36.50  123.1   1.24

IC  OL CT NS H1    1.24    121.9   160.0  121.2  0.9933
IC  OL CT NS H2    1.24    121.9    0.00  121.2  0.9933

IMPR CT CS NS OL
IMPR CT NS CS OL
IMPR NS CT H2 H1
IMPR NS CT H1 H2
```

Parameters for **6** (added to parameters for DNA/proteins, par_all27_prot_na.inp)

```
ANGLES
!beta linker

H  NH1  H     40.000     118.2      !A.E.L.

DIHEDRALS

CT3  C   NH1  H      2.5    2    180.0

CT3  C   NH1  CT2    1.6    1     0.00
CT3  C   NH1  CT2    2.5    2    180.0

HA  CT3  C    O      0.00   3    180.0
HA  CT3  C    NH1    0.00   3     0.00
C   NH1  CT2  HA     0.00   3     0.00

C   NH1  CT2  CT2    0.25   2      0.0
C   NH1  CT2  CT2    0.3    1      0.0

O   C    NH1   H     2.5    2     180.0
O   C    NH1  CT2    2.5    2     180.0
NH1  CT2 CT2  HA     0.1950 3      0.00
NH1  CT2 CT2  CC     0.1950 3      0.00
```

```
H    NH1 CT2  HA     0.00   3      0.00
H    NH1 CT2  CT2    0.000  1      0.00

CT2  CT2 CC    O      0.0   2      0.00
CT2  CT2 CC    NH2    1.0   1      0.0
CT2  CT2 CC    NH2    0.3   3      0.00

!CT2  CT2 CC    NH2    1.0  1      0.0

HA   CT2 CT2  HA     0.1950 3      0.00
HA   CT2 CT2  CC     0.1950 3      0.00
CT2  CC   NH2  H     2.5    2     180.0
HA   CT2 CC    O     0.00   3     180.0
HA   CT2 CC    NH2   0.00   3     0.00
O   CC   NH2   H     2.5    2     180.0
```

Topology and parameters for **netropsin and distamycin** (streamed in; used with parameters for DNA/proteins, par_all27_prot_na.inp)

```
RESI NETR    2.00  !netropsin
!1st pyrrole ring
GROUP
ATOM HG   HP     0.115 !
ATOM CG   CY    -0.145 !
ATOM CD2  CPT   -0.020 !
ATOM CD1  CA     0.035 !
ATOM HD1  HP     0.115 !
ATOM NE1  NY    -0.230 !
!ATOM HE1  H      0.380 !
ATOM CE2  CPT    0.130 !
GROUP
ATOM CH   CT3   -0.27  !
ATOM H1   HA     0.09
ATOM H2   HA     0.09
ATOM H3   HA     0.09
GROUP
ATOM N2    NH1    -0.47
ATOM HM     H      0.235
ATOM CB     C      0.745
ATOM O2     O     -0.51
!-----------------------------------------------
!2nd pyrrole ring
GROUP
ATOM HG1   HP     0.115 !
ATOM CG1   CY    -0.145 !
ATOM CD3  CPT   -0.020 !
ATOM CD4  CA     0.035 !
ATOM HD4  HP     0.115 !
ATOM NE2  NY    -0.230 !
ATOM CE3  CPT    0.130 !
GROUP
ATOM CI   CT3   -0.27
ATOM HI1   HA    0.09
ATOM HI2   HA    0.09
ATOM HI3   HA    0.09
```

```
GROUP
ATOM NK    NH1    -0.47
ATOM HN     H      0.235
ATOM CK     C      0.745
ATOM OK     O     -0.51
GROUP
ATOM NZ    NH1    -0.47
ATOM HZ     H      0.235
ATOM CZ     C      0.745
ATOM OZ     O     -0.51
!ATOM HZ     H      0.235
!----------------------------------------
GROUP
ATOM CV   CT2   -0.18
ATOM HV1  HA     0.09
!ATOM HR2  HA     0.09
ATOM HV3  HA     0.09

GROUP
ATOM CC   CT2    -0.18
ATOM HC1  HA      0.09
ATOM HC2  HA      0.09
!----------------------------------------
GROUP
ATOM CZ2  C       0.64
ATOM NH1  NC2     -0.74
ATOM HH11 HC      0.46
ATOM HH12 HC      0.46
ATOM NH2  NC2     -0.74
ATOM HH22 HC      0.46
ATOM HH21 HC      0.46
!----------------------------------------
GROUP
ATOM CC2   CT2    -0.18
ATOM HC12  HA      0.09
ATOM HC14  HA      0.09
ATOM NC    NC2    -0.47
ATOM HC    HC       0.47
ATOM CZ3  C       0.64
ATOM NH4  NC2     -0.74
ATOM HH14 HC      0.46
ATOM HH16 HC      0.46
ATOM NH5  NC2     -0.74
ATOM HH25 HC      0.46
ATOM HH27 HC      0.46

!bonds for 1st pyrrole ring
BOND CD2 HG  CD2 CG  NE1 CD1
BOND NE1 CE2
BOND CD1 HD1
DOUBLE CD1 CG  CE2 CD2
BOND CH NE1
BOND CH H1  CH H2  CH H3
BOND CE2 CK
DOUBLE CB O2
BOND N2 CG  N2 HM  N2 CB
!----------------------------------------
```

```
!bonds for 2nd pyrrole ring
BOND CD3 HG1  CD3 CG1  NE2 CD4
BOND NE2 CE3
BOND CD4 HD4
DOUBLE CD4 CG1  CE3 CD3
BOND CI NE2
BOND CI HI1  CI HI2  CI HI3
BOND CZ CE3
DOUBLE CK OK
BOND NK CG1  NK HN  NK CK
!-------------------------------------------
!1st tail amidino group
DOUBLE CZ OZ
BOND HV1 CV  HV3 CV  CV NZ  CZ NZ
BOND CC CV  CC HC1  CC HC2  NZ HZ
BOND CC CZ2  CZ2 NH2
DOUBLE CZ2 NH1
BOND NH1 HH12  NH1 HH11
BOND NH2 HH22  NH2 HH21
!-------------------------------------------
!2nd tail guanidino group
BOND CB CC2  CC2 HC12  CC2 HC14
BOND CC2 NC  NC HC
BOND NC CZ3  CZ3 NH5
DOUBLE CZ3 NH4
BOND NH4 HH14  NH4 HH16
BOND NH5 HH25  NH5 HH27

!-------------------------------------------
!ic tables
IC   CZ3   NC   CC2   CB       1.47  105.82  -180.0  109.5  1.54

ic  HM  N2  CB  CC2    0.9933 119.23    0.  116.25  1.47
ic  CG  N2  CB  CC2    1.40   119.23 -180.0 116.25  1.47
ic  N2  CB  CC2 NC     1.3418 116.25 -180.  109.5   1.47
ic  CG  N2  CB  O2     1.38   122.0   0.00  123.5   1.2233

IC  CD2  CG  N2  CB    1.40   126.4  -180.0 130.3   1.46
ic  CD1  CG  N2  CB    1.37   126.4     0.0 122.1   1.46
ic  N2   CB  CC2 HC12  1.36   115.0  -60.0  109.5  1.07
ic  N2   CB  CC2 HC14  1.36   115.0  60.0  109.5   1.07
ic  O2   CB  CC2 HC12  1.22   123.0 -114.0 109.5   1.07
ic  O2   CB  CC2 HC14  1.22   123.0  115.0 109.5   1.07

!from arginine
IC NC   CB   *CC2  HC14   1.5034 107.0900  120.6900 109.4100  1.1143
IC NC   CB   *CC2  HC12   1.5034 107.0900 -119.0400 111.5200  1.1150
IC CB   CC2   NC  CZ3    1.5384 107.0900  -180.0000 123.0500  1.3401
IC CZ3  CC2   *NC  HC    1.3401 123.0500  -180.0000 113.1400  1.0065
IC CC2   NC  CZ3   NH4   1.5034 123.0500     0.0000 118.0600  1.3311
IC NC   CZ3   NH4  HH14  1.3401 118.0600  10.2800 120.6100  0.9903
IC HH14 CZ3   *NH4 HH16  0.9903 120.6100  180.1900 116.2900  1.0023
IC NH4  NC   *CZ3  NH5   1.3311 118.0600  160.6400 122.1400  1.3292
IC NC   CZ3   NH5  HH25  1.3401 122.1400 170.1400 119.9100  0.9899
IC HH25 CZ3   *NH5 HH27  0.9899 119.9100  70.1600 116.8800  0.9914
IC CG   CD1  NE1  CE2   0.0000  0.0000  0.0000  0.0000  0.0000
IC CD2  CB   *CG  CD1   0.0000  0.0000 180.0000  0.0000  0.0000
```

```
IC CD1  CG   CD2  CE2   0.0000  0.0000  0.0000  0.0000  0.0000
IC CD2  CG   CD1  NE1   0.0000  0.0000  0.0000  0.0000  0.0000
IC CD1  CE2  *NE1 CH    0.0000  0.0000 -180.0000  127.0  1.46
IC CG   NE1  *CD1 HD1   0.0000  0.0000 180.0000  0.0000  0.0000
IC CE2  CG   CD2  HG    0.0000  0.0000 -180.0000  126.9  1.08

!ic's for methyl group
IC  CD1 NE1 CH H1          1.37    123.2   0.00    109.5   1.08
IC  CD1 NE1 CH H2          1.37    123.2  -120.0   109.5   1.08
IC  CD1 ND1 CH H3          1.37    123.2   120.    109.5   1.08
ic  CE2 NE1 CH H3          1.37    123.2   -60.0   109.5   1.08

ic  HN  NK  CK  CE2    0.9933 119.23    0.  116.25  1.47
ic  CG1 NK  CK  CE2    1.40   119.23 -180.0 116.25  1.47
ic  NK  CK  CE2 NE1    1.3418 116.25  180.  122.1  1.38
ic  NK  CK  CE2 CD2    1.3418 116.25    0.  130.3  1.40
ic  NE1 CE2 CK  OK     1.38   122.0   0.00  123.5  1.2233
ic  CD2 CE2 CK  OK     1.40   130.3   180.0 123.5  1.2233
IC  CG  CD2 CE2 CK     1.40   107.1  -180.0 130.3  1.46
ic  CD1 NE1 CE2 CK     1.37   109.8   180.0 122.1  1.46
ic  CH  NE1 CE2 CK     1.46   127.0   0.00  122.1  1.46

!ic's for amide group on CG
ic  O2  CB  N2  HM      1.2233 122.84  180.  119.23  0.9933
ic  HM  N2  CB  O2      0.9933 119.23  180.   112.25  1.23
ic  O2  CB  N2  CG      1.2233 122.84    0.   125.57  1.40
ic  CB  N2  CG  CD2     1.3418 122.57  180.   124.9  1.41
ic  CB  N2  CG  CD1     1.3418 122.57    0.   127.2  1.40
ic  N2  CG  CD2 CE2     1.40   124.9   180.   107.1  1.40
ic  N2  CG  CD1 NE1     1.40   127.2  -180.0  107.6  1.37
!-------------------------------------------------------
!ic's for 2nd pyrrole ring
IC CG1   CD4  NE2  CE3   0.0000  0.0000  0.0000  0.0000  0.0000
IC CD4  CG1   CD3  CE3   0.0000  0.0000  0.0000  0.0000  0.0000
IC CD3  CG1   CD4  NE2   0.0000  0.0000  0.0000  0.0000  0.0000
IC CD4  CE3  *NE2 CI   0.0000  0.0000 -180.0000  127.0  1.46
IC CG1   NE2  *CD4 HD4   0.0000  0.0000 180.0000  0.0000  0.0000
IC CE3  CG1   CD3  HG1    0.0000  0.0000 -180.0000  126.9  1.08

!ic's for methyl group
IC  CD4 NE2 CI HI1          1.37    123.2   0.00    109.5   1.08
IC  CD4 NE2 CI HI2          1.37    123.2  -120.0   109.5   1.08
IC  CD4 NC1 CI HI3          1.37    123.2   120.    109.5   1.08
ic  CE3 NE2 CI HI3          1.37    123.2   -60.0   109.5   1.08

ic  HZ  NZ  CZ  CE3    0.9933 119.23    0.  116.25  1.47
ic  CG2  NZ  CZ  CE3   1.40   119.23 -180.0 116.25  1.47
ic  NZ  CZ  CE3 NE2    1.3418 116.25  180.  122.1  1.38
ic  NZ  CZ  CE3 CD3    1.3418 116.25    0.  130.3  1.40
ic  NE2 CE3 CZ  OZ     1.38   122.0   0.00  123.5  1.2233
ic  CD3 CE3 CZ  OZ     1.40   130.3   180.0 123.5  1.2233
IC  CG1  CD3 CE3 CZ     1.40   107.1  -180.0 130.3  1.46
ic  CD4 NE2 CE3 CZ    1.37   109.8   180.0 122.1  1.46
ic  CI  NE2 CE3 CZ     1.46   127.0   0.00  122.1  1.46

ic  OK  CK  NK  HN      1.2233 122.84  -180.  119.23  0.9933
ic  HN  NK  CK  CE2      0.9933 119.23    0.   112.25  1.47 !!
```

```
ic  OK  CK  NK  CG1      1.2233 122.84    0.   125.57  1.40
ic  CK  NK  CG1  CD3     1.3418 122.57  180.   124.9  1.41
ic  CK  NK  CG1  CD4     1.3418 122.57    0.   127.2  1.40
IC  NK  CG1  CD3 CE3     1.40   124.9   180.   107.1  1.40
ic  NK  CG1  CD4 NE2     1.40   127.2  -180.0  107.6  1.37

!amide group and 1st tail
ic  OZ  CZ  NZ  HZ      1.2233 122.84  -180.  119.23  0.9933
ic  HZ  NZ  CZ  CE3      0.9933 119.23    0.   112.25  1.47 !!
ic  CV  NZ  CZ  CE3     1.4700  120.0   -180.0 120.0   1.47
ic  OZ  CZ  NZ  CV      1.2233 122.84    0.   125.57  1.40
ic  CZ  NZ  CV  CC      1.3418 122.57  -180.   124.9  1.41
ic  NZ  CV  CC  CZ2     1.40   124.9   60.   107.1  1.40
ic  NZ  CV  CC  HC1     1.40   127.2  -60.0   107.6  1.07
ic  NZ  CV  CC  HC2     1.40   127.2   60.0   107.6  1.07
IC  CV  CC  CZ2 NH1     1.54   109.5  -120.0  120.0  1.33
IC  CV  CC  CZ2 NH2     1.54   109.5   120.0   120.0  1.33
IC  CZ  NZ  CV  HV1     1.47   120.6   100.0   109.5  1.07
IC  CZ  NZ  CV  HV3     1.47   120.6  -18.0   109.5  1.07
IC  CC  CZ2 NH2 HH21    1.54   120.0   170.0  109.5  0.99
IC  CC  CZ2 NH2 HH22    1.54   120.0   -68.0  109.5  0.99
IC  CC  CZ2 NH1 HH11    1.54   120.0   25.0   109.5  0.99
IC  CC  CZ2 NH1 HH12    1.54   120.0  -145.0  109.5  0.99

IMPR OK CG1 NK CK
impr CZ3 NH4 NH5 NC
impr O2 NC CC2 CB
impr CZ2 NH2 NH1 CC

RESI DIST        1.00 ! distamycin
GROUP                 !
ATOM CB    CC     0.55 !
ATOM N2    NH2   -0.55 !
ATOM HM    H      0.25 !
ATOM O2    O     -0.55 !
ATOM H     H      0.30 !
!1st pyrrole ring
GROUP
ATOM HG   HP     0.115 !
ATOM CG   CY    -0.145 !
ATOM CD2  CPT   -0.020 !
ATOM CD1  CA     0.035 !
ATOM HD1  HP     0.115 !
ATOM NE1  NY    -0.230 !
ATOM CE2  CPT    0.130 !
GROUP
ATOM CH   CT3   -0.27  !
ATOM H1   HA     0.09
ATOM H2   HA     0.09
ATOM H3   HA     0.09
GROUP
ATOM NK    NH1    -0.47
ATOM HN     H      0.235
ATOM CK     C      0.745
ATOM OK     O     -0.51
!2nd pyrrole ring
GROUP
```

```
ATOM HG1   HP     0.115 !
ATOM CG1   CY    -0.145 !
ATOM CD3  CPT   -0.020 !
ATOM CD4  CA     0.035 !
ATOM HD4  HP     0.115 !
ATOM NE2  NY    -0.230 !
!ATOM HE1  H      0.380 !
ATOM CE3  CPT    0.130 !
GROUP
ATOM CI   CT3    -0.27  !
ATOM HI1   HA     0.09
ATOM HI2   HA     0.09
ATOM HI3   HA     0.09

GROUP
ATOM NZ    NH1    -0.47
ATOM HZ     H      0.235
ATOM CZ     C      0.745
ATOM OZ     O     -0.51
!3rd pyrrole ring
GROUP
ATOM HG2   HP     0.115 !
ATOM CG2   CY    -0.145 !
ATOM CD5  CPT    -0.020 !
ATOM CD6  CA      0.035 !
ATOM HD6  HP      0.115 !
ATOM NE3  NY     -0.230 !
!ATOM HE1  H      0.380 !
ATOM CE4  CPT     0.130 !
GROUP
ATOM CF   CT3     -0.27  !
ATOM HF1   HA     0.09
ATOM HF2   HA     0.09
ATOM HF3   HA     0.09
ATOM CP    C      0.51
ATOM OP    O     -0.51
ATOM NP    NH1   -0.47
ATOM HP    H      0.47
ATOM CR   CT2    -0.18
ATOM HR1  HA      0.09
ATOM HR3  HA      0.09
!GROUP
ATOM CS  CT2     -0.18
ATOM HS1  HA      0.09
ATOM HS2  HA      0.09
GROUP
ATOM CZ2  C       0.64
ATOM NH1  NC2     -0.74
ATOM HH11 HC      0.46
ATOM HH12 HC      0.46
ATOM NH2  NC2     -0.74
ATOM HH22 HC      0.46
ATOM HH21 HC      0.46
!bonds
BOND  CB N2   N2 HM  CB H  N2 CG
DOUBLE  CB O2
!bonds for 1st pyrrole ring
```

```
BOND CD2 HG  CD2 CG  NE1 CD1
BOND NE1 CE2
BOND CD1 HD1
DOUBLE CD1 CG  CE2 CD2
BOND CH NE1
BOND CH H1  CH H2  CH H3
BOND CE2 CK

!bonds for 2nd pyrrole ring
BOND CD3 HG1  CD3 CG1  NE2 CD4
BOND NE2 CE3
BOND CD4 HD4
DOUBLE CD4 CG1  CE3 CD3
BOND CI NE2
BOND CI HI1  CI HI2  CI HI3
BOND CZ CE3
DOUBLE CK OK
BOND NK CG1  NK HN  NK CK

!bonds for 3rd pyrrole ring
BOND CD5 HG2  CD5 CG2  NE3 CD6
BOND NE3 CE4
BOND CD6 HD6
DOUBLE CD6 CG2  CE4 CD5
BOND CF NE3
BOND CF HF1  CF HF2  CF HF3
BOND CP CE4
DOUBLE CZ OZ
BOND NZ CG2  NZ HZ  NZ CZ

!bonds for tail
BOND CP NP  NP CR  NP HP
BOND CR HR1   CR HR3  CR CS
DOUBLE CP OP
BOND CS HS1  CS HS2
BOND CS CZ2  CZ2 NH2
DOUBLE CZ2 NH1
BOND NH1 HH12  NH1 HH11
BOND NH2 HH22  NH2 HH21

!ic table
IMPR  CB H N2 O2   CB N2 H O2
IMPR  N2 CB CG HM   N2 CB HM CG
IC  O2   CB   N2   CG     1.2012 124.70  -18.0  118.63 1.40   !
IC  O2   CB   N2   HM     1.2012 124.70  -162.  120.92  0.9930 !
IC  CG   N2   CB   H      1.40   118.63   163.  115.65  1.34   !
IC  HM   N2   CB   H      1.0015 118.63  19.7   115.54  1.34

!ic's for 1st pyrrole ring
IC CG   CD1  NE1  CE2   0.0000  0.0000  0.0000  0.0000  0.0000
IC CD2  CB   *CG  CD1   0.0000  0.0000 180.0000  0.0000  0.0000
IC CD1  CG   CD2  CE2   0.0000  0.0000  0.0000  0.0000  0.0000
IC CD2  CG   CD1  NE1   0.0000  0.0000  0.0000  0.0000  0.0000
IC CD1  CE2  *NE1 CH   0.0000  0.0000 -180.0000  127.0  1.46
IC CG   NE1  *CD1 HD1   0.0000  0.0000 180.0000  0.0000  0.0000
IC CE2  CG   CD2  HG    0.0000  0.0000 -180.0000  126.9  1.08
```

```
!ic's for methyl group
IC  CD1 NE1 CH H1          1.37    123.2   0.00    109.5   1.08
IC  CD1 NE1 CH H2          1.37    123.2  -120.0   109.5   1.08
IC  CD1 ND1 CH H3          1.37    123.2   120.    109.5   1.08
ic  CE2 NE1 CH H3          1.37    123.2   -60.0   109.5   1.08

!ic's for amide group on CE2
!ic  O1  CA  N1  HA     1.2233 122.84  180.  119.23  0.9933
!ic  O1  CA  N1  HB     1.2233 122.84  0.00  119.23  0.9933
ic  HN  NK  CK  CE2    0.9933 119.23    0.  116.25  1.47
ic  CG1 NK  CK  CE2    1.40  119.23 -180.0 116.25  1.47
ic  NK  CK  CE2 NE1    1.3418 116.25  180.  122.1  1.38
ic  NK  CK  CE2 CD2    1.3418 116.25    0.  130.3  1.40
ic  NE1 CE2 CK  OK     1.38   122.0   0.00  123.5  1.2233
ic  CD2 CE2 CK  OK     1.40   130.3   180.0 123.5  1.2233
IC  CG  CD2 CE2 CK     1.40   107.1  -180.0 130.3  1.46
ic  CD1 NE1 CE2 CK     1.37   109.8   180.0 122.1  1.46
ic  CH  NE1 CE2 CK     1.46   127.0   0.00  122.1  1.46

!ic's for amide group on CG
ic  O2  CB  N2  HM      1.2233 122.84  -180.  119.23  0.9933
ic  HM  N2  CB  H       0.9933 119.23    0.   112.25  1.10
ic  O2  CB  N2  CG      1.2233 122.84    0.   125.57  1.40
ic  CB  N2  CG  CD2     1.3418 122.57  180.   124.9  1.41
ic  CB  N2  CG  CD1     1.3418 122.57    0.   127.2  1.40
IC  N2  CG  CD2 CE2     1.40   124.9   180.   107.1  1.40
ic  N2  CG  CD1 NE1     1.40   127.2  -180.0  107.6  1.37

!---------------------------------------------------
!ic's for 2nd pyrrole ring
IC CG1   CD4  NE2  CE3   0.0000  0.0000  0.0000  0.0000  0.0000
!IC CD2  CB   *CG  CD1   0.0000  0.0000 180.0000  0.0000  0.0000
IC CD4  CG1   CD3  CE3   0.0000  0.0000  0.0000  0.0000  0.0000
IC CD3  CG1   CD4  NE2   0.0000  0.0000  0.0000  0.0000  0.0000
IC CD4  CE3  *NE2 CI   0.0000  0.0000 -180.0000  127.0  1.46
IC CG1   NE2  *CD4 HD4   0.0000  0.0000 180.0000  0.0000  0.0000
IC CE3  CG1   CD3  HG1    0.0000  0.0000 -180.0000  126.9  1.08

!ic's for methyl group
IC  CD4 NE2 CI HI1          1.37    123.2   0.00    109.5   1.08
IC  CD4 NE2 CI HI2          1.37    123.2  -120.0   109.5   1.08
IC  CD4 NC1 CI HI3          1.37    123.2   120.    109.5   1.08
ic  CE3 NE2 CI HI3          1.37    123.2   -60.0   109.5   1.08

!ic's for amide group on CE3
!ic  OJ  CJ  NA  HC     1.2233 122.84  180.  119.23  0.9933
!ic  OJ  CJ  NA  HD     1.2233 122.84  0.00  119.23  0.9933
ic  HZ  NZ  CZ  CE3    0.9933 119.23    0.  116.25  1.47
ic  CG2  NZ  CZ  CE3    1.40 119.23 -180.0 116.25  1.47
ic  NZ  CZ  CE3 NE2    1.3418 116.25  180.  122.1  1.38
ic  NZ  CZ  CE3 CD3    1.3418 116.25    0.  130.3  1.40
ic  NE2 CE3 CZ  OZ     1.38   122.0   0.00  123.5  1.2233
ic  CD4 CE3 CZ  OZ     1.40   130.3   180.0 123.5  1.2233
IC  CG1  CD3 CE3 CZ     1.40   107.1  -180.0 130.3  1.46
ic  CD4 NE2 CE3 CZ    1.37   109.8   180.0 122.1  1.46
ic  CI  NE2 CE3 CZ     1.46   127.0   0.00  122.1  1.46
```

```
!ic's for amide group on CG1
ic  OK  CK  NK  HN      1.2233 122.84  -180.  119.23  0.9933
ic  HN  NK  CK  CE2      0.9933 119.23    0.   112.25  1.10 !!
ic  OK  CK  NK  CG1      1.2233 122.84    0.   125.57  1.40
ic  CK  NK  CG1  CD3     1.3418 122.57  180.   124.9  1.41
ic  CK  NK  CG1  CD4     1.3418 122.57    0.   127.2  1.40
IC  NK  CG1  CD3 CE3     1.40   124.9   180.   107.1  1.40
ic  NK  CG1  CD4 NE2     1.40   127.2  -180.0  107.6  1.37
!-----------------------------------------------------------
!ic's for 3rd pyrrole ring
IC CG2   CD6  NE3  CE4   0.0000  0.0000  0.0000  0.0000  0.0000
!IC CD2  CB   *CG  CD1   0.0000  0.0000 180.0000  0.0000  0.0000
IC CD6  CG2   CD5  CE4   0.0000  0.0000  0.0000  0.0000  0.0000
IC CD5  CG2   CD6  NE3   0.0000  0.0000  0.0000  0.0000  0.0000
IC CD6  CE4  *NE3 CF   0.0000  0.0000 -180.0000  127.0  1.46
IC CG2   NE3  *CD6 HD6   0.0000  0.0000 180.0000  0.0000  0.0000
IC CE4  CG2   CD5  HG2    0.0000  0.0000 -180.0000  126.9  1.08


!ic's for methyl group
IC  CD6 NE3 CF HF1          1.37    123.2   0.00    109.5   1.08
IC  CD6 NE3 CF HF2          1.37    123.2  -120.0   109.5   1.08
IC  CD6 NF1 CF HF3          1.37    123.2   120.    109.5   1.08
ic  CE4 NE3 CF HF3          1.37    123.2   -60.0   109.5   1.08

!ic's for amide group on CE4
!ic  OL  CL  NB  HE     1.2233 122.84  180.  119.23  0.9933
!ic  OL  CL  NB  HF     1.2233 122.84  0.00  119.23  0.9933
ic  HP  NP  CP  CE4    0.9933 119.23    0.  116.25  1.47
ic  CR  NP  CP  CE4    1.40  119.23 -180.0 116.25  1.47
ic  NP  CP  CE4 NE3    1.3418 116.25  180.  122.1  1.38
ic  NP  CP  CE4 CD5    1.3418 116.25    0.  130.3  1.40
ic  NE3 CE5 CP  OP     1.38   122.0   0.00  123.5  1.2233
ic  CD6 CE4 CP  OP     1.40   130.3   180.0 123.5  1.2233
IC  CG2 CD5 CE4 CP     1.40   107.1  -180.0 130.3  1.46
ic  CD6 NE3 CE4 CP    1.37   109.8   180.0 122.1  1.46
ic  CF  NE3 CE4 CP     1.46   127.0   0.00  122.1  1.46
!ic's for amide group on CG2
ic  OZ  CZ  NZ  HZ      1.2233 122.84  -180.  119.23  0.9933
ic  HZ  NZ  CZ  CE3      0.9933 119.23    0.   112.25  1.10 !!
ic  OZ  CZ  NZ  CG2      1.2233 122.84    0.   125.57  1.40
ic  CZ  NZ  CG2  CD5     1.3418 122.57  180.   124.9  1.41
ic  CZ  NZ  CG2  CD6     1.3418 122.57    0.   127.2  1.40
IC  NZ  CG2  CD5 CE4     1.40   124.9   180.   107.1  1.40
ic  NZ  CG2  CD6 NE3     1.40   127.2  -180.0  107.6  1.37

!ic's for cationic tail
IMPR  NP  CP  CR HP
IMPR  CP  CE4 NP  OP

IC  OP   CP   NP   HP      1.2233 122.84  -166. 119.23 0.9933
IC  HP   NP   CP   CE4     0.9933 119.23   14.  116.25 1.40
IC  OP   CP   NP   CR     1.2233 122.84   -15.  117.57 1.4488
IC  NP   CP   CE4  CD5    1.3418 116.25    0.  130.0  1.39
IC  NP   CP   CE4  NE3    1.3418 116.25   180. 121.3  1.39
!IC  NP   CP   CE4  HL3    1.3418 116.25  -87. 109.3  1.109
IC  CP   NP   CR  HR1    1.3418 117.00    60.  110.7  1.113  !-180.
```

```
IC  CP   NP   CR  CS     1.3418 117.00   180.   110.7  1.50
IC  CP   NP   CR  HR3    1.3418 117.00   -60.  110.7  1.113  !-150.
IC  CE4  CP   NP  CR     1.40   115.9   164.8  117.17  1.47

IC  NP  CR  CS CZ2       1.47   109.5   180.0  109.5  1.50
IC  NP  CR  CS HS1      1.47   109.5   60.0  109.5  1.10
IC  NP  CR  CS HS2      1.47   109.5   -60.0  109.5  1.10


IC  CR  CS  CZ2 NH1     1.54   109.5  -120.0  120.0  1.33
IC  CR  CS  CZ2 NH2     1.54   109.5   120.0   120.0  1.33
IC  CS  CZ2 NH2 HH21    1.54   120.0   170.0  109.5  0.99
IC  CS  CZ2 NH2 HH22    1.54   120.0   -68.0  109.5  0.99
IC  CS  CZ2 NH1 HH11    1.54   120.0   25.0   109.5  0.99
IC  CS  CZ2 NH1 HH12    1.54   120.0  -145.0  109.5  0.99
!IMPR OK CG1 NK CK
!impr CZ3 NH4 NH5 NC
!impr O2 NC CC2 CB
impr CZ2 NH2 NH1 CS

impr CG2 CZ NZ HZ
impr NZ CE3 CZ OZ

impr CG1 CK NK HN
impr NK CE2 CK OK

impr NP CE4 CP OP
impr CR CP NP HP

impr CD1 CE2 NE1 CH
impr CD4 CE3 NE2 CI
impr CD6 CE4 NE3 CF
impr NP CP CR HP
impr CP CE4 NP OP
impr OP CE4 CP NP
impr OP NP CP CE4
impr OP NE3 CE4 CP

!parameters
read parameter card append

BONDS
!
!atom types  Kb          b0
!

!polyamide bonds
NH2  CPT   500.000     1.41 !A.E.L.
CD   CA    500.000     1.45 !A.E.L.
CT3  NY    500.000     1.46 !A.E.L.
CPT  HP    500.000     1.083 !A.E.L.
NC2  CPT   500.000     1.41   !A.E.L.
CD   NH2   500.00      1.36   !A.E.L.
OS  CC     400.000      1.35   !a.E.L.
OS  CT1    400.00       1.46   !A.E.L.

NC2  H    500.000     1.00   !A.E.L.
```

```
CC   H    500.00      1.10
NH2  CY   500.00      1.40
!-----------------------------------------
!bonds for ipy drug
!N-methyl-pyrrole
CPT  HP    500.00     1.08   !A.E.L.
CT3  NY    500.000    1.46
C    CPT   500.00     1.47
NH1  CY    500.00     1.40
C    H     500.00     1.10
OS  CT1    400.00     1.46   !A.E.L.
OS  C     400.000      1.35   !a.E.L.

!imidazole
NR1 CT3   500.00      1.46   !A.E.L.
CPH2 C    500.000     1.47
CPH1 NH1  500.00      1.40
C    H    500.00      1.10
!-------------------------------------------------

ANGLES
!
!atom types     Ktheta    Theta0   Kub     S0
!
!polyamide angles
!---------------------------------------
NH2  CC   H     50.000    120.00   !
O    CC   H     44.00     122.00
CC   NH2  CY    40.       125.2
H    NH2  CY    40.       117.2

CA   CA   NY    40.00     120.0    !A.E.L.
CA   CA   CT1   40.00     120.0    !A.E.L.
CA   CT1  CT3   40.00     108.3    !A.E.L.
CPT  CPT  NH2   40.00     126.440    !A.E.L.
CPT  CT3  HA    40.00     109.50    !A.E.L.
CY  CPT  NH2    40.00     126.370    !A.E.L.
CY  CA  CD      40.00     130.110    !A.E.L.
CA  CD  OH1     40.00     112.290   !A.E.L.
CA  CD  OB      40.00      130.070  !A.E.L.
CPT  NH2 HC     40.00      109.50  !A.E.L.
NY  CPT  CT3    40.00      121.490  !A.E.L.
CPT CPT CT3     40.00      130.090  !A.E.L.
CA  NY CA       40.00      125.590  !A.E.L.
NY  CA  CD      40.00      121.470  !A.E.L.
CPT CY  CPT     40.00      35.915   !A.E.L.
CA  NY  CT3     40.0       125.60   !A.E.L.
CPT NY  CT3     40.00      125.60   !A.E.L.
NY  CT3 HA      40.0       109.5    !A.E.L.
CPT CY  NH2     40.00      126.320  !A.E.L.
CA  CY  NH2     40.00      126.320  !A.E.L.
CY  NH2 HC      40.00      109.50   !A.E.L.
CPT CPT HP      40.00      130.110  !A.E.L.
NY  CPT HP      40.00      121.470  !A.E.L.
CT2 NH2 CT3     40.00      109.5    !A.E.L.
CT3 NH2 CT3     40.00      105.970  !A.E.L.
CPT NY CPT       40.000     108.8    !A.E.L.
```

```
CA  CD   NH2     40.000      115.0    !A.E.L.
OB  CD   NH2    40.000      121.9    !A.E.L.  !!!
CD  NH2  CPT    40.000      120.6    !A.E.L.
CD  NH2  HC     40.000      121.2    !A.E.L.
NH2  CC  OS    50.00    116.50   50.000   2.45000
O    CC  OS    15.00    121.00   50.00    2.44000
CC   OS  CT1   33.00    109.50   30.000   2.16300
OS   CT1 CT3   34.500   110.10   22.53    2.17900
CPT  NH2 CC    40.000   120.6        !A.E.L.
HC   NH2 CC    40.000   121.2        !A.E.L.

HA   CT2  NH2   40.000   115.0     !A.E.L.
CT2  CT2  NH2   40.000   108.8     !A.E.L.
CT3  NC2  CT3   40.000   109.5     !A.E.L.
CT2  NC2  CT3   40.000   109.5     !A.E.L.
CT2  NC2  H     40.000   109.5     !a.E.L.
CT3  NC2  H     40.000   109.5     !A.E.L.
CD   NH2  CT2    40.000   120.6     !A.E.L.
CT2  C    NC2   40.00    120.0     !A.E.L.
C    CT2  NC2   40.00    35.00     !A.E.L.
!-----------------------------------------------------
!angles for ipy drug
!Boc group
NH1  C  OS    50.00    116.50   50.000   2.45000
O    C  OS    15.00    121.00   50.00    2.44000
C   OS  CT1   33.00    109.50   30.000   2.16300
OS   CT1 CT3   34.500   110.10   22.53    2.17900

!N-methyl-pyrrole
CPT  CY   NH1    40.000    124.9     !A.E.L.
CA   CA   NH1    40.000    127.2
HP   CPT  CY     40.000    126.9
HP   CPT  CPT    40.000    125.9
CA   NY   CT3    40.000    123.3
CPT  NY   CT3    40.000    127.
CPT  CPT  C      40.000    130.2
NY   CPT  C      40.000    122.1
NY   CT3  HA     40.000    109.5
CPT  C    O      40.000    123.5
CPT  C    NH1    40.000    114.4
H   NH1   H      40.000    115.7
CY  NH1   H      40.000    117.2
CY  NH1   C      40.00     125.2
NH1  C    H      40.00     112.0
CA   CY   NH1    40.000    127.2
CA   CY   C      40.00     100.3
CPT  CY   C      40.000    151.9
CPT  CY   CPT    40.000    36.3
!imidazole
CPH1 CPH1 NH1    40.000   129.3  !A.E.L.
NR2  CPH1 NH1    40.000   120.0
CPH1 NR1  CT3    40.000   124.7
CPH2 NR1  CT3    40.000   127.8
NR1  CPH2 C      40.000   123.8
NR2  CPH2 C      40.000   124.8
NR1  CT3  HA     40.000   107.4
CPH2 C    O      40.000   122.6
```

```
CPH2 C    NH1    40.000   112.9
H   NH1   H      40.000   121.1
CPH1 NH1  H      40.000   115.7
CPH1 NH1  C      40.000   124.5
NH1  C    H      40.000   112.4

!cationic tail
CT2 NH3 CT3    40.000     109.5  !A.E.l.
CT3 NH3 CT3    40.00      110.8

CA  CY  CC     40.0       100.3
CPT CY  CC     40.0       151.9
!-----------------------------------------------------

DIHEDRALS
!
!atom types             Kchi    n   delta
!
!polyamide dihedrals
!-----------------------------------------------------------
CA  CA  CT1 CT3     0.2000  1  169.040 !A.E.L.
CA  NY  CA  CY      0.2000  1  180.00 !A.E.L.
CA  NY  CA  CD      0.2000  1  0.000  !A.E.L.
CA  NY  CPT CT3     0.2000  1  0.000 !A.E.L.
CA  CA  CA  CT1     0.2000  1  180.0  !A.E.L.
CA  CA  NY  CPT     0.2000  1  90.0   !A.E.L.
CA  CA  CA  NY      0.2000  1  180.00 !A.E.L.
HP  CA  CA  CT1     0.2000  1  0.000  !A.E.L.
CT1 CA  CA  NY      0.2000  1  0.000  !A.E.L.
HP  CY  CPT NH2     0.2000  1  0.000  !A.E.L.
HP  CY  CA  CD      0.2000  1  0.000  !A.E.L.
CY  CPT NH2 HC      0.2000  1  0.00 !A.E.L.  !


CPT CY  CA  CD      0.6     1  -180.00 !A.E.L.    !0.2000 1 -180.0
CPT CPT CT3 HA      0.2000  1  0.000  !A.E.L.
CA  CA  NY CA       0.2000  1  -90.0   !A.E.L.
CA  CY  CPT  NH2    0.2000  1  180.00  !A.E.L.

NY  CPT CT3  HA     0.0000  1 -180.00  !A.E.L.

CPT  CPT  NH2  HC   0.00  1    -180.0  !A.E.L. !180
CPT  CPT  NH2  HC   0.00  3     0.00


CPT  NY CA  CD      0.6000  1    180.00  !A.E.L.   !0.200  1  180.0
NH2  CPT CPT CT3    0.2000  1    0.000 !A.E.L.

CY  CA NY CT3       0.2000  1   180.0  !A.E.L.
CPT CPT NY CT3      0.2000  1   -180.0 !A.E.L.

CA  NY CT3 HA       0.20000  1    120.0 !A.E.L.

CPT NY CT3 HA       0.000   1    -60.0 !A.E.L.
CPT NY CT3 HA       0.000   1    0.00  !A.E.L.
CPT NY CT3 HA       0.000   1    -120.0 !A.E.L.
```

```
CPT CY NH2 HC       0.2000  1    -180.0 !A.E.L.
CA  CY NH2 HC       0.2000  1    -180.00 !A.E.L.
NY  CA  CY NH2      0.2000  1     180.0 !A.E.L.
CPT CPT CY NH2      0.2000  1     180.0 !A.E.L.
CA  NY  CPT HP      0.2000  1     180.0 !A.E.L.
HP  CPT NY CT3      0.2000  1       0.0     !A.E.L.
CT3 NY  CA CD       0.2000  1       0.0     !A.E.L.
CT2 CT2 NH2 CT3     0.2000  1     180.0     !A.E.L.
HA  CT2 NH2 CT3     0.2000  1      60.0     !A.E.L.


CA  CD  NH2 CPT     1.6000   1       0.0     !A.E.L.
CA  CD  NH2 CPT     2.5      2      180.0    !A.E.L.
CA  CD  NH2 HC      1.4      2        0.0     !A.E.L.  !2.5 180.0

!CY  CPT NH2 CD      0.2000   1     180.0     !A.E.L.
CY  CPT NH2 CD      1.5      2     180.0

CPT CPT NH2 CD      0.9650  1     180.00      !A.E.L.
CPT CPT NH2 CD      3.8500  2     180.0
!CPT CPT NH2 CD      .2     0      0.00


CPT NH2 CD  OB      1.5      2     0.0     !A.E.L.
HC  NH2 CD  OB      3.5      2     180.0     !A.E.L.
!------------------------------------------------------------
NH2 CC  OS  CT1     0.20       1      -180.0  !A.E.L.

O   CC  OS  CT1      0.9650      1    180.0 !A.E.L.
O   CC  OS  CT1      2.8500      2    180.0 !A.E.L.

CT3 CT1 OS CC        0.6        1     0.00  !A.E.L.

CY  CPT NH2 CC       0.200      1    180.0  !A.E.L.

CPT CPT NH2 CC       0.9650     1    180.0
CPT CPT NH2 CC       3.8500     2    180.0

CPT NH2 CC  O        1.5        2      0.0  !A.E.L.

HC  NH2 CC  O        2.5        2    180.0  !A.E.L.
HC  NH2 CC  OS       1.4        2      0.0  !a.E.L>

CPT NH2 CC OS        1.600      1    0.00
CPT NH2 CC OS        2.500      2   180.0
!---------------------------------------------------------------------
HA  CT2 NH2 HC        0.200      1.5   -120.00   !!!!
HA  CT2 NH2 HC        0.200      1     120.0

CT2 CT2 NH2 HC         1.500       1      60.000

CT2 CT2 NC2 H         0.9          3.7    0.00
CT2 CT2 NC2 H         0.0           0     60.00

NH2 CT2 CT2 CT2      1.0           1      180.0
NH2 CT2 CT2 CT2      1.3          1.5    -180.0
NH2 CT2 CT2 CT2      0.5          3       0.00
```

```
CA  CD  NH2  CT2     1.6    1    0.00
CA  CD  NH2  CT2     2.5    2   180.0

CD  NH2 CT2  HA     0.2000  1   60.0
CD  NH2 CT2  CT2    0.200   1   180.0

OB  CD  NH2  CT2    1.5     2   0.0


CY  CA  CD   NH2    2.3   2    -180.00  !A.E.L.
NY  CA  CD   NH2    2.3  2     -180.0   !A.E.L.
CY  CA  CD   OB     2.3  2      180.00  !A.E.L.
NY  CA  CD   OB     2.3  2       180.00  !A.E.L.
!-----------------------------------------------------------
!dihedrals for ipy drug
!N-methyl-pyrrole

!-----------------------------------------------
CA  NY   CT3  HA     0.00     1       0.00   !methyl group
CA  NY   CT3  HA     0.000    3     -120.0
CA  NY   CT3  HA     0.000    2     -60.0
!-----------------------------------------------
!-----------------------------------------
CPT NY   CT3  HA     0.055   3        0.0    !methyl group
CPT NY   CT3  HA     0.00    1      -120.
CPT NY   CT3  HA     0.00    2       -60.0
!-----------------------------------------

HP  CPT  CY   CA     2.0     1        180.0
HP  CPT  CY   NH1    2.8     2        180.0

CY  CA   NY   CT3    0.8    2        180.0
CY  NH1  C     O     1.5    2        180.0
CY  NH1  C     H     1.20   1        180.0
CPT CY   NH1   H     0.000  1         0.00
CPT CY   NH1   C     0.200  1        180.0

CPT  CPT  NY   CT3   0.8    2        180.0

CPT  CPT  C    O     2.5    2        180.0

CPT  CPT  C    NH1   1.5    2        180.0

CA   NY   CPT  C     2.00   1        180.0

HP   CA   CY   NH1   1.00   2        180.0

CA   CY   NH1  H     0.0    1        0.00

CA   CY   NH1  C     2.5    2       180.0
CA   CY   NH1  C     3.0    1       180.0

HP   CA   NY   CT3   0.4000  2      180.0
NY   CA   CY   NH1   2.0     2      180.0
```

```
NY   CPT  C    O     1.3   2      180.0
NY   CPT  C    O     3.7    1      180.0


NY   CPT  C    NH1   1.0     1     0.0

CPT  CPT  CY   NH1   3.0    2      180.0
CT3  NY   CPT  C     0.8     2      180.0

H    NH1  C    H     1.4     2      180.0

CPT  C    NH1   H      2.5     2      180.0
!-----------------------------------------------
!linking of boc and pyrrole
OS   C    NH1  CY     1.20   1        180.0
H    NH1  C    OS     1.4    2        180.0

CT1  OS   C    NH1    0.20   1       -180.0  !A.E.L.

O   C  OS  CT1       1.9650  1        180.0 !A.E.L.
O   C  OS  CT1       3.8500  2        180.0 !A.E.L.

CT3 CT1 OS  C         1.5     1         0.00
!-----------------------------------------------
CPT  C  NH1 CY      2.5     2     180.0
CPT  C  NH1 CT2     2.5     2     180.0
!-----------------------------------------------
!N-methyl-imidazole
!-----------------------------------------------------
CPH1 NR1  CT3  HA      0.00  1     0.00
CPH1 NR1  CT3  HA      0.000  3     -120.0     !methyl group
CPH1 NR1  CT3  HA      0.000  2    -60.0
!-----------------------------------------------------
!-----------------------------------------------
CPH2 NR1 CT3 HA        0.000   3    0.
CPH2 NR1 CT3 HA        0.0000   1   -120.0     !methyl group
CPH2 NR1 CT3 HA        0.0000   2   -60.0
!-----------------------------------------------

CPH1 CPH1 NH1 H        0.1   2   -180.0  !A.E.L.
CPH1 CPH1 NH1 H        0.00  1    0.00

CPH1 CPH1 NH1 C        1.5     1   180.0
CPH1 CPH1 NH1 C        1.0    2      00.0

HR3  CPH1 CPH1 NH1     0.5000  2    0.00

CPH1 NH1  C    O       0.2000  2    0.00
CPH1 NH1  C    O       3.000  1   180.0

CPH1 NH1  C   H        0.80000  2   -180.0
CPH1 NH1  C   H        0.5000   1     0.00
NR1  CPH1 CPH1 NH1     1.0000   1   -180.0
CPH2 NR2 CPH1 NH1      0.2000   1    -180.0
NR2 CPH1 NH1  H        1.3      2   0.0
NR2 CPH1 NH1  C        1.3      2  -180.0
H   NH1  C   H         0.2000   2    0.00
```

```
CPH1 CPH1 NR1  CT3     0.2000  1   -180.0
O  C  NH1  H          2.5000  2    180.0

!-----------------------------------------------

CPH1 NR1  CPH2 C       0.2000  1     -180.0  !
HR3  CPH1 NR1  CT3     0.2000  1    0.00    !
CPH1 NR2  CPH2 C       0.2000  1    -180.0

NR1  CPH2 C   O        5.000    2   180.0   !5.5  !5.0
NR1  CPH2 C   O        4.500    1   180.0   !3.00

NR1  CPH2 C  NH1       1.000    2   0.0  !
NR2 CPH2 C   O         1.0000   1    0.0    !
NR2 CPH2 C  NH1        0.0000   2   0.00    !

CPH2 C   NH1  H        0.2000   1    0.00
CPH2 C   NH1  H        0.2000   2   180.0

NR2 CPH2 NR1 CT3       0.2000   1  180.0
CT3 NR1  CPH2 C        0.2000   1   0.00

CPH2 C  NH1 CY         0.2000  1   0.00
CPH2 C  NH1 CY         0.2000  2   180.0
!------------------------------------------------------
!netropsin
!CY  NH1 C   CT2        0.2000  1   0.00
CY  NH1 C   CT2        1.000  1   -180.0
O   C   CT2 NC2        0.5000  1   0.000
CT2 CT2 C   NC2        0.000   1   0.00
NH1 C   CT2 NC2        0.6000  1   0.00
HA  CT2 C   NC2        0.0000  3   0.00
!------------------------------------------------------
!distamycin
CC  NH2  CY  CPT       0.2000  1   180.0
CC  NH2  CY  CA        2.5     2   180.0
CC  NH2  CY  CA        3.7     1   180.0

NH2 CY   CPT HP        2.8     2   180.0
NH2 CY   CA  HP        1.00    1   180.0

H   NH2  CC  H         1.4     2   180.0
H   NH2  CY  CPT       0.050   1   0.00

H   NH2  CY  CA        0.05    1   0.00
O   CC   NH2 CY        1.50    2   180.0
H   CC   NH2 CY        1.20    1   180.0

IMPROPER
!atom types             Kpsi        psi0

!polyamide impropers
CA   CA   CA   NY     90.000          0     -180.00 !A.E.L.
CPT  CA   NY   CA     90.000          0      180.00 !A.E.L.
CY   CPT  CPT  NH2    90.000          0      180.00 !A.E.L.
CA   CA   CA   CT1    90.000          0      180.00 !A.E.L.
OH1  CA   CD   OB     90.000          0      180.0  !A.E.L.
```

```
CD   NY   CA   CY     90.000          0      180.0  !A.E.L.
CT3  NY   CPT  CPT    90.000          0     -180.0  !A.E.L.
CA  CPT   NY   CT3    90.000          0     -180.0  !A.E.L.
CA  CPT   CY   NH2    90.000          0      180.00 !A.E.L.
CPT CY    CPT  NH2    90.000          0     -180.0  !A.E.L.
CPT NY    CPT  HP     90.000          0      180.0  !A.E.L.

OB   CA   CD    OH1   90.00           0     -180.0  !A.E.L.

OH1  OB   CD    CA    90.00           0     -180.0  !A.E.L.
OB   OH1  CD    CA    90.00           0      180.0  !A.E.L.
CY   CD   OH1   H     90.000          0      -180.0 !A.E.L.
HP   CA   CD    OH1   90.00           0       0.00  !A.E.L.
OH1  CA   CY   HP     90.000          0       0.00  !A.E.L.
OB   CA   NY   CT3    90.000          0       0.00  !A.E.L.
NH2  CT2  CT2  HA     90.000          0      -180.0 !A.E.L.
OB   NH2  CT2  CT2    90.000          0       180.0 !A.E.L.
CA  OB   CD    OH1    90.000          0       180.0 !A.E.L.
NY  CD   OH1   H      90.000          0       180.0 !A.E.L.
CY  CD   CA    NY     90.000          0      -180.0 !A.E.L.
H   CD   CA    CA     90.000          0        0.000!A.E.L.
H   CD   CA    NY     90.000          0       -180.0 !A.E.L.
CD  CY   CA    NY     90.000          0        180.0 !A.E.L.
CD  CA   CPT   CPT    90.000          0        0.00  !A.E.L.
CA  CA   NY    CPT    90.000          0       -180.0 !A.E.L.
CA  CA   CA    HP     90.000          0        180.0 !A.E.L.
CT1 CA   CA    HP     90.000          0       -180.0 !A.E.L.
HC  HC  NH2 CPT       90.000          0         -120.0 !A.E.L.
HC  CPT NH2 HC        90.000          0          120.0 !A.E.L.
CA  CPT NH2 HC        90.000          0           -20.33 !A.E.L.

NY CD  NH2 CPT        90.00           0        180.0  !A.E.L.

NH2  CD  CPT HC        20.000          0        0.00  !A.E.L.
CD   CA  NH2 OB        45.000          0        0.00  !A.E.L.
CA   CPT NH2 HC        90.000          0        0.00  !A.E.L.

CC  OS   NH2   O      45.000          0      0.00  !A.E.L.
CC  NH2  OS    O      45.000          0      0.00  !A.E.L.
OS  NH2  CC    O      90.000          0     -180.0 !A.E.L.
OS  O    CC    NH2    90.00           0      180.0 !A.E.L.
CT3 OS   CT1   CT3    90.000        0      -120.0 !A.E.L.

NH2 CC  CPT  HC       20.00         0       0.00  !A.e.L.

NH2 CD  CT2  HC      20.000         0       0.00  !A.E.L.
CPT CA  CY  HP       100.000        0       180.0
NY  CPT CPT HP       100.000        0      -180.0
HP  CPT CY  CA       100.000        0       180.0
HP  CA  CY  CPT      100.000        0      -180.0

CD  CA  NH2  OB       45.000         0         0.00
CD  NH2 CA   OB       45.000         0         0.000
!-------------------------------------------------------
!impropers for ipy drug
!Boc Binder
NH1  C   H   CY       4.00           0      0.00
```

```
!--------------------
!imid
CT3 CPH1 NR1 CPH2    90.000          0      180.0
NR1 CPH1 CPH2 CT3    90.000         0       0.00
NR1 CPH2 CPH1 CT3    90.000         0       0.00
CPH2 NR1 NR2  C      90.000         0       0.00
CPH2 NR2 NR1  C      90.000         0       0.00
O CPT C NH1          90.00          0       -180.0

CPH1 CPH2 NR1 CT3    90.000         0       180.0
O   CT2  C   NH1     90.000         0       -180.0
O   NH1  C   CT2     90.000         0       180.0
CY  C   NH1  H       90.00          0       180.0
C  CA  CY  NH1       90.000         0       0.00
H  C   NH1 CT2       90.000         0       180.0
NC2 O  C   CT2       90.000         0       0.00
C NC2 CT2  HA        90.000         0       120.0
NH1 C CT2  CT2       90.0           0        0.00
C NC2 NC2 CT2        90.00          0        0.00
!-------------------------------------------------------
!distamycin
NH2  CC  H   CY        4.0            0        0.000
NH1  CPT C   O         90.0           0        180.0
O    NH1 C   CPT       90.0           0        180.0
CT2  C   NH1 H         90.0           0        180.0
CC   H   NH2 O         120.0          0         0.00
CC   NH2  H  O         120.0          0         0.00



NONBONDED nbxmod  5 atom cdiel shift vatom vdistance vswitch -
cutnb 14.0 ctofnb 12.0 ctonnb 10.0 eps 1.0 e14fac 1.0 wmin 1.5
                !adm jr., 5/08/91, suggested cutoff scheme
!
!atom  ignored    epsilon      Rmin/2   ignored   eps,1-4
Rmin/2,1-4
!

end
```

Topology and parameters for **Dervan polyamides**; **IPY** is the imidazole-pyrrole polyamide and **HP2** is the hydroxypyrrole polyamide (streamed in; used with parameters for DNA/proteins, par_all27_prot_na.inp).

```
!------------------------------
RESI IPY       1.00
GROUP
ATOM CZ1   CPH1  0.06 !
ATOM HE1   HR3    0.09 !
ATOM CZ2  CPH1   0.22 !
ATOM NZ1  NR1   -0.36 !
ATOM CZ3  CPH2   0.25 !
ATOM NZ2  NR2   -0.36 !
ATOM HE2  HR3    0.10

GROUP
ATOM CZ4   CT3    -0.27
ATOM HZ1   HA     0.09
ATOM HZ2   HA     0.09
ATOM HZ3   HA     0.09
!----------------------------
!1st pyrrole ring
GROUP
ATOM HG   HP     0.115 !
ATOM CG   CY    -0.145 !
ATOM CD2  CPT   -0.020 !
ATOM CD1  CA     0.035 !
ATOM HD1  HP     0.115 !
ATOM NE1  NY    -0.230 !
!ATOM HE1  H      0.380 !
ATOM CE2  CPT    0.130 !

GROUP
ATOM CH   CT3   -0.27  !change C1 to CH
ATOM H1   HA     0.09
ATOM H2   HA     0.09
ATOM H3   HA     0.09
GROUP
ATOM N2    NH1    -0.47
ATOM HM     H      0.235
ATOM CB     C      0.745
ATOM O2     O     -0.51
!ATOM HZ     H      0.235
!-----------------------------------------
!2nd pyrrole ring
GROUP
ATOM HG1   HP     0.115 !
ATOM CG1   CY    -0.145 !
ATOM CD3  CPT   -0.020 !
ATOM CD4  CA     0.035 !
ATOM HD4  HP     0.115 !
ATOM NE2  NY    -0.230 !
!ATOM HE1  H      0.380 !
ATOM CE3  CPT    0.130 !
GROUP
ATOM CI   CT3   -0.27  !change C1 to CH
```

```
ATOM HI1   HA     0.09
ATOM HI2   HA     0.09
ATOM HI3   HA     0.09

GROUP
ATOM NK    NH1    -0.47
ATOM HN     H      0.235
ATOM CK     C      0.745
ATOM OK     O     -0.51
!ATOM HZ     H      0.235
!-----------------------------------------
!3rd pyrrole ring
GROUP
ATOM HG2   HP     0.115 !
ATOM CG2   CY    -0.145 !
ATOM CD5  CPT   -0.020 !
ATOM CD6  CA     0.035 !
ATOM HD6  HP     0.115 !
ATOM NE3  NY    -0.230 !
!ATOM HE1  H      0.380 !
ATOM CE4  CPT    0.130 !

GROUP
ATOM CF   CT3   -0.27  !change C1 to CH
ATOM HF1   HA     0.09
ATOM HF2   HA     0.09
ATOM HF3   HA     0.09

GROUP
ATOM CL    C      0.51
ATOM OL    O     -0.51
GROUP
ATOM NZ    NH1    -0.47
ATOM HZ     H      0.235
ATOM CZ     C      0.745
ATOM OZ     O     -0.51
!ATOM HZ     H      0.235
!-----------------------------------------
!beta-alanine
GROUP
ATOM NB    NH1   -0.47
ATOM HB    H      0.47
ATOM CV   CT2   -0.18
ATOM HV1  HA     0.09
!ATOM HR2  HA     0.09
ATOM HV3  HA     0.09

GROUP
ATOM CC   CT2    -0.18
ATOM HC1  HA      0.09
ATOM HC2  HA      0.09
GROUP
ATOM CT   C       0.51
ATOM OT   O      -0.51
!ATOM NS   NH1     -0.47
!ATOM HS1   H       0.235
!ATOM HS2   H       0.235
```

```
!----------------------------------------
!cationic tail
GROUP
ATOM NC    NH1   -0.47
ATOM HC    H      0.47
ATOM CE   CT2   -0.11
ATOM HM1  HA     0.09
ATOM HM3  HA     0.09

GROUP
ATOM CJ  CT2     -0.11
ATOM HJ1  HA      0.09
AtOM HJ2  HA      0.09
ATOM CM   CT2    -0.11
ATOM HN1  HA      0.09
ATOM HN2  HA      0.09
GROUP
ATOM NN   NH3     -0.30
ATOM HH   HC       0.33
ATOM CX   CT3      0.23
ATOM HX1  HA       0.05
ATOM HX2  HA       0.05
ATOM HX3  HA       0.05
ATOM CY   CT3      0.23
ATOM HY1  HA       0.05
ATOM HY2  HA       0.05
ATOM HY3  HA       0.05


!----------------------------------------
!bonds for imidazole
BOND NZ2 CZ2  NZ1 CZ1  CZ3 NZ1
BOND CZ1 HE1  CZ2 HE2
DOUBLE NZ2 CZ3  CZ2 CZ1
BOND NZ1 CZ4  CZ4 HZ1  CZ4 HZ2  CZ4 HZ3
BOND CZ3 CB
!-------------------------------------
!bonds for 1st pyrrole ring
BOND CD2 HG  CD2 CG  NE1 CD1
BOND NE1 CE2
BOND CD1 HD1
DOUBLE CD1 CG  CE2 CD2
BOND CH NE1
BOND CH H1  CH H2  CH H3
BOND CE2 CK
DOUBLE CB O2
BOND N2 CG  N2 HM  N2 CB
!-----------------------------------------
!bonds for 2nd pyrrole ring
BOND CD3 HG1  CD3 CG1  NE2 CD4
BOND NE2 CE3
BOND CD4 HD4
DOUBLE CD4 CG1  CE3 CD3
BOND CI NE2
BOND CI HI1  CI HI2  CI HI3
BOND CZ CE3  !NA CJ  NA HC  NA HD
DOUBLE CK OK
```

```
BOND NK CG1  NK HN  NK CK  !CB HZ
!-----------------------------------------
!bonds for 3rd pyrrole ring
BOND CD5 HG2  CD5 CG2  NE3 CD6
BOND NE3 CE4
BOND CD6 HD6
DOUBLE CD6 CG2  CE4 CD5
BOND CF NE3
BOND CF HF1  CF HF2  CF HF3
BOND CL CE4  CL NB  !NB HE  NB HF
DOUBLE CZ OZ CL OL
BOND NZ CG2  NZ HZ  NZ CZ  !CB HZ
!-----------------------------------------
!bonds for beta-alanine
BOND  NB CV
BOND  NB HB
BOND  HV1 CV  HV3 CV
DOUBLE CT OT
BOND CV CC  CC CT  !CT NS
BOND CT NC   !NS HS1  NS HS2
BOND CC HC1  CC HC2
!-----------------------------------------
!bonds for cationic tail
BOND NC HC   NC CE
!BOND CL HL1  CL HL2  CL HL3
BOND CE HM1   CE HM3  CE CJ
!DOUBLE C O
BOND CJ HJ1  CJ HJ2  CJ CM
BOND CM HN1  CM HN2  CM NN
BOND NN HH  NN CX  NN CY
BOND CX HX1  CX HX2  CX HX3
BOND CY HY1  CY HY2  CY HY3

!-----------------------------------------
!ic's for imidazole ring
! KEEPS HYDROGENS IN RING PLANE
!IMPH NZ1  CZ1  CZ3  CZ4       NZ1  CZ3  CZ1   CZ4
IMPH CZ2  CZ1   NZ2  HE2       CZ2  NZ2  CZ1   HE2
IMPH CZ3  NZ1  NZ2  CB       CZ3  NZ2  NZ1  CB
IMPH CZ1   CZ2  NZ1  HE1        CZ1   NZ1  CZ2  HE1
IC   HE1    CZ1    NZ1   CZ3    1.07    122.67  -173.67  109.79
1.2987
IC   CZ1    NZ1   CZ3   NZ2     1.2854  109.79    0.21   110.31
1.3071
IC   NZ1   CZ3   NZ2   CZ2      1.2987  110.31    0.03   105.82
1.3165
IC   CZ3   NZ2   CZ2   CZ1      1.3071  105.82   -0.23  108.68   1.3758
IC   NZ2   CZ2   CZ1   NZ1      1.3165  108.68    0.35  105.39   1.2854
IC   NZ2   CZ2   CZ1    HE1     1.3165  108.68  172.86  131.52   1.07
IC   CZ2   CZ1    NZ1   CZ3     1.3758  105.39   -0.34  109.79   1.2987
IC   CZ2   NZ2   CZ3   CB       1.3165  105.82  -180.0  124.7    1.46
IC   NZ2   CZ3   NZ1   CZ4      1.34    110.31   180.0  127.8    1.46
IC   HE1   CZ1   CZ2   HE2      1.07    131.52    0.00  129.3    1.07

!ic's for methyl group
IC  CZ1 NZ1 CZ4 HZ1          1.37    124.7   0.00    109.5   1.08
IC  CZ1 NZ1 CZ4 HZ2          1.37    124.7  -120.0   109.5   1.08
```

```
IC  CZ1 NZ1 CZ4 HZ3          1.37    124.7   120.    109.5   1.08
ic  CZ3 NZ1 CZ4 HZ3         1.37    124.7   -60.0   109.5   1.08

impr CZ4 CZ1 NZ1 CZ3  !for now
!impr CZ1 CZ3 NZ1 CZ4

!ic's for amide group on CE1
IMPR  CB CZ3 N2 O2   !for now

!ic  O3  CC  N3  HC     1.2233 122.84  -180.  119.23  0.9933
!ic  O3  CC  N3  HD     1.2233 122.84  0.00  119.23  0.9933
ic  HM  N2  CB  CZ3    0.9933 119.23    0.  116.25  1.47
ic  CG  N2  CB  CZ3    1.40   119.23  180.0 116.25  1.47
ic  N2  CB  CZ3 NZ1    1.3418 116.25  180.  122.1   1.38
ic  N2  CB  CZ3 NZ2    1.3418 116.25    0.  130.3   1.40
ic  NZ1 CZ3 CB  O2     1.38   122.0   0.00  123.5   1.2233
ic  NZ2 CZ3 CB  O2     1.40   130.3   180.0 123.5   1.2233
IC  CZ2 NZ2 CZ3 CB     1.40   107.1  -180.0 130.3   1.46
ic  CZ1  ND1 CZ3 CB    1.37   109.8   -180.0 122.1   1.46
ic  CZ4  ND1 CZ3 CB    1.46   127.0   0.00  122.1   1.46

!!impr O3 ND1 CE1 CC  !for now
!impr CC ND1 CE1 NE2
!-------------------------------------------------
!ic's for 1st pyrrole ring
IC CG   CD1  NE1  CE2   0.0000  0.0000  0.0000  0.0000  0.0000
IC CD2  CB   *CG  CD1   0.0000  0.0000 180.0000  0.0000  0.0000
IC CD1  CG   CD2  CE2   0.0000  0.0000  0.0000  0.0000  0.0000
IC CD2  CG   CD1  NE1   0.0000  0.0000  0.0000  0.0000  0.0000
IC CD1  CE2  *NE1 CH   0.0000  0.0000 -180.0000  127.0  1.46
IC CG   NE1  *CD1 HD1   0.0000  0.0000 180.0000  0.0000  0.0000
IC CE2  CG   CD2  HG    0.0000  0.0000 -180.0000  126.9  1.08

!ic's for methyl group
IC  CD1 NE1 CH H1          1.37    123.2   0.00    109.5   1.08
IC  CD1 NE1 CH H2          1.37    123.2  -120.0   109.5   1.08
IC  CD1 ND1 CH H3          1.37    123.2   120.    109.5   1.08
ic  CE2 NE1 CH H3          1.37    123.2   -60.0   109.5   1.08

!ic's for amide group on CE2
!!ic  O1  CA  N1  HA     1.2233 122.84  180.  119.23  0.9933
!!ic  O1  CA  N1  HB     1.2233 122.84  0.00  119.23  0.9933
ic  HN  NK  CK  CE2    0.9933 119.23    0.  116.25  1.47
ic  CG1 NK  CK  CE2    1.40   119.23 -180.0 116.25  1.47
ic  NK  CK  CE2 NE1    1.3418 116.25  180.  122.1  1.38
ic  NK  CK  CE2 CD2    1.3418 116.25    0.  130.3  1.40
ic  NE1 CE2 CK  OK     1.38   122.0   0.00  123.5  1.2233
ic  CD2 CE2 CK  OK     1.40   130.3   180.0 123.5  1.2233
IC  CG  CD2 CE2 CK     1.40   107.1  -180.0 130.3  1.46
ic  CD1 NE1 CE2 CK     1.37   109.8   180.0 122.1  1.46
ic  CH  NE1 CE2 CK     1.46   127.0   0.00  122.1  1.46

!ic's for amide group on CG
ic  O2  CB  N2  HM      1.2233 122.84  -180.  119.23  0.9933
ic  HM  N2  CB  O2      0.9933 119.23   180.   112.25  1.23
ic  O2  CB  N2  CG      1.2233 122.84    0.   125.57  1.40
ic  CB  N2  CG  CD2     1.3418 122.57  180.   124.9  1.41
```

```
ic  CB  N2  CG  CD1     1.3418 122.57    0.   127.2  1.40
IC  N2  CG  CD2 CE2     1.40   124.9   180.   107.1  1.40
ic  N2  CG  CD1 NE1     1.40   127.2  -180.0  107.6  1.37
!-------------------------------------------------------
!ic's for 2nd pyrrole ring
IC CG1   CD4  NE2  CE3   0.0000  0.0000  0.0000  0.0000  0.0000
!IC CD2  CB   *CG  CD1   0.0000  0.0000 180.0000  0.0000  0.0000
IC CD4  CG1   CD3  CE3   0.0000  0.0000  0.0000  0.0000  0.0000
IC CD3  CG1   CD4  NE2   0.0000  0.0000  0.0000  0.0000  0.0000
IC CD4  CE3  *NE2 CI   0.0000  0.0000 -180.0000  127.0  1.46
IC CG1   NE2  *CD4 HD4   0.0000  0.0000 180.0000  0.0000  0.0000
IC CE3  CG1   CD3  HG1    0.0000  0.0000 -180.0000  126.9  1.08

!ic's for methyl group
IC  CD4 NE2 CI HI1          1.37    123.2   0.00    109.5   1.08
IC  CD4 NE2 CI HI2          1.37    123.2  -120.0   109.5   1.08
IC  CD4 NC1 CI HI3          1.37    123.2   120.    109.5   1.08
ic  CE3 NE2 CI HI3          1.37    123.2   -60.0   109.5   1.08

!ic's for amide group on CE3
!!ic  OJ  CJ  NA  HC     1.2233 122.84  180.  119.23  0.9933
!!ic  OJ  CJ  NA  HD     1.2233 122.84  0.00  119.23  0.9933
ic  HZ  NZ  CZ  CE3    0.9933 119.23    0.  116.25  1.47
ic  CG2  NZ  CZ  CE3   1.40   119.23 -180.0 116.25  1.47
ic  NZ  CZ  CE3 NE2    1.3418 116.25  180.  122.1  1.38
ic  NZ  CZ  CE3 CD3    1.3418 116.25    0.  130.3  1.40
ic  NE2 CE3 CZ  OZ     1.38   122.0   0.00  123.5  1.2233
ic  CD3 CE3 CZ  OZ     1.40   130.3   180.0 123.5  1.2233
IC  CG1  CD3 CE3 CZ     1.40   107.1  -180.0 130.3  1.46
ic  CD4 NE2 CE3 CZ    1.37   109.8   180.0 122.1  1.46
ic  CI  NE2 CE3 CZ     1.46   127.0   0.00  122.1  1.46
!ic's for amide group on CG1
ic  OK  CK  NK  HN      1.2233 122.84  -180.  119.23  0.9933
ic  HN  NK  CK  CE2      0.9933 119.23    0.   112.25  1.47 !!
ic  OK  CK  NK  CG1      1.2233 122.84    0.   125.57  1.40
ic  CK  NK  CG1  CD3     1.3418 122.57  180.   124.9  1.41
ic  CK  NK  CG1  CD4     1.3418 122.57    0.   127.2  1.40
IC  NK  CG1  CD3 CE3     1.40   124.9   180.   107.1  1.40
ic  NK  CG1  CD4 NE2     1.40   127.2  -180.0  107.6  1.37
!-------------------------------------------------------
!ic's for 3rd pyrrole ring
IC CG2   CD6  NE3  CE4   0.0000  0.0000  0.0000  0.0000  0.0000
!IC CD2  CB   *CG  CD1   0.0000  0.0000 180.0000  0.0000  0.0000
IC CD6  CG2   CD5  CE4   0.0000  0.0000  0.0000  0.0000  0.0000
IC CD5  CG2   CD6  NE3   0.0000  0.0000  0.0000  0.0000  0.0000
IC CD6  CE4  *NE3 CF   0.0000  0.0000 -180.0000  127.0  1.46
IC CG2   NE3  *CD6 HD6   0.0000  0.0000 180.0000  0.0000  0.0000
IC CE4  CG2   CD5  HG2    0.0000  0.0000 -180.0000  126.9  1.08
!ic's for methyl group
IC  CD6 NE3 CF HF1          1.37    123.2   0.00    109.5   1.08
IC  CD6 NE3 CF HF2          1.37    123.2  -120.0   109.5   1.08
IC  CD6 NF1 CF HF3          1.37    123.2   120.    109.5   1.08
ic  CE4 NE3 CF HF3          1.37    123.2   -60.0   109.5   1.08

!ic's for amide group on CE4
!!ic  OL  CL  NB  HE     1.2233 122.84  180.  119.23  0.9933
!!ic  OL  CL  NB  HF     1.2233 122.84  0.00  119.23  0.9933
```

```
ic  HB  NB  CL  CE4    0.9933 119.23    0.  116.25  1.47
!ic  HF  NB  CL  CE4    0.9933 119.23  -180.  116.25  1.47
ic  CV  NB  CL  CE4    1.40   119.23 -180.0 116.25  1.47
ic  NB  CL  CE4 NE3    1.3418 116.25  180.  122.1  1.38
ic  NB  CL  CE4 CD5    1.3418 116.25    0.  130.3  1.40
ic  NE3 CE4 CL  OL     1.38   122.0   0.00  123.5  1.2233
ic  CD5 CE4 CL  OL     1.40   130.3   180.0 123.5  1.2233
IC  CG2 CD5 CE4 CL     1.40   107.1  -180.0 130.3  1.46
ic  CD6 NE3 CE4 CL    1.37   109.8   180.0 122.1  1.46
ic  CF  NE3 CE4 CL     1.46   127.0   0.00  122.1  1.46
!ic's for amide group on CG2
ic  OZ  CZ  NZ  HZ      1.2233 122.84  -180.  119.23  0.9933
ic  HZ  NZ  CZ  CE3      0.9933 119.23    0.   112.25  1.47 !!
ic  OZ  CZ  NZ  CG2      1.2233 122.84    0.   125.57  1.40
ic  CZ  NZ  CG2  CD5     1.3418 122.57  180.   124.9  1.41
ic  CZ  NZ  CG2  CD6     1.3418 122.57    0.   127.2  1.40
IC  NZ  CG2  CD5 CE4     1.40   124.9   180.   107.1  1.40
ic  NZ  CG2  CD6 NE3     1.40   127.2  -180.0  107.6  1.37

!-------------------------------------------------------
!ic's for beta-alanine
IMPR  NB  CL  CV HB
IMPR  CL  CE4 NB OL
IC  OL CL NB HB      1.2233 122.84  -180. 119.23 0.9933
IC  HB NB CL CE4     0.9933 119.23    0. 116.25 1.47
IC  OL CL NB CV     1.2233 122.84    0. 122.57 1.4488
IC  NB CL CE4 CD5    1.3418 116.25   0. 109.3  1.40
IC  NB CL CE4 NE3    1.3418 116.25  -180. 109.3  1.38
IC  NB CL CE4 HL3    1.3418 116.25  -60. 109.3  1.109
IC  CL NB CV  CC     1.3418 122.57  180. 110.7  1.50
IC  CL NB CV  HV1    1.3418 122.57  -60. 110.7  1.50
IC  CL NB CV  HV3    1.3418 122.57   60. 110.7  1.50

IC  NB CV CC HC1    1.4488  110.7   60.0  109.5   1.10
IC  NB CV CC HC2    1.4488  110.7   -60.0  109.5   1.10
IC  NB CV CC CT     1.4488  110.7  180.0 109.5   1.50
!IC  CV CC CT NS     1.50    109.5  180.0  115.0   1.40
IC  CV CC CT NC    1.50    109.5   -180.0 115.0   1.40
IC  CV CC CT OT    1.50    109.5    0.00  123.1   1.24

IC  OT CT NC HC    1.24    121.9   180.0  121.2  0.9933
IC  OT CT NC CE    1.24    121.9    0.00  121.2  1.47

!IC  OT CT NS HS1    1.24    121.9   180.0  121.2  0.9933
!IC  OT CT NS HS2    1.24    121.9    0.00  121.2  1.47

!IMPR CT CC NS OT
IMPR CT CC NC OT
!-------------------------------------------------
!ic's for tail
IMPR  NC  CT  CE HC
IMPR  CT  CC NC  OT

IC  OT   CT   NC   HC      1.2233 122.84  180. 119.23 0.9933
IC  HC   NC   CT   CL     0.9933 119.23    0. 116.25 1.5118
IC  OT   CT   NC   CE     1.2233 122.84    0. 122.57 1.4488
IC  NC   CT   CC  HC1    1.3418 116.25   -60. 109.3  1.109
```

```
IC  NC   CT   CC  HC2    1.3418 116.25    60. 109.3  1.109
IC  NC   CT   CC  CV    1.3418 116.25  -180. 109.3  1.50
IC  CT   NC   CE  HM1    1.3418 122.57   60. 110.7  1.113
IC  CT   NC   CE  CJ     1.3418 122.57  180. 110.7  1.50
IC  CT   NC   CE  HM3    1.3418 122.57  -60.  110.7  1.113
IC  NC  CE  CJ CM       1.47   109.5   180.0  109.5  1.50
IC  NC  CE  CJ HJ1      1.47   109.5   -60.0  109.5  1.10
IC  NC  CE  CJ HJ2      1.47   109.5    60.0  109.5  1.10
IC  CE CJ  CM HN1      1.50   109.5    60.0  109.5  1.10
IC  CE CJ  CM HN2      1.50   109.5    -60.0 109.5  1.10
IC  CE CJ  CM NN       1.50   109.5    180.0 109.5  1.40
IC  CJ CM NN HH         1.50   109.5   60.00   109.5  1.10
IC  CJ CM NN CX        1.50   109.5  -60.0  109.5  1.47
IC  CJ CM NN CY        1.50   109.5   180.0  109.5  1.47

IC  CM NN CX HX1       1.47   109.5   -60.0  109.5  1.10
IC  CM NN CX HX2       1.47   109.5    60.0  109.5  1.10
IC  CM NN CX HX3       1.47   109.5   -180.  109.5  1.10
IC  CM NN CY HY1       1.47   109.5    60.0  109.5  1.10
IC  CM N1 CY HY2       1.47   109.5    180.0 109.5  1.10
IC  CM N1 CY HY3       1.47   109.5   -60.0  109.5  1.10

!IC  H1 N1 CX HX1       1.10   109.5    -180.0  109.5  1.10
IC  HY2 CY NN HH        1.10   109.5    -60.0    109.5  1.10
IC  HY3 CY NN HH        1.10   109.5     60.0    109.5  1.10

!impr OL NE3 CE4 CL
!impr OL CE4 CL NB
impr OT CC CT NC
impr OT CV CC CT
impr OT NC CT CC
!--------------------------------

RESI HP2      1.00
!imidazole ring
GROUP
ATOM CZ1   CPH1  0.06 !
ATOM HE1   HR3    0.09 !
ATOM CZ2  CPH1   0.22 !
ATOM NZ1  NR1   -0.36 !
ATOM CZ3  CPH2   0.25 !
ATOM NZ2  NR2   -0.36 !
ATOM HE2  HR3    0.10

GROUP
ATOM CZ4   CT3    -0.27
ATOM HZ1   HA     0.09
ATOM HZ2   HA     0.09
ATOM HZ3   HA     0.09
!----------------------------
!1st pyrrole ring
GROUP
ATOM HG   HP     0.115 !
ATOM CG   CY    -0.145 !
ATOM CD2  CPT   -0.020 !
ATOM CD1  CA     0.035 !
ATOM HD1  HP     0.115 !
```

```
ATOM NE1  NY    -0.230 !
!ATOM HE1  H      0.380 !
ATOM CE2  CPT    0.130 !

GROUP
ATOM CH   CT3   -0.27
ATOM H1   HA     0.09
ATOM H2   HA     0.09
ATOM H3   HA     0.09
GROUP
ATOM N2    NH1    -0.47
ATOM HM     H      0.235
ATOM CB     C      0.745
ATOM O2     O     -0.51
!ATOM HZ     H      0.235
!-----------------------------------------
!2nd pyrrole ring--hydroxypyrrole
GROUP
!ATOM HG1   HP     0.115 !
ATOM OG1   OH1     -0.326
ATOM HG1   H        0.397
ATOM CG1   CY     -0.140 !
ATOM CD3   CPT     0.020 !
ATOM CD4   CA      0.014 !
ATOM HD4   HP      0.115 !
ATOM NE2   NY     -0.230 !
!ATOM HE1  H       0.380 !
ATOM CE3   CPT     0.150 !
GROUP
ATOM CI   CT3   -0.27  !change C1 to CH
ATOM HI1   HA     0.09
ATOM HI2   HA     0.09
ATOM HI3   HA     0.09

GROUP
ATOM NK    NH1    -0.47
ATOM HN     H      0.235
ATOM CK     C      0.745
ATOM OK     O     -0.51
!ATOM HZ     H      0.235
!-----------------------------------------
!3rd pyrrole ring
GROUP
ATOM HG2   HP     0.115 !
ATOM CG2   CY    -0.145 !
ATOM CD5  CPT    -0.020 !
ATOM CD6  CA      0.035 !
ATOM HD6  HP      0.115 !
ATOM NE3  NY     -0.230 !
!ATOM HE1  H      0.380 !
ATOM CE4  CPT     0.130 !

GROUP
ATOM CF   CT3   -0.27  !change C1 to CH
ATOM HF1   HA     0.09
ATOM HF2   HA     0.09
ATOM HF3   HA     0.09
```

```
GROUP
ATOM CL    C      0.51
ATOM OL    O     -0.51
!ATOM NB    NH1   -0.62
!ATOM HE     H     0.31
!ATOM HF     H     0.31
GROUP
ATOM NZ    NH1    -0.47
ATOM HZ     H      0.235
ATOM CZ     C      0.745
ATOM OZ     O     -0.51
!ATOM HZ     H      0.235
!-----------------------------------------
!beta-alanine
GROUP
ATOM NB    NH1   -0.47
ATOM HB    H      0.47
ATOM CV   CT2   -0.18
ATOM HV1  HA     0.09
!ATOM HR2  HA     0.09
ATOM HV3  HA     0.09

GROUP
ATOM CC   CT2    -0.18
ATOM HC1  HA      0.09
ATOM HC2  HA      0.09
GROUP
ATOM CT   C       0.51
ATOM OT   O      -0.51
!ATOM NS   NH1     -0.47
!ATOM HS1   H       0.235
!ATOM HS2   H       0.235
!----------------------------------------
!cationic tail
GROUP
ATOM NC    NH1   -0.47
ATOM HC    H      0.47
ATOM CE   CT2   -0.11
ATOM HM1  HA     0.09
ATOM HM3  HA     0.09
GROUP
ATOM CJ  CT2     -0.11
ATOM HJ1  HA      0.09
AtOM HJ2  HA      0.09
ATOM CM   CT2    -0.11
ATOM HN1  HA      0.09
ATOM HN2  HA      0.09
GROUP
ATOM NN   NH3     -0.30
ATOM HH   HC       0.33
ATOM CX   CT3      0.23
ATOM HX1  HA       0.05
ATOM HX2  HA       0.05
ATOM HX3  HA       0.05
ATOM CY   CT3      0.23
ATOM HY1  HA       0.05
ATOM HY2  HA       0.05
```

```
ATOM HY3  HA       0.05


!---------------------------------------
!bonds for imidazole
BOND NZ2 CZ2  NZ1 CZ1  CZ3 NZ1
BOND CZ1 HE1  CZ2 HE2
DOUBLE NZ2 CZ3  CZ2 CZ1
BOND NZ1 CZ4  CZ4 HZ1  CZ4 HZ2  CZ4 HZ3
BOND CZ3 CB
!---------------------------------------
!bonds for 1st pyrrole ring
BOND CD2 HG  CD2 CG  NE1 CD1
BOND NE1 CE2
BOND CD1 HD1
DOUBLE CD1 CG  CE2 CD2
BOND CH NE1
BOND CH H1  CH H2  CH H3
BOND CE2 CK
DOUBLE CB O2
BOND N2 CG  N2 HM  N2 CB
!----------------------------------------
!bonds for 2nd pyrrole ring
BOND CD3 OG1  OG1 HG1  CD3 CG1  NE2 CD4
BOND NE2 CE3
BOND CD4 HD4
DOUBLE CD4 CG1  CE3 CD3
BOND CI NE2
BOND CI HI1  CI HI2  CI HI3
BOND CZ CE3  !NA CJ  NA HC  NA HD
DOUBLE CK OK
BOND NK CG1  NK HN  NK CK  !CB HZ
!----------------------------------------
!bonds for 3rd pyrrole ring
BOND CD5 HG2  CD5 CG2  NE3 CD6
BOND NE3 CE4
BOND CD6 HD6
DOUBLE CD6 CG2  CE4 CD5
BOND CF NE3
BOND CF HF1  CF HF2  CF HF3
BOND CL CE4  CL NB  !NB HE  NB HF
DOUBLE CZ OZ CL OL
BOND NZ CG2  NZ HZ  NZ CZ  !CB HZ
!----------------------------------------
!bonds for beta-alanine
BOND  NB CV
BOND  NB HB
BOND  HV1 CV  HV3 CV
DOUBLE CT OT
BOND CV CC  CC CT  !CT NS
BOND CT NC   !NS HS1  NS HS2
BOND CC HC1  CC HC2
!----------------------------------------
!bonds for cationic tail
BOND NC HC   NC CE
!BOND CL HL1  CL HL2  CL HL3
BOND CE HM1   CE HM3  CE CJ
```

```
!DOUBLE C O
BOND CJ HJ1  CJ HJ2  CJ CM
BOND CM HN1  CM HN2  CM NN
BOND NN HH  NN CX  NN CY
BOND CX HX1  CX HX2  CX HX3
BOND CY HY1  CY HY2  CY HY3

!-----------------------------------------
!ic's for imidazole ring
! KEEPS HYDROGENS IN RING PLANE
!IMPH NZ1  CZ1  CZ3  CZ4       NZ1  CZ3  CZ1   CZ4
IMPH CZ2  CZ1   NZ2  HE2       CZ2  NZ2  CZ1   HE2
IMPH CZ3  NZ1  NZ2  CB       CZ3  NZ2  NZ1  CB
IMPH CZ1   CZ2  NZ1  HE1        CZ1   NZ1  CZ2  HE1
IC   HE1    CZ1    NZ1   CZ3    1.07    122.67  -173.67  109.79
1.2987
IC   CZ1    NZ1   CZ3   NZ2     1.2854  109.79    0.21   110.31
1.3071
IC   NZ1   CZ3   NZ2   CZ2      1.2987  110.31    0.03   105.82
1.3165
IC   CZ3   NZ2   CZ2   CZ1      1.3071  105.82   -0.23  108.68   1.3758
IC   NZ2   CZ2   CZ1   NZ1      1.3165  108.68    0.35  105.39   1.2854
IC   NZ2   CZ2   CZ1    HE1     1.3165  108.68  172.86  131.52   1.07
IC   CZ2   CZ1    NZ1   CZ3     1.3758  105.39   -0.34  109.79   1.2987
IC   CZ2   NZ2   CZ3   CB       1.3165  105.82  -180.0  124.7    1.46
IC   NZ2   CZ3   NZ1   CZ4      1.34    110.31   180.0  127.8    1.46
IC   HE1   CZ1   CZ2   HE2      1.07    131.52    0.00  129.3    1.07

!ic's for methyl group
IC  CZ1 NZ1 CZ4 HZ1          1.37    124.7   0.00    109.5   1.08
IC  CZ1 NZ1 CZ4 HZ2          1.37    124.7  -120.0   109.5   1.08
IC  CZ1 NZ1 CZ4 HZ3          1.37    124.7   120.    109.5   1.08
ic  CZ3 NZ1 CZ4 HZ3         1.37    124.7   -60.0   109.5   1.08

impr CZ4 CZ1 NZ1 CZ3


!ic's for amide group on CE1
IMPR  CB CZ3 N2 O2

!ic  O3  CC  N3  HC     1.2233 122.84  -180.  119.23  0.9933
!ic  O3  CC  N3  HD     1.2233 122.84  0.00  119.23  0.9933
ic  HM  N2  CB  CZ3    0.9933 119.23    0.  116.25  1.47
ic  CG  N2  CB  CZ3    1.40   119.23  180.0 116.25  1.47
ic  N2  CB  CZ3 NZ1    1.3418 116.25  180.  122.1   1.38
ic  N2  CB  CZ3 NZ2    1.3418 116.25    0.  130.3   1.40
ic  NZ1 CZ3 CB  O2     1.38   122.0   0.00  123.5   1.2233
ic  NZ2 CZ3 CB  O2     1.40   130.3   180.0 123.5   1.2233
IC  CZ2 NZ2 CZ3 CB     1.40   107.1  -180.0 130.3   1.46
ic  CZ1  NZ1 CZ3 CB    1.37   109.8   -180.0 122.1   1.46
ic  CZ4  NZ1 CZ3 CB    1.46   127.0   0.00  122.1   1.46

!-----------------------------------------------
!ic's for 1st pyrrole ring
IC CG   CD1  NE1  CE2   0.0000  0.0000  0.0000  0.0000  0.0000
IC CD2  CB   *CG  CD1   0.0000  0.0000 180.0000  0.0000  0.0000
IC CD1  CG   CD2  CE2   0.0000  0.0000  0.0000  0.0000  0.0000
```

```
IC CD2  CG   CD1  NE1   0.0000  0.0000  0.0000  0.0000  0.0000
IC CD1  CE2  *NE1 CH   0.0000  0.0000 -180.0000  127.0  1.46
IC CG   NE1  *CD1 HD1   0.0000  0.0000 180.0000  0.0000  0.0000
IC CE2  CG   CD2  HG    0.0000  0.0000 -180.0000  126.9  1.08

!ic's for methyl group
IC  CD1 NE1 CH H1          1.37    123.2   0.00    109.5   1.08
IC  CD1 NE1 CH H2          1.37    123.2  -120.0   109.5   1.08
IC  CD1 ND1 CH H3          1.37    123.2   120.    109.5   1.08
ic  CE2 NE1 CH H3          1.37    123.2   -60.0   109.5   1.08

!ic's for amide group on CE2
!!ic  O1  CA  N1  HA     1.2233 122.84  180.  119.23  0.9933
!!ic  O1  CA  N1  HB     1.2233 122.84  0.00  119.23  0.9933
ic  HN  NK  CK  CE2    0.9933 119.23    0.  116.25  1.47  !look at this
section
ic  CG1 NK  CK  CE2    1.40   119.23 -180.0 116.25  1.47
ic  NK  CK  CE2 NE1    1.3418 116.25  180.  122.1  1.38
ic  NK  CK  CE2 CD2    1.3418 116.25    0.  130.3  1.40
ic  NE1 CE2 CK  OK     1.38   122.0   0.00  123.5  1.2233
ic  CD2 CE2 CK  OK     1.40   130.3   180.0 123.5  1.2233
IC  CG  CD2 CE2 CK     1.40   107.1  -180.0 130.3  1.46
ic  CD1 NE1 CE2 CK     1.37   109.8   180.0 122.1  1.46
ic  CH  NE1 CE2 CK     1.46   127.0   0.00  122.1  1.46

!ic's for amide group on CG
ic  O2  CB  N2  HM      1.2233 122.84  -180.  119.23  0.9933
ic  HM  N2  CB  O2      0.9933 119.23   180.   112.25  1.23
ic  O2  CB  N2  CG      1.2233 122.84    0.   125.57  1.40
ic  CB  N2  CG  CD2     1.3418 122.57  180.   124.9  1.41
ic  CB  N2  CG  CD1     1.3418 122.57    0.   127.2  1.40
IC  N2  CG  CD2 CE2     1.40   124.9   180.   107.1  1.40
ic  N2  CG  CD1 NE1     1.40   127.2  -180.0  107.6  1.37
!-------------------------------------------------------
!ic's for 2nd pyrrole ring
IC CG1   CD4  NE2  CE3   0.0000  0.0000  0.0000  0.0000  0.0000
!IC CD2  CB   *CG  CD1   0.0000  0.0000 180.0000  0.0000  0.0000
IC CD4  CG1   CD3  CE3   0.0000  0.0000  0.0000  0.0000  0.0000
IC CD3  CG1   CD4  NE2   0.0000  0.0000  0.0000  0.0000  0.0000
IC CD4  CE3  *NE2 CI   0.0000  0.0000 -180.0000  127.0  1.46
IC CG1   NE2  *CD4 HD4   0.0000  0.0000 180.0000  0.0000  0.0000
IC CE3  CG1   CD3  OG1    0.0000  0.0000 -180.0000  126.9  1.08
IC CE3  CD3   OG1  HG1   0.000    0.00   0.00     0.00    0.00

!ic's for methyl group
IC  CD4 NE2 CI HI1          1.37    123.2   0.00    109.5   1.08
IC  CD4 NE2 CI HI2          1.37    123.2  -120.0   109.5   1.08
IC  CD4 NC1 CI HI3          1.37    123.2   120.    109.5   1.08
ic  CE3 NE2 CI HI3          1.37    123.2   -60.0   109.5   1.08

!ic's for amide group on CE3
!!ic  OJ  CJ  NA  HC     1.2233 122.84  180.  119.23  0.9933
!!ic  OJ  CJ  NA  HD     1.2233 122.84  0.00  119.23  0.9933
ic  HZ  NZ  CZ  CE3    0.9933 119.23    0.  116.25  1.47
ic  CG2  NZ  CZ  CE3   1.40   119.23 -180.0 116.25  1.47
ic  NZ  CZ  CE3 NE2    1.3418 116.25  180.  122.1  1.38
ic  NZ  CZ  CE3 CD3    1.3418 116.25    0.  130.3  1.40
```

```
ic  NE2 CE3 CZ  OZ     1.38   122.0   0.00  123.5  1.2233
ic  CD3 CE3 CZ  OZ     1.40   130.3   180.0 123.5  1.2233
IC  CG1  CD3 CE3 CZ     1.40   107.1  -180.0 130.3  1.46
ic  CD4 NE2 CE3 CZ    1.37   109.8   180.0 122.1  1.46
ic  CI  NE2 CE3 CZ     1.46   127.0   0.00  122.1  1.46
!ic's for amide group on CG1
ic  OK  CK  NK  HN      1.2233 122.84  -180.  119.23  0.9933
ic  HN  NK  CK  CE2      0.9933 119.23    0.   112.25  1.47 !!
ic  OK  CK  NK  CG1      1.2233 122.84    0.   125.57  1.40
ic  CK  NK  CG1  CD3     1.3418 122.57  180.   124.9  1.41
ic  CK  NK  CG1  CD4     1.3418 122.57    0.   127.2  1.40
IC  NK  CG1  CD3 CE3     1.40   124.9   180.   107.1  1.40
ic  NK  CG1  CD4 NE2     1.40   127.2  -180.0  107.6  1.37
!-------------------------------------------------------
!ic's for 3rd pyrrole ring
IC CG2   CD6  NE3  CE4   0.0000  0.0000  0.0000  0.0000  0.0000
!IC CD2  CB   *CG  CD1   0.0000  0.0000 180.0000  0.0000  0.0000
IC CD6  CG2   CD5  CE4   0.0000  0.0000  0.0000  0.0000  0.0000
IC CD5  CG2   CD6  NE3   0.0000  0.0000  0.0000  0.0000  0.0000
IC CD6  CE4  *NE3 CF   0.0000  0.0000 -180.0000  127.0  1.46
IC CG2   NE3  *CD6 HD6   0.0000  0.0000 180.0000  0.0000  0.0000
IC CE4  CG2   CD5  HG2    0.0000  0.0000 -180.0000  126.9  1.08
!ic's for methyl group
IC  CD6 NE3 CF HF1          1.37    123.2   0.00    109.5   1.08
IC  CD6 NE3 CF HF2          1.37    123.2  -120.0   109.5   1.08
IC  CD6 NF1 CF HF3          1.37    123.2   120.    109.5   1.08
ic  CE4 NE3 CF HF3          1.37    123.2   -60.0   109.5   1.08

!ic's for amide group on CE4
!!ic  OL  CL  NB  HE     1.2233 122.84  180.  119.23  0.9933
!!ic  OL  CL  NB  HF     1.2233 122.84  0.00  119.23  0.9933
ic  HB  NB  CL  CE4    0.9933 119.23    0.  116.25  1.47
!ic  HF  NB  CL  CE4    0.9933 119.23  -180.  116.25  1.47
ic  CV  NB  CL  CE4    1.40   119.23 -180.0 116.25  1.47
ic  NB  CL  CE4 NE3    1.3418 116.25  180.  122.1  1.38
ic  NB  CL  CE4 CD5    1.3418 116.25    0.  130.3  1.40
ic  NE3 CE4 CL  OL     1.38   122.0   0.00  123.5  1.2233
ic  CD5 CE4 CL  OL     1.40   130.3   180.0 123.5  1.2233
IC  CG2 CD5 CE4 CL     1.40   107.1  -180.0 130.3  1.46
ic  CD6 NE3 CE4 CL    1.37   109.8   180.0 122.1  1.46
ic  CF  NE3 CE4 CL     1.46   127.0   0.00  122.1  1.46
!ic's for amide group on CG2
ic  OZ  CZ  NZ  HZ      1.2233 122.84  -180.  119.23  0.9933
ic  HZ  NZ  CZ  CE3      0.9933 119.23    0.   112.25  1.47 !!
ic  OZ  CZ  NZ  CG2      1.2233 122.84    0.   125.57  1.40
ic  CZ  NZ  CG2  CD5     1.3418 122.57  180.   124.9  1.41
ic  CZ  NZ  CG2  CD6     1.3418 122.57    0.   127.2  1.40
IC  NZ  CG2  CD5 CE4     1.40   124.9   180.   107.1  1.40
ic  NZ  CG2  CD6 NE3     1.40   127.2  -180.0  107.6  1.37

!-------------------------------------------------------
!ic's for beta-alanine
IMPR  NB  CL  CV HB
IMPR  CL  CE4 NB OL
IC  OL CL NB HB      1.2233 122.84  -180. 119.23 0.9933
IC  HB NB CL CE4     0.9933 119.23    0. 116.25 1.47
IC  OL CL NB CV     1.2233 122.84    0. 122.57 1.4488
```

```
IC  NB CL CE4 CD5    1.3418 116.25   0. 109.3  1.40
IC  NB CL CE4 NE3    1.3418 116.25  -180. 109.3  1.38
IC  NB CL CE4 HL3    1.3418 116.25  -60. 109.3  1.109
IC  CL NB CV  CC     1.3418 122.57  180. 110.7  1.50
IC  CL NB CV  HV1    1.3418 122.57  -60. 110.7  1.50
IC  CL NB CV  HV3    1.3418 122.57   60. 110.7  1.50

IC  NB CV CC HC1    1.4488  110.7   60.0  109.5   1.10
IC  NB CV CC HC2    1.4488  110.7   -60.0  109.5   1.10
IC  NB CV CC CT     1.4488  110.7  180.0 109.5   1.50
!IC  CV CC CT NS     1.50    109.5  180.0  115.0   1.40
IC  CV CC CT NC    1.50    109.5   -180.0 115.0   1.40
IC  CV CC CT OT    1.50    109.5    0.00  123.1   1.24

IC  OT CT NC HC    1.24    121.9   180.0  121.2  0.9933
IC  OT CT NC CE    1.24    121.9    0.00  121.2  1.47

!IC  OT CT NS HS1    1.24    121.9   180.0  121.2  0.9933
!IC  OT CT NS HS2    1.24    121.9    0.00  121.2  1.47

!IMPR CT CC NS OT
IMPR CT CC NC OT
!---------------------------------------------------
!ic's for tail
IMPR  NC  CT  CE HC
IMPR  CT  CC NC  OT

IC  OT   CT   NC   HC      1.2233 122.84  180. 119.23 0.9933
IC  HC   NC   CT   CL     0.9933 119.23    0. 116.25 1.5118
IC  OT   CT   NC   CE     1.2233 122.84    0. 122.57 1.4488
IC  NC   CT   CC  HC1    1.3418 116.25   -60. 109.3  1.109
IC  NC   CT   CC  HC2    1.3418 116.25    60. 109.3  1.109
IC  NC   CT   CC  CV    1.3418 116.25  -180. 109.3  1.50
IC  CT   NC   CE  HM1    1.3418 122.57   60. 110.7  1.113
IC  CT   NC   CE  CJ     1.3418 122.57  180. 110.7  1.50
IC  CT   NC   CE  HM3    1.3418 122.57  -60.  110.7  1.113
IC  NC  CE  CJ CM       1.47   109.5   180.0  109.5  1.50
IC  NC  CE  CJ HJ1      1.47   109.5   -60.0  109.5  1.10
IC  NC  CE  CJ HJ2      1.47   109.5    60.0  109.5  1.10
IC  CE CJ  CM HN1      1.50   109.5    60.0  109.5  1.10
IC  CE CJ  CM HN2      1.50   109.5    -60.0 109.5  1.10
IC  CE CJ  CM NN       1.50   109.5    180.0 109.5  1.40
IC  CJ CM NN HH         1.50   109.5   60.00   109.5  1.10
IC  CJ CM NN CX        1.50   109.5  -60.0  109.5  1.47
IC  CJ CM NN CY        1.50   109.5   180.0  109.5  1.47

IC  CM NN CX HX1       1.47   109.5   -60.0  109.5  1.10
IC  CM NN CX HX2       1.47   109.5    60.0  109.5  1.10
IC  CM NN CX HX3       1.47   109.5   -180.  109.5  1.10
IC  CM NN CY HY1       1.47   109.5    60.0  109.5  1.10
IC  CM N1 CY HY2       1.47   109.5    180.0 109.5  1.10
IC  CM N1 CY HY3       1.47   109.5   -60.0  109.5  1.10

!IC  H1 N1 CX HX1       1.10   109.5    -180.0  109.5  1.10
IC  HY2 CY NN HH        1.10   109.5    -60.0    109.5  1.10
IC  HY3 CY NN HH        1.10   109.5     60.0    109.5  1.10
```

```
impr OT CC CT NC
impr OT CV CC CT
impr OT NC CT CC


end


read parameter card append

BONDS
!
!atom types  Kb          b0
!

!sulfate bond
OC   S     400.000     1.43

!polyamide bonds
NH2  CPT   500.000     1.41 !A.E.L.
CD   CA    500.000     1.45 !A.E.L.
CT3  NY    500.000     1.46 !A.E.L.
CPT  HP    500.000     1.083 !A.E.L.
NC2  CPT   500.000     1.41   !A.E.L.
CD   NH2   500.00      1.36   !A.E.L.
OS  CC     400.000      1.35   !a.E.L.
OS  CT1    400.00       1.46   !A.E.L.

!NH2 CT2   500.000     1.46   !A.E.L.
NC2  H    500.000     1.00   !A.E.L.
CC   H    500.00      1.10
NH2  CY   500.00      1.40
!-------------------------------------------
!bonds for ipy drug
!N-methyl-pyrrole

CPT  HP    500.00     1.08   !A.E.L.
CT3  NY    500.000    1.46
C    CPT   500.00     1.47
NH1  CY    500.00     1.40
C    H     500.00     1.10
OS  CT1    400.00     1.46   !A.E.L.
OS  C     400.000      1.35   !a.E.L.
CPT OH1   500.000     1.37   !ael; for hydroxypyrrole

!imidazole
NR1 CT3   500.00      1.46   !A.E.L.
CPH2 C    500.000     1.47
CPH1 NH1  500.00      1.40
C    H    500.00      1.10
!-------------------------------------------------

ANGLES
!
!atom types     Ktheta    Theta0   Kub     S0
!
```

```
!sulfate angle
OC   S    OC    40.00     109.5

!polyamide angles
!---------------------------------------
NH2  CC   H     50.000    120.00   !from formamide
O    CC   H     44.00     122.00
CC   NH2  CY    40.       125.2
H    NH2  CY    40.       117.2
H    OH1  CPT   40.0      106.4
OH1  CPT  CY    40.0      123.1
OH1  CPT  CPT   40.00     129.0

CA   CA   NY    40.00     120.0    !A.E.L.
CA   CA   CT1   40.00     120.0    !A.E.L.
CA   CT1  CT3   40.00     108.3    !A.E.L.
CPT  CPT  NH2   40.00     126.440    !A.E.L.
CPT  CT3  HA    40.00     109.50    !A.E.L.
CY  CPT  NH2    40.00     126.370    !A.E.L.
CY  CA  CD      40.00     130.110    !A.E.L.
CA  CD  OH1     40.00     112.290   !A.E.L.
CA  CD  OB      40.00      130.070  !A.E.L.
CPT  NH2 HC     40.00      109.50  !A.E.L.
NY  CPT  CT3    40.00      121.490  !A.E.L.
CPT CPT CT3     40.00      130.090  !A.E.L.
CA  NY CA       40.00      125.590  !A.E.L.
NY  CA  CD      40.00      121.470  !A.E.L.
CPT CY  CPT     40.00      35.915   !A.E.L.
CA  NY  CT3     40.0       125.60   !A.E.L.
CPT NY  CT3     40.00      125.60   !A.E.L.
NY  CT3 HA      40.0       109.5    !A.E.L.
CPT CY  NH2     40.00      126.320  !A.E.L.
CA  CY  NH2     40.00      126.320  !A.E.L.
CY  NH2 HC      40.00      109.50   !A.E.L.
CPT CPT HP      40.00      130.110  !A.E.L.
NY  CPT HP      40.00      121.470  !A.E.L.
!HA  CT2 NH2     40.00      109.5    !A.E.L.
!CT2 CT2 NH2     40.00      109.5    !A.E.L.
!CT2 NH2 HC      40.00      109.5    !A.E.L.
CT2 NH2 CT3     40.00      109.5    !A.E.L.
CT3 NH2 CT3     40.00      105.970  !A.E.L.
CPT NY CPT       40.000     108.8    !A.E.L.
CA  CD   NH2     40.000      115.0    !A.E.L.
OB  CD   NH2    40.000      121.9    !A.E.L.  !!!
CD  NH2  CPT    40.000      120.6    !A.E.L.
CD  NH2  HC     40.000      121.2    !A.E.L.
!NH2 CD   OB     40.000      121.9    !A.E.L.
NH2  CC  OS    50.00    116.50   50.000   2.45000
O    CC  OS    15.00    121.00   50.00    2.44000
CC   OS  CT1   33.00    109.50   30.000   2.16300
OS   CT1 CT3   34.500   110.10   22.53    2.17900
CPT  NH2 CC    40.000   120.6        !A.E.L.
HC   NH2 CC    40.000   121.2        !A.E.L.

HA   CT2  NH2   40.000   115.0     !A.E.L.
CT2  CT2  NH2   40.000   108.8     !A.E.L.
CT3  NC2  CT3   40.000   109.5     !A.E.L.
```

```
CT2  NC2  CT3   40.000   109.5     !A.E.L.
CT2  NC2  H     40.000   109.5     !a.E.L.
CT3  NC2  H     40.000   109.5     !A.E.L.
CD   NH2  CT2    40.000   120.6     !A.E.L.
CT2  C    NC2   40.00    120.0     !A.E.L.
C    CT2  NC2   40.00    35.00     !A.E.L.
!-------------------------------------------------
!angles for ipy drug
!Boc group
NH1  C  OS    50.00    116.50   50.000   2.45000
O    C  OS    15.00    121.00   50.00    2.44000
C   OS  CT1   33.00    109.50   30.000   2.16300
OS   CT1 CT3   34.500   110.10   22.53    2.17900

!N-methyl-pyrrole
CPT  CY   NH1    40.000    124.9     !A.E.L.
CA   CA   NH1    40.000    127.2
HP   CPT  CY     40.000    126.9
HP   CPT  CPT    40.000    125.9
CA   NY   CT3    40.000    123.3
CPT  NY   CT3    40.000    127.
CPT  CPT  C      40.000    130.2
NY   CPT  C      40.000    122.1
NY   CT3  HA     40.000    109.5
CPT  C    O      40.000    123.5
CPT  C    NH1    40.000    114.4
H   NH1   H      40.000    115.7
CY  NH1   H      40.000    117.2
CY  NH1   C      40.00     125.2
NH1  C    H      40.00     112.0
CA   CY   NH1    40.000    127.2
CA   CY   C      40.00     100.3
CPT  CY   C      40.000    151.9
CPT  CY   CPT    40.000    36.3

!imidazole
CPH1 CPH1 NH1    40.000   129.3  !A.E.L.
NR2  CPH1 NH1    40.000   120.0
CPH1 NR1  CT3    40.000   124.7
CPH2 NR1  CT3    40.000   127.8
NR1  CPH2 C      40.000   123.8
NR2  CPH2 C      40.000   124.8
NR1  CT3  HA     40.000   107.4
CPH2 C    O      40.000   122.6
CPH2 C    NH1    40.000   112.9
H   NH1   H      40.000   121.1
CPH1 NH1  H      40.000   115.7
CPH1 NH1  C      40.000   124.5
NH1  C    H      40.000   112.4

!cationic tail
CT2 NH3 CT3    40.000     109.5  !A.E.l.
CT3 NH3 CT3    40.00      110.8

CA  CY  CC     40.0       100.3
CPT CY  CC     40.0       151.9
!-------------------------------------------------
```

```
DIHEDRALS
!
!atom types             Kchi    n   delta
!
!polyamide dihedrals
!-------------------------------------------------------
CA  CA  CT1 CT3     0.2000  1  169.040 !A.E.L.
CA  NY  CA  CY      0.2000  1  180.00 !A.E.L.
CA  NY  CA  CD      0.2000  1  0.000  !A.E.L.
CA  NY  CPT CT3     0.2000  1  0.000 !A.E.L.
CA  CA  CA  CT1     0.2000  1  180.0  !A.E.L.
CA  CA  NY  CPT     0.2000  1  90.0   !A.E.L.
CA  CA  CA  NY      0.2000  1  180.00 !A.E.L.
HP  CA  CA  CT1     0.2000  1  0.000  !A.E.L.
CT1 CA  CA  NY      0.2000  1  0.000  !A.E.L.
HP  CY  CPT NH2     0.2000  1  0.000  !A.E.L.
HP  CY  CA  CD      0.2000  1  0.000  !A.E.L.
CY  CPT NH2 HC      0.2000  1  0.00 !A.E.L.  !


CPT CY  CA  CD      0.6     1  -180.00 !A.E.L.
CPT CPT CT3 HA      0.2000  1  0.000  !A.E.L.
CA  CA  NY CA       0.2000  1  -90.0   !A.E.L.
CA  CY  CPT  NH2    0.2000  1  180.00  !A.E.L.

NY  CPT CT3  HA     0.0000  1 -180.00  !A.E.L.

CPT  CPT  NH2  HC   0.00  1    -180.0  !A.E.L.
CPT  CPT  NH2  HC   0.00  3     0.00


CPT  NY CA  CD      0.6000  1    180.00  !A.E.L.
NH2  CPT CPT CT3    0.2000  1    0.000 !A.E.L.

CY  CA NY CT3       0.2000  1   180.0  !A.E.L.
CPT CPT NY CT3      0.2000  1   -180.0 !A.E.L.

CA  NY CT3 HA       0.20000  1    120.0 !A.E.L.

CPT NY CT3 HA       0.000   1    -60.0 !A.E.L.
CPT NY CT3 HA       0.000   1    0.00  !A.E.L.
CPT NY CT3 HA       0.000   1    -120.0 !A.E.L.

CPT CY NH2 HC       0.2000  1    -180.0 !A.E.L.
CA  CY NH2 HC       0.2000  1    -180.00 !A.E.L.
NY  CA  CY NH2      0.2000  1     180.0 !A.E.L.
CPT CPT CY NH2      0.2000  1     180.0 !A.E.L.
CA  NY  CPT HP      0.2000  1     180.0 !A.E.L.
HP  CPT NY CT3      0.2000  1       0.0     !A.E.L.
CT3 NY  CA CD       0.2000  1       0.0     !A.E.L.
CT2 CT2 NH2 CT3     0.2000  1     180.0     !A.E.L.
HA  CT2 NH2 CT3     0.2000  1      60.0     !A.E.L.


CA  CD  NH2 CPT     1.6000   1       0.0     !A.E.L.
CA  CD  NH2 CPT     2.5      2      180.0    !A.E.L.
```

```
CA  CD  NH2 HC      1.4      2        0.0     !A.E.L.  !2.5 180.0

!CY  CPT NH2 CD      0.2000   1     180.0     !A.E.L.
CY  CPT NH2 CD      1.5      2     180.0

CPT CPT NH2 CD      0.9650  1     180.00      !A.E.L.
CPT CPT NH2 CD      3.8500  2     180.0
!CPT CPT NH2 CD      .2     0      0.00


CPT NH2 CD  OB      1.5      2     0.0     !A.E.L.
HC  NH2 CD  OB      3.5      2     180.0     !A.E.L.
!------------------------------------------------------------
NH2 CC  OS  CT1     0.20       1      -180.0  !A.E.L.

O   CC  OS  CT1      0.9650      1    180.0 !A.E.L.
O   CC  OS  CT1      2.8500      2    180.0 !A.E.L.

CT3 CT1 OS CC        0.6        1     0.00  !A.E.L.

CY  CPT NH2 CC       0.200      1    180.0  !A.E.L.

CPT CPT NH2 CC       0.9650     1    180.0
CPT CPT NH2 CC       3.8500     2    180.0

CPT NH2 CC  O        1.5        2      0.0  !A.E.L.

HC  NH2 CC  O        2.5        2    180.0  !A.E.L.
HC  NH2 CC  OS       1.4        2      0.0  !a.E.L>

CPT NH2 CC OS        1.600      1    0.00
CPT NH2 CC OS        2.500      2   180.0
!--------------------------------------------------------------------
HA  CT2 NH2 HC        0.200      1.5   -120.00   !!!!
HA  CT2 NH2 HC        0.200      1     120.0

CT2 CT2 NH2 HC         1.500       1      60.000

CT2 CT2 NC2 H         0.9          3.7    0.00
CT2 CT2 NC2 H         0.0           0     60.00

NH2 CT2 CT2 CT2      1.0           1      180.0
NH2 CT2 CT2 CT2      1.3          1.5    -180.0
NH2 CT2 CT2 CT2      0.5          3       0.00


CA  CD  NH2  CT2     1.6    1    0.00
CA  CD  NH2  CT2     2.5    2   180.0

CD  NH2 CT2  HA     0.2000  1   60.0
CD  NH2 CT2  CT2    0.200   1   180.0

OB  CD  NH2  CT2    1.5     2   0.0


CY  CA  CD   NH2    2.3   2    -180.00  !A.E.L.
NY  CA  CD   NH2    2.3  2     -180.0   !A.E.L.
```

```
CY  CA  CD   OB     2.3  2      180.00  !A.E.L.
NY  CA  CD   OB     2.3  2       180.00  !A.E.L.
!-------------------------------------------------------
!dihedrals for ipy drug
!N-methyl-pyrrole

!------------------------------------------------
CA  NY   CT3  HA     0.00     1       0.00   !methyl group
CA  NY   CT3  HA     0.000    3     -120.0
CA  NY   CT3  HA     0.000    2     -60.0
!------------------------------------------------
!-----------------------------------------
CPT NY   CT3  HA     0.055   3        0.0    !methyl group
CPT NY   CT3  HA     0.00    1      -120.
CPT NY   CT3  HA     0.00    2       -60.0
!-----------------------------------------

HP  CPT  CY   CA     2.0     1        180.0

HP  CPT  CY   NH1    2.8     2        180.0

CY  CA   NY   CT3    0.8    2        180.0
CY  NH1  C     O     1.5    2        180.0
CY  NH1  C     H     1.20   1        180.0
CPT CY   NH1   H     0.000  1         0.00
CPT CY   NH1   C     0.200  1        180.0

CPT  CPT  NY   CT3   0.8    2        180.0

CPT  CPT  C    O     2.5    2        180.0

CPT  CPT  C    NH1   1.5    2        180.0

CA   NY   CPT  C     2.00   1        180.0

HP   CA   CY   NH1   1.00   2        180.0

CA   CY   NH1  H     0.0    1        0.00

CA   CY   NH1  C     2.5    2       180.0
CA   CY   NH1  C     3.0    1       180.0

HP   CA   NY   CT3   0.4000  2      180.0
NY   CA   CY   NH1   2.0     2      180.0

NY   CPT  C    O     1.3   2      180.0
NY   CPT  C    O     3.7    1      180.0


NY   CPT  C    NH1   1.0     1     0.0

CPT  CPT  CY   NH1   3.0    2      180.0
CT3  NY   CPT  C     0.8     2      180.0

H    NH1  C    H     1.4     2      180.0

CPT  C    NH1   H      2.5     2      180.0
```

```
!-------------------------------------------------
!-----------------------------------------------
CPT  C  NH1 CY      2.5     2     180.0
CPT  C  NH1 CT2     2.5     2     180.0
!-----------------------------------------------
!N-methyl-imidazole
!-------------------------------------------------------
CPH1 NR1  CT3  HA      0.00  1     0.00
CPH1 NR1  CT3  HA      0.000  3     -120.0     !methyl group
CPH1 NR1  CT3  HA      0.000  2    -60.0
!-------------------------------------------------------
!-----------------------------------------------
CPH2 NR1 CT3 HA        0.000   3    0.
CPH2 NR1 CT3 HA        0.0000   1   -120.0     !methyl group
CPH2 NR1 CT3 HA        0.0000   2   -60.0
!-------------------------------------------------

CPH1 CPH1 NH1 H        0.1   2   -180.0  !A.E.L.
CPH1 CPH1 NH1 H        0.00  1    0.00

CPH1 CPH1 NH1 C        1.5     1   180.0
CPH1 CPH1 NH1 C        1.0    2      00.0

HR3  CPH1 CPH1 NH1     0.5000  2    0.00

CPH1 NH1  C    O       0.2000  2    0.00
CPH1 NH1  C    O       3.000  1   180.0

CPH1 NH1  C   H        0.80000  2   -180.0
CPH1 NH1  C   H        0.5000   1     0.00

NR1  CPH1 CPH1 NH1     1.0000   1   -180.0
CPH2 NR2 CPH1 NH1      0.2000   1    -180.0
NR2 CPH1 NH1  H        1.3      2   0.0
NR2 CPH1 NH1  C        1.3      2  -180.0
H   NH1  C   H         0.2000   2    0.00
CPH1 CPH1 NR1  CT3     0.2000  1   -180.0
O  C  NH1  H           2.5000  2    180.0

!-------------------------------------------------

CPH1 NR1  CPH2 C       0.2000  1     -180.0  !
HR3  CPH1 NR1  CT3     0.2000  1    0.00    !
CPH1 NR2  CPH2 C       0.2000  1    -180.0

NR1  CPH2 C   O        5.000    2   180.0
NR1  CPH2 C   O        4.500    1   180.0

NR1  CPH2 C  NH1       1.000    2   0.0  !
NR2 CPH2 C   O         1.0000   1    0.0    !
NR2 CPH2 C  NH1        0.0000   2   0.00    !

CPH2 C   NH1  H        0.2000   1    0.00
CPH2 C   NH1  H        0.2000   2   180.0

NR2 CPH2 NR1 CT3       0.2000   1  180.0
CT3 NR1  CPH2 C        0.2000   1   0.00
```

```
CPH2 C  NH1 CY         0.2000  1   0.00
CPH2 C  NH1 CY         0.2000  2   180.0
!-------------------------------------------------------
!netropsin
!CY  NH1 C   CT2        0.2000  1   0.00
CY  NH1 C   CT2        1.000  1   -180.0

O   C   CT2 NC2        0.5000  1   0.000
CT2 CT2 C   NC2        0.000   1   0.00
NH1 C   CT2 NC2        0.6000  1   0.00
HA  CT2 C   NC2        0.0000  3   0.00
!-------------------------------------------------------
!distamycin
CC  NH2  CY  CPT       0.2000  1   180.0
CC  NH2  CY  CA        2.5     2   180.0
CC  NH2  CY  CA        3.7     1   180.0

NH2 CY   CPT HP        2.8     2   180.0
NH2 CY   CA  HP        1.00    1   180.0

H   NH2  CC  H         1.4     2   180.0
H   NH2  CY  CPT       0.050   1   0.00

H   NH2  CY  CA        0.05    1   0.00
O   CC   NH2 CY        1.50    2   180.0
H   CC   NH2 CY        1.20    1   180.0
!hydroxypyrrole
OH1 CPT  CY  CA        3.1     2   180.0

OH1 CPT  CY  NH1       0.20    2   180.0

H  OH1  CPT  CY        0.00    2   90.0
H  OH1  CPT  CY        4.0     1   0.00

H  OH1  CPT  CPT       0.50    2   180.0
H  OH1  CPT  CPT       0.30    1    0.00

IMPROPER
!atom types             Kpsi        psi0
!polyamide impropers
CA   CA   CA   NY     90.000          0     -180.00 !A.E.L.
CPT  CA   NY   CA     90.000          0      180.00 !A.E.L.
CY   CPT  CPT  NH2    90.000          0      180.00 !A.E.L.
CA   CA   CA   CT1    90.000          0      180.00 !A.E.L.
OH1  CA   CD   OB     90.000          0      180.0  !A.E.L.
CD   NY   CA   CY     90.000          0      180.0  !A.E.L.
CT3  NY   CPT  CPT    90.000          0     -180.0  !A.E.L.
CA  CPT   NY   CT3    90.000          0     -180.0  !A.E.L.
CA  CPT   CY   NH2    90.000          0      180.00 !A.E.L.
CPT CY    CPT  NH2    90.000          0     -180.0  !A.E.L.
CPT NY    CPT  HP     90.000          0      180.0  !A.E.L.

OB   CA   CD    OH1   90.00           0     -180.0  !A.E.L.

OH1  OB   CD    CA    90.00           0     -180.0  !A.E.L.
OB   OH1  CD    CA    90.00           0      180.0  !A.E.L.
```

```
CY   CD   OH1   H     90.000          0      -180.0 !A.E.L.
HP   CA   CD    OH1   90.00           0       0.00  !A.E.L.
OH1  CA   CY   HP     90.000          0       0.00  !A.E.L.
OB   CA   NY   CT3    90.000          0       0.00  !A.E.L.
NH2  CT2  CT2  HA     90.000          0      -180.0 !A.E.L.
OB   NH2  CT2  CT2    90.000          0       180.0 !A.E.L.
CA  OB   CD    OH1    90.000          0       180.0 !A.E.L.
NY  CD   OH1   H      90.000          0       180.0 !A.E.L.
CY  CD   CA    NY     90.000          0      -180.0 !A.E.L.
H   CD   CA    CA     90.000          0        0.000!A.E.L.
H   CD   CA    NY     90.000          0       -180.0 !A.E.L.
CD  CY   CA    NY     90.000          0        180.0 !A.E.L.
CD  CA   CPT   CPT    90.000          0        0.00  !A.E.L.
CA  CA   NY    CPT    90.000          0       -180.0 !A.E.L.
CA  CA   CA    HP     90.000          0        180.0 !A.E.L.
CT1 CA   CA    HP     90.000          0       -180.0 !A.E.L.
HC  HC  NH2 CPT       90.000          0       -120.0 !A.E.L.
HC  CPT NH2 HC        90.000          0        120.0 !A.E.L.
CA  CPT NH2 HC        90.000          0       -20.33 !A.E.L.

NY CD  NH2 CPT        90.00           0        180.0  !A.E.L.

NH2  CD  CPT HC        20.000          0        0.00  !A.E.L.
CD   CA  NH2 OB        45.000          0        0.00  !A.E.L.
CA   CPT NH2 HC        90.000          0        0.00  !A.E.L.

CC  OS   NH2   O      45.000          0      0.00  !A.E.L.
CC  NH2  OS    O      45.000          0      0.00  !A.E.L.
OS  NH2  CC    O      90.000          0     -180.0 !A.E.L.
OS  O    CC    NH2    90.00           0      180.0 !A.E.L.
CT3 OS   CT1   CT3    90.000         0      -120.0 !A.E.L.

NH2 CC  CPT  HC       20.00          0       0.00  !A.e.L.

NH2 CD  CT2  HC      20.000          0       0.00  !A.E.L.
CPT CA  CY  HP       100.000         0       180.0
NY  CPT CPT HP       100.000         0      -180.0
HP  CPT CY  CA       100.000         0       180.0
HP  CA  CY  CPT      100.000         0      -180.0

CD  CA  NH2  OB       45.000         0         0.00
CD  NH2 CA   OB       45.000         0         0.000
!-------------------------------------------------------
!impropers for ipy drug
!
NH1  C   H   CY       4.00            0      0.00
!--------------------
!imid
CT3 CPH1 NR1 CPH2    90.000          0      180.0
NR1 CPH1 CPH2 CT3    90.000         0       0.00
NR1 CPH2 CPH1 CT3    90.000         0       0.00
CPH2 NR1 NR2  C      90.000         0       0.00
CPH2 NR2 NR1  C      90.000         0       0.00
O CPT C NH1          90.00          0       -180.0

CPH1 CPH2 NR1 CT3    90.000         0       180.0
O   CT2  C   NH1     90.000         0       -180.0
```

```
O   NH1  C   CT2     90.000         0       180.0
CY  C   NH1  H       90.00          0       180.0
C  CA  CY  NH1       90.000         0       0.00
H  C   NH1 CT2       90.000         0       180.0
NC2 O  C   CT2       90.000         0       0.00
C NC2 CT2  HA        90.000         0       120.0
NH1 C CT2  CT2       90.0           0        0.00
C NC2 NC2 CT2        90.00          0        0.00
!-------------------------------------------------------
!distamycin
NH2  CC  H   CY        4.0            0        0.000
NH1  CPT C   O         90.0           0        180.0
O    NH1 C   CPT       90.0           0        180.0
CT2  C   NH1 H         90.0           0        180.0
CC   H   NH2 O         120.0          0         0.00
CC   NH2  H  O         120.0          0         0.0


NONBONDED nbxmod  5 atom cdiel shift vatom vdistance vswitch -
cutnb 14.0 ctofnb 12.0 ctonnb 10.0 eps 1.0 e14fac 1.0 wmin 1.5
                !adm jr., 5/08/91, suggested cutoff scheme
!
!atom  ignored    epsilon      Rmin/2   ignored   eps,1-4
Rmin/2,1-4
!
!Butadiene
CC1A   0.0   -0.0680     2.0900 !
CC1B   0.0   -0.0680     2.0900 !
CC2    0.0   -0.0640     2.0800 !
!sulfate
!OC     2.1400  -0.6469   1.6000
!S      0.3400   -0.0430  1.890
!SCH1, SCH2
NS1    0.000000  -0.200000     1.850000 !N for deprotonated Schiff's
base
NS2    0.000000  -0.200000     1.850000 !N for protonated Schiff's base

end
```

Appendix B

This appendix contains the initial CHARMM parameters for the retinoid model compounds, CHARMM atom names and numbering for the retinoids used in the protein/retinoid simulations, and distances measured from the protein/retinoid simulations. The retinoid topology and parameters are at the end of this appendix.

Initial CHARMM surfaces for retinoid model compounds

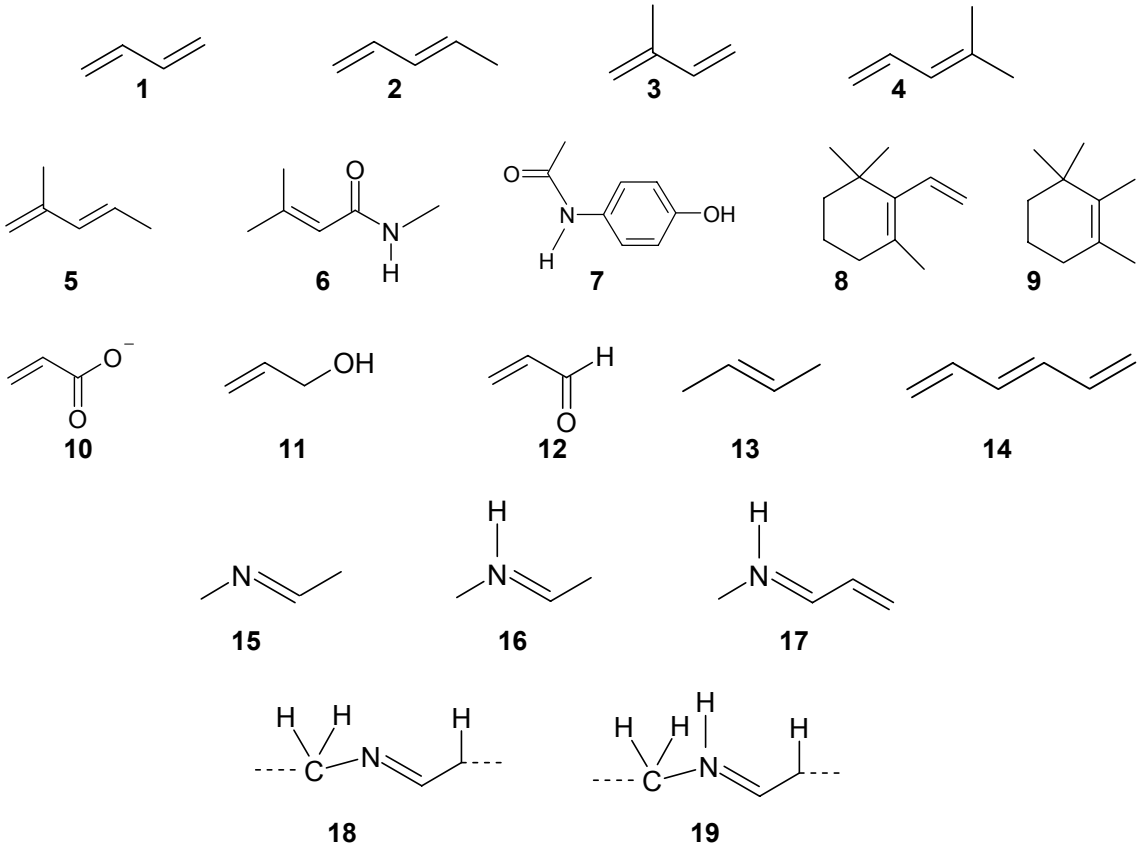


Figure B.1 Model compounds and patch residues for retinoids (some hydrogens not shown for clarity). Models **1**-**17** are the model compounds, and **18** and **19** are used to create patches for forming Schiff bases.

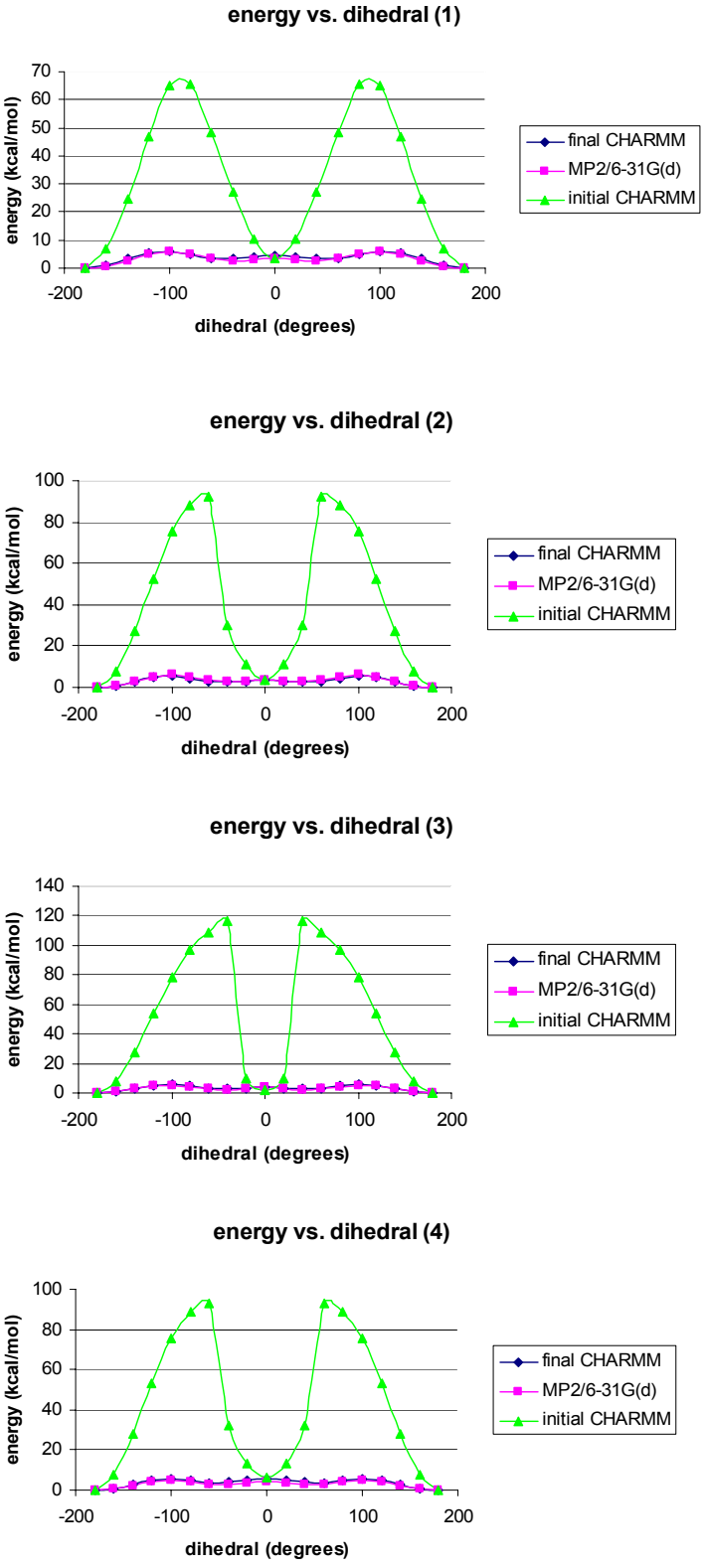


Figure B.2 Initial CHARMM surfaces for model compounds **1-4**.

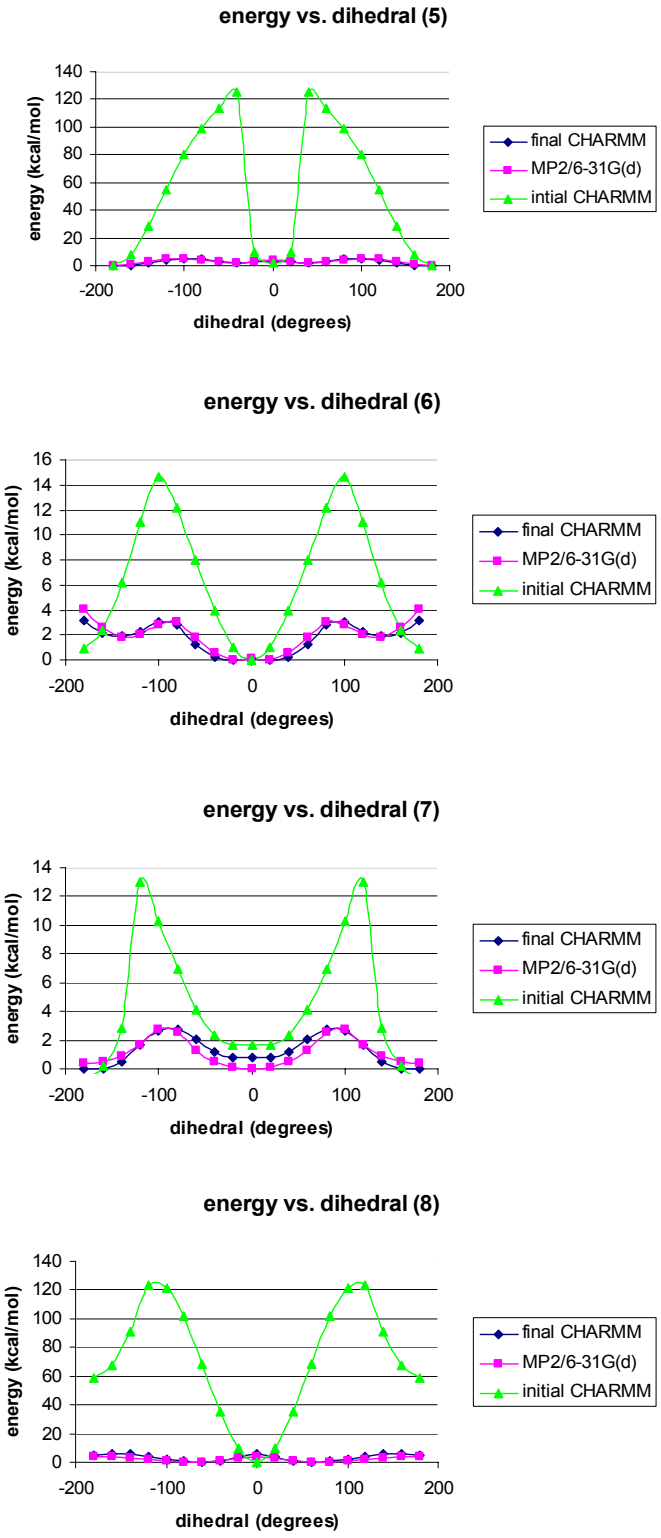


Figure B.3 Initial CHARMM surfaces for model compounds **5**-**8**.

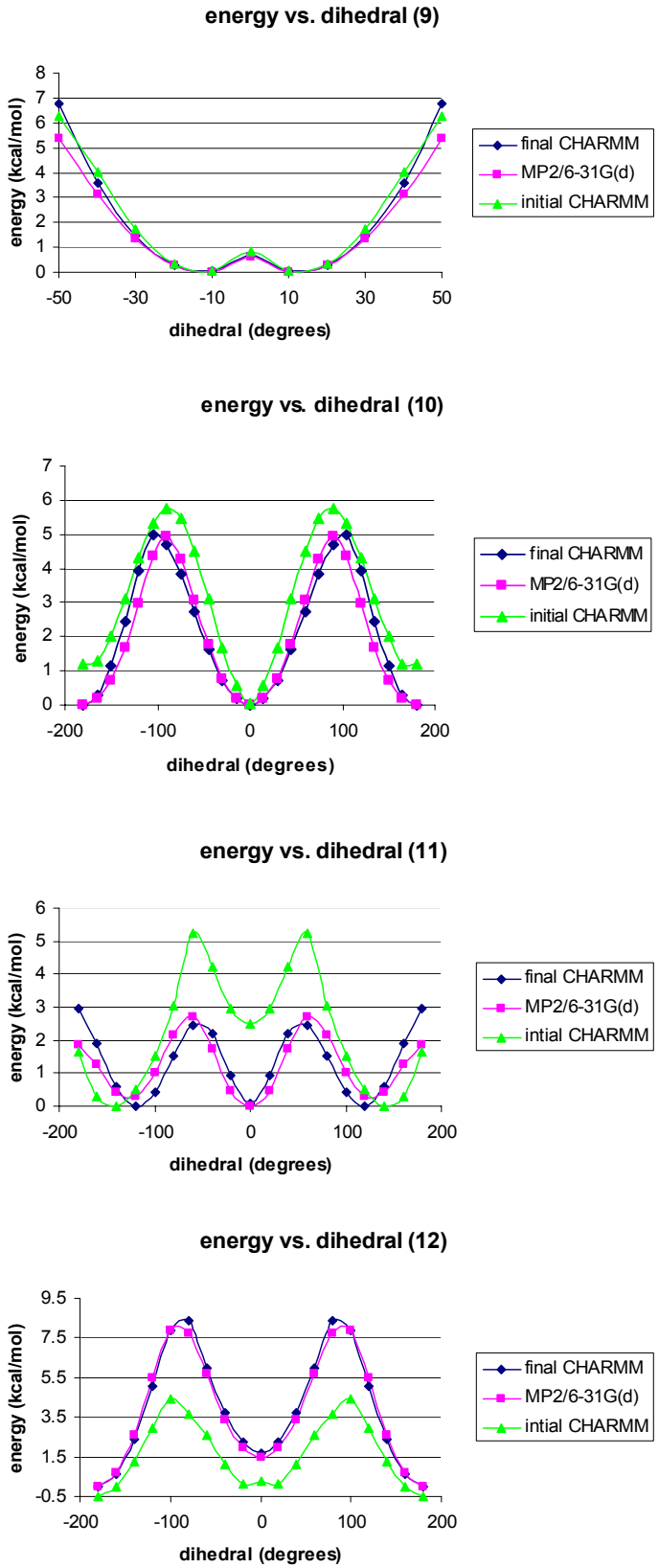


Figure B.4 Initial CHARMM surfaces for model compounds **9**-**12**.

Additional tables for comparison of CHARMM energy minimized structures and experimental data

Table B.1 Experimental, QM, and MM comparison of *p*-acetamide phenol (**7**)[a]

Bond (Å)	x-ray	MP2	CHARMM	Error (MP2 & x-ray)	Error (CHARMM & x-ray)	Error (MP2 & CHARMM)
C22-C27	1.39	1.40	1.41	0.01	0.02	0.01
C27-C26	1.39	1.39	1.40	0	0.01	0.01
C26-C25	1.38	1.40	1.40	0.02	0.02	0
C25-C24	1.38	1.40	1.40	0.02	0.02	0
C24-C23	1.39	1.40	1.41	0.01	0.02	0.01
C25-O28	1.38	1.38	1.41	0	0.03	0.03
C22-N21	1.42	1.41	1.42	0.01	0	0.01
N21-C15	1.34	1.38	1.34	0.04	0	0.04
C15-O29	1.22	1.23	1.22	0.01	0	0.01
C15-C14	1.51	1.52	1.48	0.01	0.03	0.04
Angle (°)						
C22-C27-C26	119.4	119.4	120.5	0	1.1	1.1
C27-C26-C25	120.7	121	119.9	0.3	0.8	1.1
C26-C25-C24	120.2	119.6	120	0.6	0.2	0.4
C25-C24-C23	119.5	119.7	120	0.2	0.5	0.3
C24-C23-C22	120.4	120.9	120.6	0.5	0.2	0.3
C23-C22-C27	119.8	119.3	119	0.5	0.8	0.3
C26-C25-O28	118.9	123.3	120.4	4.4	1.5	2.9
C24-C25-O28	120.8	117.1	119.6	3.7	1.2	2.5
C23-C22-N21	116.5	117.2	115.9	0.7	0.6	1.3
C27-C22-N21	123.8	123.5	125.1	0.3	1.3	1.6
C22-N21-C15	129.8	128.7	129.2	1.1	0.6	0.5
N21-C15-O29	122.7	124.2	124.8	1.5	2.1	0.6
N21-C15-C14	114.6	113.8	115.2	0.8	0.6	1.4
O29-C15-C14	122.7	122	119.9	0.7	2.8	2.1
Dihedral (°)						
C22-C27-C26-C25	0.45	0.09	0.26	0.36	0.19	0.17
C27-C26-C5-C24	2.23	0.13	0	2.1	2.23	0.13
C26-C25-C24-C23	-2.65	-0.11	-0.04	2.54	2.61	0.07
C25-C24-C23-C22	0.4	0.07	0.18	0.33	0.22	0.11
C24-C23-C22-C27	2.27	0.03	0.43	2.24	1.84	0.4
C23-C22-C27-C26	-2.68	-0.05	-0.48	2.63	2.2	0.43
O28-C25-C26-C27	-178.2	-180	-179.9	1.8	1.7	0.1
O28-C25-C24-C23	177.8	180	179.9	2.2	2.1	0.1
C23-C22-N21-C15	-164.3	-179.7	-164.4	15.4	0.1	15.3
C27-C22-N21-C15	17.7	0.2	14.2	17.5	3.5	14
C22-N21-C15-O29	0.57	0.9	3.39	0.33	2.82	2.49
C22-N21-C15-C14	-178.8	-177	-179.4	1.8	0.6	2.4

[a]Table C.1 is continued from Table 4.5 in Chapter 4.

Table B.2 Comparison of CHARMM (energy-minimized structure) and x-ray data for retinal

Bond (Å)	**x-ray**	**CHARMM**	**error**
C1-C2	1.55	1.55	0
C1-C6	1.54	1.53	0.01
C1-C17	1.51	1.55	0.04
C1-C18	1.54	1.55	0.01
C2-C3	1.42	1.53	0.11
C3-C4	1.49	1.53	0.04
C4-C5	1.51	1.51	0
C5-C6	1.33	1.37	0.04
C5-C16	1.51	1.51	0
C6-C7	1.48	1.5	0.02
C7-C8	1.32	1.35	0.03
C8-C9	1.47	1.48	0.01
C9-C10	1.35	1.35	0
C9-C19	1.49	1.51	0.02
C10-C11	1.44	1.48	0.04
C11-C12	1.34	1.35	0.01
C12-C13	1.45	1.48	0.03
C13-C14	1.34	1.35	0.01
C13-C20	1.5	1.51	0.01
C14-C15	1.46	1.48	0.02
C15-O	1.2	1.2	0
Angle (°)			
C2-C1-C6	110.2	112	1.8
C1-C6-C5	122.8	121	1.8
C6-C5-C4	122.6	122.8	0.2
C5-C4-C3	115.3	114.4	0.9
C4-C3-C2	115.5	109.9	5.6
C3-C2-C1	115.8	111.9	3.9
C4-C5-C16	113.1	113.9	0.8
C6-C5-C16	124.3	123.2	1.1
C2-C1-C18	104.7	107.1	2.4
C2-C1-C17	112	109.2	2.8
C6-C1-C18	110.3	110.8	0.5
C6-C1-C17	110.2	109.4	0.8
C1-C6-C7	115	116.1	1.1
C5-C6-C7	122.2	122	0.2
C6-C7-C8	124.5	127.2	2.7
C7-C8-C9	126.5	125.7	0.8
C8-C9-C10	118.3	121.6	3.3
C8-C9-C19	118.4	115	3.4
C10-C9-C19	123.3	123.3	0
C9-C10-C11	127.2	125.7	1.5

Table B.2 (cont'd)

C10-C11-C12	123.6	123.8	0.2
C11-C12-C13	126	124.9	1.1
C12-C13-C14	118.3	123.3	5
C12-C13-C20	118.3	112.7	5.6
C14-C13-C20	123.4	123.5	0.1
C13-C14-C15	125.9	121.9	4
C14-C15-O	123.2	122.8	0.4
Dihedral (°)			
C1-C2-C3-C4	50	59.9	9.9
C2-C3-C4-C5	-30.8	-45.4	14.6
C3-C4-C5-C6	6.83	21.3	14.47
C4-C5-C6-C1	-2.15	-9.54	7.39
C5-C6-C1-C2	18.7	22.6	3.9
C6-C1-C2-C3	-42.6	-48.2	5.6
C16-C5-C4-C3	-172.3	-162.8	9.5
C16-C5-C6-C1	176.9	175	1.9
C16-C5-C6-C7	-2.37	-6.5	4.13
C18-C1-C6-C5	133.9	142	8.1
C18-C1-C6-C7	-46.8	-48.9	2.1
C18-C1-C2-C3	-161.2	-169.8	8.6
C17-C1-C6-C5	-105.5	-98.7	6.8
C17-C1-C6-C7	73.9	70.5	3.4
C17-C1-C2-C3	80.5	73.2	7.3
C5-C6-C7-C8	-58.3	-58.7	0.4
C1-C6-C7-C8	122.4	132.3	9.9
C6-C7-C8-C9	-179.3	-179.7	0.4
C7-C8-C9-C10	175.8	134.2	41.6
C7-C8-C9-C19	-4.47	-43.1	38.63
C8-C9-C10-C11	179.1	178.5	0.6
C9-C10-C11-C12	-179.1	-177	2.1
C19-C9-C10-C11	-0.64	-1.46	0.82
C10-C11-C12-C13	175.5	174.6	0.9
C11-C12-C13-C14	-178.4	-63.4	115
C11-C12-C13-C20	-0.63	-124.9	124.27
C12-C13-C14-C15	177.7	179.5	1.8
C20-C13-C14-C15	0	-8.71	8.71
C13-C14-C15-O	179.3	178.4	0.9

Table B.3 Comparison of CHARMM (energy-minimized structure) and x-ray data for retinoic acid (triclinic form)

Bond (Å)	**x-ray**	**CHARMM**	**error**
C1-C2	1.54	1.55	0.01
C1-C6	1.54	1.53	0.01
C1-C16	1.53	1.54	0.01
C1-C17	1.51	1.54	0.03
C2-C3	1.43	1.53	0.1
C3-C4	1.5	1.53	0.03
C4-C5	1.51	1.51	0
C5-C6	1.34	1.37	0.03
C5-C18	1.51	1.51	0
C6-C7	1.47	1.49	0.02
C7-C8	1.34	1.35	0.01
C8-C9	1.45	1.48	0.03
C9-C10	1.35	1.35	0
C9-C19	1.51	1.51	0
C10-C11	1.44	1.48	0.04
C11-C12	1.34	1.35	0.01
C12-C13	1.46	1.49	0.03
C13-C14	1.34	1.35	0.01
C13-C20	1.49	1.51	0.02
C14-C15	1.47	1.49	0.02
C15-O29	1.31	1.26	0.05
C15-O30	1.21	1.26	0.05
O29-O30	2.66	2.22	0.44
Angle (°)			
C2-C1-C6	110.2	112.5	2.3
C2-C1-C16	105.4	106.9	1.5
C2-C1-C17	110.9	109.1	1.8
C6-C1-C16	111.5	111.9	0.4
C6-C1-C17	109.1	108	1.1
C16-C1-C17	109.8	108.3	1.5
C1-C2-C3	116.3	112.4	3.9
C2-C3-C4	112.5	108.9	3.6
C3-C4-C5	113.6	113.5	0.1
C4-C5-C6	123	122.8	0.2
C4-C5-C18	112.3	114.5	2.2
C6-C5-C18	124.6	122.6	2
C1-C6-C5	122.1	121.9	0.2
C1-C6-C7	115	117.1	2.1
C5-C6-C7	122.9	121	1.9
C6-C7-C8	127.7	125.9	1.8
C7-C8-C9	125.6	124.8	0.8
C8-C9-C10	118.6	123.5	4.9
C8-C9-C19	118.3	112.7	5.6

Table B.3 (cont'd).

C10-C9-C19	123.2	123.7	0.5
C9-C10-C11	126.2	125.7	0.5
C10-C11-C12	124.5	123.9	0.6
C11-C12-C13	126.1	126.3	0.2
C12-C13-C14	118.4	121.2	2.8
C12-C13-C20	117.9	115.5	2.4
C14-C13-C20	123.7	123.4	0.3
C13-C14-C15	126.5	129.9	3.4
C14-C15-O30	112.3	114.8	2.5
C14-C15-O30	126	121.7	4.3
O29-C15-O30	121.7	123	1.3
Dihedral (°)			
C1-C2-C3-C4	-60	-60.8	0.8
C2-C3-C4-C5	-45	48.7	93.7
C3-C4-C5-C6	-11.5	-20.3	8.8
C4-C5-C6-C1	9.74	1.45	8.29
C5-C6-C1-C2	-11	-12.3	1.3
C6-C1-C2-C3	43.4	42.3	1.1
C18-C5-C4-C3	175	177.9	2.9
C18-C5-C6-C1	165.4	163	2.4
C18-C5-C6-C7	-4	-1.69	2.31
C16-C1-C6-C5	102.5	108.2	5.7
C16-C1-C6-C7	-108	-72.2	35.8
C16-C1-C2-C3	-77	-77.5	0.5
C17-C1-C6-C5	-126	-132.7	6.7
C17-C1-C6-C7	45.4	46.9	1.5
C17-C1-C2-C3	168.1	165.6	2.5
C1-C6-C7-C8	132.9	115.5	17.4
C5-C6-C7-C8	-63.7	-64.7	1
C6-C7-C8-C9	-180	-170.3	9.7
C7-C8-C9-C19	0	-38.3	38.3
C7-C8-C9-C10	178.5	140.3	38.2
C8-C9-C10-C11	-177.7	-178.1	0.4
C19-C9-C10-C11	0.66	-1.85	2.51
C9-C10-C11-C12	179.2	178.3	0.9
C10-C11-C12-C13	-179.5	-176.6	2.9
C11-C12-C13-C20	1.2	-33.8	35
C11-C12-C13-C14	-180	-146.4	33.6
C12-C13-C14-C15	-178	-179.4	1.4
C20-C13-C14-C15	0.68	0.3	0.38
C13-C14-C15-O29	-7.61	-11.7	4.09
C13-C14-C15-O30	176.4	175.5	0.9

Table B.4 Comparison of CHARMM (energy-minimized structure) and x-ray data for retinoic acid (monoclinic form)

Bond (Å)	**x-ray**	**CHARMM**	**error**
C1-C2	1.55	1.55	0
C1-C6	1.53	1.53	0
C1-C16	1.54	1.54	0
C1-C17	1.53	1.54	0.01
C2-C3	1.5	1.53	0.03
C3-C4	1.51	1.53	0.02
C4-C5	1.5	1.51	0.01
C5-C6	1.36	1.37	0.01
C5-C18	1.52	1.51	0.01
C6-C7	1.46	1.49	0.03
C7-C8	1.34	1.35	0.01
C8-C9	1.45	1.48	0.03
C9-C10	1.35	1.35	0
C9-C19	1.5	1.51	0.01
C10-C11	1.44	1.48	0.04
C11-C12	1.34	1.35	0.01
C12-C13	1.45	1.49	0.04
C13-C14	1.34	1.35	0.01
C13-C20	1.5	1.51	0.01
C14-C15	1.47	1.49	0.02
C15-O29	1.32	1.26	0.06
C15-O30	1.22	1.26	0.04
O29-O30	2.67	2.22	0.45
Angle (°)			
C2-C1-C6	110.1	112.5	2.4
C2-C1-C16	104.5	106.9	2.4
C2-C1-C17	110.1	109.1	1
C6-C1-C16	111.1	111.9	0.8
C6-C1-C17	111.3	108	3.3
C16-C1-C17	109.6	108.3	1.3
C1-C2-C3	112.2	112.4	0.2
C2-C3-C4	109.5	108.9	0.6
C3-C4-C5	113.6	113.5	0.1
C4-C5-C6	123.7	122.8	0.9
C4-C5-C18	112.5	114.5	2
C6-C5-C18	123.7	122.6	1.1
C1-C6-C5	121.6	121.9	0.3
C1-C6-C7	120.7	117.1	3.6
C5-C6-C7	117.7	121	3.3
C6-C7-C8	131.5	125.9	5.6
C7-C8-C9	125	124.8	0.2
C8-C9-C10	119.3	123.5	4.2
C8-C9-C19	118.4	112.7	5.7

Table B.4 (cont'd)

C10-C9-C19	122.3	123.7	1.4
C9-C10-C11	127.3	125.7	1.6
C10-C11-C12	123.4	123.9	0.5
C11-C12-C13	126.6	126.3	0.3
C12-C13-C14	117.4	121.2	3.8
C12-C13-C20	118.5	115.5	3
C14-C13-C20	124.1	123.4	0.7
C13-C14-C15	128.8	129.9	1.1
Dihedral (°)			
C14-C15-O29	112.1	114.8	2.7
C14-C15-O30	125.9	121.7	4.2
O29-C15-O30	122	123	1
C1-C2-C3-C4	-63.3	-60.8	2.5
C2-C3-C4-C5	43.2	48.7	5.5
C3-C4-C5-C6	-10.4	-20.3	9.9
C4-C5-C6-C1	-4.8	1.45	6.25
C5-C6-C1-C2	-13.4	-12.3	1.1
C6-C1-C2-C3	47.5	42.3	5.2
C18-C5-C6-C1	175.1	177.9	2.8
C18-C5-C4-C3	169.6	163	6.6
C18-C5-C6-C7	-3.1	-1.69	1.41
C16-C1-C6-C5	108.9	108.2	0.7
C16-C1-C6-C7	-72.9	-72.2	0.7
C16-C1-C2-C3	-75.5	-77.5	2
C17-C1-C6-C5	-128.7	-132.7	4
C17-C1-C6-C7	49.5	46.9	2.6
C17-C1-C2-C3	167	165.6	1.4
C1-C6-C7-C8	15.9	115.5	99.6
C5-C6-C7-C8	-165.8	-64.7	101.1
C6-C7-C8-C9	179.6	179.6	0
C7-C8-C9-C10	176.2	140.3	35.9
C7-C8-C9-C19	-1.9	-38.1	36.2
C8-C9-C10-C11	-178.2	-178.1	0.1
C19-C9-C10-C11	-0.13	-1.85	1.72
C9-C10-C11-C12	172	178.3	6.3
C10-C11-C12-C13	-177.8	-176.6	1.2
C11-C12-C13-C14	179.4	146	33.4
C11-C12-C13-C20	1.9	33.8	31.9
C12-C13-C14-C15	-179.8	-179.4	0.4
C15-C14-C13-C20	-2.5	0.3	2.8
O30-C15-C14-C13	-6.9	-11.7	4.8
O29-C15-C14-C13	173.7	175.5	1.8

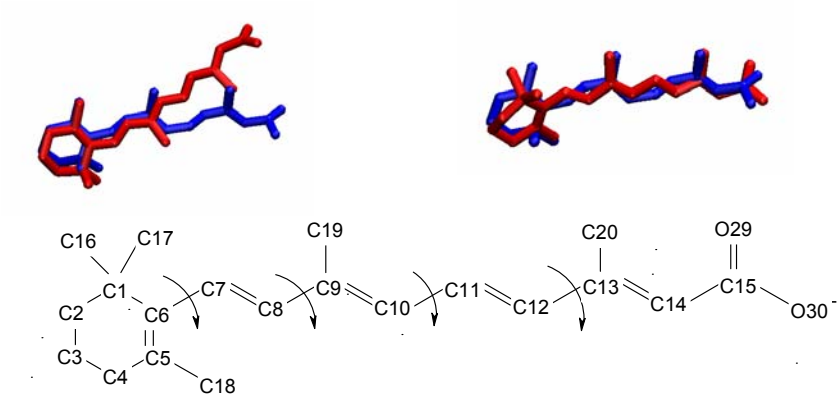


Figure B.5 Overlay of cyclohexene rings (left) and overlay of chains (right). X-ray structure is shown in red; average structure from the simulation is shown in blue. Bottom picture shows the CHARMM atom names of the carbons in the retinoic acid ligand, and it shows the dihedrals that contribute to the large average dihedral error.

The chain rotates with respect to the cyclohexene ring during the simulation, which causes large deviations in the torsion angles that are involved around the connection of the ring to the chain (rotation about the C6-C7 bond). The other sources of large RMSD differences involve parts of the chain (rotation about C8-C9, C10-C11, and C12-C13). These differences are seen during the simulation because these are single bonds, and the ligand will probably try to orient itself to optimize contacts with the protein side chains, and this does not appear to cause any major distortion of the surrounding protein residues. The cyclohexene ring retains the C2'-*endo* conformation during the simulation.

The protein-ligand contact time series plots not shown in Chapter 4 are shown below in Figures B.6 and B.7.

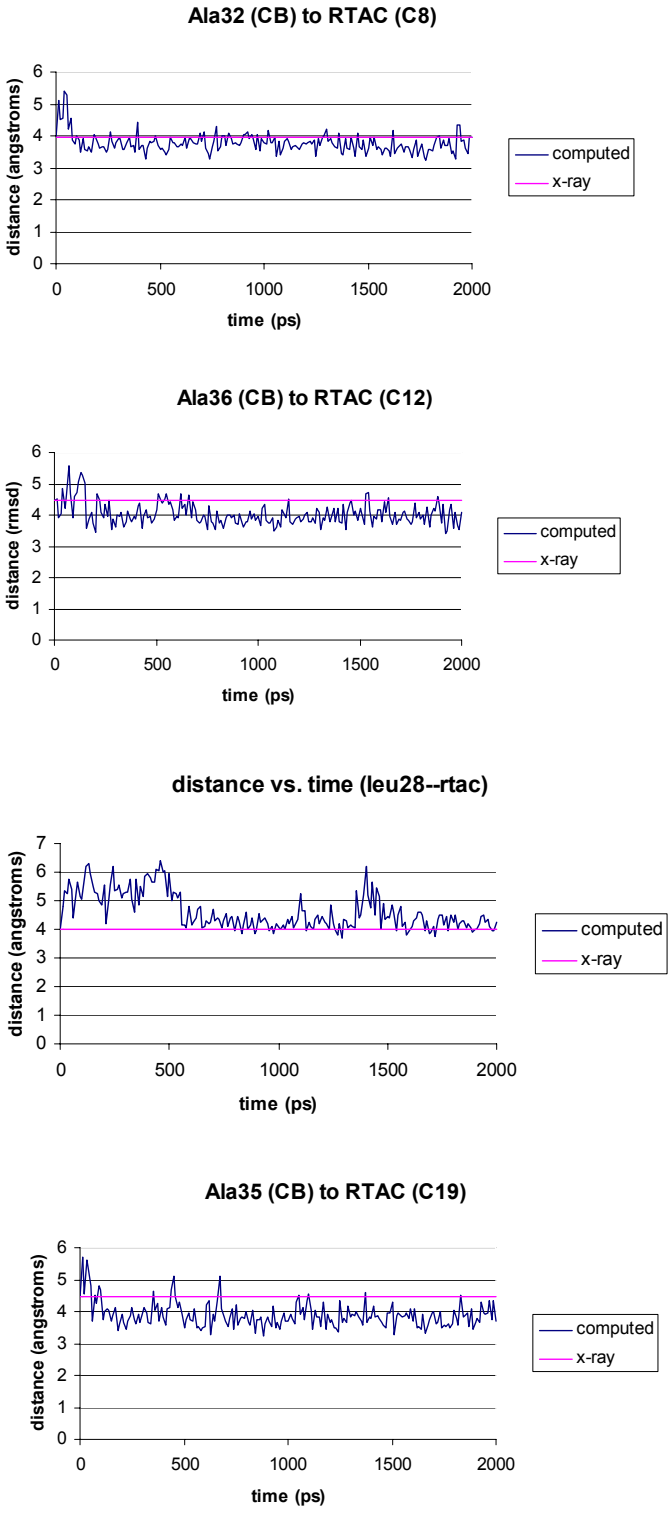


Figure B.6 Retinoic acid-protein distances.

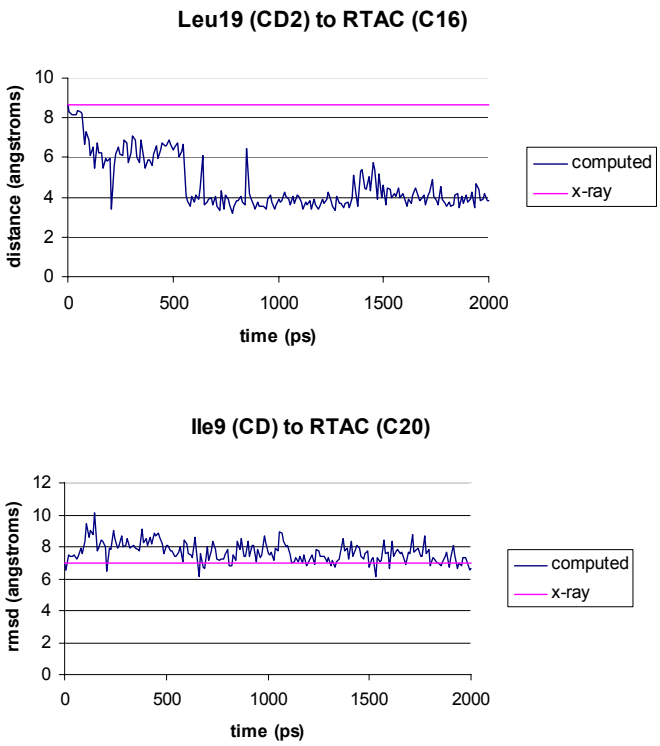


Figure B.7 Retinoic acid-protein distances.

The protein-protein contacts not shown in Chapter 4 are shown in Figure B.8.

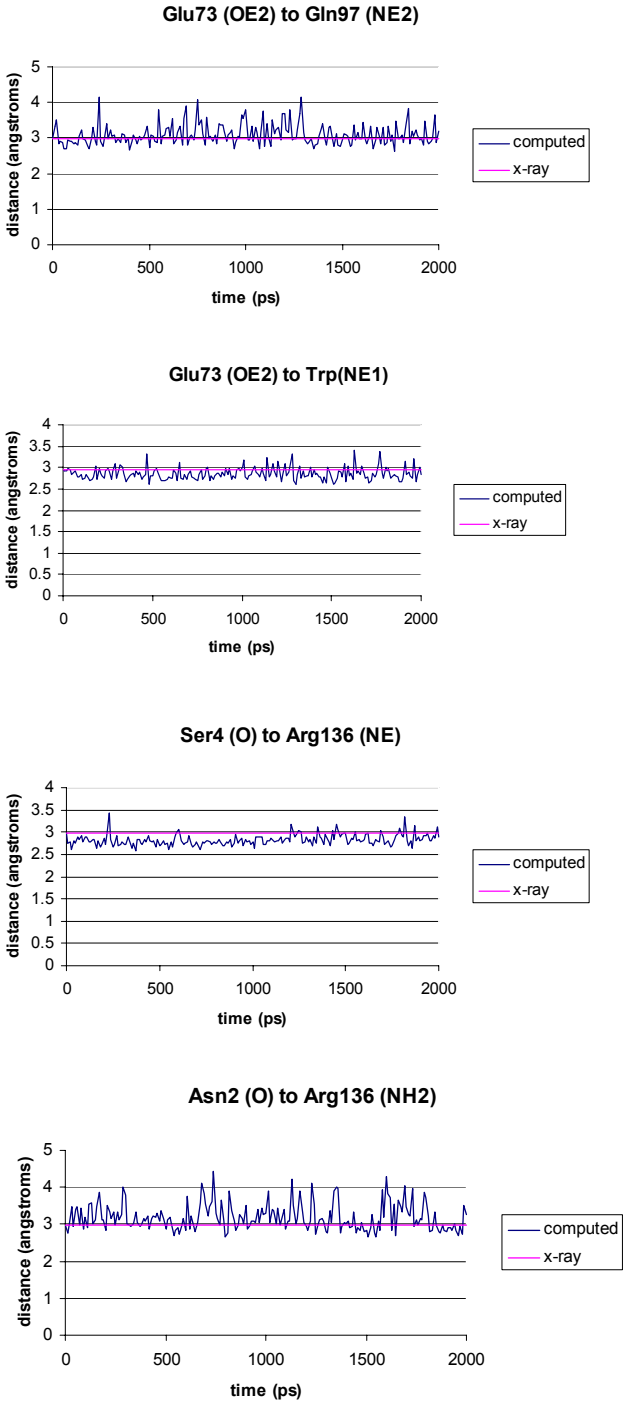


Figure B.8 Protein-protein distances from CRABPII/retinoic acid simulation.

Additional information and plots for CRBP/retinol simulation

The protein-ligand distance plots not shown in Chapter 4 are shown in Figures B.9-B.12.

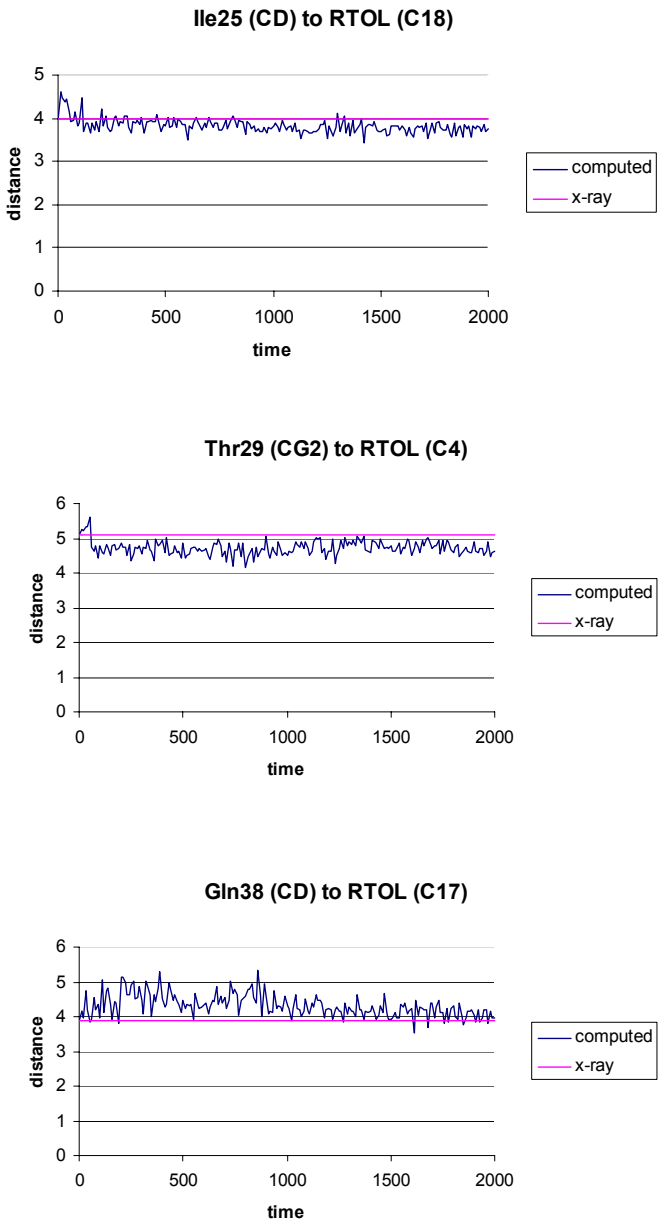


Figure B.9 Protein-ligand distances from CRBP/retinol simulation.

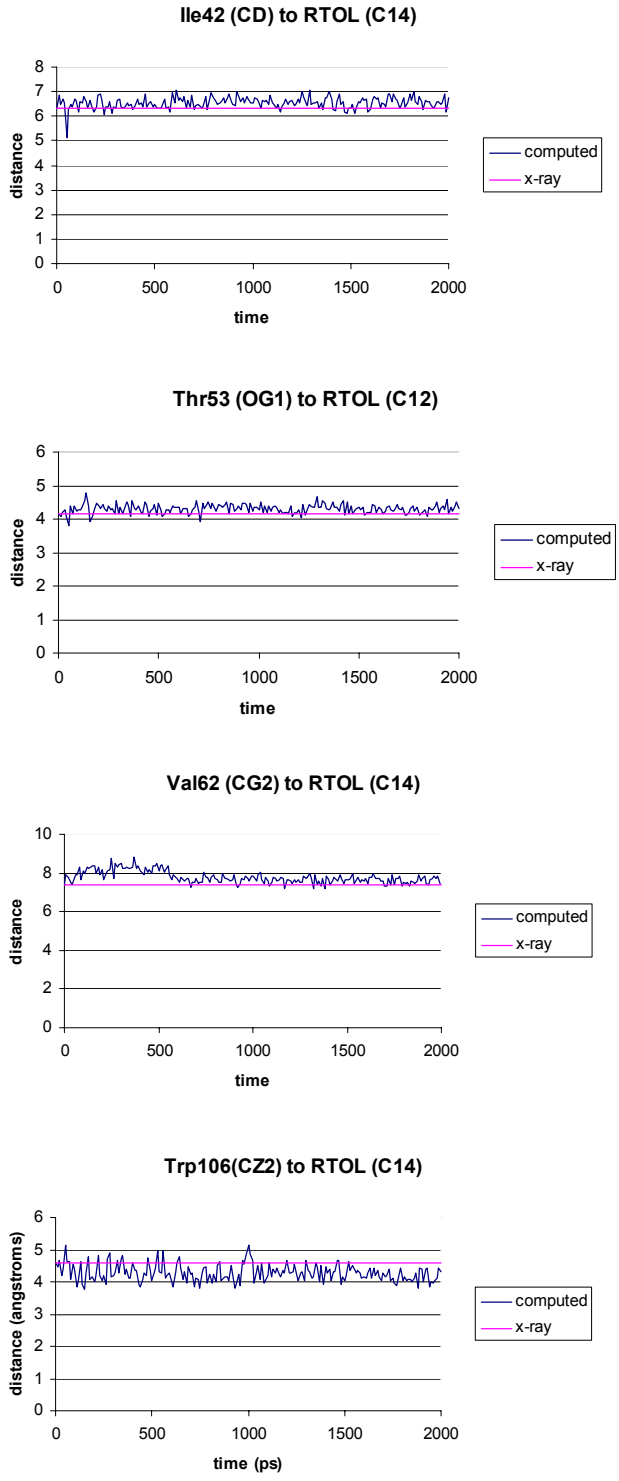


Figure B.10 Protein-ligand distances from CRBP/retinol simulation.

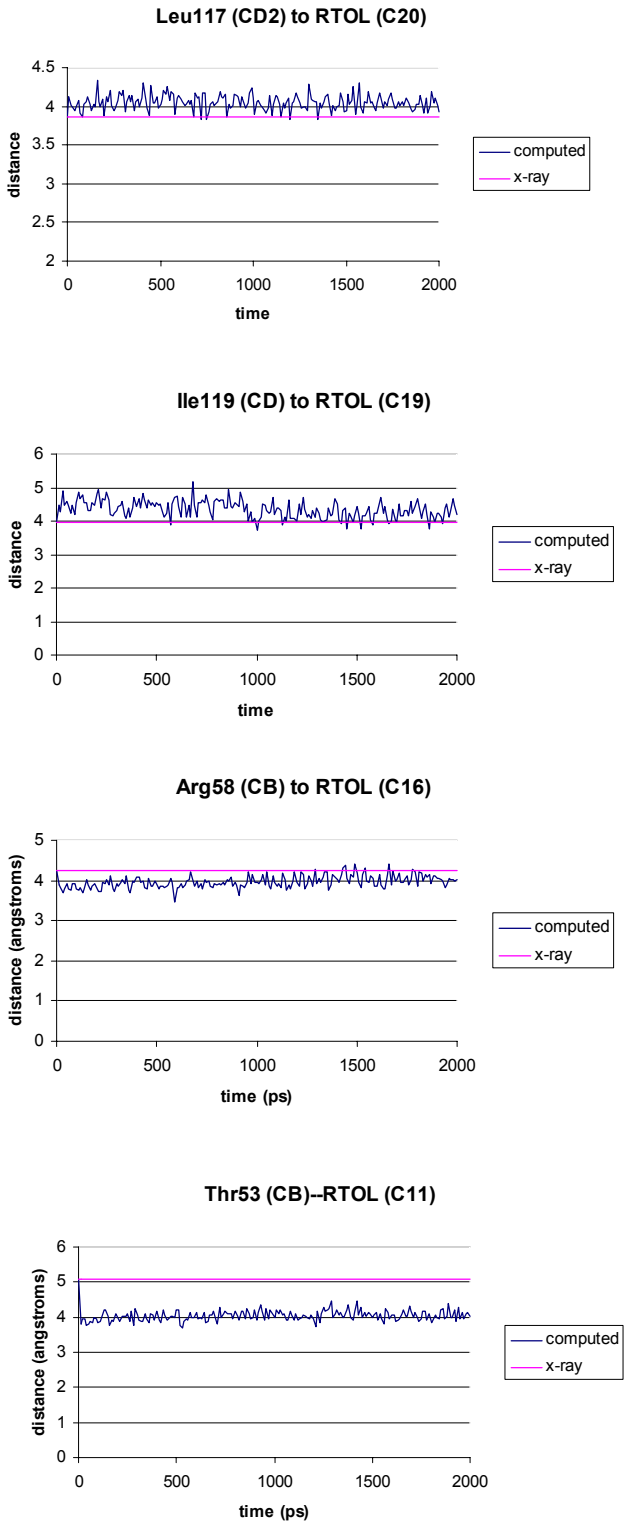


Figure B.11 Protein-ligand distances from CRBP/retinol simulation.

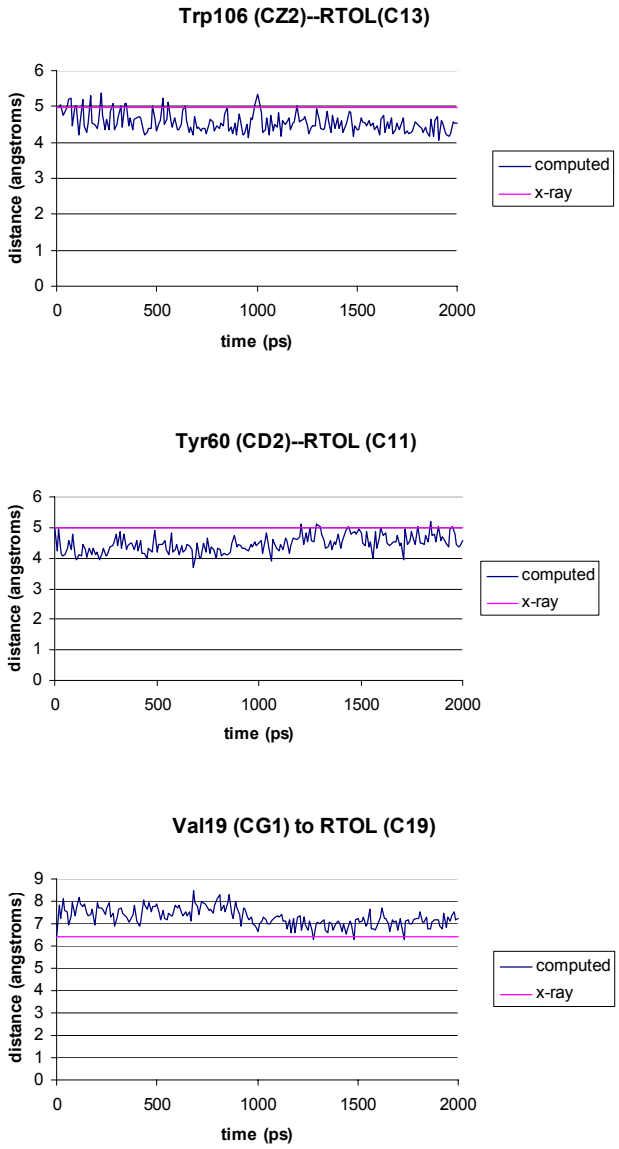


Figure B.12 Protein-ligand distances from CRBP/retinol simulation.

Additional information and plots for RBP/fenretinide simulation

The plots shown are for the ligand-protein distances not shown in Chapter 4 (B.13-B.17).

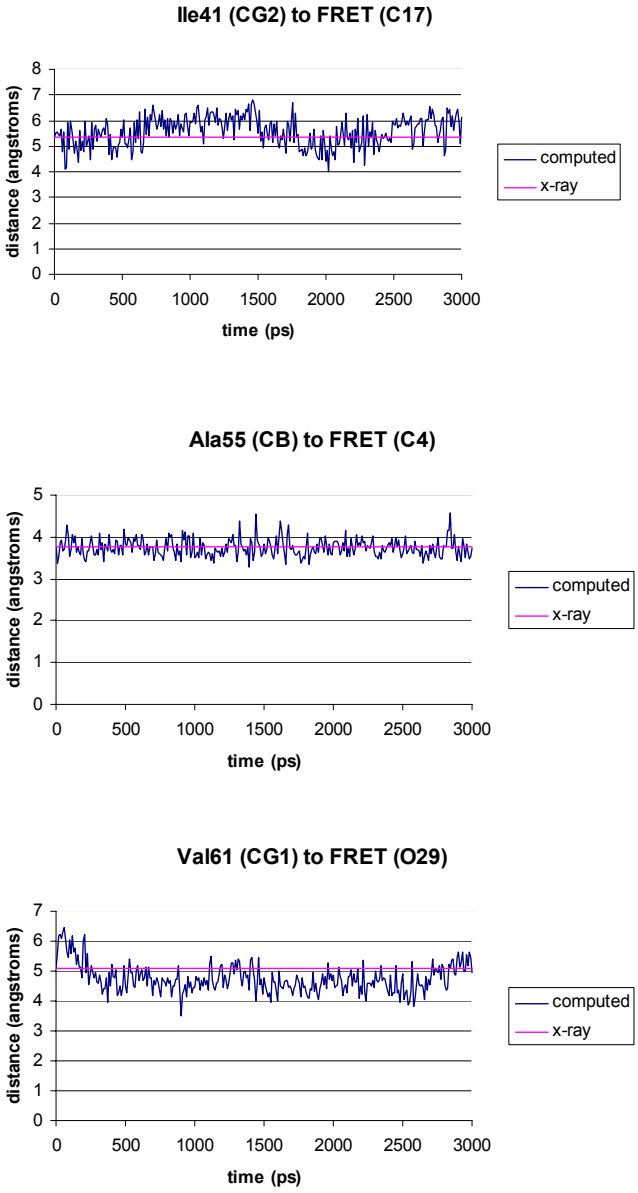


Figure B.13 Protein-ligand distances from RBP/fenretinide simulation.

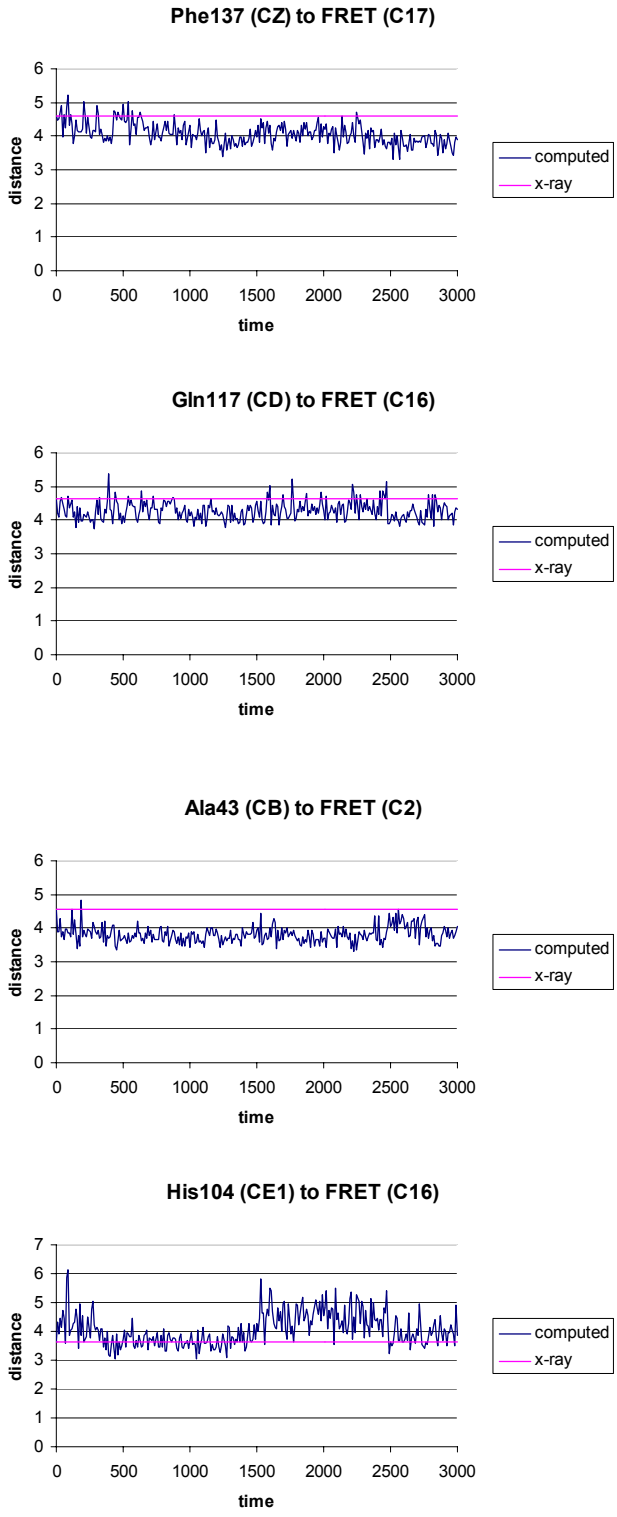
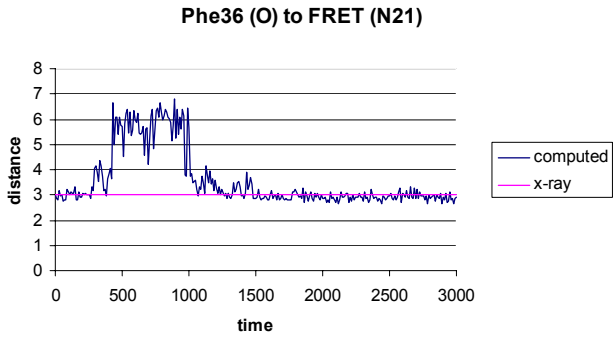
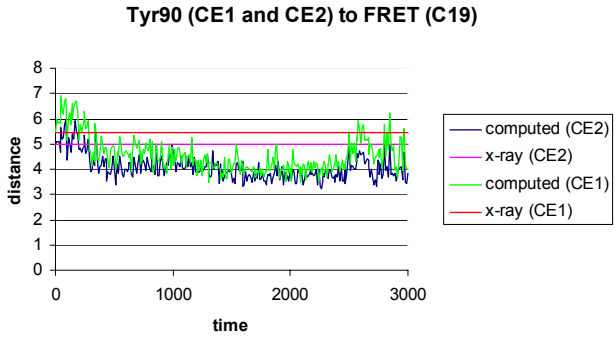
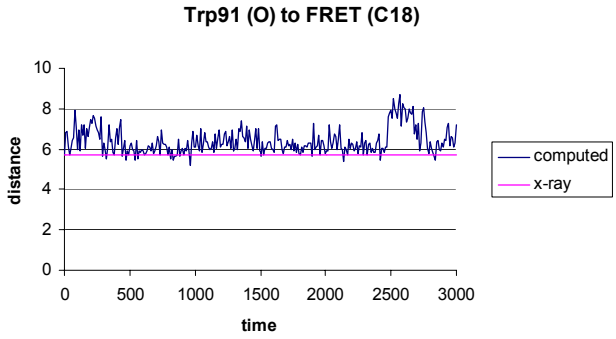


Figure B.14 Protein-ligand distances from RBP/fenretinide simulation.







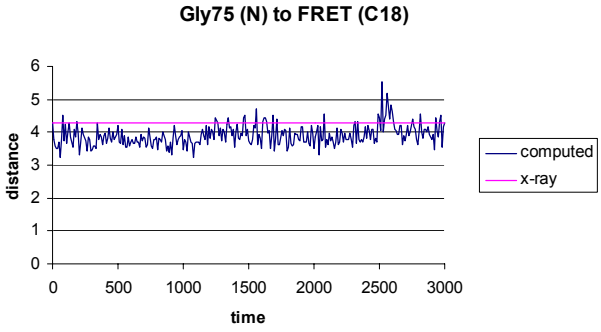


Figure B.15 Protein-ligand distances from RBP/fenretinide simulation.

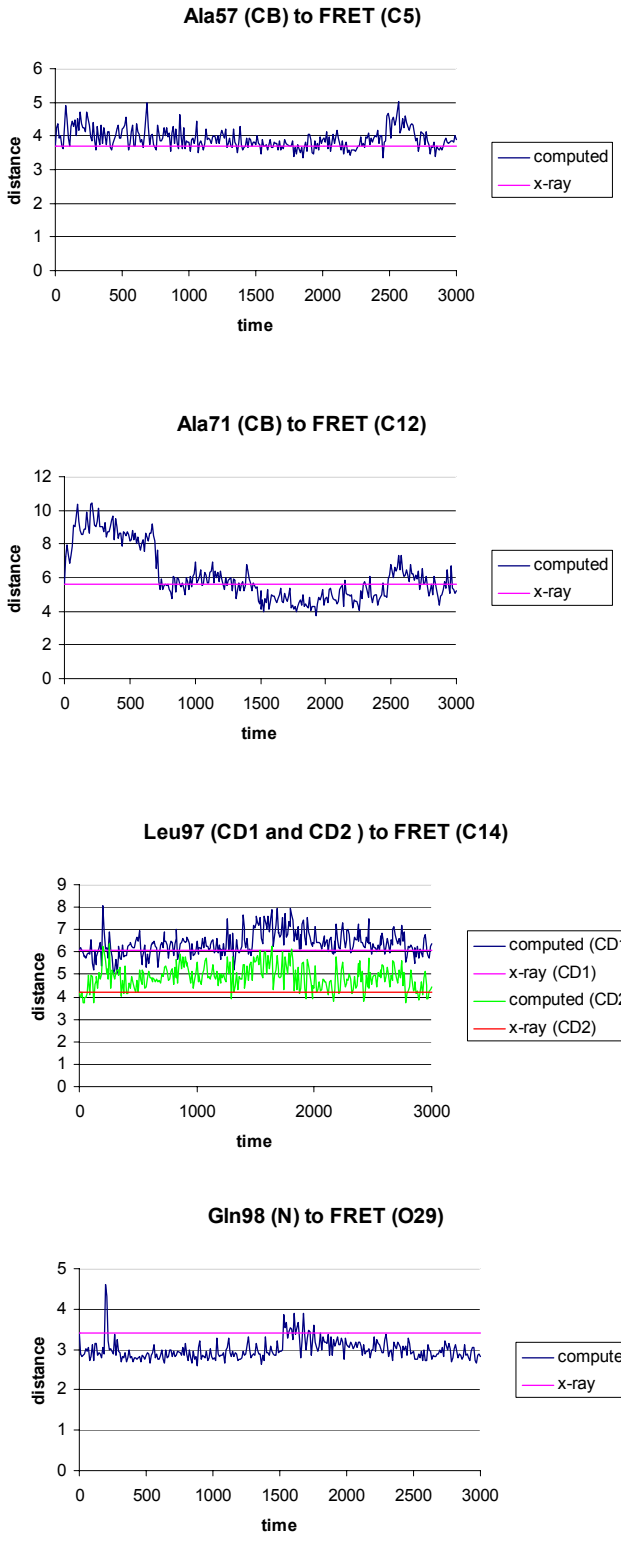


Figure B.16 Protein-ligand distances from RBP/fenretinide simulation.

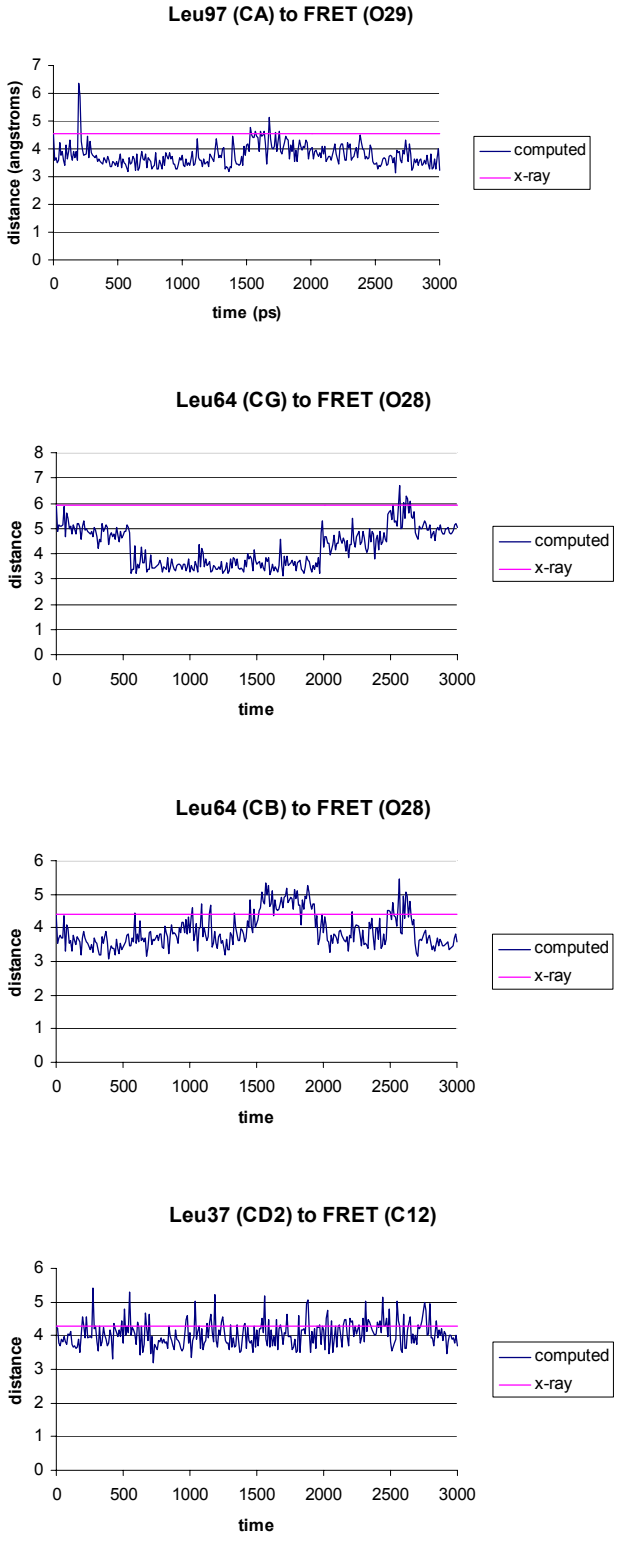


Figure B.17 Protein-ligand distances from RBP/fenretinide simulation.

Additional information and plots for SRII/retinal simulation

The plots below are the ligand-protein distances not shown in Chapter 4.

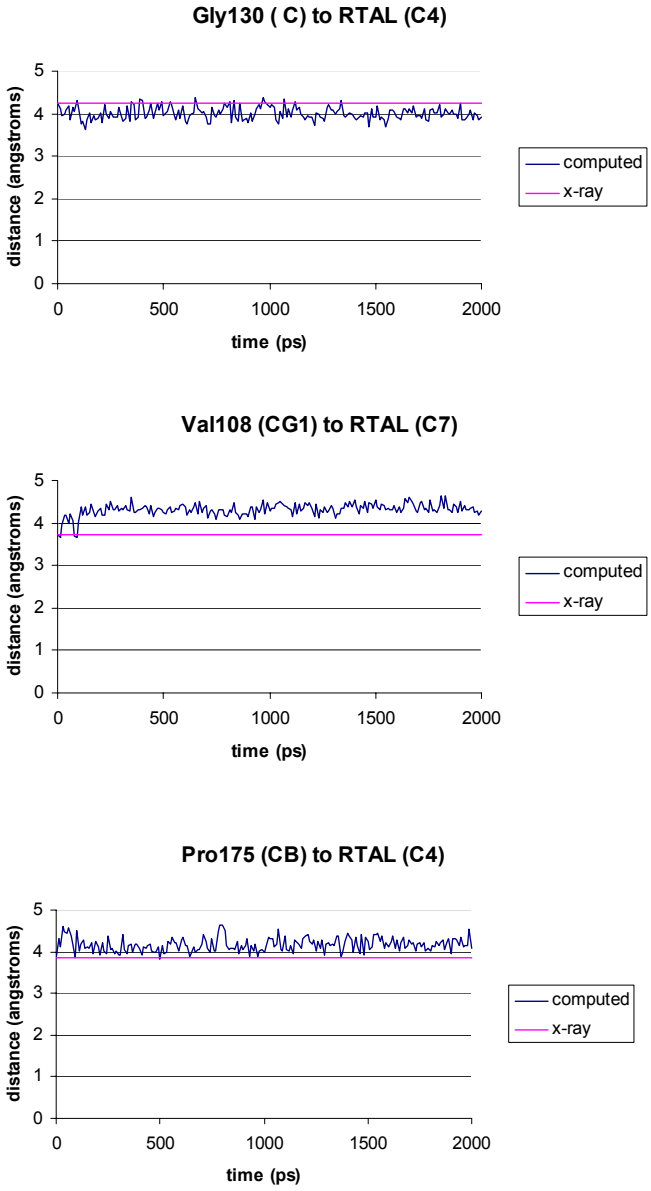


Figure B.18 Protein-ligand distances from SRII/retinal simulation.

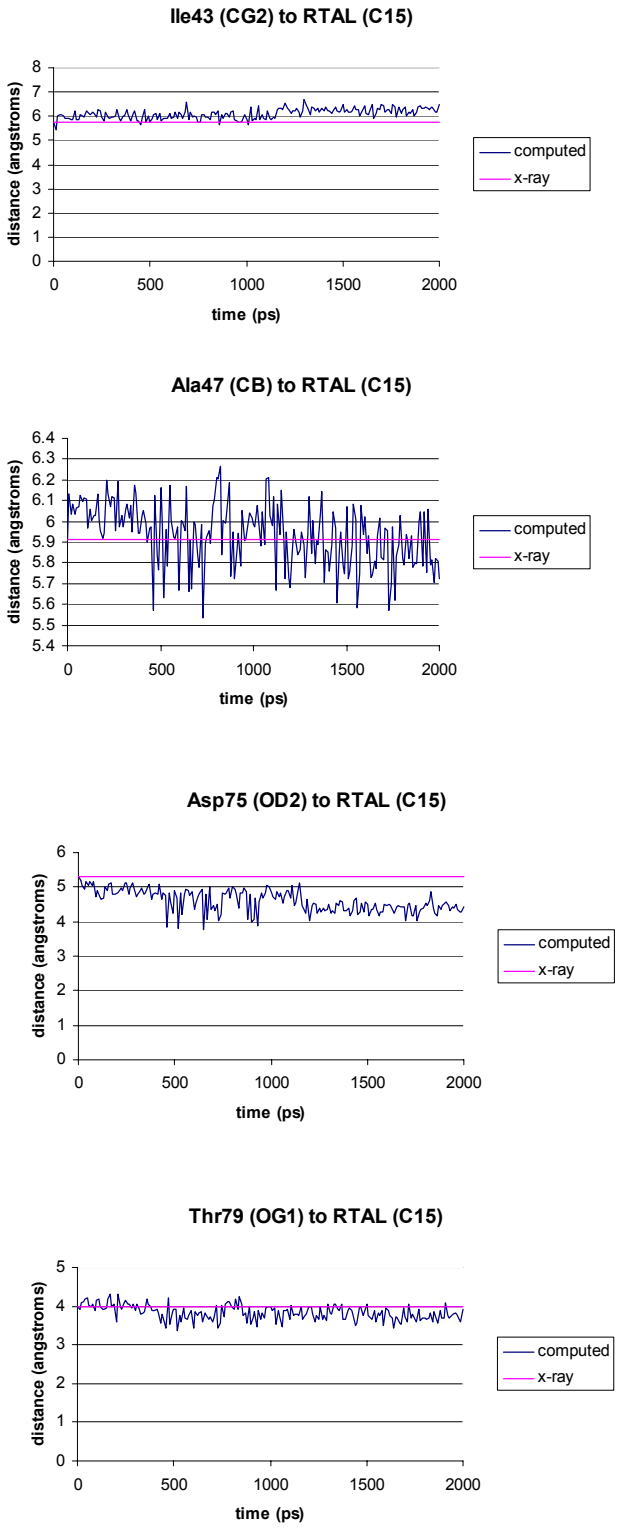
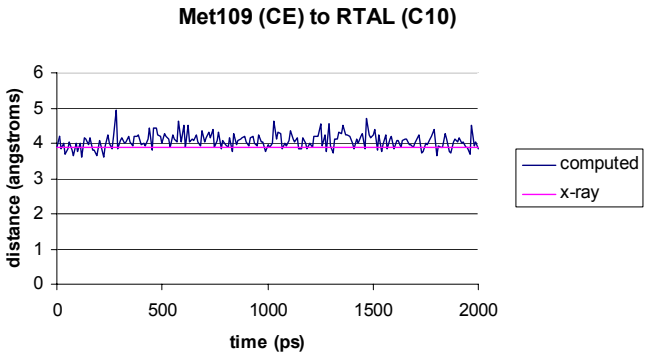
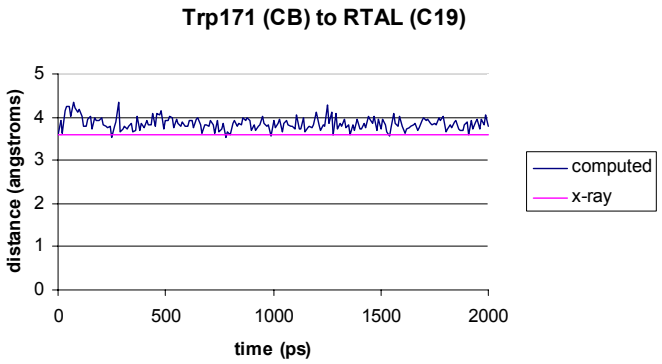


Figure B.19 Protein-ligand distances from SRII/retinal simulation.





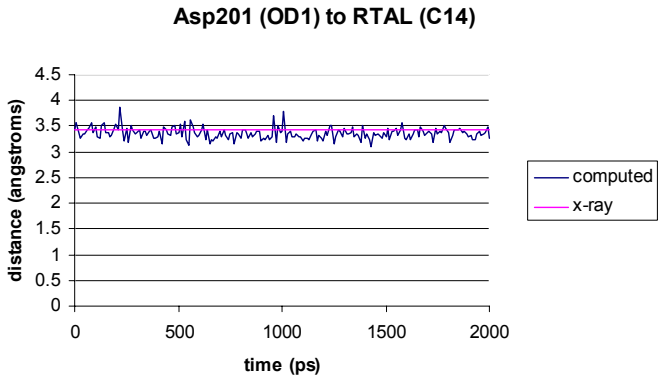


Figure B.20 Protein-ligand distances from SRII/retinal simulation.

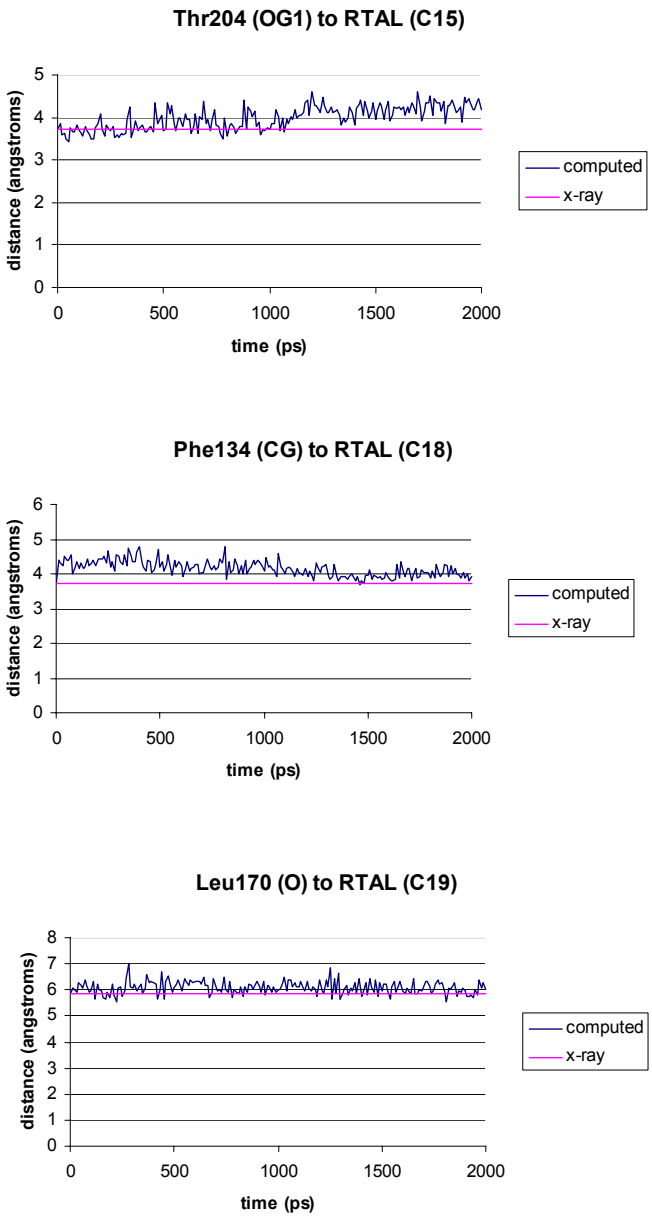


Figure B.21 Protein-ligand distances from SRII/retinal simulation.

Additional information and plots for BR/retinal simulation

The plots below are the ligand-protein distances not shown in Chapter 4 (B.22-B.23).

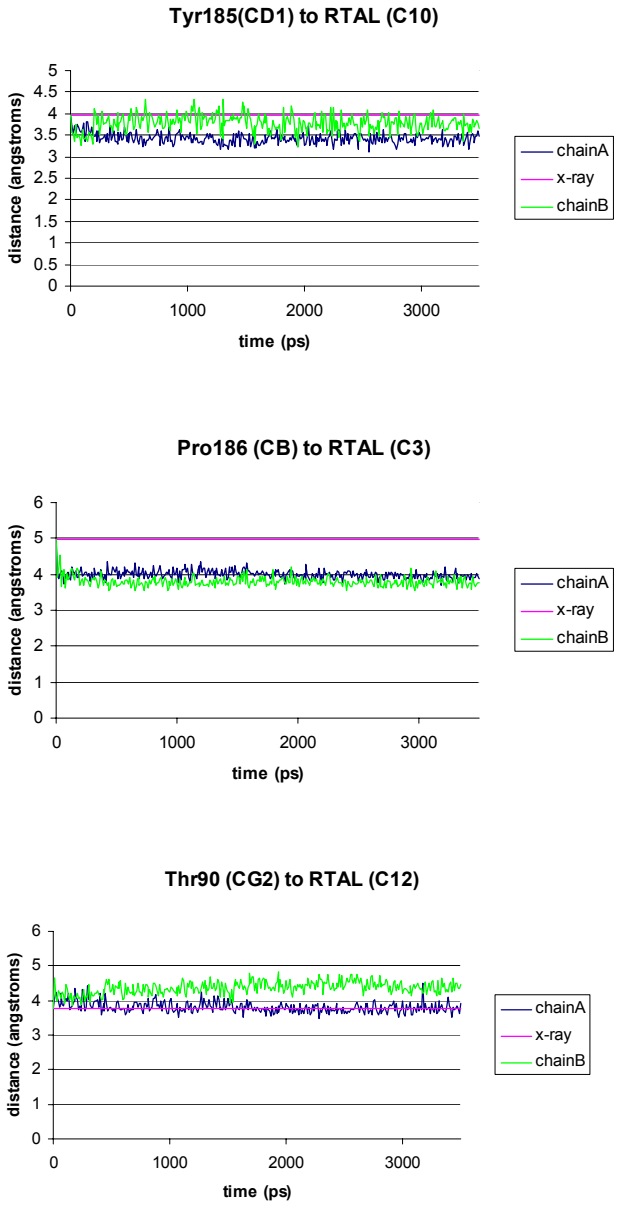


Figure B.22 Protein-ligand distances from BR/retinal simulation.

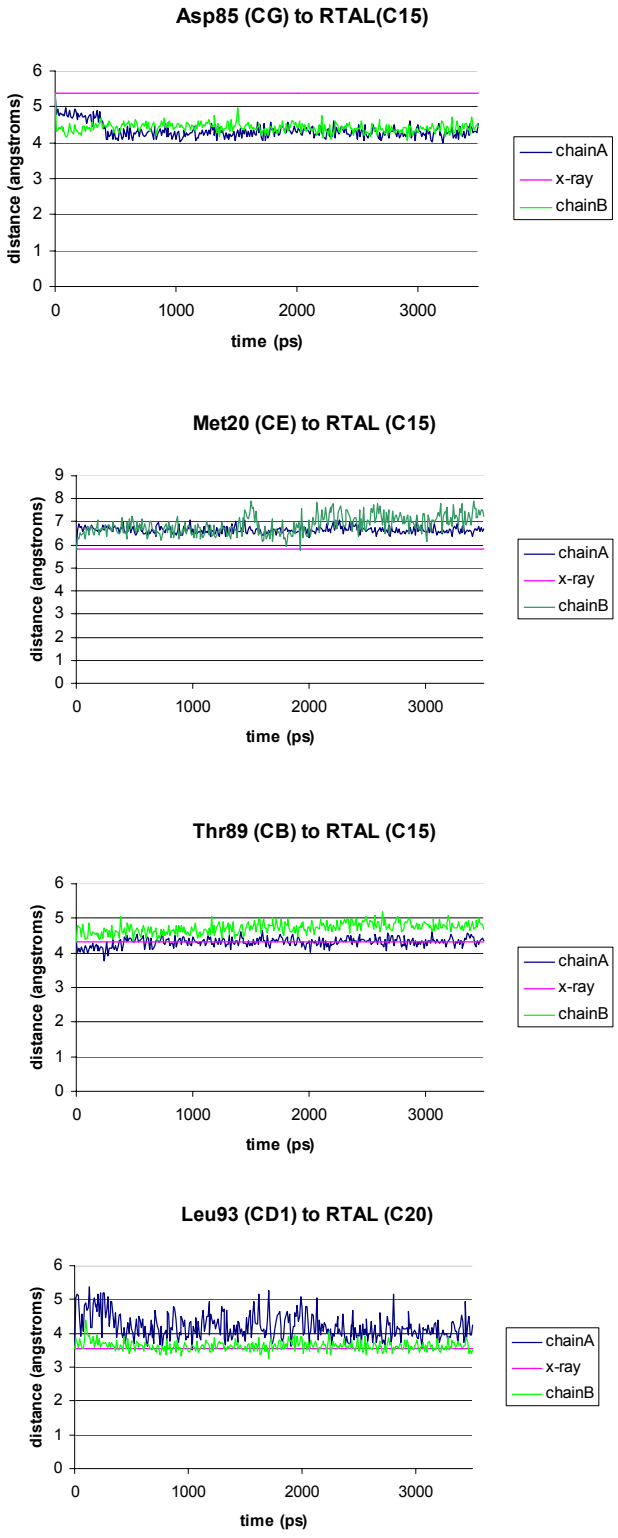


Figure B.23 Protein-ligand distances from BR/retinal simulation.

CHARMM Topology and Parameters for Retinoids and Retinoid Model Compounds

Topology and parameters for retinoids were streamed into topology and parameters for proteins (top_all27_prot_na.inp and par_all27_prot_na.inp with CMAP backbone dihedral correction). The CHARMM atom names and numbering used in plots and tables can also be found in the topology.

Topology for **1**

```
RESI 13DB       0.00 ! 1,3-dibutene
!
GROUP
ATOM C1   CC2  -0.42 ! H11     H21
ATOM H11  HE2   0.21 !   \    /
ATOM H12  HE2   0.21 !    C1=C2      H41
GROUP                !   /    \     /
ATOM C2   CC1A -0.15 ! H12     C3=C4
ATOM H21  HE1   0.15 !        /     \
ATOM C3   CC1B -0.15 !       H31     H42
ATOM H31  HE1   0.15 !
GROUP                !
ATOM C4   CC2  -0.42 !
ATOM H41  HE2   0.21 !
ATOM H42  HE2   0.21 !

BOND C1   C2   C2   C3   C3   C4
BOND C1   H11  C1   H12  C2   H21   C3   H31
BOND C4   H41  C4   H42

IC  C1   C2   C3   C4   0.00 0.00  180.0 0.00 0.00
IC  C2   C3   C4   H41  0.00 0.00  180.0 0.00 0.00
IC  C3   C2   C1   H11  0.00 0.00  180.0 0.00 0.00
IC  C2   H11  *C1  H12  0.00 0.00  180.0 0.00 0.00
IC  C3   H41  *C4  H42  0.00 0.00  180.0 0.00 0.00
IC  C1   C3   *C2  H21  0.00 0.00  180.0 0.00 0.00
IC  C2   C4   *C3  H31  0.00 0.00  180.0 0.00 0.00
PATCH FIRST NONE LAST NONE
```

Topology for **2**

```
RESI 13DP       0.00 ! 1,3-dipentene
!
GROUP
ATOM C1   CC2  -0.42 !
ATOM H11  HE2   0.21 ! H11     H21
ATOM H12  HE2   0.21 !    \    /
ATOM C2   CC1A -0.15 !    C1=C2      H41
ATOM H21  HE1   0.15 !    /    \     /
GROUP                ! H12      C3=C4   H51
ATOM C3   CC1B -0.15 !          /    \  /
ATOM H31  HE1   0.15 !        H31     C5-H52
ATOM C4   CC1B -0.15 !                 \
ATOM H41  HE1   0.15 !                  H53
GROUP                !
```

```
ATOM C5   CT3  -0.27 !
ATOM H51  HA    0.09 !
ATOM H52  HA    0.09 !
ATOM H53  HA    0.09 !

BOND C1   C2   C2   C3   C3   C4    C4   C5
BOND C1   H11  C1   H12  C2   H21   C3   H31
BOND C4   H41  C5   H51  C5   H52   C5   H53

IC  C1   C2   C3   C4   0.00 0.00  180.0 0.00 0.00
IC  C2   C3   C4   C5   0.00 0.00  180.0 0.00 0.00
IC  C3   C4   C5   H51  0.00 0.00  180.0 0.00 0.00
IC  C3   C2   C1   H11  0.00 0.00  180.0 0.00 0.00
IC  C2   H11  *C1  H12  0.00 0.00  180.0 0.00 0.00
IC  C1   C3   *C2  H21  0.00 0.00  180.0 0.00 0.00
IC  C2   C4   *C3  H31  0.00 0.00  180.0 0.00 0.00
IC  C3   C5   *C4  H41  0.00 0.00  180.0 0.00 0.00
IC  C4   H51  *C5  H52  0.00 0.00  120.0 0.00 0.00
IC  C4   H51  *C5  H53  0.00 0.00 -120.0 0.00 0.00
PATCH FIRST NONE LAST NONE
```

Topology for **3**

```
RESI DMB1       0.00 ! 2-methyl-1,3-butadiene
!
GROUP
ATOM C1   CC2  -0.42 !
ATOM H11  HE2   0.21 !     H2M1 H2M2
ATOM H12  HE2   0.21 !        \  /
GROUP                ! H11     C2M-H2M3
ATOM C2   CC1A  0.00 !    \    /
ATOM C2M  CT3  -0.27 !    C1=C2      H41
ATOM H2M1 HA    0.09 !    /    \    /
ATOM H2M2 HA    0.09 ! H12     C3=C4
ATOM H2M3 HA    0.09 !         /    \
GROUP                !       H31     H42
ATOM C3   CC1B -0.15 !
ATOM H31  HE1   0.15 !
GROUP                !
ATOM C4   CC2  -0.42 !
ATOM H41  HE2   0.21 !
ATOM H42  HE2   0.21 !

BOND C1   C2   C2   C3   C3   C4    C2   C2M
BOND C1   H11  C1   H12  C3   H31   C4   H41  C4   H42
BOND C2M  H2M1 C2M  H2M2 C2M  H2M3
IC  C1   C2   C3   C4   0.00 0.00  180.0 0.00 0.00
IC  C2   C3   C4   H41  0.00 0.00  180.0 0.00 0.00
IC  C3   C2   C1   H11  0.00 0.00  180.0 0.00 0.00
IC  C2   H11  *C1  H12  0.00 0.00  180.0 0.00 0.00
IC  C1   C3   *C2  C2M  0.00 0.00  180.0 0.00 0.00
IC  C2   C4   *C3  H31  0.00 0.00  180.0 0.00 0.00
IC  C3   H41  *C4  H42  0.00 0.00  180.0 0.00 0.00
IC  C1   C2   C2M  H2M1 0.00 0.00  180.0 0.00 0.00
IC  C2   H2M1 *C2M H2M2 0.00 0.00  120.0 0.00 0.00
IC  C2   H2M1 *C2M H2M3 0.00 0.00 -120.0 0.00 0.00
```

```
PATCH FIRST NONE LAST NONE
```

Topology for **4**

```
RESI DMP1       0.00 ! 4-methyl-1,3-pentadiene
!
GROUP
ATOM C1   CC2  -0.42 !
ATOM H11  HE2   0.21 ! H11     H21 H4M1 H4M2
ATOM H12  HE2   0.21 !    \    /    \   /
GROUP                !    C1=C2      C4M-H4M3
ATOM C2   CC1A -0.15 !    /    \     /
ATOM H21  HE1   0.15 ! H12      C3=C4   H51
ATOM C3   CC1B -0.15 !         /     \  /
ATOM H31  HE1   0.15 !       H31      C5-H52
GROUP                !                 \
ATOM C4   CC1B  0.00 !                  H53
ATOM C4M  CT3  -0.27 !
ATOM H4M1 HA    0.09 !
ATOM H4M2 HA    0.09 !
ATOM H4M3 HA    0.09 !
GROUP
ATOM C5   CT3  -0.27 !
ATOM H51  HA    0.09 !
ATOM H52  HA    0.09 !
ATOM H53  HA    0.09 !

BOND C1   C2   C2   C3   C3   C4    C4   C5   C4   C4M
BOND C1   H11  C1   H12  C2   H21   C3   H31
BOND C4M  H4M1 C4M  H4M2 C4M  H4M3
BOND C5   H51  C5   H52  C5   H53

IC  C1   C2   C3   C4   0.00 0.00  180.0 0.00 0.00
IC  C2   C3   C4   C5   0.00 0.00  180.0 0.00 0.00
IC  C3   C4   C5   H51  0.00 0.00  180.0 0.00 0.00
IC  C3   C2   C1   H11  0.00 0.00  180.0 0.00 0.00
IC  C2   H11  *C1  H12  0.00 0.00  180.0 0.00 0.00
IC  C1   C3   *C2  H21  0.00 0.00  180.0 0.00 0.00
IC  C2   C4   *C3  H31  0.00 0.00  180.0 0.00 0.00
IC  C3   C5   *C4  C4M  0.00 0.00  180.0 0.00 0.00
IC  C3   C4   C4M  H4M1 0.00 0.00  180.0 0.00 0.00
IC  C4   H4M1 *C4M H4M2 0.00 0.00  120.0 0.00 0.00
IC  C4   H4M1 *C4M H4M3 0.00 0.00 -120.0 0.00 0.00
IC  C4   H51  *C5  H52  0.00 0.00  120.0 0.00 0.00
IC  C4   H51  *C5  H53  0.00 0.00 -120.0 0.00 0.00
PATCH FIRST NONE LAST NONE
```

Topology for **5**

```
RESI DMP2       0.00 ! 2-methyl-1,3-pentadiene
!
GROUP                !     H2M1 H2M2
ATOM C1   CC2  -0.42 !        \  /
ATOM H11  HE2   0.21 ! H11     C2M-H2M3
ATOM H12  HE2   0.21 !    \    /
GROUP                !    C1=C2      H41
ATOM C2   CC1A  0.00 !    /    \     /
ATOM C2M  CT3  -0.27 ! H12      C3=C4   H51
ATOM H2M1 HA    0.09 !          /    \  /
ATOM H2M2 HA    0.09 !         H31    C5-H52
ATOM H2M3 HA    0.09 !                 \
ATOM C3   CC1B -0.15 !                  H53
ATOM H31  HE1   0.15 !
GROUP                !
ATOM C4   CC1B -0.15 !
ATOM H41  HE1   0.15 !
ATOM C5   CT3  -0.27 !
ATOM H51  HA    0.09 !
ATOM H52  HA    0.09 !
ATOM H53  HA    0.09 !

BOND C1   C2   C2   C3   C3   C4    C4   C5   C2  C2M
BOND C1   H11  C1   H12  C3   H31   C4   H41
BOND C2M  H2M1 C2M  H2M2 C2M  H2M3
BOND C5   H51  C5   H52  C5   H53

IC  C1   C2   C3   C4   0.00 0.00  180.0 0.00 0.00
IC  C2   C3   C4   C5   0.00 0.00  180.0 0.00 0.00
IC  C3   C4   C5   H51  0.00 0.00  180.0 0.00 0.00
IC  C3   C2   C1   H11  0.00 0.00  180.0 0.00 0.00
IC  C2   H11  *C1  H12  0.00 0.00  180.0 0.00 0.00
IC  C1   C3   *C2  C2M  0.00 0.00  180.0 0.00 0.00
IC  C2   C4   *C3  H31  0.00 0.00  180.0 0.00 0.00
IC  C3   C5   *C4  H41  0.00 0.00  180.0 0.00 0.00
IC  C1   C2   C2M  H2M1 0.00 0.00  180.0 0.00 0.00
IC  C2   H2M1 *C2M H2M2 0.00 0.00  120.0 0.00 0.00
IC  C2   H2M1 *C2M H2M3 0.00 0.00 -120.0 0.00 0.00
IC  C4   H51  *C5  H52  0.00 0.00  120.0 0.00 0.00
IC  C4   H51  *C5  H53  0.00 0.00 -120.0 0.00 0.00
PATCH FIRST NONE LAST NONE
```

Topology for **6**

```
RESI CROT       0.00 ! 2-buteneamide,N,3-dimethyl (crotonamide)
                     ! numbering as in fennretinide
! note different atom types for N21 and C15: this is consistent
! with differences in NMA and acetamide
GROUP
ATOM C12  CT3  -0.27 !
ATOM H121 HA    0.09 !       H202 H203
ATOM H122 HA    0.09 !          \  /
ATOM H123 HA    0.09 !     H201-C20
GROUP                !           |
ATOM C13  CC1A  0.00 !H121-C12--C13     O29
ATOM C14  CC1A -0.15 !    /  |    \\     ||
ATOM H141 HE1   0.15 !H122 H123    C14--C15
GROUP                !            /        \
ATOM C15  C     0.51 !         H141         N21-H211 (cis)
ATOM O29  O    -0.51 !                       |
GROUP                !                 H132-C16-H131
ATOM N21  NH1  -0.47 !                       |
ATOM H211 H     0.28 !                      H133
ATOM C16  CT3  -0.08
ATOM H131 HA    0.09
ATOM H132 HA    0.09
ATOM H133 HA    0.09
GROUP                !
ATOM C20  CT3  -0.27 !
ATOM H201 HA    0.09 !
ATOM H202 HA    0.09 !
ATOM H203 HA    0.09 !

BOND C12  C13   C13  C14   C14  C15  C15   N21  N21  C16
BOND C13  C20   C15  O29   C14  H141 N21  H211
BOND C16  H131  C16  H132  C16  H133
BOND C12  H121  C12  H122  C12  H123
BOND C20  H201  C20  H202  C20  H203
IC C12   C13   C14   C15    0.0000    0.00  180.00    0.00   0.0000
IC C13   C14   C15   N21    0.0000    0.00  180.00    0.00   0.0000
IC C14   C15   N21   C16    0.0000    0.00  180.00    0.00   0.0000
IC H121  C13   *C12  H122   0.0000    0.00  120.00    0.00   0.0000
IC H121  C13   *C12  H123   0.0000    0.00 -120.00    0.00   0.0000
IC H121  C12   C13   C14    0.0000    0.00  180.00    0.00   0.0000
IC C14   C12   *C13  C20    0.0000    0.00  180.00    0.00   0.0000
IC C15   C13   *C14  H141   0.0000    0.00  180.00    0.00   0.0000
IC N21   C14   *C15  O29    0.0000    0.00  180.00    0.00   0.0000
IC C16   C15   *N21  H211   0.0000    0.00  180.00    0.00   0.0000
IC C15   N21   C16   H131   0.0000    0.00  180.00    0.00   0.0000
IC H131  N21   *C16  H132   0.0000    0.00  120.00    0.00   0.0000
IC H131  N21   *C16  H133   0.0000    0.00 -120.00    0.00   0.0000
IC C12   C13   C20   H201   0.0000    0.00  180.00    0.00   0.0000
IC H201  C13   *C20  H202   0.0000    0.00  120.00    0.00   0.0000
IC H201  C13   *C20  H203   0.0000    0.00 -120.00    0.00   0.0000

PATCH FIRST NONE LAST NONE
```

Topology for 7

```
RESI PACP       0.00 ! p-acetamide-phenol
                     ! numbering as in fennretinide
                     !
ATOM C14  CT3  -0.27 !           H142
ATOM H141 HA    0.09 !            |
ATOM H142 HA    0.09 !       H141-C14-H143
ATOM H143 HA    0.09 !            |
GROUP                !            C15=O29
ATOM C15  C     0.52 !            |
ATOM O29  O    -0.52 !            N21-H211
GROUP                !            |
ATOM N21  NH1  -0.47 !           C22
ATOM H211 H     0.33 !          //  \
ATOM C22  CA    0.14 !   H231-C23   C27-H271
GROUP                !         |     ||
ATOM C23  CA   -0.115!   H241-C24   C26-H261
ATOM H231 HP    0.115!         \\   /
GROUP                !           C25
ATOM C24  CA   -0.115!            |
ATOM H241 HP    0.115!           O28
GROUP                !             \
ATOM C26  CA   -0.115!              H281
ATOM H261 HP    0.115
GROUP
ATOM C27  CA   -0.115
ATOM H271 HP    0.115
GROUP
ATOM C25  CA    0.11
ATOM O28  OH1  -0.54
ATOM H281 H     0.43
BOND C14  C15  C15  N21  N21  C22  C22  C23  C15  O29
BOND C23  C24  C24  C25  C25  C26  C26  C27  C27  C22
BOND C25  O28  O28  H281 N21  H211
BOND C14  H141 C14  H142 C14  H143
BOND C23  H231 C24  H241 C26  H261 C27  H271
IC C14   C15   N21   C22    0.0000    0.00  180.00    0.00   0.0000
IC C15   N21   C22   C27    0.0000    0.00    0.00    0.00   0.0000
IC N21   C22   C23   C24    0.0000    0.00  180.00    0.00   0.0000
IC C22   C23   C24   C25    0.0000    0.00    0.00    0.00   0.0000
IC C23   C24   C25   O28    0.0000    0.00  180.00    0.00   0.0000
IC H141  C15   *C14  H142   0.0000    0.00  120.00    0.00   0.0000
IC H141  C15   *C14  H143   0.0000    0.00 -120.00    0.00   0.0000
IC H141  C14   C15   N21    0.0000    0.00  180.00    0.00   0.0000
IC N21   C14   *C15  O29    0.0000    0.00  180.00    0.00   0.0000
IC C22   C15   *N21  H211   0.0000    0.00  180.00    0.00   0.0000
IC C27   N21   *C22  C23    0.0000    0.00  180.00    0.00   0.0000
IC C24   C22   *C23  H231   0.0000    0.00  180.00    0.00   0.0000
IC C25   C23   *C24  H241   0.0000    0.00  180.00    0.00   0.0000
IC O28   C24   *C25  C26    0.0000    0.00  180.00    0.00   0.0000
IC C27   C25   *C26  H261   0.0000    0.00  180.00    0.00   0.0000
IC C26   C22   *C27  H271   0.0000    0.00  180.00    0.00   0.0000
IC C24   C25   O28   H281   0.0000    0.00  180.00    0.00   0.0000
```

```
PATCH FIRST NONE LAST NONE
```
Topology for **8**

```
RESI MECH       0.00 ! 1,6,6-trimethyl-2-ethene-cyclohexene
                     ! using retinol numbering
GROUP
ATOM C1   CT    0.00 !H162 H163  H171 H172
ATOM C2   CT2  -0.18 !    \  |    |  /
ATOM H21  HA    0.09 !H161-C16    C17-H173  H81
ATOM H22  HA    0.09 !        \  /           |
ATOM C3   CT2  -0.18 !  H21    C1           C8-H82
ATOM H31  HA    0.09 !     \  /  \         //
ATOM H32  HA    0.09 !  H22-C2    C6------C7
ATOM C4   CT2  -0.18 !      |     ||        \
ATOM H41  HA    0.09 !  H31-C3    C5   H181  H71
ATOM H42  HA    0.09 !     /  \  /  \  /
ATOM C5   CC1A  0.00 !  H32    C4    C18-H18
ATOM C6   CC1A  0.00 !        /  \     \
GROUP                !      H41  H42   H183
ATOM C7   CC1B -0.15 !
ATOM H71  HE1   0.15 !
ATOM C8   CC2  -0.42 !
ATOM H81  HE2   0.21 !
ATOM H82  HE2   0.21 !
GROUP               !
ATOM C16  CT3  -0.27 !
ATOM H161 HA    0.09 !
ATOM H162 HA    0.09 !
ATOM H163 HA    0.09 !
GROUP                !
ATOM C17  CT3  -0.27 !
ATOM H171 HA    0.09 !
ATOM H172 HA    0.09 !
ATOM H173 HA    0.09 !
GROUP                !
ATOM C18  CT3  -0.27 !
ATOM H181 HA    0.09 !
ATOM H182 HA    0.09 !
ATOM H183 HA    0.09 !

BOND C1   C2   C2   C3   C3   C4   C4   C5   C5   C6   C6   C1
BOND C6   C7   C7   C8   C1   C16  C1   C17  C5   C18
BOND C2   H21  C2   H22  C3   H31  C3   H32  C4   H41  C4   H42
BOND C16  H161 C16  H162 C16  H163 C17  H171 C17  H172 C17  H173
BOND C7   H71  C8   H81  C8   H82  C18  H181 C18  H182 C18  H183

!initial ring geometry is planar
IC  C1   C2   C3   C4   0.00 0.00   63.0 0.00 0.00
IC  C2   C3   C4   C5   0.00 0.00  -46.0 0.00 0.00
IC  C3   C4   C5   C6   0.00 0.00   15.0 0.00 0.00
IC  C1   C3   *C2  H21  0.00 0.00  120.0 0.00 0.00
IC  C1   C3   *C2  H22  0.00 0.00 -120.0 0.00 0.00
IC  C2   C4   *C3  H31  0.00 0.00  120.0 0.00 0.00
IC  C2   C4   *C3  H32  0.00 0.00 -120.0 0.00 0.00
IC  C3   C5   *C4  H41  0.00 0.00  120.0 0.00 0.00
IC  C3   C5   *C4  H42  0.00 0.00 -120.0 0.00 0.00
```

```
IC  C1   C5   *C6  C7   0.00 0.00  180.0 0.00 0.00
IC  C5   C6   C7   C8   0.00 0.00   60.0 0.00 0.00
IC  C6   C8   *C7  H71  0.00 0.00  180.0 0.00 0.00
IC  C6   C7   C8   H81  0.00 0.00  180.0 0.00 0.00
IC  C7   H81  *C8  H82  0.00 0.00  180.0 0.00 0.00
IC  C6   C2   *C1  C16  0.00 0.00  120.0 0.00 0.00
IC  C6   C2   *C1  C17  0.00 0.00 -120.0 0.00 0.00
IC  C2   C1   C16  H161 0.00 0.00  180.0 0.00 0.00
IC  C1   H161 *C16 H162 0.00 0.00  120.0 0.00 0.00
IC  C1   H161 *C16 H163 0.00 0.00 -120.0 0.00 0.00
IC  C2   C1   C17  H171 0.00 0.00  180.0 0.00 0.00
IC  C1   H171 *C17 H172 0.00 0.00  120.0 0.00 0.00
IC  C1   H171 *C17 H173 0.00 0.00 -120.0 0.00 0.00
IC  C4   C6   *C5  C18  0.00 0.00  180.0 0.00 0.00
IC  C4   C5   C18  H181 0.00 0.00  180.0 0.00 0.00
IC  C5   H181 *C18 H182 0.00 0.00  120.0 0.00 0.00
IC  C5   H181 *C18 H183 0.00 0.00 -120.0 0.00 0.00
PATCH FIRST NONE LAST NONE
```

Topology for **9**

```
RESI TMCH       0.00 ! 1,2,6,6-tetramethylcyclohexene
                     ! using retinol numbering
GROUP
ATOM C1   CT    0.00 !H162 H163  H171 H172
ATOM C2   CT2  -0.18 !    \  |    |  /
ATOM H21  HA    0.09 !H161-C16    C17-H173
ATOM H22  HA    0.09 !        \  /
ATOM C3   CT2  -0.18 !  H21    C1        H71 H72
ATOM H31  HA    0.09 !     \  /  \        | /
ATOM H32  HA    0.09 !  H22-C2    C6------C7-H73
ATOM C4   CT2  -0.18 !      |     ||
ATOM H41  HA    0.09 !  H31-C3    C5   H181
ATOM H42  HA    0.09 !     /  \  /  \  /
ATOM C5   CC1A  0.00 !  H32    C4    C18-H18
ATOM C6   CC1A  0.00 !        /  \     \
GROUP                !      H41  H42   H183
ATOM C7   CT3  -0.27 !
ATOM H71  HA    0.09 !
ATOM H72  HA    0.09 !
ATOM H73  HA    0.09 !
GROUP                !
ATOM C16  CT3  -0.27 !
ATOM H161 HA    0.09 !
ATOM H162 HA    0.09 !
ATOM H163 HA    0.09 !
GROUP                !
ATOM C17  CT3  -0.27 !
ATOM H171 HA    0.09 !
ATOM H172 HA    0.09 !
ATOM H173 HA    0.09 !
GROUP                !
ATOM C18  CT3  -0.27 !
ATOM H181 HA    0.09 !
ATOM H182 HA    0.09 !
ATOM H183 HA    0.09 !
```

```
BOND C1   C2   C2   C3   C3   C4   C4   C5   C5   C6   C6   C1
BOND C6   C7   C1   C16  C1   C17  C5   C18
BOND C2   H21  C2   H22  C3   H31  C3   H32  C4   H41  C4   H42
BOND C16  H161 C16  H162 C16  H163 C17  H171 C17  H172 C17  H173
BOND C7   H71  C7   H72  C7   H73  C18  H181 C18  H182 C18  H183

!initial geometry is planar
IC  C1   C2   C3   C4   0.00 0.00  -63.0 0.00 0.00
IC  C2   C3   C4   C5   0.00 0.00   46.0 0.00 0.00
IC  C3   C4   C5   C6   0.00 0.00  -13.0 0.00 0.00
IC  C1   C3   *C2  H21  0.00 0.00  120.0 0.00 0.00
IC  C1   C3   *C2  H22  0.00 0.00 -120.0 0.00 0.00
IC  C2   C4   *C3  H31  0.00 0.00  120.0 0.00 0.00
IC  C2   C4   *C3  H32  0.00 0.00 -120.0 0.00 0.00
IC  C3   C5   *C4  H41  0.00 0.00  120.0 0.00 0.00
IC  C3   C5   *C4  H42  0.00 0.00 -120.0 0.00 0.00
IC  C1   C5   *C6  C7   0.00 0.00  180.0 0.00 0.00
IC  C5   C6   C7   H71  0.00 0.00  180.0 0.00 0.00
IC  C6   H71  *C7  H72  0.00 0.00  120.0 0.00 0.00
IC  C6   H71  *C7  H73  0.00 0.00 -120.0 0.00 0.00
IC  C6   C2   *C1  C16  0.00 0.00  120.0 0.00 0.00
IC  C6   C2   *C1  C17  0.00 0.00 -120.0 0.00 0.00
IC  C2   C1   C16  H161 0.00 0.00  180.0 0.00 0.00
IC  C1   H161 *C16 H162 0.00 0.00  120.0 0.00 0.00
IC  C1   H161 *C16 H163 0.00 0.00 -120.0 0.00 0.00
IC  C2   C1   C17  H171 0.00 0.00  180.0 0.00 0.00
IC  C1   H171 *C17 H172 0.00 0.00  120.0 0.00 0.00
IC  C1   H171 *C17 H173 0.00 0.00 -120.0 0.00 0.00
IC  C4   C6   *C5  C18  0.00 0.00  180.0 0.00 0.00
IC  C4   C5   C18  H181 0.00 0.00  180.0 0.00 0.00
IC  C5   H181 *C18 H182 0.00 0.00  120.0 0.00 0.00
IC  C5   H181 *C18 H183 0.00 0.00 -120.0 0.00 0.00
PATCH FIRST NONE LAST NONE
```

Topology for **10**

```
RESI PRAC      -1.00 ! 2-propenoic acid
GROUP
ATOM C1  CC2   -0.42 !
ATOM H11 HE2    0.21 !
ATOM H12 HE2    0.21 !  H11    OD1
GROUP               !   \     ||
ATOM C2  CC1A  -0.25 !   C1    CG
ATOM H21 HE1    0.15 !   / \\ / \
ATOM CG  CC     0.62 ! H12  C2  OD2 (-)
ATOM OD1 OC    -0.76 !      |
ATOM OD2 OC    -0.76 !     H21

BOND C1  H11  C1  H12  C1  C2   C2   H21
BOND C2  CG   CG  OD1  CG  OD2
IC C1    C2    CG    OD1    0.0000    0.00  180.00    0.00   0.0000
IC OD1   C2    *CG   OD2    0.0000    0.00  180.00    0.00   0.0000
IC H11   C1    C2    CG     0.0000    0.00  180.00    0.00   0.0000
IC H11   C2    *C1   H12    0.0000    0.00  180.00    0.00   0.0000
IC CG    C1    *C2   H21    0.0000    0.00  180.00    0.00   0.0000
```

```
PATCH FIRST NONE LAST NONE
```

Topology for **11**

```
RESI PROL       0.00 ! 3-propenol (1-hydroxy-2-propene)
!
GROUP
ATOM C1  CC2   -0.42 !
ATOM H11 HE2    0.21 !  H11    H31
ATOM H12 HE2    0.21 !   \      |
ATOM C2  CC1A  -0.15 !   C1    C3-OR
ATOM H21 HE1    0.15 !   / \\ / |   \
GROUP               ! H12  C2 H32  HR
ATOM C3  CT2    0.05 !       |
ATOM H31 HA     0.09 !      H21
ATOM H32 HA     0.09
ATOM OR  OH1   -0.66
ATOM HR  H      0.43
BOND C1   C2   C2   C3   C3   OR   OR   HR
BOND H11  C1   H12  C1   H21  C2   H31  C3
BOND H32  C3

IC C1    C2    C3    OR     0.0000    0.00    0.00    0.00   0.0000
IC C2    C3    OR    HR     0.0000    0.00   60.00    0.00   0.0000
IC H11   C1    C2    C3     0.0000    0.00  180.00    0.00   0.0000
IC H11   C2    *C1   H12    0.0000    0.00  180.00    0.00   0.0000
IC C3    C1    *C2   H21    0.0000    0.00  180.00    0.00   0.0000
IC OR    C2    *C3   H31    0.0000    0.00  120.00    0.00   0.0000
IC OR    C2    *C3   H32    0.0000    0.00 -120.00    0.00   0.0000

PATCH FIRST NONE LAST NONE
```

Topology for **12**

```
RESI PRAL       0.00 ! 1-propenal
GROUP
ATOM C1  CC2   -0.42 !
ATOM H11 HE2    0.21 !  H11    OD1
ATOM H12 HE2    0.21 !   \     ||
ATOM C2  CC1A  -0.01 !   C1    CG
ATOM H21 HE1    0.15 !   / \\ / \
ATOM CG  CD     0.19 ! H12  C2   H
ATOM OD1 O     -0.43 !      |
ATOM H   HR1    0.10 !     H21

BOND C1  H11  C1  H12  C1  C2  C2  H21
BOND C2  CG   CG  OD1  CG  H
IMPR CG  C2   OD1 H

IC C1    C2    CG    OD1    0.0000    0.00  180.00    0.00   0.0000
IC H11   C1    C2    CG     0.0000    0.00  180.00    0.00   0.0000
IC H11   C2    *C1   H12    0.0000    0.00  180.00    0.00   0.0000
IC CG    C1    *C2   H21    0.0000    0.00  180.00    0.00   0.0000
IC OD1   C2    *CG   H      0.0000    0.00  180.00    0.00   0.0000
```

Topology for **13**

```
RESI BTE2       0.00 ! 2-Butene, adm jr.
GROUP
ATOM C1  CT3  -0.27 !  H12 H13
ATOM H11 HA    0.09 !    \ |
ATOM H12 HA    0.09 ! H13-C1      H31
ATOM H13 HA    0.09 !       \     /
GROUP               !        C2=C3
ATOM C2  CC1A -0.15 !       /     \
ATOM H21 HE1   0.15 !     H21      C4-H41
GROUP               !              | \
ATOM C3  CC1A -0.15 !            H43 H42
ATOM H31 HE1   0.15 !
GROUP
ATOM C4  CT3  -0.27
ATOM H41 HA    0.09
ATOM H42 HA    0.09
ATOM H43 HA    0.09

BOND C1 H11  C1 H12  C1 H13 C1 C2   C2 H21
DOUBLE C2 C3
BOND C3 H31  C3 C4  C4 H41  C4 H42  C4 H43

IC C1    C2    C3    C4     0.0000    0.00  180.00    0.00   0.0000
IC H11   C2    *C1   H12    0.0000    0.00  120.00    0.00   0.0000
IC H11   C2    *C1   H13    0.0000    0.00 -120.00    0.00   0.0000
IC H11   C1    C2    C3     0.0000    0.00  180.00    0.00   0.0000
IC C3    C1    *C2   H21    0.0000    0.00  180.00    0.00   0.0000
IC C4    C2    *C3   H31    0.0000    0.00  180.00    0.00   0.0000
IC C2    C3    C4    H41    0.0000    0.00  180.00    0.00   0.0000
IC H41   C3    *C4   H42    0.0000    0.00  120.00    0.00   0.0000
IC H41   C3    *C4   H43    0.0000    0.00 -120.00    0.00   0.0000

PATCH FIRS NONE LAST NONE
```

Topology for 14

```
RESI HEP3       0.00 ! 1,3,5-heptatriene
!
GROUP
ATOM C1   CC2  -0.42 !
ATOM H11  HE2   0.21 ! H11     H21
ATOM H12  HE2   0.21 !    \    /
ATOM C2   CC1A -0.15 !    C1=C2      H41
ATOM H21  HE1   0.15 !    /    \     /
GROUP                ! H12      C3=C4      H61
ATOM C3   CC1B -0.15 !          /    \    /
ATOM H31  HE1   0.15 !        H31     C5=C6
ATOM C4   CC1B -0.15 !                /    \
ATOM H41  HE1   0.15 !             H51     H62
GROUP                !
ATOM C5   CC1A -0.15
ATOM H51  HE1   0.15
```

```
GROUP
ATOM C6   CC2  -0.42
ATOM H61  HE2   0.21
ATOM H62  HE2   0.21

BOND C1   C2   C2   C3   C3   C4    C4   C5   C5   C6
BOND C1   H11  C1   H12  C2   H21   C3   H31
BOND C4   H41  C5   H51  C6   H61   C6   H62

IC C1    C2    C3    C4     0.0000    0.00  180.00    0.00   0.0000
IC C2    C3    C4    C5     0.0000    0.00  180.00    0.00   0.0000
IC C3    C4    C5    C6     0.0000    0.00  180.00    0.00   0.0000
IC H11   C2    *C1   H12    0.0000    0.00  180.00    0.00   0.0000
IC H11   C1    C2    C3     0.0000    0.00  180.00    0.00   0.0000
IC C3    C1    *C2   H21    0.0000    0.00  180.00    0.00   0.0000
IC C4    C2    *C3   H31    0.0000    0.00  180.00    0.00   0.0000
IC C5    C3    *C4   H41    0.0000    0.00  180.00    0.00   0.0000
IC C6    C4    *C5   H51    0.0000    0.00  180.00    0.00   0.0000
IC C4    C5    C6    H61    0.0000    0.00  180.00    0.00   0.0000
IC H61   C5    *C6   H62    0.0000    0.00  180.00    0.00   0.0000

PATCH FIRST NONE LAST NONE
```

Topology for **15**

```
RESI SCH1       0.00 ! Schiff's base model compound 1, deprotonated
!                    ! new atom type NS1
GROUP
ATOM C1  CT3   -0.05 !   H11         H31
ATOM H11 HA     0.09 !      \        /
ATOM H12 HA     0.09 !  H12--C1-N2=C3    H41
ATOM H13 HA     0.09 !      /        \  /
ATOM N2  NS1   -0.60 !   H13          C4--H42
ATOM C3  CC1B   0.23 !                  \
ATOM H31 HE1    0.15 !                   H43
GROUP               !
ATOM C4  CT3   -0.27 !
ATOM H41 HA     0.09 !
ATOM H42 HA     0.09
ATOM H43 HA     0.09

BOND C1  H11  C1  H12  C1  H13
BOND C1  N2   N2  C3   C3  H31
BOND C3  C4   C4  H41  C4  H42  C4  H43

IC C1    N2    C3    C4     0.0000    0.00  180.00    0.00   0.0000
IC N2    C3    C4    H41    0.0000    0.00  180.00    0.00   0.0000
IC C3    N2    C1    H11    0.0000    0.00    0.00    0.00   0.0000
IC H11   N2    *C1   H12    0.0000    0.00  120.00    0.00   0.0000
IC H11   N2    *C1   H13    0.0000    0.00 -120.00    0.00   0.0000
IC C4    N2    *C3   H31    0.0000    0.00  180.00    0.00   0.0000
IC H41   C3    *C4   H42    0.0000    0.00  120.00    0.00   0.0000
IC H41   C3    *C4   H43    0.0000    0.00 -120.00    0.00   0.0000
```

Topology for **16**

```
RESI SCH2       1.00 ! Schiff's base model compound 2, protonated
!                    ! new atom type, NS2
GROUP
ATOM C1  CT3    0.18 !   H11   H21    H31
ATOM H11 HA     0.09 !      \   |    /
ATOM H12 HA     0.09 !  H12--C1-N2=C3    H41
ATOM H13 HA     0.09 !      /        \  /
ATOM N2  NS2   -0.40 !   H13          C4--H42
ATOM H21 HC     0.38 !                  \
ATOM C3  CC1B   0.37 !                   H43
ATOM H31 HR1    0.20 ! hydrogen with intermediate VDW radius
GROUP               !
ATOM C4  CT3   -0.27 !
ATOM H41 HA     0.09 !
ATOM H42 HA     0.09
ATOM H43 HA     0.09

BOND C1  H11  C1  H12  C1  H13
BOND C1  N2   N2  H21  N2  C3   C3  H31
BOND C3  C4   C4  H41  C4  H42  C4  H43

IC C1    N2    C3    C4     0.0000    0.00  180.00    0.00   0.0000
IC N2    C3    C4    H41    0.0000    0.00  180.00    0.00   0.0000
IC C1    C3    *N2   H21    0.0000    0.00  180.00    0.00   0.0000
IC C3    N2    C1    H11    0.0000    0.00    0.00    0.00   0.0000
IC H11   N2    *C1   H12    0.0000    0.00  120.00    0.00   0.0000
IC H11   N2    *C1   H13    0.0000    0.00 -120.00    0.00   0.0000
IC C4    N2    *C3   H31    0.0000    0.00  180.00    0.00   0.0000
IC H41   C3    *C4   H42    0.0000    0.00  120.00    0.00   0.0000
IC H41   C3    *C4   H43    0.0000    0.00 -120.00    0.00   0.0000
```

Topology for **17**

```
RESI SCH3       1.00 ! Schiff's base model compound 3, protonated
!                   ! new atom type, NS2
GROUP
ATOM C1  CT3    0.18 !   H11   H21    H31
ATOM H11 HA     0.09 !      \   |    /
ATOM H12 HA     0.09 !  H12--C1-N2=C3      H51
ATOM H13 HA     0.09 !      /        \     /
ATOM N2  NS2   -0.40 !   H13          C4=C5
ATOM H21 HC     0.38 !               /     \
ATOM C3  CC1B   0.37 !             H41     H52
ATOM H31 HR1    0.20 ! hydrogen with intermediate VDW radius
GROUP               !
ATOM C4  CC1A  -0.15
ATOM H41 HE1    0.15
ATOM C5  CC2   -0.42
ATOM H51 HE2    0.21
ATOM H52 HE2    0.21
```

```
BOND C1  H11  C1  H12  C1  H13
BOND C1  N2   N2  H21  N2  C3   C3  H31
BOND C3  C4   C4  H41
BOND C4  C5   C5  H51  C5  H52

IC C1    N2    C3    C4     0.0000    0.00  180.00    0.00   0.0000
IC N2    C3    C4    C5     0.0000    0.00  180.00    0.00   0.0000
IC H11   C1    N2    C3     0.0000    0.00  180.00    0.00   0.0000
IC H11   N2    *C1   H12    0.0000    0.00  120.00    0.00   0.0000
IC H11   N2    *C1   H13    0.0000    0.00 -120.00    0.00   0.0000
IC C3    C1    *N2   H21    0.0000    0.00  180.00    0.00   0.0000
IC C4    N2    *C3   H31    0.0000    0.00  180.00    0.00   0.0000
IC C5    C3    *C4   H41    0.0000    0.00  180.00    0.00   0.0000
IC C3    C4    C5    H51    0.0000    0.00  180.00    0.00   0.0000
IC H51   C4    *C5   H52    0.0000    0.00  180.00    0.00   0.0000
```

Topology for **18**

```
PRES SCK0       0.00 ! patch to link lysine with retinal to form a
                     ! deprotonated schiff's base
                     ! follow with AUTOgenerate ANGLes DIHEdrals
command
! residue 1: lysine
! residue 2: retinal
!
DELETE ATOM 1HZ1
DELETE ATOM 1HZ2
DELETE ATOM 1HZ3
DELETE ATOM 2O15
GROUP                ! HE1 HE2
ATOM 1CE  CT2   0.04 !   \ /
ATOM 1HE1 HA    0.09 ! --1CE1       2H15
ATOM 1HE2 HA    0.09 !       \       /
ATOM 1NZ  NS1  -0.60 !       1NZ=2C15
ATOM 2C15 CC1B  0.23 !       /       \
ATOM 2H15 HE1   0.15 !    2HZ1        2C14==
GROUP                !              /
ATOM 2C14 CC1A -0.15 !            2HE14
ATOM 2H141 HE1  0.15

BOND 1NZ 2C15

IC 1CD    1CE    1NZ    2C15    0.0000    0.00  180.00    0.00   0.0000
IC 1CE    1NZ    2C15   2C14    0.0000    0.00  180.00    0.00   0.0000
IC 1NZ    2C15   2C14   2C13    0.0000    0.00  180.00    0.00   0.0000
IC 2C14   1NZ    *2C15  2H15    0.0000    0.00  180.00    0.00   0.0000
IC 2C5    2C6    2C1    2C2     0.0000    0.00    0.00    0.00   0.0000
! required to build retinal
```

Topology for **19**

```
PRES SCK1       1.00 ! patch to link lysine with retinal to form a
                     ! protonated schiff's base
                     ! follow with AUTOgenerate ANGLes DIHEdrals
command
! residue 1: lysine
! residue 2: retinal
!
DELETE ATOM 1HZ2
DELETE ATOM 1HZ3
DELETE ATOM 2O15
GROUP                !  HE1 HE2
ATOM 1CE  CT2   0.27 !    \ /
ATOM 1HE1 HA    0.09 !  --1CE1       2H15
ATOM 1HE2 HA    0.09 !        \       /
ATOM 1NZ  NS2  -0.40 !        1NZ=2C15
ATOM HZ1  HC    0.38 !        /       \
ATOM 2C15 CC1B  0.37 !     2HZ1        2C14==
ATOM 2H15 HR1   0.20 !                 /
GROUP                !             2HE14
ATOM 2C14 CC1A -0.15
ATOM 2H141 HE1  0.15

BOND 1NZ 2C15

IC 1CD    1CE    1NZ    2C15    0.0000    0.00  180.00    0.00   0.0000
IC 1CE    1NZ    2C15   2C14    0.0000    0.00  180.00    0.00   0.0000
IC 1NZ    2C15   2C14   2C13    0.0000    0.00  180.00    0.00   0.0000
IC 2C15   1CE    *1NZ   1HZ1    0.0000    0.00  180.00    0.00   0.0000
IC 2C14   1NZ    *2C15  2H15    0.0000    0.00  180.00    0.00   0.0000
IC 2C5    2C6    2C1    2C2     0.0000    0.00    0.00    0.00   0.0000
! required to build retinal

end
```

Topology for **fenretinide**

```
RESI FRET       0.00 ! fenretinide, nomenclature consistent with
retinol
!
!
!
GROUP
ATOM C1   CT    0.00
ATOM C2   CT2  -0.18
ATOM H21  HA    0.09
ATOM H22  HA    0.09
ATOM C3   CT2  -0.18
ATOM H31  HA    0.09
ATOM H32  HA    0.09
ATOM C4   CT2  -0.18
ATOM H41  HA    0.09
```

```
ATOM H42  HA    0.09
ATOM C5   CC1A  0.00
ATOM C6   CC1A  0.00
GROUP
ATOM C7   CC1B -0.15
ATOM H71  HE1   0.15
ATOM C8   CC1B -0.15
ATOM H81  HE1   0.15
ATOM C9   CC1A  0.00
ATOM C10  CC1A -0.15
ATOM H101 HE1   0.15
GROUP
ATOM C11  CC1B -0.15 !H162 H163  H171 H172
ATOM H111 HE1   0.15 !    \  |    |  /
ATOM C12  CC1B -0.15 !H161-C16    C17-H173            H191 H192
ATOM H121 HE1   0.15 !        \  /                      \   /
ATOM C13  CC1A  0.00 !  H21    C1        H71    H81      C19
ATOM C14  CC1A -0.15 !     \  /  \        |      |       /  \
ATOM H141 HE1   0.15 !  H22-C2    C6------C7=====C8----C9    H193
GROUP                !      |     ||                     ||
ATOM C16  CT3  -0.27 !  H31-C3    C5   H181             C10-H101
ATOM H161 HA    0.09 !     /  \  /  \  /                |
ATOM H162 HA    0.09 !  H32    C4    C18-H182          C11-H111
ATOM H163 HA    0.09 !        /  \     \                ||
GROUP                !      H41  H42   H183            C12-H121 H201
ATOM C17  CT3  -0.27 !                                 |       /
ATOM H171 HA    0.09 !                                 C13---C20-H202
ATOM H172 HA    0.09 !                                 ||      \
ATOM H173 HA    0.09 !                                 C14-H141 H203
GROUP                !                                 |
ATOM C18  CT3  -0.27 !                                 C15=O29
ATOM H181 HA    0.09 !                                 |
ATOM H182 HA    0.09 !                                 N21-H211
ATOM H183 HA    0.09 !                                 |
GROUP                !                                C22
ATOM C19  CT3  -0.27 !                               //  \
ATOM H191 HA    0.09 !                          H231-C23   C27-H271
ATOM H192 HA    0.09 !                                |     ||
ATOM H193 HA    0.09 !                          H241-C24   C26-H261
GROUP                !                                \\  /
ATOM C20  CT3  -0.27 !                                 C25
ATOM H201 HA    0.09 !                                  |
ATOM H202 HA    0.09 !                                 O28
ATOM H203 HA    0.09 !                                   \
GROUP                !                                    H281
ATOM C15  C     0.51 ! charges from CROT
ATOM O29  O    -0.51 !
GROUP
ATOM N21  NH1  -0.47 !
ATOM H211 H     0.33 ! charges from PACP
ATOM C22  CA    0.14 !
GROUP
ATOM C23  CA   -0.115
ATOM H231 HP    0.115
GROUP
ATOM C24  CA   -0.115
ATOM H241 HP    0.115
```

```
GROUP
ATOM C26  CA   -0.115
ATOM H261 HP    0.115
GROUP
ATOM C27  CA   -0.115
ATOM H271 HP    0.115
GROUP
ATOM C25  CA    0.11
ATOM O28  OH1  -0.54
ATOM H281 H     0.43
BOND C1   C2   C2   C3   C3   C4   C4   C5   C5   C6   C6   C1
BOND C6   C7   C7   C8   C8   C9   C9   C10  C10  C11  C11  C12
BOND C12  C13  C13  C14  C14  C15  C15  N21  N21  C22  C22  C23
BOND C23  C24  C24  C25  C25  C26  C26  C27  C27  C22
BOND C1   C16  C1   C17  C5   C18  C9   C19  C13  C20  C15  O29
BOND C25  O28  O28  H281 N21  H211
BOND C2   H21  C2   H22  C3   H31  C3   H32  C4   H41  C4   H42
BOND C7   H71  C8   H81  C10  H101 C11  H111 C12  H121 C14  H141
BOND C16  H161 C16  H162 C16  H163 C17  H171 C17  H172 C17  H173
BOND C18  H181 C18  H182 C18  H183 C19  H191 C19  H192 C19  H193
BOND C20  H201 C20  H202 C20  H203
BOND C23  H231 C24  H241 C26  H261 C27  H271
IC C1    C2    C3    C4     0.0000    0.00    0.00    0.00   0.0000
IC C2    C3    C4    C5     0.0000    0.00    0.00    0.00   0.0000
IC C6    C1    C2    C3     0.0000    0.00    0.00    0.00   0.0000
IC C1    C6    C7    C8     0.0000    0.00  180.00    0.00   0.0000
IC C6    C7    C8    C9     0.0000    0.00  180.00    0.00   0.0000
IC C7    C8    C9    C10    0.0000    0.00  180.00    0.00   0.0000
IC C2    C6    *C1   C16    0.0000    0.00  120.00    0.00   0.0000
IC C2    C6    *C1   C17    0.0000    0.00 -120.00    0.00   0.0000
IC C3    C1    *C2   H21    0.0000    0.00  120.00    0.00   0.0000
IC C3    C1    *C2   H22    0.0000    0.00 -120.00    0.00   0.0000
IC C4    C2    *C3   H31    0.0000    0.00  120.00    0.00   0.0000
IC C4    C2    *C3   H32    0.0000    0.00 -120.00    0.00   0.0000
IC C5    C3    *C4   H41    0.0000    0.00  120.00    0.00   0.0000
IC C5    C3    *C4   H42    0.0000    0.00 -120.00    0.00   0.0000
IC C5    C1    *C6   C7     0.0000    0.00  180.00    0.00   0.0000
IC C8    C6    *C7   H71    0.0000    0.00  180.00    0.00   0.0000
IC C9    C7    *C8   H81    0.0000    0.00  180.00    0.00   0.0000
IC C10   C8    *C9   C19    0.0000    0.00  180.00    0.00   0.0000
IC C8    C9    C10   C11    0.0000    0.00  180.00    0.00   0.0000
IC C11   C9    *C10  H101   0.0000    0.00  180.00    0.00   0.0000
IC C9    C10   C11   C12    0.0000    0.00  180.00    0.00   0.0000
IC C12   C10   *C11  H111   0.0000    0.00  180.00    0.00   0.0000
IC C10   C11   C12   C13    0.0000    0.00  180.00    0.00   0.0000
IC C13   C11   *C12  H121   0.0000    0.00  180.00    0.00   0.0000
IC C11   C12   C13   C14    0.0000    0.00  180.00    0.00   0.0000
IC C14   C12   *C13  C20    0.0000    0.00  180.00    0.00   0.0000
IC C12   C13   C14   C15    0.0000    0.00  180.00    0.00   0.0000
IC C15   C13   *C14  H141   0.0000    0.00  180.00    0.00   0.0000
IC C6    C1    C16   H161   0.0000    0.00  180.00    0.00   0.0000
IC H161  C1    *C16  H162   0.0000    0.00  120.00    0.00   0.0000
IC H161  C1    *C16  H163   0.0000    0.00 -120.00    0.00   0.0000
IC C6    C1    C17   H171   0.0000    0.00  180.00    0.00   0.0000
IC H171  C1    *C17  H172   0.0000    0.00  120.00    0.00   0.0000
IC H171  C1    *C17  H173   0.0000    0.00 -120.00    0.00   0.0000
IC C6    C4    *C5   C18    0.0000    0.00  180.00    0.00   0.0000
```

```
IC C4    C5    C18   H181   0.0000    0.00  180.00    0.00   0.0000
IC H181  C5    *C18  H182   0.0000    0.00  120.00    0.00   0.0000
IC H181  C5    *C18  H183   0.0000    0.00 -120.00    0.00   0.0000
IC C8    C9    C19   H191   0.0000    0.00  180.00    0.00   0.0000
IC H191  C9    *C19  H192   0.0000    0.00  120.00    0.00   0.0000
IC H191  C9    *C19  H193   0.0000    0.00 -120.00    0.00   0.0000
IC C12   C13   C20   H201   0.0000    0.00  180.00    0.00   0.0000
IC H201  C13   *C20  H202   0.0000    0.00  120.00    0.00   0.0000
IC H201  C13   *C20  H203   0.0000    0.00 -120.00    0.00   0.0000
IC C13   C14   C15   N21    0.0000    0.00  180.00    0.00   0.0000
IC N21   C14   *C15  O29    0.0000    0.00  180.00    0.00   0.0000
IC C14   C15   N21   C22    0.0000    0.00  180.00    0.00   0.0000
IC C22   C15   *N21  H211   0.0000    0.00  180.00    0.00   0.0000
IC C15   N21   C22   C27    0.0000    0.00  180.00    0.00   0.0000
IC C27   N21   *C22  C23    0.0000    0.00  180.00    0.00   0.0000
IC N21   C22   C23   C24    0.0000    0.00  180.00    0.00   0.0000
IC C24   C22   *C23  H231   0.0000    0.00  180.00    0.00   0.0000
IC C22   C23   C24   C25    0.0000    0.00    0.00    0.00   0.0000
IC C25   C23   *C24  H241   0.0000    0.00  180.00    0.00   0.0000
IC C23   C24   C25   O28    0.0000    0.00  180.00    0.00   0.0000
IC O28   C24   *C25  C26    0.0000    0.00  180.00    0.00   0.0000
IC C27   C25   *C26  H261   0.0000    0.00  180.00    0.00   0.0000
IC C26   C22   *C27  H271   0.0000    0.00  180.00    0.00   0.0000
IC C24   C25   O28   H281   0.0000    0.00  180.00    0.00   0.0000

PATCH FIRST NONE LAST NONE
```

Topology for **retinol**

```
RESI RTOL       0.00 ! retinol, nomenclature from PDB
!
GROUP
ATOM C1   CT    0.00
ATOM C2   CT2  -0.18
ATOM H21  HA    0.09
ATOM H22  HA    0.09
GROUP
ATOM C3   CT2  -0.18
ATOM H31  HA    0.09
ATOM H32  HA    0.09
GROUP
ATOM C4   CT2  -0.18
ATOM H41  HA    0.09
ATOM H42  HA    0.09
ATOM C5   CC1A  0.00
ATOM C6   CC1A  0.00
GROUP
ATOM C7   CC1B -0.15
ATOM H71  HE1   0.15
ATOM C8   CC1B -0.15
ATOM H81  HE1   0.15
GROUP
ATOM C9   CC1A  0.00
ATOM C10  CC1A -0.15
ATOM H101 HE1   0.15
GROUP
```

```
ATOM C11  CC1B -0.15
ATOM H111 HE1   0.15
GROUP
ATOM C12  CC1B -0.15
ATOM H121 HE1   0.15
GROUP
ATOM C13  CC1A  0.00
ATOM C14  CC1A -0.15 !
ATOM H141 HE1   0.15 !H162 H163  H171 H172
GROUP                !    \  |    |  /
ATOM C15  CT2   0.05 !H161-C16    C17-H173            H191 H192
ATOM H151 HA    0.09 !        \  /                      \   /
ATOM H152 HA    0.09 !  H21    C1        H71    H81      C19
ATOM OR   OH1  -0.66 !     \  /  \        |      |       /  \
ATOM HR   H     0.43 !  H22-C2    C6------C7=====C8----C9    H193
GROUP                !      |     ||                   ||
ATOM C16  CT3  -0.27 !  H31-C3    C5   H181            C10-H101
ATOM H161 HA    0.09 !     /  \  /  \  /               |
ATOM H162 HA    0.09 !  H32    C4    C18-H182          C11-H111
ATOM H163 HA    0.09 !        /  \     \               ||
GROUP                !      H41  H42   H183            C12-H121 H201
ATOM C17  CT3  -0.27 !                                 |       /
ATOM H171 HA    0.09 !                                 C13---C20-H202
ATOM H172 HA    0.09 !                                 ||      \
ATOM H173 HA    0.09 !                                 C14-H141 H203
GROUP                !                                 |
ATOM C18  CT3  -0.27 !                           H151-C15-H152
ATOM H181 HA    0.09 !                                 |
ATOM H182 HA    0.09 !                                 OR
ATOM H183 HA    0.09 !                                   \
GROUP                !                                    HR
ATOM C19  CT3  -0.27 !
ATOM H191 HA    0.09 !
ATOM H192 HA    0.09 !
ATOM H193 HA    0.09 !
GROUP                !
ATOM C20  CT3  -0.27 !
ATOM H201 HA    0.09 !
ATOM H202 HA    0.09 !
ATOM H203 HA    0.09 !

BOND C1   C2   C2   C3   C3   C4   C4   C5   C5   C6   C6   C1
BOND C6   C7   C7   C8   C8   C9   C9   C10  C10  C11  C11  C12
BOND C12  C13  C13  C14  C14  C15  C15  OR
BOND C1   C16  C1   C17  C5   C18  C9   C19  C13  C20
BOND C2   H21  C2   H22  C3   H31  C3   H32  C4   H41  C4   H42
BOND C7   H71  C8   H81  C10  H101 C11  H111 C12  H121 C14  H141
BOND C15  H151 C15  H152 OR   HR
BOND C16  H161 C16  H162 C16  H163 C17  H171 C17  H172 C17  H173
BOND C18  H181 C18  H182 C18  H183 C19  H191 C19  H192 C19  H193
BOND C20  H201 C20  H202 C20  H203
IC C1    C2    C3    C4     0.0000    0.00    0.00    0.00   0.0000
IC C6    C1    C2    C3     0.0000    0.00    0.00    0.00   0.0000
IC C2    C3    C4    C5     0.0000    0.00    0.00    0.00   0.0000
IC C1    C6    C7    C8     0.0000    0.00  180.00    0.00   0.0000
IC C6    C7    C8    C9     0.0000    0.00  180.00    0.00   0.0000
IC C7    C8    C9    C10    0.0000    0.00  180.00    0.00   0.0000
```

```
IC C2    C6    *C1   C16    0.0000    0.00  120.00    0.00   0.0000
IC C2    C6    *C1   C17    0.0000    0.00 -120.00    0.00   0.0000
IC C3    C1    *C2   H21    0.0000    0.00  120.00    0.00   0.0000
IC C3    C1    *C2   H22    0.0000    0.00 -120.00    0.00   0.0000
IC C4    C2    *C3   H31    0.0000    0.00  120.00    0.00   0.0000
IC C4    C2    *C3   H32    0.0000    0.00 -120.00    0.00   0.0000
IC C5    C3    *C4   H41    0.0000    0.00  120.00    0.00   0.0000
IC C5    C3    *C4   H42    0.0000    0.00 -120.00    0.00   0.0000
IC C5    C1    *C6   C7     0.0000    0.00  180.00    0.00   0.0000
IC C8    C6    *C7   H71    0.0000    0.00  180.00    0.00   0.0000
IC C9    C7    *C8   H81    0.0000    0.00  180.00    0.00   0.0000
IC C10   C8    *C9   C19    0.0000    0.00  180.00    0.00   0.0000
IC C8    C9    C10   C11    0.0000    0.00  180.00    0.00   0.0000
IC C11   C9    *C10  H101   0.0000    0.00  180.00    0.00   0.0000
IC C9    C10   C11   C12    0.0000    0.00  180.00    0.00   0.0000
IC C12   C10   *C11  H111   0.0000    0.00  180.00    0.00   0.0000
IC C10   C11   C12   C13    0.0000    0.00  180.00    0.00   0.0000
IC C13   C11   *C12  H121   0.0000    0.00  180.00    0.00   0.0000
IC C11   C12   C13   C14    0.0000    0.00  180.00    0.00   0.0000
IC C14   C12   *C13  C20    0.0000    0.00  180.00    0.00   0.0000
IC C12   C13   C14   C15    0.0000    0.00  180.00    0.00   0.0000
IC C15   C13   *C14  H141   0.0000    0.00  180.00    0.00   0.0000
IC C13   C14   C15   OR     0.0000    0.00  180.00    0.00   0.0000
IC OR    C14   *C15  H151   0.0000    0.00  120.00    0.00   0.0000
IC OR    C14   *C15  H152   0.0000    0.00 -120.00    0.00   0.0000
IC C14   C15   OR    HR     0.0000    0.00  180.00    0.00   0.0000
IC C6    C1    C16   H161   0.0000    0.00  180.00    0.00   0.0000
IC H161  C1    *C16  H162   0.0000    0.00  120.00    0.00   0.0000
IC H161  C1    *C16  H163   0.0000    0.00 -120.00    0.00   0.0000
IC C6    C1    C17   H171   0.0000    0.00  180.00    0.00   0.0000
IC H171  C1    *C17  H172   0.0000    0.00  120.00    0.00   0.0000
IC H171  C1    *C17  H173   0.0000    0.00 -120.00    0.00   0.0000
IC C6    C4    *C5   C18    0.0000    0.00  180.00    0.00   0.0000
IC C4    C5    C18   H181   0.0000    0.00  180.00    0.00   0.0000
IC H181  C5    *C18  H182   0.0000    0.00  120.00    0.00   0.0000
IC H181  C5    *C18  H183   0.0000    0.00 -120.00    0.00   0.0000
IC C8    C9    C19   H191   0.0000    0.00  180.00    0.00   0.0000
IC H191  C9    *C19  H192   0.0000    0.00  120.00    0.00   0.0000
IC H191  C9    *C19  H193   0.0000    0.00 -120.00    0.00   0.0000
IC C12   C13   C20   H201   0.0000    0.00  180.00    0.00   0.0000
IC H201  C13   *C20  H202   0.0000    0.00  120.00    0.00   0.0000
IC H201  C13   *C20  H203   0.0000    0.00 -120.00    0.00   0.0000

PATCH FIRST NONE LAST NONE
```

Topology for **retinal**

```
RESI RTAL       0.00 ! retinal, nomenclature from PDB based on retinol
!
GROUP
ATOM C1   CT    0.00
ATOM C2   CT2  -0.18
ATOM H21  HA    0.09
ATOM H22  HA    0.09
ATOM C3   CT2  -0.18
ATOM H31  HA    0.09
ATOM H32  HA    0.09
ATOM C4   CT2  -0.18
ATOM H41  HA    0.09
ATOM H42  HA    0.09
ATOM C5   CC1A  0.00
ATOM C6   CC1A  0.00
GROUP
ATOM C7   CC1B -0.15
ATOM H71  HE1   0.15
ATOM C8   CC1B -0.15
ATOM H81  HE1   0.15
ATOM C9   CC1A  0.00
ATOM C10  CC1A -0.15
ATOM H101 HE1   0.15
GROUP
ATOM C11  CC1B -0.15
ATOM H111 HE1   0.15
ATOM C12  CC1B -0.15
ATOM H121 HE1   0.15 !
ATOM C13  CC1A  0.00 !H162 H163  H171 H172
GROUP                !    \  |    |  /
ATOM C14  CC1A -0.01 !H161-C16    C17-H173            H191 H192
ATOM H141 HE1   0.15 !        \  /                      \   /
ATOM C15  CD    0.19 !  H21    C1        H71    H81      C19
ATOM O15  O    -0.43 !     \  /  \        |      |       /  \
ATOM H15  HR1   0.10 !  H22-C2    C6------C7=====C8----C9    H193
GROUP                !      |     ||                    ||
ATOM C16  CT3  -0.27 !  H31-C3    C5   H181            C10-H101
ATOM H161 HA    0.09 !     /  \  /  \  /               |
ATOM H162 HA    0.09 !  H32    C4    C18-H182          C11-H111
ATOM H163 HA    0.09 !       /  \     \                ||
GROUP                !     H41  H42   H183            C12-H121 H201
ATOM C17  CT3  -0.27 !                                 |       /
ATOM H171 HA    0.09 !                                 C13---C20-H202
ATOM H172 HA    0.09 !                                 ||      \
ATOM H173 HA    0.09 !                                 C14-H141 H203
GROUP                !                                 |
ATOM C18  CT3  -0.27 !                                 C15
ATOM H181 HA    0.09 !                                /  \\
ATOM H182 HA    0.09 !                             H15   O15
ATOM H183 HA    0.09 !
GROUP                !
ATOM C19  CT3  -0.27 !
ATOM H191 HA    0.09 !
ATOM H192 HA    0.09 !
```

```
ATOM H193 HA    0.09 !
GROUP                !
ATOM C20  CT3  -0.27 !
ATOM H201 HA    0.09 !
ATOM H202 HA    0.09 !
ATOM H203 HA    0.09 !

BOND C1   C2   C2   C3   C3   C4   C4   C5   C5   C6   C6   C1
BOND C6   C7   C7   C8   C8   C9   C9   C10  C10  C11  C11  C12
BOND C12  C13  C13  C14  C14  C15  C15  O15  C15  H15
BOND C1   C16  C1   C17  C5   C18  C9   C19  C13  C20
BOND C2   H21  C2   H22  C3   H31  C3   H32  C4   H41  C4   H42
BOND C7   H71  C8   H81  C10  H101 C11  H111 C12  H121 C14  H141
BOND C16  H161 C16  H162 C16  H163 C17  H171 C17  H172 C17  H173
BOND C18  H181 C18  H182 C18  H183 C19  H191 C19  H192 C19  H193
BOND C20  H201 C20  H202 C20  H203
IC C1    C2    C3    C4     0.0000    0.00    0.00    0.00   0.0000
IC C2    C3    C4    C5     0.0000    0.00    0.00    0.00   0.0000
IC C6    C1    C2    C3     0.0000    0.00    0.00    0.00   0.0000
IC C1    C6    C7    C8     0.0000    0.00  180.00    0.00   0.0000
IC C6    C7    C8    C9     0.0000    0.00  180.00    0.00   0.0000
IC C7    C8    C9    C10    0.0000    0.00  180.00    0.00   0.0000
IC C8    C9    C10   C11    0.0000    0.00  180.00    0.00   0.0000
IC C2    C6    *C1   C16    0.0000    0.00  120.00    0.00   0.0000
IC C2    C6    *C1   C17    0.0000    0.00 -120.00    0.00   0.0000
IC C3    C1    *C2   H21    0.0000    0.00  120.00    0.00   0.0000
IC C3    C1    *C2   H22    0.0000    0.00 -120.00    0.00   0.0000
IC C4    C2    *C3   H31    0.0000    0.00  120.00    0.00   0.0000
IC C4    C2    *C3   H32    0.0000    0.00 -120.00    0.00   0.0000
IC C5    C3    *C4   H41    0.0000    0.00  120.00    0.00   0.0000
IC C5    C3    *C4   H42    0.0000    0.00 -120.00    0.00   0.0000
IC C5    C1    *C6   C7     0.0000    0.00  180.00    0.00   0.0000
IC C8    C6    *C7   H71    0.0000    0.00  180.00    0.00   0.0000
IC C9    C7    *C8   H81    0.0000    0.00  180.00    0.00   0.0000
IC C10   C8    *C9   C19    0.0000    0.00  180.00    0.00   0.0000
IC C11   C9    *C10  H101   0.0000    0.00  180.00    0.00   0.0000
IC C9    C10   C11   C12    0.0000    0.00  180.00    0.00   0.0000
IC C12   C10   *C11  H111   0.0000    0.00  180.00    0.00   0.0000
IC C10   C11   C12   C13    0.0000    0.00  180.00    0.00   0.0000
IC C13   C11   *C12  H121   0.0000    0.00  180.00    0.00   0.0000
IC C11   C12   C13   C14    0.0000    0.00  180.00    0.00   0.0000
IC C14   C12   *C13  C20    0.0000    0.00  180.00    0.00   0.0000
IC C12   C13   C14   C15    0.0000    0.00  180.00    0.00   0.0000
IC C15   C13   *C14  H141   0.0000    0.00  180.00    0.00   0.0000
IC C13   C14   C15   O15    0.0000    0.00  180.00    0.00   0.0000
IC O15   C14   *C15  H15    0.0000    0.00  180.00    0.00   0.0000
IC C6    C1    C16   H161   0.0000    0.00  180.00    0.00   0.0000
IC H161  C1    *C16  H162   0.0000    0.00  120.00    0.00   0.0000
IC H161  C1    *C16  H163   0.0000    0.00 -120.00    0.00   0.0000
IC C6    C1    C17   H171   0.0000    0.00  180.00    0.00   0.0000
IC H171  C1    *C17  H172   0.0000    0.00  120.00    0.00   0.0000
IC H171  C1    *C17  H173   0.0000    0.00 -120.00    0.00   0.0000
IC C6    C4    *C5   C18    0.0000    0.00  180.00    0.00   0.0000
IC C4    C5    C18   H181   0.0000    0.00  180.00    0.00   0.0000
IC H181  C5    *C18  H182   0.0000    0.00  120.00    0.00   0.0000
IC H181  C5    *C18  H183   0.0000    0.00 -120.00    0.00   0.0000
IC C8    C9    C19   H191   0.0000    0.00  180.00    0.00   0.0000
```

```
IC H191  C9    *C19  H192   0.0000    0.00  120.00    0.00   0.0000
IC H191  C9    *C19  H193   0.0000    0.00 -120.00    0.00   0.0000
IC C12   C13   C20   H201   0.0000    0.00  180.00    0.00   0.0000
IC H201  C13   *C20  H202   0.0000    0.00  120.00    0.00   0.0000
IC H201  C13   *C20  H203   0.0000    0.00 -120.00    0.00   0.0000

PATCH FIRST NONE LAST NONE
```

Topology for **retinoic acid**

```
RESI RTAC      -1.00 ! all-trans-retinoic acid, nomenclature from PDB
!
! retinoic acid is depronated form which exists at physiological pH
GROUP
ATOM C1   CT    0.00
ATOM C16  CT3  -0.27
ATOM H161 HA    0.09
ATOM H162 HA    0.09
ATOM H163 HA    0.09
ATOM C17  CT3  -0.27
ATOM H171 HA    0.09
ATOM H172 HA    0.09
ATOM H173 HA    0.09
GROUP
ATOM C2   CT2  -0.18
ATOM H21  HA    0.09
ATOM H22  HA    0.09
ATOM C3   CT2  -0.18
ATOM H31  HA    0.09
ATOM H32  HA    0.09
ATOM C4   CT2  -0.18
ATOM H41  HA    0.09
ATOM H42  HA    0.09
GROUP
ATOM C5   CC1A  0.00
ATOM C6   CC1A  0.00
GROUP
ATOM C18  CT3  -0.27
ATOM H181 HA    0.09
ATOM H182 HA    0.09
ATOM H183 HA    0.09
GROUP
ATOM C7   CC1B -0.15
ATOM H71  HE1   0.15
ATOM C8   CC1B -0.15
ATOM H81  HE1   0.15
GROUP
ATOM C9   CC1A  0.00
ATOM C19  CT3  -0.27 !H162 H163  H171 H172
ATOM H191 HA    0.09 !    \  |    |  /
ATOM H192 HA    0.09 !H161-C16    C17-H173            H191 H192
ATOM H193 HA    0.09 !        \  /                      \   /
ATOM C10  CC1A -0.15 !  H21    C1        H71    H81      C19
ATOM H101 HE1   0.15 !     \  /  \        |      |       /  \
GROUP                !  H22-C2    C6------C7=====C8----C9    H193
ATOM C11  CC1B -0.15 !      |     ||                    ||
```

```
ATOM H111 HE1   0.15 !  H31-C3    C5   H181           C10-H101
ATOM C12  CC1B -0.15 !     /  \  /  \  /               |
ATOM H121 HE1   0.15 !  H32    C4    C18-H182         C11-H111
GROUP                !        /  \     \              ||
ATOM C13  CC1A  0.00 !      H41  H42   H183           C12-H121 H201
ATOM C20  CT3  -0.27 !                                |       /
ATOM H201 HA    0.09 !                                C13---C20-H202
ATOM H202 HA    0.09 !                                ||      \
ATOM H203 HA    0.09 !                                C14-H141 H203
GROUP                !                                |
ATOM C14  CC1A -0.25 !                                C15=O29
ATOM H141 HE1   0.15 !                                |   (-)
ATOM C15  CC    0.62 !                                O29B
ATOM O29  OC   -0.76 !
ATOM O29B OC   -0.76

BOND C1   C2   C2   C3   C3   C4   C4   C5   C5   C6   C6   C1
BOND C6   C7   C7   C8   C8   C9   C9   C10  C10  C11  C11  C12
BOND C12  C13  C13  C14  C14  C15  C15  O29  C15  O29B
BOND C1   C16  C1   C17  C5   C18  C9   C19  C13  C20
BOND C2   H21  C2   H22  C3   H31  C3   H32  C4   H41  C4   H42
BOND C7   H71  C8   H81  C10  H101 C11  H111 C12  H121 C14  H141
BOND C16  H161 C16  H162 C16  H163 C17  H171 C17  H172 C17  H173
BOND C18  H181 C18  H182 C18  H183 C19  H191 C19  H192 C19  H193
BOND C20  H201 C20  H202 C20  H203
IC C1    C2    C3    C4     0.0000    0.00    0.00    0.00   0.0000
IC C2    C3    C4    C5     0.0000    0.00    0.00    0.00   0.0000
IC C6    C1    C2    C3     0.0000    0.00    0.00    0.00   0.0000
IC C1    C6    C7    C8     0.0000    0.00  180.00    0.00   0.0000
IC C6    C7    C8    C9     0.0000    0.00  180.00    0.00   0.0000
IC C7    C8    C9    C10    0.0000    0.00  180.00    0.00   0.0000
IC C2    C6    *C1   C16    0.0000    0.00  120.00    0.00   0.0000
IC C2    C6    *C1   C17    0.0000    0.00 -120.00    0.00   0.0000
IC C6    C1    C16   H161   0.0000    0.00  180.00    0.00   0.0000
IC H161  C1    *C16  H162   0.0000    0.00  120.00    0.00   0.0000
IC H161  C1    *C16  H163   0.0000    0.00 -120.00    0.00   0.0000
IC C6    C1    C17   H171   0.0000    0.00  180.00    0.00   0.0000
IC H171  C1    *C17  H172   0.0000    0.00  120.00    0.00   0.0000
IC H171  C1    *C17  H173   0.0000    0.00 -120.00    0.00   0.0000
IC C3    C1    *C2   H21    0.0000    0.00  120.00    0.00   0.0000
IC C3    C1    *C2   H22    0.0000    0.00 -120.00    0.00   0.0000
IC C4    C2    *C3   H31    0.0000    0.00  120.00    0.00   0.0000
IC C4    C2    *C3   H32    0.0000    0.00 -120.00    0.00   0.0000
IC C5    C3    *C4   H41    0.0000    0.00  120.00    0.00   0.0000
IC C5    C3    *C4   H42    0.0000    0.00 -120.00    0.00   0.0000
IC C6    C4    *C5   C18    0.0000    0.00  180.00    0.00   0.0000
IC C4    C5    C18   H181   0.0000    0.00  180.00    0.00   0.0000
IC H181  C5    *C18  H182   0.0000    0.00  120.00    0.00   0.0000
IC H181  C5    *C18  H183   0.0000    0.00 -120.00    0.00   0.0000
IC C5    C1    *C6   C7     0.0000    0.00  180.00    0.00   0.0000
IC C8    C6    *C7   H71    0.0000    0.00  180.00    0.00   0.0000
IC C9    C7    *C8   H81    0.0000    0.00  180.00    0.00   0.0000
IC C10   C8    *C9   C19    0.0000    0.00  180.00    0.00   0.0000
IC C8    C9    C19   H191   0.0000    0.00  180.00    0.00   0.0000
IC H191  C9    *C19  H192   0.0000    0.00  120.00    0.00   0.0000
IC H191  C9    *C19  H193   0.0000    0.00 -120.00    0.00   0.0000
IC C8    C9    C10   C11    0.0000    0.00  180.00    0.00   0.0000
```

```
IC C11   C9    *C10  H101   0.0000    0.00  180.00    0.00   0.0000
IC C9    C10   C11   C12    0.0000    0.00  180.00    0.00   0.0000
IC C12   C10   *C11  H111   0.0000    0.00  180.00    0.00   0.0000
IC C10   C11   C12   C13    0.0000    0.00  180.00    0.00   0.0000
IC C13   C11   *C12  H121   0.0000    0.00  180.00    0.00   0.0000
IC C11   C12   C13   C14    0.0000    0.00  180.00    0.00   0.0000
IC C14   C12   *C13  C20    0.0000    0.00  180.00    0.00   0.0000
IC C12   C13   C20   H201   0.0000    0.00  180.00    0.00   0.0000
IC H201  C13   *C20  H202   0.0000    0.00  120.00    0.00   0.0000
IC H201  C13   *C20  H203   0.0000    0.00 -120.00    0.00   0.0000
IC C12   C13   C14   C15    0.0000    0.00  180.00    0.00   0.0000
IC C15   C13   *C14  H141   0.0000    0.00  180.00    0.00   0.0000
IC C13   C14   C15   O29    0.0000    0.00  180.00    0.00   0.0000
IC O29   C14   *C15  O29B   0.0000    0.00  180.00    0.00   0.0000

PATCH FIRST NONE LAST NONE
```

Parameters for retinoids

```
read parameter card append

BONDS
!
!atom types  Kb          b0
!
!BTE2, 2-butene
HE1   CC1A  360.500     1.100
HE1   CC1B  360.500     1.100
HE2   CC2   365.000     1.100
CC1A  CC1A  440.000     1.340
CC1B  CC1B  440.000     1.340
!13DB, Butadiene
CC1A  CC2   500.000     1.342
CC1B  CC2   500.000     1.342
CC1A  CC1B  300.000     1.470
!13DP, 1,3-Pentadiene
CC1B  CT3   383.000     1.504
!MECH
CC1A  CT2   365.000     1.502
CC1A  CT    365.000     1.502
CC1A  CT3   383.000     1.504
!TMCH/MECH
CT   CT2    222.500     1.538
CT   CT3    222.500     1.538
!PACP
CA    NH1   305.000     1.414
!PRAL
CC1A  CD    300.000     1.4798
CD    O     720.000     1.205
CD    HR1   330.000     1.110
!PRAC
CC1A  CC    440.000     1.489
!CROT
CC1A  C     440.000     1.489
!SCH1, Schiff's base model compound 1, deprotonated
```

```
CT3   NS1   310.000     1.440
NS1   CC1B  500.000     1.276
!SCH2, Schiff's base model compound 2, protonated
HR1   CC1B  360.500     1.100
CT3   NS2   300.000     1.453
NS2   CC1B  470.000     1.283
NS2   HC    455.000     1.000
!SCK0, deprotonated Schiff's base, lysine retinal patch
CT2   NS1   310.000     1.440
!SCK1, , Schiff's base, lysine retinal patch
CT2   NS2   300.000     1.453

ANGLES
!
!atom types     Ktheta    Theta0   Kub     S0
!
!BTE2, 2-butene
CC1A CC1A CT3    48.00     123.50
HA   CT2  CC1A   45.00     111.50
HA   CT3  CC1A   42.00     111.50
HA   CT3  CC1B   42.00     111.50
HE1  CC1A CC1A   42.00     119.00
HE1  CC1B CC1B   42.00     119.00
HE1  CC1A CT3    42.00     117.50
HE1  CC1B CT3    22.00     117.50
!13DB, 1,3-Butadiene
HE1  CC1A CC2    42.00     118.00
HE1  CC1B CC2    42.00     118.00
HE2  CC2  CC1A   45.00     120.50
HE2  CC2  CC1B   45.00     120.50
HE1  CC1B CC1A   42.00     118.00
HE1  CC1A CC1B   42.00     118.00
CC1B CC1A CC2    48.00     123.50
CC1A CC1B CC2    48.00     123.50
HE2  CC2  HE2    19.00     119.00
!13DP, Pentadiene
CC1A CC1B CC1B   48.00     123.50
CC1B CC1B CT3    48.00     123.50
!MECH
CC1A CT   CT3    32.00     112.20
CC1A CT   CT2    32.00     112.20
CC1A CT2  CT2    32.00     112.20
CC1A CC1A CC1B   48.00     123.50
CC1A CC1A CT     48.00     123.50
CC1A CC1A CT2    48.00     123.50
CC1B CC1A CT     48.00     123.50
CT2  CC1A CT3    48.00     123.50
!DMB1, 2-methyl-1,3-butadiene
CC2  CC1A CT3    48.00     123.50
CC1B CC1A CT3    48.00     113.00
!DMP1, 4-methyl-1,3-pentadiene
CT3  CC1B CT3    47.00     113.00
!TMCH/MECH
CT2  CT   CT3    58.35     113.50   11.16   2.561
CT3  CT   CT3    58.35     113.50   11.16   2.561
CT   CT2  CT2    58.35     113.50   11.16   2.561
CT   CT2  HA     26.50     110.10   22.53   2.179
```

```
CT   CT3  HA     33.43     110.10   22.53   2.179
CT   CC1A CT3    48.00     123.50
!PACP
CA   NH1  C      50.00     120.00
H    NH1  CA     34.00     117.00
NH1  CA   CA     40.00     120.00   35.00   2.4162
!CROT
CT3  CC1A CT3    47.00     125.20
CC1A CC1A C      48.00     123.50
HE1  CC1A C      52.00     119.50
O    C    CC1A   80.00     122.50
NH1  C    CC1A   80.00     116.50
! PROL
HE1  CC1A CT2    40.00     116.00
OH1  CT2  CC1A   75.70     110.10
CT2  CC1A CC2    28.00     126.00
!PRAL
CC2  CC1A CD     60.00     120.00
CC1A CC1A CD     60.00     120.00
HE1  CC1A CD     32.00     122.00 ! sum=242
CC1A CD   O      75.00     124.00
CC1A CD   HR1    15.00     115.00
O    CD   HR1    75.00     121.00 ! fixed
!PRAC
HE1  CC1A CC     52.00     119.50 !
CC2  CC1A CC     40.00     119.00   35.00   2.5267
CC1A CC   OC     40.00     118.00   50.00   2.3880
CC1A CC1A CC     48.00     123.50 !
!SCH1, Schiff's base model compound 1, deprotonated
CT3  NS1  CC1B   67.00     111.00
NS1  CC1B CT3    52.00     123.00 !sum=242.5
NS1  CC1B HE1    38.00     119.50 !
HA   CT3  NS1    42.00     113.50
!SCH2, Schiff's base model compound 2, protonated
CT3  NS2  CC1B   67.00     123.60
CT3  NS2  HC     38.00     117.40
CC1B NS2  HC     38.00     118.80
NS2  CC1B CT3    40.00     125.60
NS2  CC1B HR1    38.00     114.00
CT3  CC1B HR1    42.00     120.40
HA   CT3  NS2    42.00     110.10
!SCH3, Schiff's base model compound 3, protonated
NS2  CC1B CC1A   40.00     125.60
HR1  CC1B CC1A   42.00     120.40
!SCK0, deprotonated Schiff's base, lysine retinal patch
CT2  CT2  NS1    67.70     110.00 !from lysine
HA   CT2  NS1    42.00     113.50
CT2  NS1  CC1B   67.00     111.00
NS1  CC1B CC1A   40.00     123.00
!SCK1, protonated Schiff's base, lysine retinal patch
CT2  NS2  CC1B   67.00     123.60
CT2  NS2  HC     38.00     117.40
CT2  CT2  NS2    67.70     110.00 !from lysine
HA   CT2  NS2    42.00     110.10

DIHEDRALS
!
```

```
!atom types             Kchi    n   delta
!
!BTE2, 2-butene
HE1  CC1A CC1A HE1      5.2000  2   180.00
CT3  CC1A CC1A HE1      5.2000  2   180.00
CT3  CC1B CC1B HE1      5.2000  2   180.00
CC1A CC1A CT3  HA       0.3000  3     0.00
HE1  CC1A CT3  HA       0.3000  3     0.00
HE1  CC1B CT3  HA       0.3000  3     0.00
CT3  CC1A CC1A CT3     10.0000  2   180.00 !double bond
CT2  CC1A CC1A CT3     10.0000  2   180.00 !double bond
!13DB, 1,3-Butadiene
CC2  CC1A CC1B CC2      0.4000  1   180.00 !
CC2  CC1A CC1B CC2      0.4000  2   180.00 !
CC2  CC1A CC1B CC2      1.3000  3     0.00 !
CC2  CC1A CC1B HE1      1.0000  2   180.00 !
HE1  CC1A CC1B CC2      1.0000  2   180.00 !
HE1  CC1A CC1B CC1B     1.0000  2   180.00 !
HE2  CC2  CC1A CC1B     5.0000  2   180.00 !double bond
HE2  CC2  CC1B CC1A     5.0000  2   180.00 !double bond
HE1  CC1B CC1B CC1A     5.2000  2   180.00 !double bond
HE1  CC1B CC1A HE1      0.0000  2   180.00 !
!HEP3, 1,3,5-heptatriene
CC1A CC1A CC1B CC2      0.5000  1   180.00 !
CC1A CC1A CC1B CC2      2.0000  2     0.00 !
CC1A CC1A CC1B CC2      1.0000  3     0.00 !
CC1A CC1A CC1B CC1B     0.5000  1   180.00 !
CC1A CC1A CC1B CC1B     2.0000  2     0.00 !
CC1A CC1A CC1B CC1B     1.0000  3     0.00 !
CC1A CC1B CC1B CC1A     0.5600  1   180.00 !double bond
CC1A CC1B CC1B CC1A     7.0000  2   180.00 !double bond
CC1B CC1A CC1A CC1B     0.5600  1   180.00 !double bond
CC1B CC1A CC1A CC1B     7.0000  2   180.00 !double bond
HE1  CC1A CC2  HE2      5.2000  2   180.00 !double bond
HE1  CC1B CC2  HE2      5.2000  2   180.00 !double bond
HE1  CC1A CC1A CC1B     1.0000  2   180.00 !double bond
!13DP, 1,3-Pentadiene
CC1B CC1B CT3  HA       0.3000  3     0.00
HE1  CC1B CC1B HE1      0.0000  2   180.00 !
CC1A CC1B CC1B CT3      0.5600  1   180.00 !double bond
CC1A CC1B CC1B CT3      7.0000  2   180.00 !double bond
CC2  CC1A CC1B CC1B     0.6000  1   180.00 ! also in DMB1, DMP1, DMP2
CC2  CC1A CC1B CC1B     0.5000  2   180.00 ! comprise values obtained
CC2  CC1A CC1B CC1B     1.0000  3     0.00 !
!DMB1, 2-methyl-1,3-butadiene
CC1B CC1A CT3  HA       0.3000  3   180.00
CT3  CC1A CC2  HE2      1.0000  2   180.00
CT3  CC1A CC1B HE1      1.0000  2   180.00
CT3  CC1A CC1B CC2      1.1000  1   180.00 !see DMP1
CT3  CC1A CC1B CC2      0.7000  2   180.00
!DMP1, 4-methyl-1,3-pentadiene, no additional terms versus 13DP
CT3  CC1B CT3  HA       0.3000  3     0.00
!DMP2, 2-methyl-1,3-pentadiene
CT3  CC1A CC1B CC1B     1.1000  1   180.00 !see DMP1
CT3  CC1A CC1B CC1B     0.7000  2   180.00
CC2  CC1A CT3  HA       0.3000  3     0.00
!TMCH
```

```
CT   CC1A CC1A CT2     10.0000  2   180.00 ! double bond
CT   CC1A CC1A CT3     10.0000  2   180.00 ! double bond
CC1A CC1A CT   CT2      0.5000  2     0.00 ! c2-c1-c6=c5
CC1A CC1A CT   CT2      0.3000  3     0.00 ! c2-c1-c6=c5
CC1A CC1A CT   CT3      0.5000  2     0.00 ! c16/c17-c1-c6=c5
CC1A CC1A CT   CT3      0.4000  3     0.00 ! c16/c17-c1-c6=c5
CC1A CC1A CT2  CT2      0.5000  2     0.00 ! c3-c4-c5=c6
CC1A CC1A CT2  CT2      0.3000  3     0.00 ! c3-c4-c5=c6
CT2  CT2  CC1A CT3      0.1900  3     0.00 ! c3-c4-c5-c18
CT2  CT   CC1A CT3      0.4000  3     0.00 ! c2-c1-c6-c7
CT3  CT   CC1A CT3      0.4000  3     0.00 ! c16/c17-c1-c6-c7
CC1A CT2  CT2  CT2      0.1900  3     0.00 !
CT2  CT2  CT   CC1A     0.2000  3     0.00 !
CT2  CT2  CT   CT3      0.2000  3     0.00 !
CT3  CC1A CT2  HA       0.1900  3     0.00
CC1A CT   CT2  HA       0.1900  3     0.00
CT3  CT   CT2  HA       0.1900  3     0.00
CC1A CT2  CT2  HA       0.1900  3     0.00
CT2  CC1A CT3  HA       0.1600  3     0.00
CT   CC1A CT3  HA       0.1600  3     0.00
CT2  CT   CT3  HA       0.1600  3     0.00
CC1A CT   CT3  HA       0.1600  3     0.00
CT3  CT   CT3  HA       0.1600  3     0.00
!MECH
CT   CC1A CC1B CC2      0.9000  1     0.00
CT   CC1A CC1B CC2      2.1000  2   180.00
CT   CC1A CC1B CC2      0.2200  3     0.00
CT   CC1A CC1B CC2      0.2500  5   180.00
CT   CC1A CC1B CC2      0.1000  6     0.00
CT   CC1A CC1B CC1B     0.9000  1     0.00
CT   CC1A CC1B CC1B     2.1000  2   180.00
CT   CC1A CC1B CC1B     0.2200  3     0.00
CT   CC1A CC1B CC1B     0.2500  5   180.00
CT   CC1A CC1B CC1B     0.1000  6     0.00
CC1B CC1A CT   CT2      0.3000  3     0.00
CC1B CC1A CT   CT3      0.3000  3     0.00
CC1B CC1A CC1A CT2      0.5600  1   180.00 ! double bond
CC1B CC1A CC1A CT2      7.0000  2   180.00 ! double bond
CC1B CC1A CC1A CT3      0.5600  1   180.00 ! double bond
CC1B CC1A CC1A CT3      7.0000  2   180.00 ! double bond
CT   CC1A CC1B HE1      1.0000  2   180.00
CC1A CC1A CC1B HE1      1.0000  2   180.00
!PACP
O    C    NH1  CA       2.5000  2   180.00 !
HP   CA   CA   NH1      4.2000  2   180.00 !
CA   CA   CA   NH1      3.1000  2   180.00 !
C    NH1  CA   CA       1.2000  2   180.00 !
C    NH1  CA   CA       1.0000  3   180.00 !
H    NH1  CA   CA       0.5000  2   180.00 !
CA   NH1  C    CT3      2.5000  2   180.00 !
!CROT
CT3  CC1A CT3  HA       0.3000  3     0.00 !
CC1B CC1A CC1A C        0.5600  1   180.00 !double bond
CC1B CC1A CC1A C        7.0000  2   180.00 !double bond
O    C    CC1A CC1A     0.7000  1   180.00 !
O    C    CC1A CC1A     1.2000  2   180.00 !
O    C    CC1A CC1A     0.1000  3   180.00 !
```

```
O    C    CC1A CC1A     0.2000  4     0.00 !
NH1  C    CC1A CC1A     0.7000  1     0.00 !
NH1  C    CC1A CC1A     1.2000  2   180.00 !
NH1  C    CC1A CC1A     0.1000  3     0.00 !
NH1  C    CC1A CC1A     0.1500  4     0.00 !
HE1  CC1A C    NH1      0.3000  3   180.00 !
HE1  CC1A C    O        0.3000  3   180.00 !
CC1A C    NH1  H        2.5000  2   180.00
CC1A C    NH1  CT3      1.6000  1     0.00
CC1A C    NH1  CT3      2.5000  2   180.00
CC1A C    NH1  CA       1.6000  1     0.00 !FRET
CC1A C    NH1  CA       2.5000  2   180.00 !FRET
C    CC1A CC1A CT3      0.5600  1   180.00 !double bond
C    CC1A CC1A CT3      7.0000  2   180.00 !double bond
!PROL
CC2  CC1A CT2  OH1      1.9000  1   180.00
CC2  CC1A CT2  OH1      0.4000  2   180.00
CC2  CC1A CT2  OH1      0.6000  3   180.00
CC1A CC1A CT2  OH1      1.9000  1   180.00
CC1A CC1A CT2  OH1      0.4000  2   180.00
CC1A CC1A CT2  OH1      0.6000  3   180.00
CC1A CT2  OH1  H        1.3000  1     0.00
CC1A CT2  OH1  H        0.7000  2     0.00
CC1A CT2  OH1  H        0.5000  3     0.00
HE1  CC1A CT2  OH1      0.2000  3     0.00
HE1  CC1A CT2  HA       0.2000  3     0.00
CC2  CC1A CT2  HA       0.0300  3     0.00
CC1A CC1A CT2  HA       0.0300  3     0.00
HE2  CC2  CC1A CT2      5.2000  2   180.00
!PRAL
CC2  CC1A CD   O        1.0000  2  180.00 !
CC1A CC1A CD   O        1.0000  2  180.00 !
CC2  CC1A CD   HR1      3.2000  2  180.00
CC1A CC1A CD   HR1      3.2000  2  180.00
HE1  CC1A CD   O        0.0000  2  180.00
HE1  CC1A CD   HR1      0.0000  2  180.00
HE2  CC2  CC1A CD       3.2000  2  180.00
CT3  CC1A CC1A CD       0.5600  1  180.00 ! double bond
CT3  CC1A CC1A CD       0.5000  2  180.00 ! db
CC1B CC1A CC1A CD       0.5600  1  180.00 ! db
CC1B CC1A CC1A CD       7.0000  2  180.00 ! db
!PRAC
CC2  CC1A CC   OC       1.3000  2  180.00
CC1A CC1A CC   OC       1.3000  2  180.00
HE2  CC2  CC1A CC       4.2000  2  180.00
HE1  CC1A CC   OC       0.0000  2  180.00
CC1B CC1A CC1A CC       0.5600  1  180.00 !double bond
CC1B CC1A CC1A CC       7.0000  2  180.00 !double bond
CC   CC1A CC1A CT3      0.5600  1  180.00 !double bond, rtac
CC   CC1A CC1A CT3      7.0000  2  180.00 !double bond, rtac
!SCH1, Schiff's base model compound 1, deprotonated
CT3  NS1  CC1B HE1      8.5000  2   180.00
CT3  NS1  CC1B CT3     12.0000  2   180.00
NS1  CC1B CT3  HA       0.1000  3   180.00
CC1B NS1  CT3  HA       0.3500  3   180.00
!SCH2, Schiff's base model compound 2, protonated
CT3  NS2  CC1B HR1      8.5000  2   180.00
```

```
CT3  NS2  CC1B CT3      7.0000  2   180.00
NS2  CC1B CT3  HA       0.1500  3   180.00
HR1  CC1B CT3  HA       0.1500  3     0.00
CC1B NS2  CT3  HA       0.1500  3   180.00
HA   CT3  NS2  HC       0.1100  3     0.00
HC   NS2  CC1B HR1      5.0000  2   180.00
HC   NS2  CC1B CT3      5.0000  2   180.00
!SCH3, Schiff's base model compound 3, protonated
CT3  NS2  CC1B CC1A     7.0000  2   180.00
HC   NS2  CC1B CC1A     5.0000  2   180.00
NS2  CC1B CC1A HE1      1.0000  2   180.00
NS2  CC1B CC1A CC2      0.5000  1     0.00 !opt
NS2  CC1B CC1A CC2      2.2000  2   180.00 !opt
NS2  CC1B CC1A CC2      1.1000  3     0.00 !opt
NS2  CC1B CC1A CC2      0.6000  4     0.00 !opt
HR1  CC1B CC1A HE1      0.0000  2   180.00
HR1  CC1B CC1A CC2      1.0000  2   180.00
!SCK0, deprotonated Schiff's base, lysine retinal patch
CT2  CT2  NS1  CC1B     0.1000  3     0.00
CT2  NS1  CC1B CC1A    12.0000  2   180.00 !from CT3  NS1  CC1B CT3
CT2  NS1  CC1B HE1      8.5000  2   180.00 !from CT3  NS1  CC1B HE1
HA   CT2  NS1  CC1B     0.1000  3     0.00
NS1  CC1B CC1A CC1A     0.5000  1     0.00 !from NS2  CC1B CC1A CC2
NS1  CC1B CC1A CC1A     2.2000  2   180.00 !N=C-C=C dihedral not
explicitly optimized
NS1  CC1B CC1A CC1A     1.1000  3     0.00 !
NS1  CC1B CC1A CC1A     0.6000  4     0.00 !
NS1  CC1B CC1A HE1      1.0000  2   180.00
!SCK1, Schiff's base, lysine retinal patch
NS2  CC1B CC1A CC1A     0.5000  1     0.00 !from NS2  CC1B CC1A CC2
NS2  CC1B CC1A CC1A     2.2000  2   180.00 !
NS2  CC1B CC1A CC1A     1.1000  3     0.00 !
NS2  CC1B CC1A CC1A     0.6000  4     0.00 !
HR1  CC1B CC1A CC1A     1.0000  2   180.00
CT2  CT2  NS2  CC1B     0.1000  3     0.00
CT2  NS2  CC1B CC1A     7.0000  2   180.00
CT2  NS2  CC1B HR1      8.5000  2   180.00
HA   CT2  NS2  CC1B     0.1000  3     0.00
CT2  CT2  NS2  HC       0.1000  3     0.00
HA   CT2  NS2  HC       0.1000  3     0.00

IMPROPER
!atom types             Kpsi        psi0
!PRAL
CD   CC1A O   HR1      14.0000  0     0.00 !

NONBONDED nbxmod  5 atom cdiel shift vatom vdistance vswitch -
cutnb 14.0 ctofnb 12.0 ctonnb 10.0 eps 1.0 e14fac 1.0 wmin 1.5
               !adm jr., 5/08/91, suggested cutoff scheme
!
!atom  ignored    epsilon      Rmin/2   ignored   eps,1-4
Rmin/2,1-4
!
!Butadiene
CC1A   0.0   -0.0680     2.0900 !
CC1B   0.0   -0.0680     2.0900 !
CC2    0.0   -0.0640     2.0800 !
```

```
!SCH1, SCH2
NS1    0.000000  -0.200000     1.850000 !N for deprotonated Schiff's
base
NS2    0.000000  -0.200000     1.850000 !N for protonated Schiff's base

end
```