Molecular Dynamics of the Hammerhead Ribozyme

Vasant T. Gandhi

Loyola University Chicago

Follow this and additional works at: https://ecommons.luc.edu/luc_theses

Part of the Chemistry Commons

Recommended Citation

https://ecommons.luc.edu/luc_theses/4217

This Thesis is brought to you for free and open access by the Theses and Dissertations at Loyola eCommons. It has been accepted for inclusion in Master's Theses by an authorized administrator of Loyola eCommons. For more information, please contact ecommons@luc.edu.

This work is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 License.

Copyright © 1997 Vasant T. Gandhi
LOYOLA UNIVERSITY CHICAGO

MOLECULAR DYNAMICS OF THE HAMMERHEAD RIBOZYME

A THESIS SUBMITTED TO
THE FACULTY OF THE GRADUATE SCHOOL
IN CANDIDACY FOR THE DEGREE OF
MASTER OF SCIENCE
DEPARTMENT OF CHEMISTRY

BY
VASANT T. GANDHI
CHICAGO, ILLINOIS
JANUARY 1997
Copyright by Vasant Gandhi, 1996
All rights reserved.
ACKNOWLEDGMENTS

I would like to thank all who gave their time, advice, and help to me in the completion of this thesis. I would like to give special thanks to Dr. Kenneth W. Olsen, Dr. David S. Crumrine, Dr. Willetta Greene-Johnson, Dr. Albert W. Herlinger, and Dr. Stephen F. Pavkovic who challenged, taught, inspired, and believed with their hearts and minds.
# TABLE OF CONTENTS

ACKNOWLEDGMENTS iii
LIST OF ILLUSTRATIONS v
LIST OF TABLES vii
LIST OF ABBREVIATIONS viii

Chapter
1. INTRODUCTION 1
2. THE HAMMERHEAD RIBOZYME 3
   The Hammerhead Cleavage Reaction Mechanism 3
   Structure Elucidation 4
3. FOLDING OF NUCLEIC ACIDS 13
4. METAL IONS AND RNA STRUCTURE 16
   Counterionic Cloud 16
5. PROPERTIES OF NUCLEIC ACIDS 19
   Conformational Statistics 19
   Thermodynamics 20
   Helix-Coil Transitions 20
   The Molten Globule State 21
   Molecular Stress 25
6. ATOMIC SIMULATIONS 31
   Molecular Dynamics Theory 31
   Full Newtonian Simulation or Deterministic Model of Atomic Motion 33
   Stochastic Models of Atomic Motion 34
7. POTENTIAL ENERGY FUNCTIONS 42
8. SIMULATIONS 46
   Molecular Complex Preparation 46
   Energy Minimization 46
Solvent Representation
Dynamics
9. TRAJECTORY ANALYSIS
   Comparison of Simulation Trajectories
   Final Structures
   In-Line Cleavage Mechanism
   RMSD Evaluation
   RMSD Calculations with respect to the Final Structure
   Phosphate Backbone Changes
   Base (Side-chain) Evolution
   RMSD Calculations with respect to the Average Structure
   Phosphate Backbone Flexibility
   Base (Side-chain) Flexibility
10. CONCLUSION
APPENDIX A
VITA
LIST OF ILLUSTRATIONS

Figure 1. (a) The hammerhead ribozyme catalytic pocket and location of the scissile bond (b) Proton abstraction in the hammerhead ribozyme cleavage mechanism (c) Ribozyme Engineering (d) Hammerhead ribozyme cleavage mechanism

Figure 2. (a) Pley structure (b) Tuschl structure

Figure 3. (a) Scott core structure (b) Tuschl core structure (c) Pley core structure

Figure 4. Scott structures showing nucleotide backbone

Figure 5. (a) In vitro pathway of protein folding illustrating possible interconversions between unfolded (U) and partially folded intermediates (I) which eventually fold to the compact intermediate (CI) state before reaching the native state (N) and a competing aggregation step (formation of occlusion bodies) for misfolded structures. (b) Pathway from unfolded (U) state to properly folded or native (N) state showing transition through intermediate states including compact intermediate and passage through a high energy state. (c) Model for structure of compact intermediates. (d) Energy landscape folding model

Figure 6. Minimized Structure showing Mg$^{2+}$ ions (small squares)

Figure 7. Backbone Side-Chain Correlation

Figure 8. Temperature versus Time

Figure 9. Potential Energy, Kinetic Energy, and Total Energy versus Time

Figure 10. Radius of Gyration versus Time

Figure 11. Hydrogen Bonds versus Time

Figure 12. Dipole Moment versus Time

Figure 13. Final Structures at 600 ps

Figure 14. O2'-P-O5' Bond Angle versus Time

Figure 15. Minimized ribozyme structure: (a) ribbon and (b) ladder views. Average Ribozyme structure over 550 ps dynamics: (c) ribbon and (d) ladder views

Figure 16. Average Structure of Ribozyme showing Mg$^{2+}$ ions

Figure 17. Ribbon and ladder views of ribozyme

Figure 18. RMSD with respect to Final

Figure 19. RMSD with respect to Average Structure
LIST OF TABLES

Table 1. Comparison of Average Molecular Properties Between Newtonian and Langevin Dynamics Simulations at 300K and 500K 61
Table 2. Backbone (Phosphate Atom) Evolution towards Final Structure 65
Table 3. Side-Chain (Nucleotide Base) Evolution towards Final Structure 66
Table 4. Backbone (Phosphate Atom) Flexibility with respect to Average Structure 67
Table 5. Side-Chain (Base) Flexibility with respect to Average Structure 68
**LIST OF ABBREVIATIONS**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Adenosine</td>
</tr>
<tr>
<td>BPTI</td>
<td>Bovine Pancreatic Trypsin Inhibitor</td>
</tr>
<tr>
<td>C</td>
<td>Cytosine</td>
</tr>
<tr>
<td>CHARMm</td>
<td>Chemistry Harvard Macromolecular Mechanics</td>
</tr>
<tr>
<td>D</td>
<td>Daltons</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic Acid</td>
</tr>
<tr>
<td>FRET</td>
<td>Fluorescence Resonance Energy Transfer</td>
</tr>
<tr>
<td>G</td>
<td>Guanine</td>
</tr>
<tr>
<td>Mg$^{2+}$</td>
<td>Magnesium Ion</td>
</tr>
<tr>
<td>RMSD</td>
<td>Root-Mean-Square Deviation</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic Acid</td>
</tr>
<tr>
<td>U</td>
<td>Uracil</td>
</tr>
</tbody>
</table>
CHAPTER 1

INTRODUCTION

Conformations of biological macromolecules result from their interrelated structural, functional, and environmental attributes. This thesis examines structural and energetic changes occurring in a specific type of biological macromolecule, nucleic acids, under simulated environmental (temperature, solvent) conditions. The computational biophysical approach through molecular dynamics simulations may help define knowledge essential to applications in molecular biology, pharmacology, medicine, nanotechnology and computer engineering.

Three candidate molecules for study were originally identified in the literature; each candidate molecule presented a different degree of nucleic acid structure: a synthetic oligoribonucleotide double helix, a transfer ribonucleic acid, and a catalytic ribonucleic acid, the hammerhead ribozyme. The hammerhead ribozyme is examined in this investigation.

The primary purpose of this thesis is to investigate and characterize the dynamics, structure, and thermodynamics of a catalytic ribonucleic acid, the hammerhead ribozyme, as the molecule searches conformational space at different temperatures and under different dynamical methods. A secondary purpose is to investigate the presence of a specific structural and energetic intermediate state, the “molten globule” state, in a nucleic acid’s search of conformational space during folding or unfolding. This study uses molecular
dynamics techniques to analyze the internal motions and thermodynamics of nucleic acids. These techniques present "an atomic picture of high resolution with respect to space, energy or time" and are valuable "probes" of molecular structure and function.

As noted by one researcher, "[T]he characterization of a biomolecular system at the atomic level in terms of structure, mobility, dynamics, and energetics is incomplete * * * This incomplete molecular picture makes it difficult to establish the link between molecular structure, mobility, dynamics, and interactions on the one hand, and biological function on the other." 

---

1. For example, a change in pH will produce a change in the electrical charge associated with the side chain of an amino acid. This change in charge causes a change in the amino acid and overall protein conformation yielding a change in function.


CHAPTER 2

THE HAMMERHEAD RIBOZYME

The self-cleaving hammerhead ribozyme has been the subject of intensive recent examination. The hammerhead ribozyme, a catalytic RNA, is an interesting example of an RNA molecule which is rich in tertiary structure. The molecule’s X-ray coordinates were received from the Protein Data Bank. The ribozyme consists of three base-paired stems or helices and a core of non-complementary nucleotides. This molecule “promotes” a magnesium ion-dependent site-specific cleavage of RNA by intermolecular or intramolecular reactions. Upon the addition of hydrogens and magnesium ions, the molecular system in this study consists of 1078 fragments, 651 ring bonds, 386 rotatable bonds, and 1331 atoms and possesses a molecular weight of 13,662.61 D. We characterize the temporal behavior of the system consisting of the molecule and an ion and solvent environment at different temperatures using various dynamical methods. The structural motif consists of three helical regions surrounding a conserved core. When viewed in two-dimensions, the hammerhead ribozyme secondary structure resembles a “hammerhead.”

The Hammerhead Cleavage Reaction Mechanism

The self-cleavage or auto-cleavage reaction results from the nucleophilic attack by a core nucleotide’s 2’-hydroxyl upon the adjacent phosphodiester bond. Figure 1(a) shows the general architecture of the hammerhead ribozyme motif and the locations of the scissile bond and catalytic pocket. The reaction mechanism yields a 2’,3’-cyclic phosphate terminal.
and a 5'-hydroxyl terminal. Significant inquiry is directed currently at the cleavage mechanism generally and the role of magnesium ions in proton abstraction (Figure 1(b)). Studies have also been conducted on the in vitro activity of the hammerhead ribozyme. Through so-called "ribozyme engineering", ribozymes may be used to "edit" or repair messenger RNA molecules (Figure 1(c)). For example, a ribozyme may be used as an "antisense" nucleotide designed to bind with and subsequently cleave deleterious RNA and DNA sequences. Figure 1(d) presents ribozyme engineering and the mechanism of hammerhead ribozyme cleavage.

The hammerhead ribozyme is one of several structurally distinct classes of catalytic RNA differing in terms of their reaction products and nucleophile. Most are metalloribozymes requiring Mg$^{2+}$ or Pb$^{2+}$ as electrophiles for their cleavage reactions. These ribozymes may be further classified as either splicing or self-cleaving.

Structure Elucidation

Recent efforts by researchers at several institutions have produced structural models for the hammerhead ribozyme. The research groups adopted differing approaches to structure solution in terms of the biophysical technique utilized or the molecular system investigated. Pley, et al., have produced a structure based on X-ray crystallography of a hammerhead RNA-DNA ribozyme-inhibitor complex to a resolution of 2.6A. Tuschi, et al., have utilized fluorescence resonance energy transfer of the hammerhead ribozyme in solution to produce a three-dimensional model. More recently, Scott, et al., have
elucidated the X-ray crystal structure of an all-RNA hammerhead ribozyme.\textsuperscript{14} Figure 2 presents the Pley, Tuschl, and Scott structures.

Comparisons of the determined structures of the hammerhead ribozyme yield useful insight into the efficacy of the method used and provide a basis for critical study of the biophysical methodology. We examined the structures obtained by FRET (Tuschl) and X-ray crystallography (Scott, Pley) to determine whether model building with the judicious application of distance constraints based upon physical measurements is sufficient to define the structure of small RNA molecules such as the hammerhead ribozyme.\textsuperscript{15}

Figure 2(d) presents the Scott structure with two ribozyme molecules within the asymmetric unit. The molecular coordinate file (revised) as deposited in the Brookhaven Protein Databank was modified to produce separate coordinate files for each molecule in the Scott structure. The Pley structure contained three molecules within the asymmetric unit; one molecule was excised from the coordinate file. The Scott structure and the Pley structure were compared with the Tuschl structure.

In order to correctly compare the methods used to generate each structure, a meaningful basis for comparison must be established, i.e. alignment of similar sequence residues. Conveniently, for our purposes, each hammerhead ribozyme molecule possesses a conserved core of nucleotides articulated into two structural domains. Because of the different hammerhead ribozyme complexes studied, a frame of reference is critical.

We define the common core in the following manner: the nucleotide sequence C U G A (the first domain) followed by U followed by the nucleotide sequence G A G C G A A
A C (the second domain). The core consists of 14 nucleotides or groups, 302 atoms, and 307 bonds. Schematically, the consensus structure of each molecule is shown in Figure 3 as a backbone representation.

The Pley and Tuschl structures exhibited a RMSD of 8.500 Å. The Scott and Tuschl structures exhibited a RMSD of 8.531 Å and the Scott and Pley structures showed a RMSD of 0.904 Å. In each case, the calculations encompassed 4 fragments, 380 ring bonds, and 160 rotatable bonds within 2 groups. 302 matches were conducted as part of the calculation. These results indicate that the FRET structure deviates strongly from either of the X-ray structures. The X-ray structures were in close agreement with each other. It is reasonable to conclude that FRET does not yet provide sufficient structural information in the case of small macromolecules to be a viable substitute for X-ray crystallography. Additionally presented are differing representations of the Scott structure, including backbone, ladder, and ball-and-stick depictions. Such representations are useful in assessing the degree of structure present in the hammerhead ribozyme.
Figure 1. (a) The hammerhead ribozyme catalytic pocket and location of the scissile bond (b) Proton abstraction in the hammerhead ribozyme cleavage mechanism (c) Ribozyme Engineering (d) Hammerhead ribozyme cleavage mechanism.
Figure 2. (a) Pley structure (b) Tuschl structure (c) Scott structure (d) Scott structure with two ribozyme molecules within asymmetric unit.
Figure 3. (a) Scott core structure  (b) Tuschl core structure  
(c) Pley core structure.
Figure 4. Scott structures showing nucleotide backbone

2 A nucleotide (a phosphate ester of a nucleoside) consists of a sugar (ribose or deoxyribose), a nitrogen heterocyclic purine or pyrimidine base, and a phosphate group. The bases are connected to the sugar-phosphate backbone. The helical backbone is comprised of a linear series of sugars which are linked through bonds from the 3' OH of the sugar to a phosphate group to the 5' OH of the next sugar; the nitrogenous base is attached to the backbone by its N to the C1 of the sugar. Base pairing between adenine and uracil consists of 2 hydrogen bonds; base pairing between cytosine and guanine consists of 3 hydrogen bonds. The orientation of the bases with respect to each other is not perfectly coplanar. The actual twisting which occurs is often termed "propeller twist." Stryer at 650.


CHAPTER 3
FOLDING OF NUCLEIC ACIDS

The architecture of folding patterns in nucleic acids\(^1\) is currently the subject of extensive research particularly as the discovery of additional structures frees nucleic acids from the historically confining designation as “carrier[s] of genetic information”.\(^2\) One central problem is that of defining the “arrangement of RNA structural elements in three-dimensional space” in a system analogous to that developed for proteins.\(^3\)

In proteins, it is commonplace to describe polypeptide folding patterns in terms of structural motifs such as helices, sheets, barrels, etc.\(^4\) The units of RNA folding\(^5\) are generally viewed as the base-paired double helix secondary structure, pseudoknots, loops which cap helices, loops within helices, RNA mispairing regions, nucleoside triple interactions, quadruplexes, and U-turns\(^6\) as well as multiplexes, junctions, and hairpins.\(^7\) Tertiary structures form, in a manner analogous to protein folding, by the “condensation” of “individual blocks of secondary structure.”\(^8\) The strength of the analogy is limited, however, by the view that “RNA secondary structural elements are very stable and capable of forming independently of tertiary structure. This has led to the concept that RNA secondary structure forms rapidly and precedes the packaging of RNA into tertiary structure, i.e. that base pairing interactions occur prior to helix or domain formation.”\(^9,10\)

Nonetheless, the foundation for analyzing polynucleotides and polypeptides in similar fashion is a sound one. As noted by one author,
“From the energy landscape perspective it is natural to suggest that the considerations that lead to the theoretical developments of protein folding should also apply to RNA folding. In general terms, the requirements for RNA folding are analogous to those of protein folding. As is true for polypeptides, the number of conformations in the fully denatured state (the Levinthal limit) is large. For RNA sequences, the kinetic problem consists of forming the correct secondary structure, that is, Watson-Crick base pairs between complementary sequences, and achieving the correct three-dimensional organization of the structural elements.”

A comparison of the kinetic folding pathways of proteins and RNA shows clear differences in the folding pathway. The theoretical time scale for RNA unfolding is significantly large so as to exceed the boundaries of the present investigation. However, one of the purposes of this project, to characterize the short-term dynamics of the molecule, is not diminished by the realization that significant large-scale structural transitions will not likely occur before 1 millisecond.

6Pyle, A.M. and Green, Justin B. (1995) RNA folding. 5 Current Opinion in Structural Biology 303. For example, “[p]seudoknots are interlocked regions of coaxially stacked helices that are commonly involved in RNA binding and folding. Pseudoknotting is a motif with a high capacity for molecular recognition **”. Pyle, A.M. and Green, Justin B. (1995) RNA folding. 5 Current Opinion in Structural Biology 303. See also Abrahams,


CHAPTER 4
METAL IONS AND RNA STRUCTURE

The presence of metal ions contributes greatly to the folding of RNA into various structures. The structural and dynamic effects of the binding of metal ions to RNA may be specific or non-specific. Thermal denaturation experiments have shown that specific (site) binding of metal ions, as with tertiary RNA structures, produces a linear relationship between the reciprocal of the RNA melting temperature and magnesium ion concentration. Non-specific RNA binding by metal ions, as in the case of charge screening and duplex formation, produces a sigmoidal plot.

It has been concluded that metal ions perform three distinct functions in RNA folding: non-specific binding, specific binding or coordination using the Mg$^{2+}$ ion, and "binding to high-affinity sites through charge density or outer-sphere coordination rather than through the formation of direct metal contacts." An important concept is the counterionic cloud within the context of the Debye-Huckel theory. Although there is evidence for the Mg$^{2+}$-independent folding of the hammerhead ribozyme, one author questions the experimental conditions. Specifically the 2'-deoxyribose substitution of the cleavage site to prevent cleavage of substrate may have disrupted ion binding.

Counterionic Cloud

The presence of ions produces, as an average property, a "counterionic cloud" which moves with the molecule. The molecule's size increases due to the cloud's presence
which alters the molecule’s electrical and hydrodynamic properties. The Debye-Huckel theory explains the charge distribution of the cloud and the resultant effect upon a charged macromolecule’s electrical potential $V$ at a given location $r$. Calculation of counterion density, i.e. position-dependent concentration, requires the following conditions: (1) non-constant concentration of ions due to thermal fluctuations, (2) equal concentrations of positive and negative counterions at remote distances from the molecule, and (3) a greater concentration of oppositely charged counterions than same charge as the molecule at local distances from the molecule.

Additionally, the Boltzmann distribution is assumed to govern charge distribution. The counterions assume positions based upon the Boltzmann distribution of their respective potential energies, the net charge density $\rho$ equals the difference between the concentration of positive and negative counterions. The relationship between net charge density and potential as a function of position is governed by the Poisson equation, $\nabla^2 V = -4\Pi \rho/D$, where $\rho$ is the net charge density and $D$ is the dielectric constant of the medium (solvent). The presence of a counterionic cloud decreases the potential energy of the molecular system in comparison to systems without a counterionic cloud (ionic atmosphere).

---


11 The concentrations of negative and positive counterions close to the molecule, respectively, is given by $e^{-U/k_B T}$ where $U$ is the electrical potential energy of the counterion at a given distance from the molecule, $k_B$ is the Boltzmann constant, and $T$ is the absolute temperature. Gabler, R. (1978). Electrical Interactions in Molecular Biophysics at 224.


14 Given that $W_{ab} = D_{ab}q$ and that there is zero potential at infinity, the potential energy, $U$, of a given counterion is $U = W = Vq$ where $V$ equals the potential at the counterion's final location and $q$ equals the counterion's charge. Ion concentration, $C$, is equal to $C_0 e^{-qV/k_BT}$ where $C_0$ is the bulk concentration of ions. Gabler, R. (1978). Electrical Interactions in Molecular Biophysics at 225.

CHAPTER 5
PROPERTIES OF NUCLEIC ACIDS

Conformational Statistics

Polymeric chains, such as polypeptides and polynucleotides, may be viewed as linear systems made of independent elements or statistical segments. In the case of polynucleotides, the relevant unit is the nucleotide. The flexibility of biopolymers, due to rotation around single bonds, produces a very large set or ensemble of possible conformations. Analysis of the potential energy of the biopolymer chain and its functional dependence upon the multitude of possible internal angles of rotation underlies the conformational statistical approach.

Internal motions of biopolymers may be functionally classified as local motions (atomic fluctuations, sidechain motions, loop motions), rigid-body motions (helix motions, domains or hinge-bending motions, and subunit), and larger-scale motions (helix-coil transitions, dissociation/association and coupled structural changes, opening and distortional fluctuations, and folding and unfolding transitions).

The rotations $x$ about the glycosidic bond are of two types: syn where $x = 0$ and anti where $x = 210$. The preferred state is a function of the degree of sugar "puckering" present. Puckering refers to the relative spatial position of the 5 atoms constituting the ribose ring. Four atoms are in the same plane; the fifth, C2' or C3', is above or below the plane. An endo conformation exists when the displacement occurs on the same side as
CS'; exo occurs when the displacement occurs on the opposite side.\(^7\) There are six rotational angles for each backbone unit.\(^8\)

By considering factors such as steric (electronic) hindrance or bad (forbidden) contacts, the possible conformations may be limited by first-order interactions (between atoms whose interatomic distance is dependent upon a single rotation angle.)\(^9\) Second-order interactions are those interactions whose separation distance is simultaneously dependent upon two adjacent angles.\(^10\)

**Thermodynamics**

Using molecular dynamics techniques, thermodynamic properties, such as entropy, may be ascertained from the system’s thermally accessible structural conformations.\(^11\) The most stable conformation of a biological macromolecule is found at its lowest energy structure.

**Helix-Coil Transitions**

The helical structure of protein or polypeptide chains is produced by intermolecular hydrogen bonding. In the case of polynucleotides, however, intramolecular hydrogen bonding between base pairs provides helical structure.\(^12\) The polyelectrolytic nature of nucleic acids due, in part, to the presence of negatively charged phosphate groups, plays a role as well in the structure of these biopolymers.\(^13\)

It is well known that polynucleotides and polypeptides undergo helix-coil transitions under conditions of increasing temperature (thermal denaturation) and changing pH of solution.\(^14\) Development of a partition function or free energy expression for a
polynucleotide system is essential to an understanding of the helix-coil transition. An example of such a relationship is

Equation 1

\[ \Delta G(T) = \Delta H(T) - T \Delta S(T) = \Delta H(T_R) + \Delta C_p(T-T_R) - T \Delta S(T_R) + \Delta C_p \ln(T/T_R) \]

where \( \Delta G(T) \) represents the difference between the molar Gibbs free energy function of any state and a reference state, \( \Delta H(T) \) represents the state’s relative enthalpy, \( \Delta S(T) \) represents the entropy in a two-state transition and equals \( \Delta H_{T_m}/T_m \); \( \Delta C_p \) represents the heat capacity at constant pressure and equals \( d\Delta H/dT \); and \( T_R \) represents an appropriate reference temperature.

The Molten Globule State

The existence of the molten globule state as an intermediate in protein folding or unfolding pathways has been defined and discussed by several researchers. A question arises whether such a state exists in the folding or unfolding pathways of other biological molecules such as nucleic acids.

Molten globules, for proteins, have been characterized as compact, mobile structures with large amounts of secondary structure, but diminished tertiary contacts relative to the native state. The molten globule state may be a common early intermediate during folding and may be an equilibrium intermediate under a variety of unfolding conditions such as high temperature, extreme pH, presence of organic solvents, or removal of stabilizing ions. Researchers have also found the molten globule state to be under kinetic control.
Daggett and Levitt have performed extensive molecular dynamics simulations of temperature-induced protein unfolding using bovine pancreatic trypsin inhibitor (BPTI).\textsuperscript{20} Their experimental approach began with the X-ray crystallographic structure of BPTI. All atoms were explicitly present during simulations. The protein and fragments were immersed in a box of water molecules and counterions were present to yield an electrically neutral system.\textsuperscript{21} Daggett \textit{et al.} observed the molten globule state through a structural analysis of protein parameters. The researchers started with BPTI in its native, folded state and observed the first transition from native state to molten globule. A second transition was observed from the molten globule to the unfolded state.

The results of phase 1 from the native state to the molten globule state were presented in terms of structural level from global to semi-local to local levels. The researchers followed time-dependencies of properties which show characteristic differences between the X-ray structure and molten globule states. The parameters and characteristics examined included global structural deviations from crystal structure (conformational sampling, size, packing interactions, hydrophobic core, and presence of secondary structural motifs) and semi-local and local structural deviations (secondary structure, turn formation, correlated motion).

Although the researchers found it "difficult to estimate numerically the deviation that one would expect for a molten globule or other partially unfolded intermediates:, the values obtained "seem[ed] reasonable for a molten globule, which is native-like in many respects, and is not expected to deviate considerably from the native state." \textsuperscript{22}
Additional studies have observed the molten globule state in cytochrome c at low pH by examining changes in the molecule's thermodynamic properties. Figure 5 presents a useful diagram illustrating the pathways. Methods of identifying compact nonnative structures from the native globular protein structures have also been investigated. The molten globule has been termed the third thermodynamical state of protein molecules.

The transitions from the native folded state (N) to the molten globule state (MG) and from MG to the denatured unfolded state (U) have been determined to follow the “all or none” mechanism due to the absence of equilibrium intermediates between the states.

It has been concluded that “all proteins, under the appropriate conditions, will form [compact intermediates]” As recognized by several researchers, it is valuable to distinguish between the varying types of partially folded intermediates. The following diagram attempts to show the relevant distinctions:

```
Partially Folded Intermediates
/               |
Compact Intermediate   Molten Globule   Compact Denatured
```

At the outset, it is important to recognize that “[t]he term compact intermediates encompasses a broad range of conformations and degrees of folding and compactness: compact intermediates have no single, unique conformation, but rather a whole plethora of structures that range from being very similar to the native state to being substantially expanded and significantly unfolded.”
The compact intermediate generally possesses the following qualities: [a] substantial secondary structure with slight tertiary structure, [b] compact collapsed state with respect to the unfolded state as measured in terms of hydrodynamic radius, [c] exposed hydrophobic surfaces, [d] heat capacity similar to the unfolded state\textsuperscript{31}, [e] absence of functional properties, and [f] less cooperative unfolding transition.\textsuperscript{32}

With respect to compact denatured states of proteins in general: (a) An individual protein may possess several stable compact intermediate states depending on the individual protein character and observation conditions; (b) structurally, two types of compact intermediates predominate: "native state-like regions of secondary structure connected by disordered regions of polypeptide, but still retaining a relatively native state-like topology, or a core of native-like structure";\textsuperscript{33} and (c) it is difficult to isolate compact intermediates from compact substrates as "the energetic difference between relatively expanded compact intermediates and compact unfolded states resides mostly in the entropy and not \textit{sic} the enthalpy."\textsuperscript{34} The compact intermediate may also be an ensemble of states before passing through a high energy state, and bypassing a competing aggregation step, to the native state.\textsuperscript{35} An energy landscape folding model is a valuable method of visualizing the folding pathway of biological macromolecules (Figure 5(d)).\textsuperscript{36} The concern has been raised that the molten globule and other non-native states may in fact be incorrectly refolded states produced as a result of the observation conditions, which may be inappropriate for folding, instead of true folding intermediates: "Does [the protein or nucleic acid] fold and unfold through an intermediate state that is stable under some conditions and does not represent an
artifact of folding, i.e. is not a misfolded state trapped in a potential well?" Such a state would represent "topological frustration" for the molecule.

Molecular Stress

For purposes of stress analysis on the molecular level, the oligonucleotide chain may be viewed as a basic unit. The molecules behave as entropic springs. Assume a polymer, e.g. a polynucleotide, of length L is subjected to a tensile force F. Macroscopic thermodynamic theory holds that

\[
F = \frac{\partial A(L,T)}{\partial L} \bigg|_T
\]

where \( A(L,T) \) is the Helmholtz free energy at absolute temperature T.

\[
A(L,T) = U(L,T) - TS(L,T)
\]

where \( U \) is the internal energy and \( S \) is the entropy of the system.

\[
F = \frac{\partial U(L,T)}{\partial L} \bigg|_T - T \frac{\partial S(L,T)}{\partial L} \bigg|_T
\]

The force necessary for polymer extension produces changes, to varying degrees, in both internal energy and entropy. In the case of constrained macromolecules, the behavior is that of a molecular entropic spring in tension. The concept of intrinsic monomer stress (IMS) has been developed to explain stress in polymeric systems at the atomic level.
Figure 5. (a) *In vitro* pathway of protein folding illustrating possible interconversions between unfolded (U) and partially folded intermediates (I) which eventually fold to the compact intermediate (CI) state before reaching the native state (N) and a competing aggregation step (formation of occlusion bodies) for misfolded structures. (b) Pathway from unfolded (U) state to properly folded or native (N) state showing transition through intermediate states including compact intermediate and passage through a high energy state. (c) Model for structure of compact intermediates. (d) Energy landscape folding model.


Cantor and Schimmel (1980) at 313.

Cantor and Schimmel (1980) at 313.

Cantor and Schimmel (1980) at 314.

Cantor and Schimmel (1980) at 314.

Cantor and Schimmel (1980) at 314.

Cantor and Schimmel (1980) at 314.

Cantor and Schimmel (1980) at 314.

Cantor and Schimmel (1980) at 314.

Cantor and Schimmel (1980) at 314.

McCammon and Harvey at 68-73.


29 However, see Haynie, D. and Freire, E. (1993). Structural Energetics of the Molten Globule State. 12 PROTEINS: Structure, Function, and Genetics 115-140. (detailing recent protein studies showing differences in enthalpy and heat capacity between the molten globule state and the unfolded state).


CHAPTER 6
ATOMISTIC SIMULATIONS

Molecular Dynamics Theory

Molecular dynamics simulations are used to determine individual molecular motions within solids, liquids, or gases.\(^1\) Molecular dynamics simulations investigate temporally-related sequences in the positions of molecules, or trajectories, through the numerical solution of the equations of motion and may be thus viewed as a purely deterministic method.\(^2\)

It should be noted that the complexity of biomolecular systems prohibits solution of Schrodinger's equation due to the large number of atoms and interactions involved.\(^3\) The use of classical mechanics theory and semi-empirical or effective interaction functions is a useful substitute for the exact quantum mechanical solutions.\(^4\) This approach substitutes an analysis of the electronic degrees of freedom for atomic degrees of freedom.\(^5\) Molecular mechanics theory treats electrons implicitly rather than explicitly.\(^6\) The uncertainty principle requires the use of advanced quantum mechanical techniques because momentum and position cannot both be simultaneously known with certainty with respect to an atom.

There are three common stages to a molecular dynamics simulation: [a] model construction, [b] trajectory calculations, and [c] analysis of trajectories.\(^7\) The model consists of the molecule under investigation, based upon defined atomic coordinates derived, for example, through X-ray analysis\(^8\), and, generally, a fixed environment or
volume of gas, solid, or liquid. The molecule and its environment comprise a system for investigation. The system is typically at thermodynamic equilibrium. Under equilibrium conditions, the physical parameters specifying a given state remain the same, e.g. pressure, volume, or temperature.

The thermodynamic state of the system is a function of the number of molecules (N), the volume (V), and the total energy (E). Successive integrations of Newton’s and Langevin’s motion equations yield sequential time-related atomic or molecular positions. The integrations produce trajectories or molecular configurations as a function of time. These time-dependent position vectors are combined with time-dependent momentum vectors, formed in response to atomic interactions, to produce a 6N-dimensional hyperspace or phase space. Phase space represents the combination of 3N-dimensional configuration space and 3N-dimensional momentum space.

Three general concerns are relevant to the reliability of a molecular model: [a] explicit representation of the electronic or atomic degrees of freedom “essential” to the representation of the phenomenon of inquiry; [b] the interaction function; [c] a choice of the appropriate set of equations for the molecular motions based upon the available degrees of freedom. Equation sets commonly used in simulations include Schrodinger (quantum mechanical), Newton or Lagrange (classical mechanical) and Langevin (stochastic).

Molecular dynamics of proteins and nucleic acids has been extensively investigated. Specific internal motions of interest in nucleic acids include the relative
vibration of bonded atoms, longitudinal motions of bases in double helices, lateral motions of bases in double helices, global stretching and twisting, elastic vibration of globular region, sugar repuckering, torsional libration of buried groups, relative motion of different globular regions (hinge bending), global bending, allosteric transitions, and local denaturation.\textsuperscript{14} The parameters of interest associated with each of these motions include spatial extent and amplitude.\textsuperscript{15} Molecular dynamics simulations of helix denaturation have been extensively examined for proteins.\textsuperscript{17}

The method of molecular dynamics is also relevant. The commonly used methods include adiabatic dynamics,\textsuperscript{18} used in the current study, as well as annealed dynamics,\textsuperscript{19} canonical (TVN) dynamics,\textsuperscript{20} and temperature-damped dynamics.\textsuperscript{21} Under the simulated annealing or "slow cooling" approach, the atoms lose their thermal mobility.\textsuperscript{22} These methods may be modified further by the use of "impulse" dynamics where the initial directional velocities of certain atoms are preset to overcome translation barriers and permitting relaxation or "quenched dynamics" where the molecular structure is minimized periodically and saved, following dynamic stages, allowing a search for low energy structures.\textsuperscript{23}

Full Newtonian Simulation or Deterministic Model of Atomic Motion

In a full Newtonian simulation, the complete equations of motion (Newton's second law)
Equation 5
\[ \ddot{r}_i = \frac{F_{\text{int} \cdot ra}}{m} \]

are integrated with respect to time for all atoms including hydrogens and, if applicable, solvent atoms. The trajectory \( r_i(t) \) of every atom is thus determined for a given set of initial conditions \( r_i(0) \). The force exerted on each atom is the negative gradient of the potential energy function:

Equation 6
\[ F_{\text{int} \cdot ra} = -\frac{\partial V}{\partial r_i}. \]

In integrating the equations of motion, a time step which is sufficiently short with respect to the smallest time scale of the atomic motions must be employed. The time scale of the atomic motion decreases as the atomic mass is reduced. Therefore, the time step of the simulation and, hence, the overall computational cost is determined by the motion of the lighter (hydrogen) atoms. This makes the computational cost of a full Newtonian simulation prohibitive. In the present study, in order to reduce the computational cost, the CHARMM-based SHAKE algorithm was employed to constrain the lengths of the hydrogen bonds during both the Newtonian and Langevin simulations. The harmonic motions represented by the fluctuation of the hydrogen bonds are weakly coupled to other atomic motions and, hence, may be constrained to within a certain interatomic distance.

Stochastic Models of Atomic Motion

Besides the motion due to the interatomic forces, another important component of the atomic motion is the harmonic vibration which corresponds to the thermodynamic
equilibrium of individual atoms with a temperature bath. The time scale of this motion is much shorter than the time scale of the motion which corresponds to the interatomic forces. Therefore, any attempt to accurately follow this motion during a full Newtonian simulation would make the computational cost of the simulation prohibitive. The representation of this vibration with a stochastic (random) motion offers an economical alternative.

The transition from the original deterministic system to the stochastic one may be better understood by introducing the concept of the probability of recurrence, i.e. the probability that a particular state (configuration) will reoccur after a finite time has elapsed. A physical system which follows Newton's laws of motion is deterministic. Therefore, every realizable state of the system can be expected to occur after a finite time with a probability equal to 1. If the expected time of recurrence is much longer than the time scale of the physical phenomena under investigation, then the system can be assumed to be random.

The state of a physical system which is stochastic in nature may be completely described by a probability density function defined in the appropriate phase space. The modeling of a physical system is made substantially easier by introducing a model system which does not obey the same deterministic laws as the original system under investigation but which is described by the same probability density function. For example, it can be shown that the probability density function which describes a model stochastic system which is governed by Brownian kinematics changes in the same way as the probability
density function of a physical system which is governed by gradient diffusion. The two systems are then said to be statistically equivalent.

Statistical equivalence is the foundation from which stochastic models of otherwise deterministic physical phenomena (such as Brownian or Langevin models) are derived. In essence, the trajectory \( r(t) \) of an atom is now assumed to be a superposition of a deterministic and a random motion

\[
\mathbf{r}_i(t) = \mathbf{r}_i^{\text{det}}(t) + \mathbf{r}_i^{\text{ran}}(t).
\]

Equation 7

The deterministic component of the atomic motion \( \mathbf{r}_i^{\text{det}}(t) \) is computed in the same manner as in the full Newtonian simulation, i.e. based on Newton’s second law. The interatomic force is again computed as the gradient of a potential energy field. However, only the motions due to the interatomic forces are being followed while the harmonic motion which corresponds to the thermodynamic equilibrium with the temperature bath is represented by the random term \( \mathbf{r}_i^{\text{ran}}(t) \). The choice of this random term (closure assumption) is dictated by the requirements of statistical equivalence, i.e. that the effect that this random motion has on the probability density distribution function of the position of the atoms is the same as the effect of the actual harmonic vibration.

The simplest choice for the random component of the atomic motion is a Brownian (random walk) model. In the Brownian model of motion, the state variable (in the present case, position, \( \mathbf{r}_i \) is assumed to change according to
Equation 8
\[ \text{d} \mathbf{r}^{\text{ran}}_{i} = \text{d} \mathbf{W}_{i} \]

where \( \text{d} \mathbf{W}_{i} \) is a random displacement, e.g. a random walk. The key assumption about the statistical properties of \( \text{d} \mathbf{W}_{i} \) is that they depend only on the present state \( (\mathbf{r}_{i}) \) of the system, i.e. \( \text{d} \mathbf{W}_{i} \) is independent of any previous states. Naturally, this assumption of uncorrelated changes of state is not consistent with the deterministic character of the underlying physical laws. However, this assumption is acceptable when the correlation time of these changes is much shorter than the time scale of the physical processes which are being modeled.

When the correlation time of these changes of state cannot reasonably be neglected, a Langevin model may be adopted instead. In the Langevin formulation, the assumption of negligible correlation time is shifted from position changes to velocity changes. The random velocity

Equation 9
\[ \mathbf{v}^{\text{ran}}_{i} = \frac{\text{d} \mathbf{r}^{\text{ran}}_{i}}{\text{d} t} \]

of the system is assumed to change according to

Equation 10
\[ \text{d} \mathbf{v}^{\text{ran}}_{i} = A \left( \mathbf{v}^{\text{ran}}_{i}, \mathbf{r}^{\text{ran}}_{i}, t \right) \text{d} t + \text{d} \mathbf{W}_{i} \]

where \( A \left( \mathbf{v}^{\text{ran}}_{i}, \mathbf{r}^{\text{ran}}_{i}, t \right) \) is a deterministic acceleration and \( \text{d} \mathbf{W}_{i} \) is a random velocity change for which the same assumption of negligible correlation time is maintained. The function \( A \left( \mathbf{v}^{\text{ran}}_{i}, \mathbf{r}^{\text{ran}}_{i}, t \right) \) determines the correlation profile of the displacements \( \text{d} \mathbf{r}^{\text{ran}}_{i} \) which were
assumed to be uncorrelated in the Brownian model. It can be shown that, at the asymptotic limit of zero correlation time, the Langevin model reduces to the Brownian model.

In the present study, a Langevin model of the form

\begin{equation}
\frac{d\mathbf{\mathbf{r}}^\text{ran}}{dt} = -\frac{\zeta}{m} \mathbf{\mathbf{r}}^\text{ran} + \frac{\mathfrak{F}_\text{ran}^\text{inter}}{m} dt
\end{equation}

was adopted. The feedback function \( A\left(\mathbf{\mathbf{r}}^\text{ran},\mathbf{\mathbf{r}}^\text{ran}, t\right) \) has been chosen so that it has the same effect as a viscous damping force, \(-\zeta \dot{\mathbf{r}}\), where \( \zeta = m b \) and \( b \) is the friction coefficient of the particular atom. The random acceleration, \( dW_i \), is the effect of a random interatomic force \( \mathfrak{F}_i^\text{inter} \) with the following statistical properties

\begin{equation}
<\mathfrak{F}_i^\text{intra}(t)> = 0
\end{equation}

\begin{equation}
<\mathfrak{F}_i^\text{intra}(t)\mathfrak{F}_i^\text{intra}(0)> = 2k_b T_o b m \delta(t)
\end{equation}

where \( k_b \) is Boltzmann's constant and \( T_o \) is the absolute temperature of the temperature bath. It can be shown that the present choice of the feedback function \( A\left(\mathbf{\mathbf{r}}^\text{ran},\mathbf{\mathbf{r}}^\text{ran}, t\right) \) corresponds to an exponential temporal correlation of the random atomic displacements with an autocorrelation time \( \tau_e = m/\zeta \).

The final form of the equation describing atomic motion in the case of the stochastic simulation (compare with Equation 6) is
Equation 13
\[ d\mathbf{r}_i = -\frac{\zeta\mathbf{r}_i}{m} \, dt + \frac{F_{\text{int er}}^i}{m} \, dt + \frac{F_{\text{int ra}}^i}{m} \, dt \]


2 In a purely stochastic methodology, such as Metropolis Monte Carlo, a molecular configuration is dependent solely upon the immediately preceding configuration, i.e. a Markov chain. The spectrum between purely stochastic and purely deterministic simulation methods is composed of hybrid methods with elements, to varying degrees, of both methods: Force-Biased Monte Carlo, Brownian Dynamics, and General Langevin Dynamics in order of increasing deterministic properties. Haile, J. (1992) at 13.


9 \[ F_i(t) = m \dot{r}_i(t) = - \frac{\partial \mu(r_N)}{\partial r_i} \] Haile, J. (1992) at 15. Newtonian dynamics, which is used in the current study, should be contrasted with Hamiltonian dynamics, based upon Newton's second law, wherein a time-independent position and velocity function, the Hamiltonian, is used. Haile, J. (1992) at 40-42.


11 Haile, J. (1992) at 43. In 3N-dimensional configuration space, the coordinate axes are components of position vectors, \( \mathbf{r}(t) \); the coordinate axes are components of momentum vectors, \( \mathbf{p}(t) \) in 3N-dimensional momentum space. A single point in space represents the positions and momenta of the entire N-atom system at a single point in time.


16 McCammon (1989) at 29.


18 In this method, the molecular structure's temperature is maintained within a specified range by periodic scaling of the atomic velocities to constrain the kinetic energy. Molecular Simulations, Inc. (6/93). CERIUS Minimizer/Dynamics Document Update.18-7.
The molecular structure's energy is gradually minimized to avoid local minimum "trapping". The temperature is incrementally increased, as a function of time, to a designated temperature and back to the starting temperature. The lowest energy structure within each temperature cycle is minimized. Molecular Simulations, Inc. (6/93). CERIUS Minimizer/Dynamics Document Update.18-7.

In this isothermal method, temperature, volume, and number of atoms is held constant in isothermal fashion by exchanging energy with a heat bath. The degree of thermal coupling between the system and the heat bath may be regulated by modification of the relaxation time. The heat bath may provide energy needed to overcome rotational barriers to conformations. Molecular Simulations, Inc. (6/93). CERIUS Minimizer/Dynamics Document Update.18-8.


CHAPTER 7

POTENTIAL ENERGY FUNCTIONS

As noted by one author, "[t]he form of the potential function and parameters [force constants] are generally chosen with the goal of mimicking the physics of interatomic interactions."

The correct choice of a reliable potential energy function is based upon considerations of intramolecular energy, charge distributions, hydrogen bonding, dispersion coefficients and other factors.

The essential components of a force field include atom types, auto-typing rules, functions for energy terms, parameters, and an unspecified parameters generator.

The potential energy function or force field equation expresses the functional dependence of potential energy upon individual atomic position.

Three general interaction models are used in molecular simulations. The Hooke’s Law model treats the atoms or molecules as connected to their neighbors by springs. The hard sphere model treats particle interactions as “billiard ball” interactions with an interaction, i.e. bouncing off, taking place only when the particles are within a certain distance of each other. The Lennard-Jones model is characterized by “forces that are strongly repulsive at very short interparticle distances, attractive at larger distances, and extremely weak attractive at very large distances.”

Typically, the model for the intermolecular potential, $\mu(r_N)$, is considered pairwise additive and accordingly represents the interaction energy among $N$ atoms as the sum of
isolated two-body contributions: $\mu(r_N) = \sum \sum u(r_{ij})$ where $i < j$. The Lennard-Jones model for soft-sphere pair potential is generally modified, to varying degrees, in producing force fields to encompass the energy terms required for different classes of molecules, including biomolecules. Molecular mechanics, or optimum geometries, and molecular or motional dynamics are evaluated from the forces calculated for two-, three-, and four-body interactions. Using truncated potentials, minimum image criterion, and neighbor lists as timesaving measures often optimizes calculations of atomic forces.

The potential function generally includes terms for bonded interactions (e.g. harmonic restoring forces between bonded nearest neighbors, penalties for angle deformations, and dihedral torsional potentials for hindered rotation of groups about bonds) and nonbonded interactions between separated atoms (including repulsive van der Waals forces, dispersion attraction, and partial charge electrostatic interactions). Parameter data is obtained experimentally.

The governing relationship between potential energy, atomic positions and atomic velocities is given by

\[
-\frac{dV}{dx} = F = ma = m\frac{d^2x}{dt^2}
\]

where $V$ equals the potential energy, $x$ represents atomic positions, $t$ represents time, $m$ equals mass, and $F$ equals force. The following is an example of a typical biomolecular force field or effective atomic interaction system:
The first term represents the covalent bond-stretching interaction, a harmonic potential, along bond $b$. The bond lengths and force constants, $K_b$, depend upon the bond type. The second term represents the three-body interaction or bond-angle bending. The third term represents dihedral-angle or four-body interactions and the fourth term represents the effective non-bonded interactions as a sum over all pairs of atoms (van der Waals and Coulomb interactions).

It should be noted, however, that “the actual behavior of molecules is governed by the nature of the free energy hyperspace, not that for the potential energy.”

The calculation of the relative free energy of systems has traditionally been computationally prohibitive. Modern free energy perturbation methods such as compositional or conformational (potential of mean force) calculations and increasingly powerful computational methods will result in more accurate simulations.

---


Haile, J. (1992) at 190-197.


CHAPTER 8

SIMULATIONS

Molecular Complex Preparation

The molecule was prepared for dynamics in several steps. Initially, the X-ray crystallographic coordinates of the hammerhead ribozyme, as deposited in the Brookhaven Protein Databank, were imported into the QUANTA\textsuperscript{1} molecular modeling program. Hydrogen atoms were added to the ribozyme and a sufficient number of magnesium ions were added to the set of coordinates to create an electrically neutral molecule-ion complex. The placement of magnesium ions in each complex was essentially random although an attempt was made to place the ions in close proximity to the phosphate groups on each molecule under the expectation that the positively charged magnesium ions would move towards the negatively charged phosphate groups during minimization and dynamics. The complex was "solvated" through the use of a distance-dependent dielectric constant.

Energy Minimization

Minimization of the complex was accomplished through the CHARMm\textsuperscript{2} program for molecular mechanics and dynamics within the QUANTA environment. CHARMm-based nucleic acid topology and potential function parameter files\textsuperscript{3} were employed in the minimization and dynamics simulations.

Minimization is used to reduce a macromolecule's potential energy to its lowest possible value before running molecular dynamics. The molecule must be "relaxed"
because "the crystallographic structure is itself a model, generated by balancing the desire for atomic positions that reproduce the observed diffraction pattern against the desire for a model with traditional values for bond lengths, bond angles, [etc]." Theoretically, the molecule will have zero kinetic energy at its potential energy minimum. The potential energy gradient, or first derivative, is brought to zero through atomic positioning and repositioning through a variety of algorithms. Minimization algorithms differentiate the potential energy function with respect to each of the molecule's x, y, and z atomic coordinates to achieve a gradient very close to zero. The various minimization methods traditionally used in simulations may differ in their use of first and second derivatives in optimizing a molecule's structure.

Steepest descents is used to quickly descend toward the global minimum potential energy. Convergence takes place under the steepest descents method by successively adjusting the molecular coordinates in the negative direction of the first derivative of the potential energy function.

The size of each successive step is determined by the immediately prior change in potential energy. A reduction in potential energy dictates a larger step size; an increase in potential energy dictates a smaller step size. Convergence is difficult to achieve using steepest descents minimization alone and is regularly achieved through a combination of methods. In general, steepest descents will "rapidly improve a very poor conformation." Additionally, a large number of iterations are generally required.
The method of conjugate gradients uses both prior and current gradients to determine the size of successive steps. Mathematically, convergence is achieved in N steps for a quadratic energy surface where N is the number of degrees of freedom in the energy.

The adopted basis Newton-Raphson minimization method iteratively solves the Newton-Raphson minimization equations by computing a derivative of the gradient using a second derivative matrix diagonalization to find an optimum step size along eigenvectors. Small or negative eigenvalues are subjected to additional energy and gradient determinations to set necessary step size and direction. Rapid convergence is often achieved through this algorithm. The adopted basis method avoids saddle points on the energy surface by numerically constructing the second derivative matrix from gradient vector changes.\(^9\)

Minimization procedures also allow normal mode analysis. The equation for a molecular potential surface of \(n\) atoms can be used to examine the vibrations around a particular minimum by transforming the potential energy expressions from Cartesian coordinate representations to normal coordinates to calculate the vibrational frequency of a normal mode. Modes with frequencies corresponding to intramolecular vibrations, molecular translation, and molecular rotation as well as harmonic normal modes can be used to describe molecular properties such as average thermal atomic motion.\(^{10}\)

The potential energy gradient of the complex was brought to 1.53 kcal per mole through the use of steepest descents, conjugate gradients, and adopted-basis Newton-
Raphson minimization techniques. The minimized structure of the hammerhead ribozyme complex is shown in Figure 6.

**Solvent Representation**

The use of a suitable solvation model is very important in macromolecular simulations as longtime dynamics of such molecules in solution are "governed by diffusion." The relevant interaction is between solute atoms and solvent atoms—the free energy of hydration—which produces conformation and thermodynamic properties. As stated by one researcher, "the enzyme plus medium must be regarded as a unit."

Among the roles of solvent molecules in simulations are [a] improvement in macromolecular packing; [b] reduction of "sizeable" cavities through minimization of free surface area due to surface tension effects; [c] satisfying unmatched hydrogen-bond relationships through hydrogen-bonding capacity, [d] exertion of shielding effect on electrical interaction between charges or dipoles in the protein or nucleic acid yielding solvent-polarity dependent changes in structure and stability and [e] viscosity of solvent.

Several methods of solvation were contemplated for the solvation of the molecule-ion complex during the simulation: full-scale immersion into a box of water molecules with implementation of periodic boundary conditions resulting in a "unit" of tiled water boxes or cells and the use of a distance-dependent dielectric constant. The use of periodic boundary conditions in a solvent simulation permits the incorporation of fewer water molecules in a simulation with no change in results from using bulk water. The Rahman method treats a rectangular box of water molecules as a period system—a constant number of neighboring
water molecules is maintained both inside and outside the cell. The departure of one water molecule from one exit is followed by the entry of the same water from the opposite direction.\textsuperscript{16}

Both methods were viewed with an eye toward reducing the computational cost of the simulations. The large number of water molecules and amount of computational time required in a periodic boundary condition water box or bulk solvent approach prompted the use of a distance-dependent dielectric constant.\textsuperscript{17}

Dynamics

After minimization, two sets of dynamics simulations were initiated by Newtonian and Langevin dynamics, using the Verlet\textsuperscript{18} algorithm, to perform initialization of atomic positions and velocities, heating and equilibration for the simulations. In the Newtonian case, heating to the desired temperatures (300K and 500K) as well as the subsequent equilibration, is achieved by scaling the initial velocities. The heating process begins at 0 K and the temperature was rapidly increased incrementally over several picoseconds. In the Langevin case, heating was accomplished through the use of a temperature bath at 300K or 500K. All molecular complex atoms came within the influence of the heatbath. The permissible temperature deviation from the final temperature for all phases of the simulations was plus or minus 10 degrees. Violation of these boundaries triggered atomic velocity scaling protocols.

Equilibration, after heating to the desired temperature, was accomplished over time in the simulations.\textsuperscript{19} The complex at the different temperatures, and under the different
methods, was subjected to a 600 picosecond molecular dynamics simulation with 2 femtosecond intervals between steps and a history file written every ten steps.

Initialization of the molecule-ion complex is accomplished by setting the initial positions, \( r_i(0) \), of the atoms to their respective X-ray, minimized, or previous molecular dynamics, for the ions, coordinates with time \( t = 0 \). The initial velocities are assigned through random number generation.\(^{20}\) The simulations were run on Silicon Graphics Indigo2 Unix Workstations under the IRIX 5.3 operating system. A total of four simulations of the molecule-ion complex were conducted: Newtonian and Langevin at 300 and 500 degrees Kelvin, respectively.

A distance-dependent dielectric constant was incorporated into the simulation. The steps of production dynamics were run and trajectory data, in the form of conformation ensembles, was collected for analysis. The analysis items include conformation-dependent (internal motions)\(^{21}\) and thermodynamic properties. The conformation-dependent properties include the radius of gyration, hydrogen bonding and dipole moment. The thermodynamic properties include temperature, potential energy, kinetic energy, and total energy. The properties are plotted as a function of time over the 600 picosecond simulations and the trajectories at each temperature are analyzed in terms of local and large-scale structural and global energy changes. Additional analysis is conducted with respect to the root-mean-square displacement of backbone (phosphate) and side-chain (base) atoms. Actual global relaxation is expected to occur on a time scale longer than that of the present simulations.
Figure 6. Minimized Structure showing Mg$^{2+}$ ions (small squares).
CHARMm is an acronym for Chemistry at Harvard-Macromolecular Mechanics. See Brooks, B.R., Brucoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., and Karplus, M. (1983). CHARMm: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. Journal of Computational Chemistry 187. Representation of objects in Cartesian space (x,y,z coordinates) is substituted for a CHARMm internal coordinate system in which CHARMm utilizes a Cartesian coordinate representation for the first 3 coordinates and bond angle, bond distance/length, and rotational (torsional) angle for the fourth coordinate. The potential energy function ("force field") is manipulated in terms of the internal coordinates.


CHARMm documentation, MINIMIZ.DOC (1991) at 4.


CHARMm documentation, MINIMIZ.DOC (1991) at 5.


14 van Gunsteren et al. (1994) at 852. See also discussion of boundaries, long-range electrostatic effects and the approximate treatment of solvent effects (853-856).


16 Daggett, V and Levitt, M. (1993). Realistic Simulations of Native-Protein Dynamics in Solution and Beyond. 22 Annu. Rev. Biophys. Biomol. Struct. at 358. Such an approach may also be necessary to avoid “edge” effects in the case of a small number of molecules where the “fraction near the boundary of the system is far greater than a real system would have.” Winn, J. (1995). Physical Chemistry 651.


18 The Verlet method is a finite-difference, third-order Störmer algorithm, used to determine positions and velocities. The addition of two Taylor expansion series of position from time $t$ forward to time $t + \Delta t$ and from time $t$ backward to time $t - \Delta t$, respectively, produces the Verlet algorithm for positions.

Forward expansion series:
$$x(t + \Delta t) = x(t) + (dx(t)/dt)(\Delta t) + \frac{1}{2} (d^2x(t)/dt^2)\Delta t^2 + \frac{1}{3!} (d^3x(t)/dt^3)\Delta t^3 + O(\Delta t^4)$$

Backward expansion series:
$$x(t - \Delta t) = x(t) - (dx(t)/dt)(\Delta t) + \frac{1}{2} (d^2x(t)/dt^2)\Delta t^2 - \frac{1}{3!} (d^3x(t)/dt^3)\Delta t^3 + O(\Delta t^4)$$

Verlet algorithm for positions:
$$x(t + \Delta t) = 2x(t) - x(t - \Delta t) + (d^2x(t)/dt^2)\Delta t^2 + O(\Delta t^4)$$

Estimation of velocities is accomplished through the use of a first-order central difference estimator, $v(t) = [x(t + \Delta t) - x(t - \Delta t)]/2\Delta t$. Acceleration is derived from Newton’s second law and intermolecular forces. Haile, J. (1992) at 158-159 and references cited therein, including Verlet, L. (1967). Computer Experiments on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. 159 Phys. Rev. 98.; Beeman,


If $x_x$, $x_y$, and $x_z$ are random numbers distributed uniformly over the interval from -1 to +1, the corresponding initial velocity coordinates are given by $v_{ix}(0) = dx_x / x$, $v_{iy}(0) = dx_y / x$, and $v_{iz}(0) = dx_z / x$. $x = [dx_x^2 + dx_y^2 + dx_z^2]^{1/2}$ Haile, J. (1992) at 202.

The analysis of the trajectory data generated in the simulations must be carefully considered. Thermodynamic (temperature, kinetic energy, potential energy, total energy) and conformational (radius of gyration, hydrogen bonds, dipole moment) variable data were collected for analysis. Each trajectory was arbitrarily separated into two phases. The first phase, Phase I, consisted of the time period from 0 to 300 ps; Phase II covered the time period from 301 to 600 ps.

The selected properties were chosen for their value in contributing to an overall characterization of the biomolecular system. The temperature of a system is directly related to the kinetic energy by

\[ \sum_{i=1}^{N_a} \frac{m_i v_i^2}{2} = \frac{N_f k_B T}{2} \]

where \( N_a \) is the number of atoms, \( m_i \) is the mass of atom \( i \), \( v_i \) is the velocity of atom \( i \), and \( N_f \) is the number of degrees of freedom. The radius of gyration reflects the root mean square displacement of the atoms from center of mass of the molecule. The number of hydrogen bonds formed through intramolecular base pairing is an indication of the relative degree of structure of the molecule. In the case of a group of distributed charges, a dipole moment exists which may be defined by
Equation 17
\[ \mu = \sum_i q_i r_i \]
or in the case of a continuous distribution of charge,

Equation 18
\[ \mu = \int \rho(r) r \, dV \]

where \( \rho(r) \) is the charge density at a given position \( r \). "Any system of charges (even if the system has a zero net charge) whose center of positive charge does not coincide with the center of negative charge will have a dipole moment."² The dipole moment is essentially an indication of the alignment of the molecule's electrical field. The dielectric constant is a measure of how much a particular substance will reduce the electric field.

Two major time frames are relevant to the analysis of the dynamics: (a) the pre-equilibration phase where heating is accomplished and equilibration occurs and (b) the post-equilibration "true" dynamics period where meaningful conclusions may be drawn regarding the molecule's search of conformational space.

The graphs on the following pages show the evolution of the properties over Phase I and Phase II. The first 50 ps of each simulation represent the results of a combined heating and equilibration protocol within QUANTA. For temperature, equilibration appears to be reached in all cases (Figure 8(a)-(f)). For energy, in the Newtonian simulations, equilibration is reached by 50 ps (Figure 9). However, in the case of the Langevin method, equilibration does not appear to have occurred until 200 ps (Figure 9). By Phase II, energy
equilibration has been reached in all simulations. The results for the radius of gyration are similar to those of the energy. In the case of hydrogen bonds, Langevin simulations (Figure 11) equilibrated more quickly than Newtonian (Figure 11) but both are fully equilibrated by 50 ps at 300K and 500K in Phase II. The dipole moment equilibrated rapidly in the Newtonian simulations at both 300K and 500K (Figure 12), but showed at least two major transitions in the Langevin simulations before equilibrium positions were reached (Figure 12).

The decrease in potential energy and contemporaneous increase in hydrogen bonding (Figure 11) during the early phases of dynamics indicate that molecular dynamics functions as a very effective energy minimization routine. This may be due to the greater conformational space being explored in the dynamics simulation as compared to the energy minimization procedure. Overall, the properties equilibrated quickly and remained relatively constant during the simulations. These observations were maintained regardless of the dynamics method or the temperature utilized. In particular, the radius of gyration and dipole moment showed the most change between dynamics methods. The radius of gyration took a longer time to equilibrate or "settle down" and achieved higher values in the Langevin simulations than in the Newtonian simulations. The dipole moment required a longer time to reach a "steady state" value in the Langevin dynamics overall and showed an increase in value in Phase II dynamics as compared to a decreasing dipole moment returned in Newtonian Phase II dynamics. Additionally, over the long run, Langevin dynamics appears to have produced conformations with somewhat less structure than Newtonian
dynamics as evidenced by the formation of fewer hydrogen bonds and possibly greater radius of gyration in the Langevin case. No major conformational changes were observed during any of the simulations; however, detailed structural analysis (RMSD), in the case of the 500K Newtonian dynamics simulation, showed variation in flexibility for different parts of the molecule.

Comparison of Simulation Trajectories

Table 1 presents the results for Phase I and Phase II dynamics. In general, the Langevin simulation maintained a slightly higher average potential energy in both phases, a correspondingly slightly higher total energy in both phases, and a significantly higher dipole moment in both phases at both temperatures. As required, the kinetic energy increased in the 500K simulations compared to the 300K ones, but there was a decrease in the potential energies at the higher temperatures. This appears to be due to the breaking of some non-covalent interactions. For example, in both the Newtonian and Langevin simulations, the number of hydrogen bonds observed decreased at the higher temperature. For the Newtonian, but not the Langevin, simulations, the radius of gyration increased at the higher temperature, possibly indicating fewer van der Waals interactions since the molecular mass would be distributed over a larger volume.
Table 1. Comparison of Average Molecular Properties Between Newtonian and Langevin Dynamics Simulations at 300K and 500K

<table>
<thead>
<tr>
<th>SIMULATION</th>
<th>PE (kcal/mole)</th>
<th>KE (kcal/mole)</th>
<th>TE (kcal/mole)</th>
<th>TEMP (K)</th>
<th>HB (A)</th>
<th>RG (A)</th>
<th>DM (Debye)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ND 300(I)</td>
<td>-6105.35</td>
<td>1081.14</td>
<td>-5024.20</td>
<td>306.34</td>
<td>51.85</td>
<td>13.45</td>
<td>167.36</td>
</tr>
<tr>
<td>LD 300 (I)</td>
<td>-5425.18</td>
<td>1052.74</td>
<td>-4372.44</td>
<td>297.78</td>
<td>53.18</td>
<td>14.77</td>
<td>448.91</td>
</tr>
<tr>
<td>ND 300 (II)</td>
<td>-6188.09</td>
<td>1095.95</td>
<td>-5092.14</td>
<td>310.53</td>
<td>53.48</td>
<td>13.21</td>
<td>150.01</td>
</tr>
<tr>
<td>LD 300 (II)</td>
<td>-5886.77</td>
<td>1053.28</td>
<td>-4833.49</td>
<td>297.94</td>
<td>57.31</td>
<td>14.37</td>
<td>364.89</td>
</tr>
<tr>
<td>ND 500 (I)</td>
<td>-5454.14</td>
<td>1793.24</td>
<td>-3660.89</td>
<td>508.11</td>
<td>44.56</td>
<td>14.71</td>
<td>300.01</td>
</tr>
<tr>
<td>LD 500 (I)</td>
<td>-4927.57</td>
<td>1754.03</td>
<td>-3173.54</td>
<td>496.16</td>
<td>44.25</td>
<td>14.34</td>
<td>425.55</td>
</tr>
<tr>
<td>ND 500 (II)</td>
<td>-5538.52</td>
<td>1848.36</td>
<td>-3690.15</td>
<td>523.73</td>
<td>48.59</td>
<td>14.22</td>
<td>300.58</td>
</tr>
<tr>
<td>LD 500 (II)</td>
<td>-5403.34</td>
<td>1754.45</td>
<td>-3648.88</td>
<td>496.26</td>
<td>47.12</td>
<td>13.43</td>
<td>342.81</td>
</tr>
</tbody>
</table>

Symbols: ND (Newtonian Dynamics), LD (Langevin Dynamics), PE (Potential Energy, kcal/mole), KE (Kinetic Energy, kcal/mole), TE (Total Energy, kcal/mole), TEMP (Temperature K), HB (Hydrogen Bonds), RG (Radius of Gyration, A), DM (Dipole Moment, Debyes)

For higher level analysis, including RMSD analysis, the 500K Newtonian dynamics simulation was chosen because the Newtonian simulations appear to reach equilibrium sooner than the Langevin ones, allowing more points for analysis in the production phase, and because the 500K simulation would explore more conformational space than the 300K ones.

Final Structures

The pictures in Figure 13 present the final structures of the hammerhead ribozyme complex at 600 ps.

In-Line Cleavage Mechanism

The hammerhead ribozyme cleavage reaction is believed to proceed via an “in-line” mechanism whereby the O5’ and P atoms of the adenine residue of the scissile bond and the
O2' atom of the cytosine residue on the other side of the scissile bond align themselves into a linear or 180 degree pattern. The 500K Newtonian dynamics simulation was analyzed to determine whether the atoms had so aligned themselves during the simulation. Figure 14 shows the achieved angle as a function of time for the Phase I and Phase II dynamics, respectively.

As shown, the angular value stays far from the 180 degree value necessitated by the "in-line" mechanism. This may be an artifact of the substitution of the 2'-OH of the cytosine residue with a 2'O-methyl group to prevent cleavage during crystallization. Although the methyl group was invisible during resolution, it may be that a "bad" angle was adopted upon substituting the hydroxyl group for the final structure. The Mg$^{2+}$ ions may be in different places in the simulation than would be required for the transition state for the proposed "in-line" mechanism. Finally, it may be that the simulation needs to be much larger to observe this intermediate.

RMSD Evaluation

As observed in one article, "[t]he most direct way to assess the stability of a [molecular dynamics] simulation over the course of time is the evaluation of the difference between the initial experimental coordinates and the generated structures, measured by the [root-mean-square difference] RMSD." RMSD analysis provides a meaningful method of assessing molecular conformational changes which arise during the course of a simulation. In the context of biological macromolecules, two major structural domains are generally
examined: the main chain or backbone and the side chains. With respect to nucleic acids, the main chain comprises the sugar-phosphate backbone and the side chains comprise the bases.

Two sets of calculations were performed using CHARMM; examples of the Fortran programs appear in Appendix A. The first set of calculations used the final structure of the ribozyme complex at 600 ps (Newtonian 500K simulation) as a reference structure with respect to which RMSD calculations for the 550 ps dynamics trajectory were produced. The trajectory comprised a concatenation of every 100th structure generated during the dynamics. These calculations permit an analysis of the structural changes which took place in the molecular system as it evolved to the final structure.

The second set of calculations used the average structure of the ribozyme complex during the 550 ps dynamics period as a reference structure with respect to which RMSD calculations for the 550 ps dynamics trajectory were produced. These calculations permit an assessment of the overall flexibility of the molecule during dynamics. Both calculations focused on the two major structural elements of biological macromolecules: the backbone (phosphate atoms) and the side-chains (using a single average value for the nucleoside or base atoms).

Figure 13 presents the minimized and average ribozyme structures and the structures of the ribozyme at various points during the dynamics, including the final
structure at 600 ps. The base-pairing representations are particularly useful in representing the degree of hydrogen bonding and overall structural disorder which occurred during the simulations.

RMSD Calculations with respect to the Final Structure

Phosphate Backbone Changes

The phosphate backbone is examined in Figure 18 which presents three-dimensional plots of phosphate atom versus RMSD with respect to the final structure versus time in the form of rotated representations to more clearly visualize the dynamical changes. Additionally presented are two-dimensional plots of the average RMSD of the phosphate atoms over time versus phosphate atom position which collectively show general trends in the conformational changes of the ribozyme backbone.

The relatively higher magnitude changes in backbone atomic positions are summarized in Table 2. The average RMSD values were 0.727 for Phase I and 0.547 for Phase II. Larger motions are associated with the terminal or near-terminal, in the case of the Guanine 18 phosphate, atoms, especially during the Phase I dynamics and with respect to the enzyme strand. As time progresses, these movements are reduced and are replaced by larger movements of the Stem 3 phosphate atoms, including some of the conserved core phosphate atoms.
Table 2. Backbone (Phosphate Atom) Evolution towards Final Structure

<table>
<thead>
<tr>
<th>Residue</th>
<th>Chain</th>
<th>Atom</th>
<th>Phase I RMSD</th>
<th>Phase II RMSD</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gl3</td>
<td>A</td>
<td>P</td>
<td>1.04</td>
<td>1.46</td>
<td>Core</td>
</tr>
<tr>
<td>Gl4</td>
<td>A</td>
<td>P</td>
<td>0.91</td>
<td>1.28</td>
<td>Near-Terminus</td>
</tr>
<tr>
<td>Cl6</td>
<td>A</td>
<td>P</td>
<td>1.37</td>
<td>0.69</td>
<td>Terminus</td>
</tr>
<tr>
<td>Gl8</td>
<td>B</td>
<td>P</td>
<td>2.65</td>
<td>1.30</td>
<td>Near-Terminus</td>
</tr>
<tr>
<td>A24</td>
<td>B</td>
<td>P</td>
<td>1.03</td>
<td>1.32</td>
<td>Core</td>
</tr>
<tr>
<td>G32</td>
<td>B</td>
<td>P</td>
<td>1.47</td>
<td>0.58</td>
<td>Stem 3</td>
</tr>
<tr>
<td>A33</td>
<td>B</td>
<td>P</td>
<td>1.64</td>
<td>0.88</td>
<td>Stem 3</td>
</tr>
</tbody>
</table>

Base (Side-chain) Evolution

Figure 18 presents the RMSD of residue atoms in two general formats. First, the RMSD of the residues are presented as a function of residue position and time. Second, the average RMSD of all the atoms comprising an individual residue is plotted versus residue position. One particular residue, the guanine 17 of the substrate strand, shows the most movement in comparison to the other residues during Phase I dynamics; this behavior, however, is smoothed out in the Phase II dynamics suggesting perhaps that the initial larger movements may be due to full equilibration not having occurred as rapidly as originally thought. Not surprisingly, the generated RMSD values for the residues are greater than those of the phosphate backbone atoms. The average RMSD values were 0.940 for Phase I and 0.795 for Phase II.

In general, the guanine and cytosine residues showed larger deviations than the adenosine and cytosine residues. Of the 10 base pairs in the original structure, 7 were guanine-cytosine base pairs. These base pair residues are located in general at the termini
of the molecule and may be more largely affected due to these positions. Further analysis of the breakdown in hydrogen bonding structure is necessary.

The following table summarizes the higher observed values for RMSD of base atoms. An average value for the atoms comprising each base was used for purposes of calculation and presentation. The movement of the bases seem to follow the same pattern as the backbone phosphate atoms, i.e. greater motions at the terminii in Phase I and greater motions within the core region in Phase II.

Table 3. Side-Chain (Nucleotide Base) Evolution towards Final Structure

<table>
<thead>
<tr>
<th>Residue</th>
<th>Chain</th>
<th>Phase I RMSD</th>
<th>Phase II RMSD</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>C6</td>
<td>A</td>
<td>2.06</td>
<td>1.87</td>
<td>Near-Terminus</td>
</tr>
<tr>
<td>C16</td>
<td>A</td>
<td>2.69</td>
<td>2.38</td>
<td>Terminus</td>
</tr>
<tr>
<td>G17</td>
<td>B</td>
<td>3.16</td>
<td>2.45</td>
<td>Terminus</td>
</tr>
<tr>
<td>G18</td>
<td>B</td>
<td>1.73</td>
<td>0.85</td>
<td>Near-Terminus</td>
</tr>
<tr>
<td>A24</td>
<td>B</td>
<td>1.09</td>
<td>1.41</td>
<td>Core</td>
</tr>
<tr>
<td>C25</td>
<td>B</td>
<td>0.99</td>
<td>1.30</td>
<td>Core</td>
</tr>
<tr>
<td>U26</td>
<td>B</td>
<td>1.02</td>
<td>1.35</td>
<td>Core</td>
</tr>
<tr>
<td>C41</td>
<td>B</td>
<td>2.74</td>
<td>0.70</td>
<td>Terminus</td>
</tr>
</tbody>
</table>

RMSD Calculations with respect to the Average Structure

Phosphate Backbone Flexibility

The phosphate backbone is examined in Figure 19 which presents three-dimensional plots of phosphate atom versus RMSD with respect to the average structure versus time, including rotated representations to more clearly visualize the dynamical changes. Additionally presented are two-dimensional plots of the average RMSD of the
phosphate atoms over time versus phosphate atoms position which show general trends in
the evolution of the ribozyme backbone towards the final structure. The values associated
with the most predominant peaks in the graphs are summarized in Table 4. The average
RMSD values were 0.393 for Phase I and 0.405 for Phase II. During the designated Phase I
dynamics, greater movement or flexibility was associated with the phosphate atoms of the
substrate strand (Chain B) as compared with the phosphate atoms of the enzyme strand
(Chain A). The near-terminus atoms in both strands showed the most flexibility in Phase I.
This behavior decreased with respect to these atoms in Phase II and was replaced, generally,
with higher motions associated with the core structure and Stem 3 atoms. The phosphate
atom associated with one of the nucleotides of the scissile bond showed significant
movement during Phase I and Phase II dynamics as well.

<table>
<thead>
<tr>
<th>Residue</th>
<th>Chain</th>
<th>Atom</th>
<th>Phase I RMSD</th>
<th>Phase II RMSD</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>G14</td>
<td>A</td>
<td>P</td>
<td>0.74</td>
<td>0.69</td>
<td>Near-Terminus</td>
</tr>
<tr>
<td>G18</td>
<td>B</td>
<td>P</td>
<td>1.28</td>
<td>1.07</td>
<td>Near-Terminus</td>
</tr>
<tr>
<td>C20</td>
<td>B</td>
<td>P</td>
<td>0.48</td>
<td>0.36</td>
<td>Near-Terminus</td>
</tr>
<tr>
<td>G21</td>
<td>B</td>
<td>P</td>
<td>0.42</td>
<td>0.56</td>
<td>Near-Terminus</td>
</tr>
<tr>
<td>A24</td>
<td>B</td>
<td>P</td>
<td>0.41</td>
<td>0.46</td>
<td>Core</td>
</tr>
<tr>
<td>C25</td>
<td>B</td>
<td>P</td>
<td>0.44</td>
<td>0.56</td>
<td>Core</td>
</tr>
<tr>
<td>G32</td>
<td>B</td>
<td>P</td>
<td>0.56</td>
<td>0.53</td>
<td>Stem 3</td>
</tr>
<tr>
<td>A33</td>
<td>B</td>
<td>P</td>
<td>0.42</td>
<td>0.58</td>
<td>Stem 3</td>
</tr>
<tr>
<td>A37</td>
<td>B</td>
<td>P</td>
<td>0.48</td>
<td>0.46</td>
<td>Scissile Bond</td>
</tr>
</tbody>
</table>

Base (Side-chain) Flexibility

Figure 19 shows changes in flexibility of the ribozyme structure with respect to the
average structure over the 550 ps production dynamics. The results are summarized for the
predominant peaks in Table 5. The average RMSD values were 0.381 for Phase I and 0.560 for Phase II. The bases showed significant flexibility as compared to the average structure during Phase I and Phase II. The greater motions are localized to residues located in the near-terminus, terminus, core, and Stem 3 areas. The values for these motions increased through Phase I and Phase II. This observation is in contrast to the flexibility of the phosphate atoms which generally evidence a decrease in flexibility as time progresses.

A strong correlation exists with respect to the average RMSD values for the backbone and base atoms in both Phases I and II (Figure 7). This suggests that the molecule is moving through conformational space as a concerted unit.

Table 5. Side-Chain (Base) Flexibility with respect to Average Structure

<table>
<thead>
<tr>
<th>Residue</th>
<th>Chain</th>
<th>Phase I RMSD</th>
<th>Phase II RMSD</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>C6</td>
<td>A</td>
<td>0.39</td>
<td>0.67</td>
<td>Near-Terminus</td>
</tr>
<tr>
<td>A9</td>
<td>A</td>
<td>0.55</td>
<td>0.80</td>
<td></td>
</tr>
<tr>
<td>G14</td>
<td>A</td>
<td>0.68</td>
<td>0.80</td>
<td>Near-Terminus</td>
</tr>
<tr>
<td>C16</td>
<td>A</td>
<td>0.65</td>
<td>0.82</td>
<td>Terminus</td>
</tr>
<tr>
<td>G17</td>
<td>B</td>
<td>0.91</td>
<td>1.16</td>
<td>Terminus</td>
</tr>
<tr>
<td>G18</td>
<td>B</td>
<td>0.99</td>
<td>1.27</td>
<td>Near-Terminus</td>
</tr>
<tr>
<td>C20</td>
<td>B</td>
<td>0.52</td>
<td>0.70</td>
<td>Near-Terminus</td>
</tr>
<tr>
<td>A24</td>
<td>B</td>
<td>0.37</td>
<td>0.70</td>
<td>Core</td>
</tr>
<tr>
<td>C25</td>
<td>B</td>
<td>0.39</td>
<td>0.66</td>
<td>Core</td>
</tr>
<tr>
<td>G32</td>
<td>B</td>
<td>0.52</td>
<td>0.74</td>
<td>Stem 3</td>
</tr>
<tr>
<td>A33</td>
<td>B</td>
<td>0.33</td>
<td>0.61</td>
<td>Stem 3</td>
</tr>
<tr>
<td>C41</td>
<td>B</td>
<td>0.43</td>
<td>0.54</td>
<td>Terminus</td>
</tr>
</tbody>
</table>
Figure 7. Backbone Side-Chain Correlation
Figure 8. Temperature versus Time
(a) 300K Newtonian Dynamics Phase I
(b) 300K Langevin Dynamics Phase I
(c) 500K Newtonian Dynamics Phase I
(d) 500K Langevin Dynamics Phase I
(e) 300K Newtonian Dynamics Phase II
(f) 300K Langevin Dynamics Phase II
(g) 500K Newtonian Dynamics Phase II
(h) 500K Langevin Dynamics Phase II
Figure 9. Potential Energy, Kinetic Energy, and Total Energy versus Time.
(a) 300K Newtonian Dynamics Phase I
(b) 300K Langevin Dynamics Phase I
(c) 500K Newtonian Dynamics Phase I
(d) 500K Langevin Dynamics Phase I
(e) 300K Newtonian Dynamics Phase II
(f) 300K Langevin Dynamics Phase II
(g) 500K Newtonian Dynamics Phase II
(h) 500K Langevin Dynamics Phase II
Figure 10. Radius of Gyration versus Time
(a) 300K Newtonian Dynamics Phase I
(b) 300K Langevin Dynamics Phase I
(c) 500K Newtonian Dynamics Phase I
(d) 500K Langevin Dynamics Phase I
(e) 300K Newtonian Dynamics Phase II
(f) 300K Langevin Dynamics Phase II
(g) 500K Newtonian Dynamics Phase II
(h) 500K Langevin Dynamics Phase II
Figure 11. Hydrogen Bonds versus Time
(a) 300K Newtonian Dynamics Phase I
(b) 300K Langevin Dynamics Phase I
(c) 500K Newtonian Dynamics Phase I
(d) 500K Langevin Dynamics Phase I
(e) 300K Newtonian Dynamics Phase II
(f) 300K Langevin Dynamics Phase II
(g) 500K Newtonian Dynamics Phase II
(h) 500K Langevin Dynamics Phase II
Figure 12. Dipole Moment versus Time
(a) 300K Newtonian Dynamics Phase I
(b) 300K Langevin Dynamics Phase I
(c) 500K Newtonian Dynamics Phase I
(d) 500K Langevin Dynamics Phase I
(e) 300K Newtonian Dynamics Phase II
(f) 300K Langevin Dynamics Phase II
(g) 500K Newtonian Dynamics Phase II
(h) 500K Langevin Dynamics Phase II
Figure 13. Final Structures at 600 ps
(a) 300K Newtonian
(b) 300K Langevin
(c) 500K Newtonian
(d) 500K Langevin
Figure 14. O2'-P-O5' Bond Angle versus Time:
(a) Phase I Dynamics  (b) Phase II Dynamics.
Figure 15. Minimized ribozyme structure: (a) ribbon and (b) ladder views. Average Ribozyme structure over 550 ps dynamics: (c) ribbon and (d) ladder views.
Figure 16. Average Structure of Ribozyme showing Mg$^{2+}$ ions
Figure 17. Ribbon and ladder views of ribozyme

(a) Ribozyme at 50 ps
(b) Ribozyme at 100 ps
(c) Ribozyme at 150 ps
(d) Ribozyme at 250 ps
(e) Ribozyme at 350 ps
(f) Ribozyme at 450 ps
(g) Ribozyme at 550 ps
(h) Ribozyme at 600 ps
Figure 18. RMSD with respect to Final Structure
(a) Phosphate Atoms (Rotated View)
(b) Average RMSD of Phosphate Atoms Over Time versus Position
(c) RMSD of Phosphate Atoms versus Time (Rotated View).
(d) RMSD of Residues versus Time (Rotated View)
(e) Average RMSD of Residues over Time
(f) Average RMSD of Cytosine Residues Over Time versus Position
(g) Average RMSD of Guanine Residues Over Time versus Position
(h) Average RMSD of Adenine Residues Over Time versus Position
(i) Average RMSD of Uracil Residues Over Time versus Position
(e) RMSD of residues (Phase IIa)

(f) Average RMSD over time vs. residue position (Phase I)

(g) Average RMSD of cytosine residues over time vs. position (Phase I)

(h) Average RMSD of guanine residues over time vs. position (Phase II)
(i) 

**Average RMSD of Adenine Residues Over Time vs. Position**

- Graph showing the average RMSD of adenine residues over time.
- Position range: 1-40.

(j) 

**Average RMSD of Uracil Residues Over Time vs. Position**

- Graph showing the average RMSD of uracil residues over time.
- Position range: 1-40.
Figure 19. RMSD with respect to Average Structure
(a) RMSD of Phosphate Atoms versus Position versus Time
(b) Average RMSD of Phosphate Atoms versus Time
(c) RMSD of Residues versus Position versus Time
(d) Average RMSD of Residues versus Time.
(a) RMSD of phosphate atoms vs. position vs. time (Phase I)

(b) Average RMSD of phosphate atoms vs. time (Phase I)

(c) RMSD of residues vs. position vs. time (Phase I)

(d) Average RMSD of residues vs. time (Phase I)


CHAPTER 10

CONCLUSION

The preliminary examination of the molecular dynamics of the hammerhead ribozyme under differing dynamic procedures and at different temperatures showed no large-scale or global conformational changes from the minimized structure throughout the 600 ps simulations or the presence of significant intermediates. The thermodynamic and conformational properties behaved as expected. Rapid equilibration was followed by a steady dynamics trajectory in each case, but the Newtonian simulations achieved equilibrium before the Langevin ones. The 500K Newtonian dynamics simulation, chosen for further RMSD analysis with respect to the final and average structures, showed localized movement in the form of slight variations in flexibility of the backbone (phosphate atoms) and the side-chains (bases).

As expected, the simulations were generally characterized by increased movement among the terminal phosphate atoms of the enzyme and substrate strands during the first phase; this behavior settles down and is replaced by increasing side-chain, conserved core, and Stem 3 movement during Phase II. The enzymatic strand showed more movement during the Phase I; this may be due to the shorter length, and decreased mass, of the enzyme strand as compared with the substrate strand. The phosphate atom associated with one of the nucleotides of the scissile bond showed significant movement during Phase I and Phase II dynamics as well.
The bases showed significant flexibility as compared to the average structure and final structure during Phase I and Phase II. The greater motions are localized to residues located in the near-terminus, terminus, core, and Stem 3 areas. The values for these motions increased through Phase I and Phase II. This observation is in contrast to the flexibility of the phosphate atoms where there is generally a decrease in flexibility as time progresses.

Additional analysis may reveal more meaningful patterns in the data including the possibility that a molten globule-like state was encountered. However, more relevant events may only be observable through a longer time frame dynamics simulation.
CHARMm (Fortran) program for RMSD calculations

* RMSD - rmsd calculations for hammerhead ribozyme simulations
*
UPPER ! case for files to write
bomblevel -5
wrnlev 0
prnlev 5
! Script to read parameter, psf, and ic files
! Generate the system using the information from Quanta
open read unit 21 card name $CHM_DATA/MASSES.RTF
read rtf unit 21 card close unit 21
open read unit 22 card name "charmmppm"
read param unit 22 card! close unit 22
open read unit 23 card name "charmppsf"
read psf unit 23 card
open read unit 24 card name "charmmic"
ic read unit 24 card
goto (color)

label yellow
!---------------------------------------------------
! Concatenation of Trajectory Files
! CHARMM versions differ in treatment of file headers
! Merge multiple trajectory files using traj and iread
open read unit 61 file name A.DCD
open read unit 62 file name B.DCD
open read unit 63 file name C.DCD
open read unit 64 file name D.DCD
open read unit 65 file name E.DCD
open read unit 66 file name F.DCD
open read unit 67 file name G.DCD
open write unit 68 file name COMPLETE.DCD
trajectory iread 61 nread 7 iwrite 68
set a 1
label start
traj read
traj write
incr a by 100
if a lt 27500 goto start
close unit 23
close unit 24
close unit 68
close unit 61
close unit 62
close unit 63
close unit 64
close unit 65
close unit 66
close unit 67

label red
!
! Calculate the Average Structure over the trajectory COMPLETE.DCD
! 1. Cartesian Coordinates
open read unit 32 file name COMPLETE.DCD
coor dyna pax firs 32 nuni 1 skip 10 begin 50010 stop 52750
close unit 32

! 2. View PAX analysis results
open read unit 32 file name COMPLETE.DCD
coor paxa firs 32 nuni 1 skip 10 begin 50010 stop 52750
close unit 32

! 3. List the rms fluctuations and write average coordinates
scal wmai show
print coor
open write unit 33 card name AVERAGE.CRD
write coor card unit 33
* 550 ps dynamics
* average coordinates for COMPLETE.DCD
*
close unit 32
close unit 33

! 4. Compute IC values to get dihedral angle references
ic fill
open read unit 24 card name ".charmmic"
read ic card unit 24

! 5. Compute and print IC averages
open read unit 32 file name COMPLETE.DCD
ic dyna aver firs 32 nuni 1 skip 10 begin 50010 stop 52750
print ic
close unit 32

! 6. Compute and print IC fluctuations about the average
open read unit 32 file name COMPLETE.DCD
ic dyna fluc firs 32 nuni 1 skip 10 begin 50010 stop 52750
print ic
close unit 32
close unit 24

label green
!--------------------------------------------------------------------------------------------------------------------
! RMSD calculations per atom - one coordinate structure set at a time
! open file for writing rmsd data sets with only one atom per residue
open write unit 57 card name RMSDP.CRD
set 2 275.5
! Fill comparison coordinate set with average coordinates
open read unit 44 card name AVERAGE.CRD
read coor card unit 44
upda cutnb 15.0 ctonnb 11.0 ctofnb 14.0 wmin 0.8
ener rdie
coor copy comp
open unit 43 read file name COMPLETE.DCD
trajectory iread 43
set 9 1
label loop
traj read
coor orient rms
coor rms sele all end
coor diff
coor dist weigh
scalar vmain average by res sele all end
print coor rms sele type P card unit 57 end
print coor rms card unit 57 sele type P end
incr 9 by 1
if 9 lt @2 goto loop

The orient command is for coordinate orientation such that the principle geometric axis coincides with the x-axis, and the next largest coincides with the y-axis. The structure is oriented about its center of geometry.

RMS keyword overlaps the main coordinates with those in the comparison set.

close unit 43

close unit 44

label purple

RMSD calculations per residue- one coordinate structure set at a time

set 2 275.5

Fill comparison coordinate set with average coordinates

open write unit 63 card name RESIDUE.CRD
set 2 275.5
open read unit 44 card name AVERAGE.CRD
read coor card unit 44
upda cutnb 15.0 ctonnb 11.0 ctofnb 14.0 wmin 0.8
ener rdie
coor copy comp
open unit 43 read file name COMPLETE.DCD
trajectory iread 43
set 9 1
label loop
traj read
coor orient rms
coor rms sele all end
coor diff
coor dist weigh
scalar wmain average by res sele all end
print coor rms sele type O5' end
incr 9 by 1
if 9 It @2 goto loop
close unit 43
close unit 44
stop
CHARMm (Fortran) programs for data manipulation

C Program table
real R(300,1500)
integer I,J
open (unit=1, file ='xl.out')
open (unit=2, file='xa.out')
Do 100 I=1,275
Do 100, J=1,39
100 read (1,10,end=300) R(I,J)
10 Format (63X,F7.5)
10 Format (63X,F7.5)
Do 200 J=1,39
200 Write (2,20) (R(I,J),I=1,125)
20 Format (125(1X,F7.5))
close (unit=1)
close (unit=2)
end

C Program table
real R(300,1500)
integer I,J
open (unit=1, file ='xl.out')
open (unit=2, file='xb.out')
Do 100 I=1,275
Do 100, J=1,39
100 read (1,10,end=300) R(I,J)
10 Format (63X,F7.5)
300 Continue
Do 200 J=1,39
200 Write (2,20) (R(I,J),I=126,275)
20 Format (150(1X,F7.5))
close (unit=1)
close (unit=2)
end

Sample Data Lines from Coordinate File

<p>| | | | | | | | | | | | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>862</td>
<td>28</td>
<td>G</td>
<td>O2P</td>
<td>20.38689</td>
<td>-9.18638</td>
<td>35.77438</td>
<td>B</td>
<td>31</td>
<td>0.27038</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>863</td>
<td>28</td>
<td>G</td>
<td>O5'</td>
<td>20.49061</td>
<td>-10.86409</td>
<td>37.65214</td>
<td>B</td>
<td>31</td>
<td>0.29958</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>864</td>
<td>28</td>
<td>G</td>
<td>C5'</td>
<td>21.65806</td>
<td>-11.12931</td>
<td>38.39593</td>
<td>B</td>
<td>31</td>
<td>0.34141</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>865</td>
<td>28</td>
<td>G</td>
<td>C4'</td>
<td>21.39165</td>
<td>-11.89022</td>
<td>39.69059</td>
<td>B</td>
<td>31</td>
<td>0.31036</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
BIBLIOGRAPHY


Kuwajima, K. (1989). The molten globules state as a clue for understanding the folding and cooperativity of globular protein structure. 6 Proteins 87-103.


Perutz, M. Protein Structure. (San Francisco: W.H. Freeman and Company).


VITA

Vasant Gandhi was born in Chicago, IL on July 27, 1962. He received undergraduate, law, and graduate taxation degrees from the University of Illinois at Urbana-Champaign. He entered Loyola University Chicago as a graduate student in the Department of Chemistry in the Fall semester of 1994.
The thesis submitted by Vasant T. Gandhi has been read and approved by the following committee:

Dr. Kenneth W. Olsen, Ph.D., Director
Professor and Chairman, Chemistry
Loyola University of Chicago

Dr. David S. Crumrine, Ph.D.
Associate Professor, Chemistry
Loyola University of Chicago

Dr. Willetta Greene-Johnson, Ph.D.
Assistant Professor, Chemistry
Loyola University of Chicago

The final copies have been examined by the director of the thesis and the signature which appears below verifies the fact that any necessary changes have been incorporated and that the thesis is now given final approval by the committee with reference to content and form.

The thesis is therefore accepted in partial fulfillment of the requirements for the degree of Master of Science.

12/9/96

Date

Kenneth W. Olsen

Research Director