

2010

Molecular Dynamics Simulations of Human Quadruplexes in K⁺ Solution and the Role of the Carboxyl Terminus of Bacteriorhodopsin in Protein Stability

Farah Rezae
Seton Hall University

Follow this and additional works at: <https://scholarship.shu.edu/dissertations>

 Part of the [Physical Chemistry Commons](#)

Recommended Citation

Rezae, Farah, "Molecular Dynamics Simulations of Human Quadruplexes in K⁺ Solution and the Role of the Carboxyl Terminus of Bacteriorhodopsin in Protein Stability" (2010). *Seton Hall University Dissertations and Theses (ETDs)*. 1303.
<https://scholarship.shu.edu/dissertations/1303>

**Molecular Dynamics Simulations of Human
Quadruplexes in K^+ Solution and the Role of the
Carboxyl Terminus of Bacteriorhodopsin in Protein
Stability**

*This Dissertation is submitted to the faculty of Department of Chemistry and
Biochemistry in the School of Arts and Sciences in Partial fulfillment of requirements for*

The degree of

Doctor of Philosophy

At

Seton Hall University
400 South Orange Avenue
South Orange, New Jersey 07079

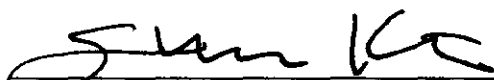
Farah Rezae

August 2010

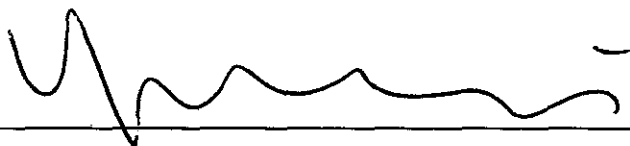
CERTIFICATION

We certify that we have read this thesis and that is in our opinion adequate in scientific scope and quality as a dissertation for the degree of Doctor of Philosophy.


Approved



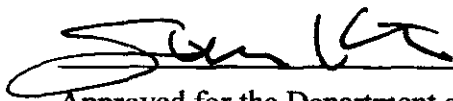
Research Advisor
Dr. Stephen P. Kelty



Member of Dissertation Committee
Dr. Yufeng Wei



Member of Dissertation Committee
Dr. Roger L. McMullen



Approved for the Department of Chemistry and Biochemistry
Chairman, department of Chemistry and Biochemistry
Dr. Stephen P. Kelty

Dedication

To My Family

Acknowledgement

This research project would not have been possible without the support of many people I would like to express my appreciation to Prof. Stephen Kelty, who was abundantly helpful and presented invaluable assistance, support and supervision. Deepest gratitude is also due to the members of the supervisory committee, Dr. Yufeng Wei and Dr. Roger McMullen. Thanks also to all my graduate friends, especially group members Mou and Jainmin. I would also like to convey thanks to the Chemistry and Biochemistry department of the Seton Hall University for providing me with the financial means and laboratory facilities. I wish to express love and appreciation to my dearly loved family for their thoughtfulness and continuous love through the length of my studies.

Table of Contents

<i>Dedication</i>	<i>iii</i>
<i>Acknowledgement</i>	<i>iv</i>
<i>List of Figures</i>	<i>viii</i>
<i>List of Tables</i>	<i>x</i>
<i>Abstract</i>	<i>x</i>
CHAPTER I	1
Molecular Dynamics simulations of Guanine Quadruplexes	1
1.1 Introduction	2
1.1.1 Nucleic acids.....	2
1.1.2 Quadruplexes.....	4
1.1.2.1 Quadruplex structures in the human genotype.....	6
1.1.2.1.1 Telomeres.....	6
1.1.2.1.2 Non -telomeric Quadruplexes.....	9
1.1.2.2. Physical Properties of Quadruplexes.....	12
1.1.2.2.1 Structural Properties.....	12
1.1.2.2.2 Cation Induced Stability.....	13
1.1.2.2.3 Molecularity and Sequence Effects.....	14
1.1.2.2.4 Strand Polarity.....	14
1.1.2.2.5 Base Orientation.....	15
1.2 Experimental Methods for Structure Determination of Quadruplexes	15
1.2.1 Motivation for the Present study.....	16
1.2.2 Objectives for our Research.....	17
1.2.3 Methodology for Quadruplex Simulations.....	17
1.3 Results and Discussion	19
1.3.1 Average Total Energy.....	21
1.3.2 Average SASA and relative stability.....	21
1.3.3 Determination of ΔC_p	24
1.3.4 Root -Mean- Square -Deviations of Different models of Quadruplexes.....	26
1.4 Concluding Remarks	27
CHAPTER II	29
Molecular Dynamics Simulations of the Bacteriorhodopsin Membrane Protein . 29	
2.1. Introduction	30
2.1.1 Membrane Proteins.....	30
2.1.2 Membrane Proteins and Lipid Bilayer.....	31
2.1.3 Alpha Helical Arrangement of Trans-membrane Proteins.....	33
2.1.4 Bacteriorhodopsin.....	34

2.1.4.1 Major differences between helices of Bacteriorhodopsin	36
2.1.4.2 Bacteriorhodopsin CT	37
2.1.4.3 bR Photocycle	37
2.1.4.4 Experimental Studies of the CT of bR	40
2.2 Objectives of Our Research	41
2.3 Methodology for simulation of the CT of bR	42
2.3.1 Methodology for Simulation of Different Models of bR	44
2.3.2 Building a Structural Model of Bacteriorhodopsin	45
2.3.3 Simulation of Bacteriorhodopsin	46
2.4 Results and Discussion	47
2.4.1 CT in solvent	47
2.4.1.1 Structure analysis	47
2.4.1.2 H-bonding	50
2.4.1.3 RMSD and Diffusion constant analysis	51
2.4.1.4 SASA analysis	51
2.4.2 Analysis of bR without CT	52
2.4.2.1 Structure analysis	52
2.4.2.2 Electrostatic Energies	53
2.4.2.3 Hydrogen Bonding	54
2.4.2.4 Salt Bridges	54
2.4.2.5 SASA analysis	54
2.4.3 Analysis of bR with the CT in Proximity (bR (CT))	54
2.4.3.1 Structure analysis	55
2.4.3.2 Electrostatic Energies	58
2.4.3.3 Hydrogen Bonding	59
2.4.3.4 Salt Bridges	60
2.4.3.5 SASA analysis	61
2.4.3.6 Heat Capacity	62
2.4.4 Analysis of bR with bound CT (bR-CT)	63
2.4.4.1 Structure analysis	63
2.4.4.2 Electrostatic Interactions	65
2.4.4.3 Hydrogen Bonding	66
2.4.4.4 Salt Bridges	67
2.4.4.5 SASA analysis	68
2.4.4.6 Heat Capacity	69
2.5 Concluding Remarks	70
Appendix I	72
I. Molecular Dynamics Simulations	73
I.A Potential Energy Functions	76
i. Intra-molecular Forcefield	76
ii. CHARMM Potential Energy Function	76
I.B Ensemble Types	78
I.C Controlling Pressure and Temperature in Ensembles	79
I.D Periodic Boundaries	80

I.E General Scheme for Running Molecular Dynamics Simulations	81
i. Initiation.....	81
ii. Equilibration	81
iii. Production phase.....	82
iv. Standard units	83
<i>Appendix II</i>	84
II. Analysis Methods	85
A. Simulation properties.....	85
i . Root-mean-square-displacement (RMSD)	85
ii. Diffusion Constant.....	85
iii. Interaction energies.....	85
iv. Ramachandran plots.....	86
v. H-Bonds	87
vi. Solvent accessible surface area (SASA).....	87
vii. Heat Capacity.....	88
viii. Salt Bridges.....	90
<i>REFERENCES</i>	91

List of Figures and flow chart

FIGURE 1: STRUCTURE OF DNA.....	3
FIGURE 2: MAJOR BASES IN DNA	4
FIGURE 3: G-TETRAD.....	6
FIGURE 4: DIFFERENT FORMS OF FOLDING OF HUMAN QUADRUPLEXES.....	13
FIGURE 5: <i>ANTI</i> AND <i>SYN</i> CONFORMATIONS IN HYBRID QUADRUPLEXES.....	15
FIGURE 6: SNAPSHOT OF PARALLEL STRANDED QUADRUPLEX (1KF1) IN 10NS OF SIMULATION.....	19
FIGURE 7: SNAPSHOT OF HYBRID I (2GKU) IN 10 NS SIMULATION.....	19
FIGURE 8: SNAPSHOT OF HYBRID II (2JPZ) IN 10NS SIMULATION.....	20
FIGURE 9: SNAPSHOT OF UNFOLDED QUADRUPLEX IN 10 NS SIMULATIONS	20
FIGURE 10: AVERAGE TOTAL ENERGIES FOR DIFFERENT MODELS OF QUADRUPLEXES	21
FIGURE 11: SASA OF DIFFERENT MODELS OF QUADRUPLEXES	22
FIGURE 12: DIFFERENT FORMS OF MEMBRANE PROTEINS.....	32
FIGURE 13: G COUPLED RECEPTOR PROTEIN	33
FIGURE 14: STRUCTURE OF BACTERIORHODOPSIN IN THE GROUND STATE.....	35
FIGURE 15: PURPLE MEMBRANE OF BACTERIORHODOPSIN	36
FIGURE 16: AMINO ACIDS OF C TERMINUS OF BACTERIORHODOPSIN.	37
FIGURE 17: PHOTOCYCLE OF BR.	39
FIGURE 18: CT OF BR	43
FIGURE 19: BR WITHOUT CARBOXYL TERMINUS	44
FIGURE 20: BR WITH CT IN PROXIMITY.....	45
FIGURE 21: BR WITH CT BOUND.....	45
FIGURE 22: SNAPSHOT OF TRAJECTORIES OF CT OF BR IN 1 AND 4M SALT CONCENTRATION IN 10 NS SIMULATION	48
FIGURE 23: RAMACHANDRAN PLOT OF CT OF BR IN 1M(A) AND 4M(B) SALT CONCENTRATION IN 10 NS SIMULATION	49
FIGURE 24: HYDROGEN BONDING OF CT IN 1 AND 4M SALT CONCENTRATION IN 10NS SIMULATION.....	50
FIGURE 25: AVERAGE RMSD PER RESIDUE FOR THE CT OF BR	51
FIGURE 26: SASA OF THE CT OF BR IN DIFFERENT SALT CONCENTRATIONS	52
FIGURE 27: RAMACHANDRAN PLOT OF BR WITHOUT CT.....	53
FIGURE 28: ELECTROSTATIC ENERGY BETWEEN MEMBRANE LIPIDS AND BR WITHOUT THE CT	53
FIGURE 29: SASA OF BR WITHOUT THE CT.....	54
FIGURE 30: RAMACHANDRAN PLOT OF BR AND THE CT IN PROXIMITY.....	56
FIGURE 31: RAMACHANDRAN PLOT OF BR WITHOUT THE CT IN BR AND THE CT IN PROXIMITY	57
FIGURE 32: RAMACHANDRAN PLOT OF THE CT IN BR AND THE CT IN PROXIMITY	58
FIGURE 33: ELECTROSTATIC ENERGIES BETWEEN MEMBRANE LIPIDS AND BR WITH CT IN PROXIMITY AND BR WITHOUT THE CT	59
FIGURE 34: AVERAGE SASA OF BR IN THE CT WITH PROXIMITY AND BR WITH NO CT.....	62
FIGURE 35: RAMACHANDRAN PLOT OF BR AND CT BOUND.....	64

FIGURE 36: RAMACHANDRAN PLOT OF BR WITHOUT CT IN BR AND CT BOUND MODEL ...	64
FIGURE 37: RAMACHANDRAN PLOT OF CT IN BR AND THE CT BOUND MODEL.....	65
FIGURE 38: ELECTROSTATIC ENERGIES BETWEEN MEMBRANE LIPIDS AND BR WITH CT ATTACHED	66
FIGURE 39: AVERAGE SASA OF BR IN THE PRESENCE AND ABSENCE OF CT	69
FIGURE I.1: PERIODIC BOUNDARY CONDITIONS	80
FLOWCHART I: GENERAL FLOWCHART FOR DESIGNING MOLECULAR DYNAMICS SIMULATIONS	83
FIGURE II.1: ALPHA AND PSI ANGLE IN PROTEIN	86
FIGURE II.2: A RAMACHANDRAN PLOT OF A PROTEIN	87
FIGURE II.3 CALCULATING SOLVENT ACCESSIBLE SURFACE AREA.....	88

List of Tables

- I. Average total energies and average solvent accessible surface areas of different models of quadruplexes in 10 ns of simulations (Page 25)**
- II. SASA of hydrophilic and hydrophobic part of different models of human telomere quadruplexes (Page 26)**
- III. Calculated ΔC_p for different isomers of human telomere quadruplexes in (Cal mol/k) in 10ns of simulations (Page 27)**
- IV. Average RMSD per Residue for different models of human telomere quadruplexes (Page 29)**
- V. # of H-bonds in bR and CT in proximity (Page 63)**
- VI. Comparison of number, location and type of salt bridges in bR alone and with the CT in proximity (Page 64)**
- VII. Average SASA for hydrophobic and hydrophilic residues and calculated ΔC_p during 10ns of simulations in 4M NaCl for bR without CT and bR with CT in proximity (Page 65)**
- VIII. Number of H-bonds in studied systems (Page 69)**
- IX. Salt bridges between CT and bR (Page 71)**
- X. Average SASA for hydrophobic and hydrophilic Residues and calculated ΔC_p for bR and bR-CT (Page 72)**
- XXI. Different kinds of ensembles (page 82)**

Abstract

Classical molecular dynamics (MD) simulations were used to study the relative stability among several known human G- quadruplex structures, find the equilibrium structure of Carboxyl Terminus (CT) of Bacteriorhodopsin (bR) and observe the interactions of CT in equilibrium with bR.

In the case of quadruplex DNA, recent experimental results indicate that the (TTAGGG)₄ quadruplex undergoes isomerization to an intermediate state upon melting in potassium solution. To help establish the likely structural state of this intermediate, we have undertaken MD simulations to determine the relative stability of several candidate isomers which differ in the secondary folding structure. Total energy, solvent accessible surface area, heat capacity and root-mean-square-displacements of the nucleic acid portion of the simulation ensemble indicate that the proposed initial state suggested in previously published experimental results is correct and that an intermediate state might exist.

Bacteriorhodopsin (bR) exists as a 7-helix trans-membrane protein with a 23-residue sequence C-terminus (CT) exposed to the aqueous intracellular environment. The presence of the CT has been shown to contribute to the overall stability and function of bR; however, to date XRD crystal data indicate that the CT lacks a well-defined equilibrium structure. All-atom MD simulations of the CT in aqueous ionic solution have been undertaken to help establish the likely structure of the CT as a function of solvent salt concentration. High temperature simulated annealing was used to mitigate initial state effects as well as to promote sampling of a broader space. All MD simulations included explicit water, a host lipid membrane matrix and solvent ions. The solvent

accessibility and root-mean-square-deviations of the CT in equilibrium are presented, which show that the overall stability of the transmembrane portion of bR is enhanced by interactions with the CT. The investigations include a comparison of CT-bR interactions with the CT bound and unbound to bR.

CHAPTER I

Molecular Dynamics simulations of Guanine Quadruplexes

1.1 Introduction

1.1.1 Nucleic acids

The most commonly recognized function of a cell's genotype is to store the code required to obtain the protein phenotype and to pass on this code along to successive generations.¹ The process of protein generation involves translating the double stranded DNA into a complementary single strand RNA that is transcribed by a ribosome into the corresponding protein.

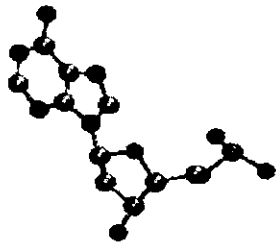
In 1953, Watson and Crick discovered the first of several three-dimensional structures of DNA.¹ Their model is a double helix with the base pairs A (Adenine)-T (Thymine) or C (Cytosine)-G (Guanine) forming the interior "steps" of a ladder and the deoxyribose-phosphate "backbone" forming the external "rungs" of the ladder (Figure 1). Watson and Crick found that the hydrogen bonded base pairs between G and C and A and T, are those that fit best within the structure.² It is important to note that three hydrogen bonds can form between G and C, but only two can form between A and T. This structure (termed the "B" form), which is the most common form under physiological conditions, is a right-handed double helix with the base pairs approximately normal to the helical axis. An "A" form can occur under conditions of low solvation in which the base pairs are canted by about 20 degrees to the helix axis. Other duplex forms of DNA have been identified but occur only under strained conditions. In all these structures the comparatively hydrophobic purine and pyrimidine bases are found in the interior of the polymer with the polar sugar-phosphate structures exteriorly located in contact with the aqueous solvent.

The structural composition of a nucleotide includes three basic components: nitrogenous base, pentose sugar, and phosphate. The structure of B DNA is shown in Figure 1.

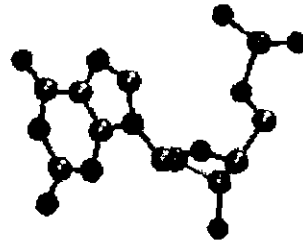


Figure 1: Structure of DNA

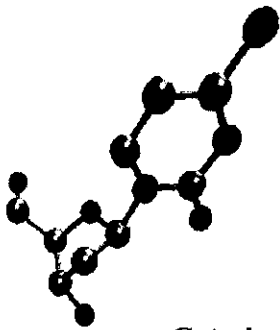
A nucleotide lacking phosphate group is named a nucleoside.¹ The nitrogenous bases belong to one of the two groups, either pyrimidine or purine. The bases and pentose sugar groups of the ordinary nucleotides are heterocyclic complexes. Both DNA and RNA contain two main purine bases (A) and (G). The pyrimidines are (C) and (T) in DNA and (C) and (U) Uracil in RNA. The structures of the four major nucleotides of DNA are shown in Figure 2.



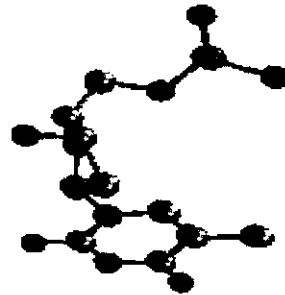
a: Adenine



b: Guanine



c: Cytosine



d: Thymine

Figure 2: Major bases in DNA

Nucleic acids have two kinds of pentoses. The recurring deoxyribonucleotide units of DNA contain deoxy-D-ribose and the ribonucleotide units of RNA contain D-ribose.

An important structural aspect of duplex DNA is that the complementary strands of the duplex run antiparallel in the sense of the standard numbering scheme of the deoxyribose subunit as shown in Figure 1.

1.1.2 Quadruplexes

In regions of a DNA sequence that contain a high fraction of Guanine residues, an alternate structural form of DNA can occur.³ These structures contain a quadruplex of

guanine residues as shown in Figure 3. The H-bonding arrangement contains strained elements not found in duplex DNA base pairs and is stabilized by the formation of several such Quadruplexes stacked together and by the presence of cations (usually Na^+ or K^+) in the tetrad center. In general, at least 2-3 such Quadruplexes are needed to stabilize the structure. The structures have been found to be more prominent in particular areas of the sequence such as toward the ends of the DNA molecule. They have been implicated in the aging process in as much that they are believed to help maintain the structural integrity of the genotype. These sequences have been observed in important areas of eukaryotic DNA such as telomeres and the regulatory segments of genes.⁴ Because pronounced formation of G-quadruplexes has been observed in tumors, anti-cancer therapeutic strategies based on G-quadruplex-drug binding is an exceedingly active focus of investigation. Hence, it is essential to fully understand the structural characteristics of G-quadruplexes if rational drug design targeting these structures is to be successful.⁵

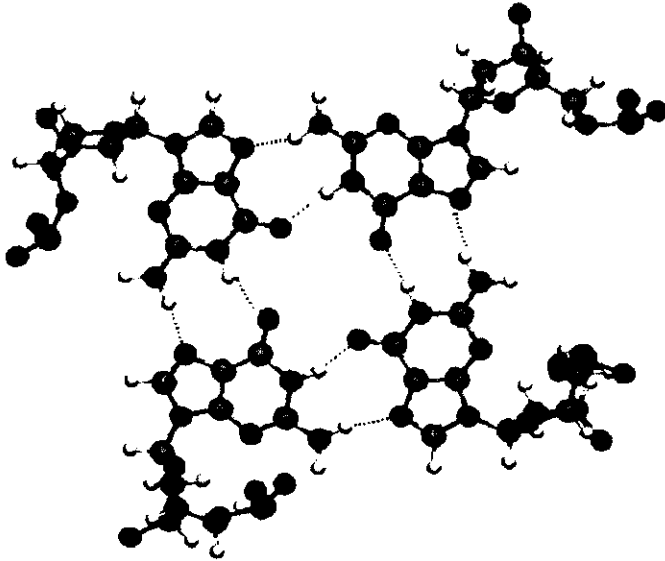


Figure 3: G-Tetrad

1.1.2.1 Quadruplex structures in the human genotype

Numerous areas in the human genome can assume G-quadruplex structures. These consist of telomeric ends, and non-telomeric sections including: immunoglobulin switch regions, mutational hot spots, and regulatory elements with oncogene promoters.⁶ Some of the salient aspects of these regions are described below.

1.1.2.1.1 Telomeres

Telomeric DNA in human cells, which is normally is 5–8 kilobases (kb) in length, was shown to contain replicates of the sequence (TTAGGG) with a single-stranded overhang of 100–200 bases.⁷ In Na⁺ solution, NMR studies indicate an antiparallel basket-type G-quadruplex structure that has diagonal and lateral TTA loops. In K⁺ solution, a propeller-type G-quadruplex was observed with Guanine columns in a uniform all-parallel design.⁸ Since the conformation in K⁺ solution is believed to be physiologically more significant, several structural studies of the telomeric G-quadruplex

in K^+ solution were completed and a number of additional structures were discovered which are not in agreement with the XRD crystal structure.⁹⁻¹⁴

The conformation observed for G-quadruplexes in K^+ solution can be further modified. For example, with platinum based cross-linking methods the basket-type structure can be formed with other quadruplex sequences in Na^+ and K^+ solution.⁹ ¹²⁵I-radioprobng results indicate that a chair-type structure is the predominant species in K^+ solution.¹⁰ More recently, sedimentation and fluorescence studies have shown that the XRD-determined basket-form crystal conformation of telomeric DNA is not the predominant species in K^+ solution and a variety of other energetically equivalent forms have been observed.¹¹ A mixture of chair-type and parallel/antiparallel hybrid structures was reported in K^+ solution.¹² Antiparallel/parallel strands with one propeller and two lateral loops in K^+ solution (hybrid-kind) were shown in at least two NMR studies.¹³⁻¹⁴ It was recently shown that human telomeric DNA can form all-parallel stranded G-quadruplexes in K^+ solution.¹⁵ Furthermore, other studies have suggested a dense stacking conformation for multimers containing a hybrid-kind and parallel-type telomeric DNA.

Telomeres are non-coding DNA (i.e., they do not contribute to the phenotype) found at the termini of linear chromosomes and are capable of forming defensive structures at these regions. In this capacity, telomers help prevent DNA from forming double-strand breaks which could result in degradation.⁷ Strict control of telomere length is vital for cell cycle control, cellular immortalization, and tumor genesis.¹⁶ Maintaining telomere length prevents the initiation of p53 and Rb tumor suppressor pathways and cellular senescence that would lead to tumor suppression.¹⁷ Telomere-induced senescence was shown to be successful in inducing apoptosis aimed at diminishing cancer

incidence,¹⁸⁻¹⁹ by inhibiting oncogene-expressing cells from progressing to malignancy.²⁰ Normal human somatic cells display weak telomerase activity, which is inadequate to keep up a constant telomere length, whereas more than 90% of human tumor cell populations have high telomerase activity and show a range of telomere lengths.²¹⁻²²

The folding and stabilization of a variety of G-quadruplexes in G-rich sequences at the end of telomeres may influence telomere length and therefore the normal regulation of telomeres in the cell cycle or other events. The effects of stabilization of these structures by small ligands have been broadly investigated over the past decade. The interactions of G-quadruplexes with ligands augment the various effects on telomere functions. For example, many small ligands capable of stabilizing G-quadruplexes are successful in telomerase inhibition, which is the most common parameter for evaluating the telomeric G-quadruplex ligands.²³⁻²⁶ Other reported effects of ligands binding to G-quadruplexes are shortening of telomere length,²⁷ induction of senescence and inhibition of cell growth.²⁸⁻³⁰ Furthermore, end-to-end fusions of chromosomes have been observed with telomeres in the presence of G-quadruplex ligands.³¹ Recently, some G-quadruplex ligands have been found to interfere with the conformation and length of the telomeric G-overhang.³² Other studies have also shown that some G-quadruplex ligands may function by dissociation of the telomere binding proteins POT1 (protection of telomeres1) and TRF2 (telomeric repeat binding factor 2), and by uncapping telomeres to make them available for extension.³²⁻³⁵

In addition, the biological function of telomeres is expected to be dependant on the variety of their diverse conformational states.¹²

1.1.2.1.2 Non -telomeric Quadruplexes

Quadruplexes were observed in double-stranded non-telomeric parts of the human genome such as in promoter and immunoglobulin switch regions and in recombination hot spot.²³ Two groups investigated the presence of quadruplex-forming sequences in the human genome³⁶⁻³⁷ and such sequences are found throughout the genome.

There is significant interest in searching for the presence of quadruplex sequences in promoter sequences due to the elevated level of G bases.³⁸ The observance of quadruplex sequences in the promoter areas of several cancer genes has motivated several structural studies. A variety of possible quadruplex sequences, which are typically located exactly upstream of the start codon site, in a series of cancer related genes are reported.³⁹

The hypothesis that quadruplex creation may supply a transcriptional regulatory signal was motivated by studies of quadruplex levels in *Escherichia coli* and other prokaryotic genomes,⁴⁰ where G₄ sequences are frequently seen in promoter sections close to the transcription start codon. Quadruplex sequences in mammalian pre-mRNA sequences were also been investigated.⁴¹ At present, little experimental information on RNA quadruplexes is available. It was suggested that the fragile X mental retardation protein (FMRP) has great affinity for G-rich mRNA which has the ability to make quadruplexes.⁴² A survey was done on all 16,654 genes in the human genome database which showed that there is a correlation between the frequency of quadruplex structures and the gene class.⁴³ Specifically, proto-oncogenes have an elevated potential for making quadruplexes but tumor suppressor genes have considerably lower potential.

In addition, potential quadruplex-forming sequences were observed in chromosomal translocations. One case for this phenomenon is the breakpoint section on

human chromosome 14.⁴⁴ The section immediately downstream of the breakpoint has G-rich tracts, which are capable of quadruplex formation. Similar G-tracts have been seen in the breakpoint section of the SHANK3 gene.⁴⁵

Non-telomeric quadruplex sequences are linked with abnormally high cell growth. Their biological role is not fully understood but they may take part in the gene expression management. Non-telomeric quadruplex structures have diversity both in topology and molecular conformation⁴⁶⁻⁴⁷ as described below.

1.1.2.1.2.1 Topology and Structure of non-Telomeric Sequences

Quadruplex development has been studied in vitro in a number of non-telomeric sequences. The NHE III1 G-rich sequence in the promoter region of the C-myc oncogene, which is responsible for 80–90% of transcriptional action, has been particularly well studied. A quadruplex in this promoter area was reported.²³⁻²⁴ Later studies recognized a connection among quadruplex stabilization in this sequence and suppression of C-myc transcriptional activities.⁴⁸ Solution phase NMR studies were used to establish the topology and full structures of several C-myc quadruplex sequences.²³⁻²⁴ Non-telomeric G-rich regions frequently have more than four uninterrupted G₄-tracts. The conformational rearrangements are different between human telomere and C-myc quadruplexes. Myc-2345 and Myc-1245 each have four G₄-tracts and form very stable quadruplexes in solution.⁴⁹ These G₄-tracts join through propeller loops, and all Guanines in the G₄ tracts have an *anti* conformation.⁴⁹ Hence, these structures have similar properties to the human telomeric crystal structure.⁸

NMR methods were used to demonstrate the existence of quadruplexes in the promoter region of the C-kit kinase gene.⁵⁰⁻⁵¹ Interestingly, NMR studies show just a

single quadruplex species in K^+ solution.⁵¹ Subsequently, another quadruplex sequence was recognized in the promoter region of the C-kit gene.⁵⁰ Both of these quadruplexes are highly preserved between vertebrate species suggesting a practical function for them.

Chemical foot printing and CD techniques were applied to distinguish quadruplex species observed in a nuclease hypersensitive sequence within the vascular endothelial growth factor (VEGF) promoter section which is necessary for basal promoter activity in human cancer cells.⁵² A quadruplex was also observed in a sequence in the hypoxia-inducible factor 1 α (HIF-1A) promoter region, based on foot printing and CD data.⁵³ This sequence has two possible single nucleotide loops. The bcl-2 oncogene has a main transcriptional promoter sequence upstream of the transcription start codon that has been identified as having quadruplex features.⁵⁴⁻⁵⁵ The suggested loop sequence in this quadruplex section has the same sequence as one of the C-kit sequences⁵¹ and this sequence was estimated to have a high incidence of occurrence.³⁷

NMR studies of the bcl-2 quadruplex⁵⁴ show that one of the topologies for this mixed parallel/anti-parallel quadruplex has two side loops and one propeller loop which is similar to one of the telomeric quadruplex topologies,¹³⁻¹⁴ but with a inverted loop order. Potential quadruplexes have also been suggested in the k-ras⁵⁶ and neuroblastoma oncogenes.⁵⁷

A recent NMR structural study of the G-quadruplex from the C-myc promoter sequence with five G₄-tracts⁵⁸ demonstrates that the existence of a fifth G₄-tract can permit surprising and important variation in topology compared to similar sequences with only four G₄-tracts.

1.1.2.2. Physical Properties of Quadruplexes

1.1.2.2.1 Structural Properties

The formation of G-quadruplexes from G-rich sequences can result in several folding patterns. This study focuses on three-layer tetrads, so discussion is limited to forms of the type (GGG)₄. The tetrads can be formed from 1-4 separate strands, which may run parallel or antiparallel along adjacent strands. Furthermore, the target materials in this study are all single strand type formed from the sequence TTAGGGTTAGGGTTAGGGTTAGGG (thereafter, (TTAGGG)₄). Basic combinatorial principles show that starting from the first TTAGGG section; the next section can bind either parallel or antiparallel in one of two adjacent corners of the tetrad or the opposite corner (2 x 3 = 6 total). The third section may bind in one of the two remaining corners (2 x 2 = 4 total) and the final section must occupy the remaining corner (2 total). This makes 6 x 4 x 2 or $\sum_{i=1}^3 2^i = 48$ possible structural states. Most of these states are far too unstable to represent a significant fraction of a physiological population and therefore are discounted. In fact, only a few have ever been experimentally observed.^{8,14,59} These studies are intended to help elucidate the results of an experimental study, which were conducted in K⁺ solution. One of the published structures was done in Na⁺ solution and is only included in these simulations for comparison. The other structures, described in greater detail below, were all found for K⁺ solutions and so are reasonable candidates for possible structures in Figure 4.

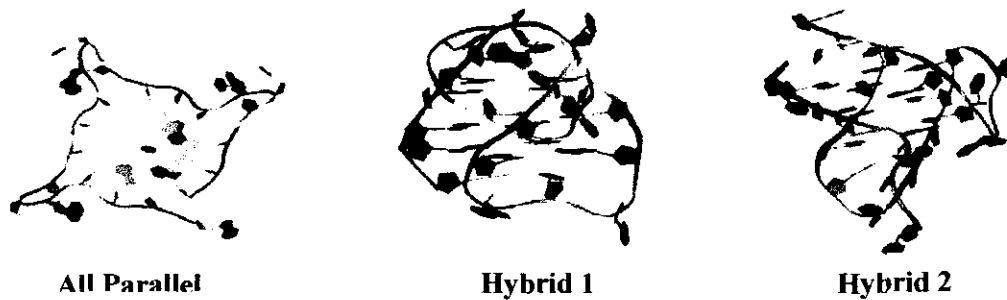


Figure 4: Different forms of folding of human quadruplexes

Beyond the essential requirement that quadruplex structures can only form in G-rich regions of the genome, a wide variety of stable structures have been observed. These structures are stabilized by the presence of cations located at the center of the tetrad. Other stabilizing characteristics include the G base orientation, the loop sequence, and the polarity (5' → 3') of the backbone in adjacent strands. The effects of these constraints impose considerable structural polymorphism compared to duplex DNA. Moreover, the local environment can also affect the stability of the quadruplexes. In particular, solvation, the presence of metal ions, ligands, or molecular crowding may also control the topology of quadruplexes. Below, these effects are examined in more detail.

1.1.2.2.2 Cation Induced Stability

G-quadruplexes are generally stabilized by the presence of cations, particularly K^+ and Na^+ . These cations are centered between layers with a typical stoichiometry of one ion per tetrad. The bounding tetrads may have ions but are typically coordinated with solvent ions.^{8,23,24,60} The center region among G-tetrad is well suited to coordinating cations of this ionic size since the two planes of adjacent tetrads are lined up with eight carbonyl O-6 atoms, which generate a central negatively charged channel inside the G-tetrad stack.

1.1.2.2.3 Molecularity and Sequence Effects

G-quadruplexes can be folded from a single G-rich sequence unimolecularly or by the intermolecular association of two (dimeric) or four (tetrameric) separate strands. Formation of a unimolecular quadruplex would necessarily impose greater constraints of the loop sequence and environmental effects than would dimeric or tetrameric quadruplexes since the loop region in the latter are more flexible.

1.1.2.2.4 Strand Polarity

Strand polarity can also give rise to structural polymorphism. For example, the polarities of the four strands in a G-quadruplex can be parallel (all 5' → 3'), three parallel and one antiparallel, adjacent parallel, or alternating parallel, resulting in different conformations. These structure types are termed all parallel, propeller (for mixed polarity), and antiparallel quadruplexes respectively.^{8,23,24,60}

Variations in strand polarity also affect the location of the loops between G-rich segments. For example, parallel G-strands need a linking loop to connect the bottom of one G-sequence with the top of the adjacent sequence, leading to propeller-type loops. Antiparallel G-strands can be associated by diagonal or lateral edgewise loops, depending on whether the strands are nearby or diagonally opposed. Overall, the sequence and dimensions of the loops help determine the topology of quadruplexes. In addition, the loop residues can form stacking and hydrogen bonding interactions further stabilizing or destabilizing G-quadruplex folds.⁶¹⁻⁶⁵ A recent molecular dynamics simulation has demonstrated that the sequences of the linking loops in quadruplexes are a key aspect in determining quadruplex flexibility and may be the possible candidate sites for drug binding.²³

1.1.2.2.5 Base Orientation

The guanine glycosidic torsion angles are another important parameter in characterizing G-quadruplexes. All parallel quadruplexes have guanine glycosidic torsion angles that are characteristic of an *anti* conformation, whereas guanine tetrads in antiparallel quadruplexes are found to adopt both *syn* and *anti* conformations (Figure5).⁶⁶

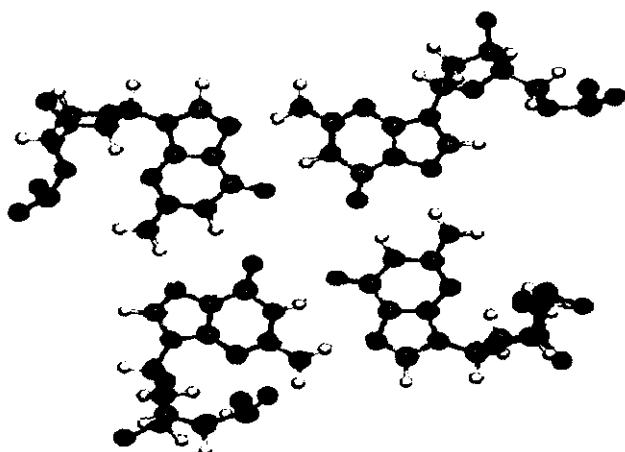


Figure 5: *Anti* and *syn* conformations in hybrid quadruplexes

1.2 Experimental Methods for Structure Determination of Quadruplexes

Circular Dichroism (CD) spectroscopy is capable of differentiating quadruplexes according to their topologies and finding dissimilarities between parallel and anti-parallel strand orientations. CD is a helpful and quick method for exploring the overall fold of a structure.³⁸ CD studies for quadruplexes have been largely used to distinguish the conditions under which parallel or antiparallel structures for a given sequence are observed.

X-ray crystallography (XRD) and high-field NMR spectroscopy provided a wealth of structural information for quadruplexes.^{8,14,54-59} Structural determinations by NMR methods require the presence of (ideally) a single thermodynamically stable

species in solution. The presence of multiple species limits the structural information that can be gained. In XRD, the species must be able to crystallize into a periodic pattern. The presence of non-periodic features in the crystal structure will not, in general, provide any structural data. Hence, the results from any NMR or crystal structure methods need to be considered carefully. Nevertheless, the availability of reliable structural data is essential for any simulation or molecular dynamics method.

1.2.1 Motivation for the Present study

The presence of telomeres in eukaryotic chromosomes is of great interest for structure-function relationships, particularly due to the observed, enhanced maintenance of these structures in cancer cells.²³ Human telomeres are known to be G-rich in (TTAGGG)₄ sequences;⁵⁹ hence, effort has been directed at determining the folding structure of this sequence under physiological conditions. Prior experimental studies by Cuesta, *et al.* and by Antonacci, *et al.* of oligomer quadruplexes with the unimolecular sequence (TTAGGG)₄ were aimed at elucidating the structural characteristics of this sequence.^{23,67} This study focused on the thermodynamic melting temperature using CD and calorimetry to determine the equilibrium state of the sequence. The authors suggested a likely equilibrium structure and the presence of an alternate intermediate folding state at higher temperature before finally melting into an unfolded final state. The key aspect of this study was to identify the low temperature (310K) ground state. In our study, we utilize these studies to hypothesize the structure of the ground and intermediate states and by using simulation methods, the likely identity of these states will be investigated.

1.2.2 Objectives for our Research

Prior thermodynamic studies on the human telomeric repeat (TTAGGG)₄ in the presence of potassium ion indicate that the transition from the initial low temperature (310K) state to the high temperature unfolded “random coil” state is accompanied by transition to an intermediate state at 310 K.⁶⁷ The nature of this intermediate state is not clearly known at present. In these studies, we utilized MD simulations to investigate the relative stability of several candidate structures. The transition from one state to another state can only occur from high stability structure to comparably lower stability structure (G_{form}) and ΔG_{form} is negative for this transition. We performed MD simulations to determine the equilibrium structure and energy in order to make comparisons among the known structures. The outcome of this study will allow the development of new drugs that specifically bind to the quadruplex once the structure is clearly identified.

1.2.3 Methodology for Quadruplex Simulations

The reader is directed to the Appendix I for background details of MD simulation techniques. In this section the simulation details, which are specific to quadruplex DNA studies are described.

The starting structures for creating the initial pdb and psf input structure and topology files were obtained from published structures for G-quadruplexes. The AGGG (TTAGGG)₃ sequence structure has been obtained using XRD structure (pdb code 1KF1⁸) and NMR structures (pdb code 2gku¹⁴, pdb code 2JPZ⁵⁴) in solution with two or three potassium ions, between the tetrads. The 1KF1 structure is an all-parallel structure. The 2gku(hybrid1) and 2JPZ (hybrid2) are hybrid structures. A schematic of the folding pattern is shown in Figure 4.

In order to provide a comparison to the energies of the structured states and an unfolded state, we prepared structure with the same primary sequence as the others but with no specific folding pattern.

In all cases, the primary sequence was the same except for the starting bases. In order to make comparisons of the same structure, we edited out any dissimilar bases so that our simulated structures all had the (TTAGGG)₄ sequence.

From the edited structure files, pdb and psf input files were prepared which contained approximately 3,800 explicit water molecules and sufficient K⁺ ions to make 100 mM [K⁺] in VMD.⁶⁸

In all cases, periodic boundary conditions were used and the Ewald Particle Mesh method was used for electrostatic potentials. All timesteps were set to 1.0 fs. The annealing processes were run under canonical ensemble conditions (constant NVT). Each production run was performed under isothermal, isobaric ensemble conditions (NPT). The simulations were run as follows:

- 1) Equilibration at T = 310K for 500 ps (500,000 steps).
- 2) Annealing with rigid bonds at T = 600K for 2 ps followed by 2 ps at T = 550K, 2 ps at 500K, 2 ps at 450K, 2 ps at 400K, 2 ps at 350K and finally 2 ps at 310K
- 3) Production runs were performed using a short 5 ps run at 310K with rigid bonds followed by 10 ns (total) at 310K without bond constraints.

All simulations were run on parallel high performance computer servers in parallel.

Typical compute timings were 2 ns per day of wall clock time.

1.3 Results and Discussion

All systems investigated were run until equilibration was achieved as determined by stabilized total energy, volume, and pressure. In each case, equilibration was determined when the fluctuations in these parameters was less than 1% of the total value. Snapshot images from the trajectories are shown Figures 6-9.



Figure 6: Snapshot of parallel stranded quadruplex (1Kf1) in 10ns of simulation

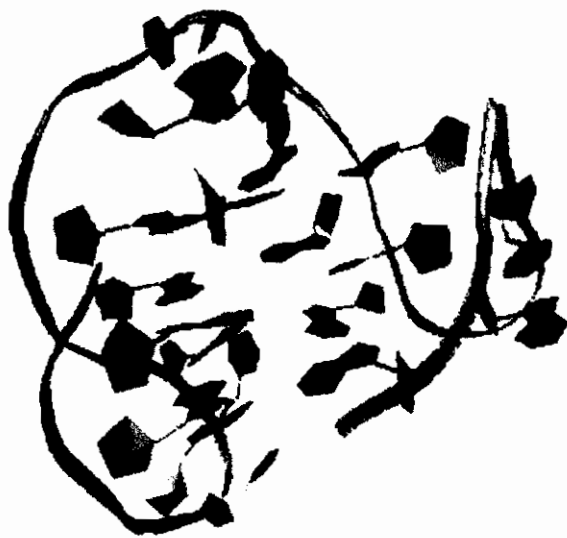


Figure 7: Snapshot of hybrid I (2gku) in 10 ns simulation



Figure 8: Snapshot of hybrid II (2JPZ) in 10ns simulation



Figure 9: Snapshot of unfolded quadruplex in 10 ns simulations

1.3.1 Average Total Energy

We calculated the average total energies for different models of human quadruplexes. As we see in Figure 10, the lowest energy is for the unfolded form of Quadruplex. The hybrid structures total energies are similar. The total energy for parallel stranded structure is lower than hybrid structures.

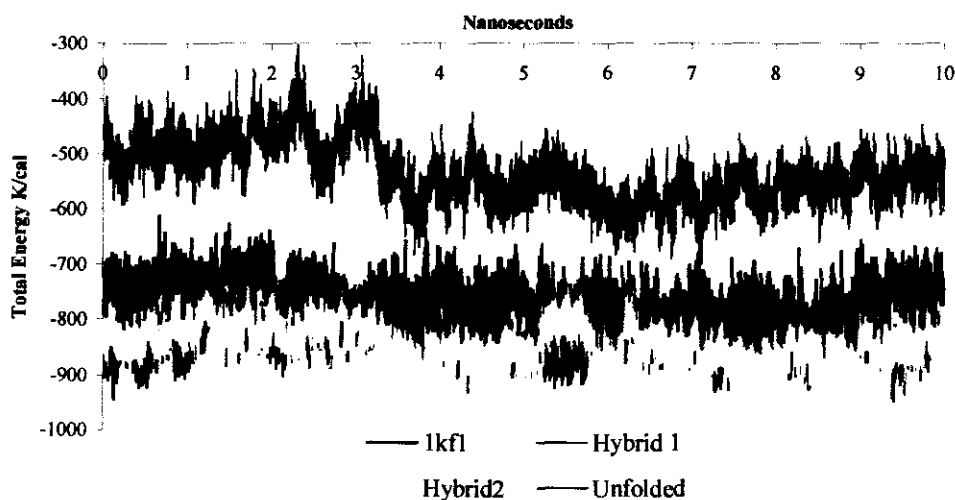


Figure 10: Average total energies for different models of quadruplexes

From the trajectory data, the average solvent accessible surface area (S), change in heat capacity, and root-mean-square-deviations were determined. These analyses are described below.

1.3.2 Average SASA and relative stability

The SASA (see Appendix II) is a measure of the surface area of the molecule that is exposed to solvent. The SASA for each of the model systems after 10 ns of simulation are shown in Figure 11. As can be seen in the Figure, the unfolded structure (uf) of

quadruplex shows the largest SASA. This is expected since the unfolded state has no G-G interactions and essentially the entire surface of the structure is exposed to the solvent. In contrast, the hybrid structures (2jtz, 2gku) have almost the same SASA. Finally, the all-parallel state (1kfl) shows an intermediate SASA compared to the hybrid and unfolded states. One likely explanation for this is that the hybrid states both exhibit “tighter” structures particularly for the loop residues whereas the all-parallel state has comparably “looser” loop residues.

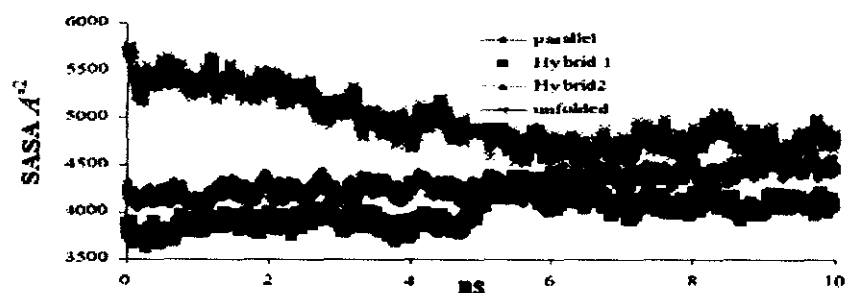


Figure 11: SASA of different models of quadruplexes

The loop residues in the all-parallel state are far more constrained since they must traverse along the diagonal length of the quadruplex and the loop residues on the hybrid states must only traverse the edge of the quadruplex. Hence, the all-parallel state has loop regions that are considerably more exposed to the solvent than those of the hybrid states. The unfolded state is essentially “all loop” and so would be expected to have a higher SASA. We also note that the unfolded state has greater variability over the trajectory. This is likely due to the greater flexibility of this state. It is important to note that since no known structure of the unfolded state has been reported from experiment, we can only treat the SASA (or any other property) for this state in a qualitative manner. Actual calculated values for this state can only be taken as indicative of a maximal value.

Shown in Table I are the total energy of each state exclusive of the solvent and ions along with the average SASA. The average total energies and average SASA with the standard deviations are shown in Table I. As we see the unfolded quadruplex has the lowest energy and highest SASA and one of the hybrid structures (2JPZ) has the highest total energy but another hybrid structure (2gku) has the lowest SASA. The average SASA values were taken from a linear time average from each trajectory. The average SASA for the unfolded states was taken from the final 4 ns of the trajectory in order to avoid the fluctuation from being too large. For the energy of each state, the more negative the energy, the more stable is the structure. Hence, the unfolded state, with the most positive energy (-560 kcal), is the least stable structure, as expected. The most stable structures are the hybrid states, and the parallel state has intermediate stability.

Table I
Average total energies and average SASA values of different models of quadruplexes in 10 ns of simulations

Structures	Type	Total energy (kcal)	Std. deviation (kcal)	Average SASA Å^2
2gku	Hybrid (h1)	-825.	35.8	4152.
2JPZ	Hybrid (h2)	-857.	31.6	4097.
1KFI	Parallel (p)	-760.	39.9	4314.
Unfolded	(uf)	-560.	39.3	4739.

We also note that the more stable hybrid structure also exhibit, smaller average SASA. We interpret this trend as resulting from the extra energy cost derived from the loop residues being strained in the parallel structure as described above. Moreover, since the parallel state loop residues have higher exposure to the solvent, they must also exhibit greater exposure of the hydrophobic regions further destabilizing the parallel state.

Hence, we find that the relative stability of the states and the average SASA values are consistent.

Although we note differences in the stability (-825 kcal and -857 kcal) as well as the corresponding SASA, the fluctuations over the trajectory as reflected in the standard deviation essentially eliminate any real difference between these two states.

1.3.3 Determination of ΔC_p

As described in Appendix II the change in the SASA from one state to another can be related to the change in heat capacity for the transition. This analysis allows for a direct comparison to experimental heat capacity studies. To do this, the total SASA for the states must be further evaluated in terms of the SASA for the polar and nonpolar portions of the molecule. In essence, the change in heat capacity is most affected by relative changes in solvent exposure to the nonpolar regions since the polar solvent must assume a greater degree of structure around exposed nonpolar residues. The reasoning follows from a consideration of the change in solvent entropy accompanying the transformation from one state to another. Shown in Table II are the SASA values for the polar and nonpolar portions of the models studied.

Table II
SASA of hydrophilic and hydrophobic part of different models of human telomere quadruplexes

Structure	Type	Nonpolar SASA Å^2	Std. dev.	Polar SASA Å^2	Std. dev.
2gku	Hybrid (h1)	2036.	64.3	2058.	39.4
2j pz	Hybrid (h2)	1990.	49.7	2055.	34.7
1kfl	Parallel (p)	2115.	79.0	2134.	48.9
unfolded	(uf)	2262.	64.6	2401.	52.2

As expected, the unfolded state has the largest nonpolar SASA. Also, consistent with the stability and total SASA described above, the nonpolar SASA for the hybrid states are measurably lower than that of the parallel state. Between the hybrid states, the 2j pz state shows the lowest nonpolar SASA. Here again, distinctions between the two hybrid states may not be realistic given the overlapping standard deviations of these two states.

Among the investigated states, there are several types of state-to-state transitions that can be studied. These include transitions between folded states and between folded states and the unfolded state. These possible candidates are then calculated in terms of the ΔC_p using the method described in Appendix I and compared with experimental values. The possible transition types among the four states (h1, h2, p, u) are shown in Table III. The value of B_{np} was 0.34 and that for B_p was 0.14 taken from the work of Spolar *et al.*⁶⁹

Table III
Calculated ΔC_p for different isomers of human telomere quadruplexes in (cal/mol/K) in 10 ns of simulations

Transition		Type	Calc ΔC_p (cal /mol/ K)
State 1	State 2		
2gku	2JPZ	h1 \rightarrow h2	-17.3
2gku	1KF1	h1 \rightarrow p	14.6
2JPZ	1KF1	h2 \rightarrow p	-31.9
2gku	Unfolded	h1 \rightarrow u	24.4
1KF1	Unfolded	p \rightarrow u	9.75
2JPZ	Unfolded	h2 \rightarrow u	41.7

Using the data in table III, we can estimate the ΔC_p of six different transitions that accommodate an intermediate and have an unfolded final state. Clearly, any such process should have an overall positive ΔC_p for the process ($\Delta C_p = d\Delta S/d\ln T > 0$). Hence we

can eliminate the process $h1 \rightarrow p \rightarrow u$ since ΔC_p for this process would be equal to -22.1 cal/mol/K. Among the remaining five, if we eliminate the ones which have a negative ΔC_p for state1 \rightarrow state2 under the assumption that each step might be expected to have $\Delta C_p > 0$ for transition to a higher temperature state, we are left with $h2 \rightarrow h1 \rightarrow u$, $p \rightarrow h2 \rightarrow u$, and $h1 \rightarrow p \rightarrow u$. If we include the total energy data from above, which indicates that the parallel state is less stable than either the h1 or h2 states, our results indicate that the most likely process is $h1 \rightarrow p \rightarrow u$. For this process we calculate $\Delta C_p = +24.4$ cal/mol/ K for the process initial state \rightarrow intermediate \rightarrow final state, and $\Delta C_p = +9.75$ cal/mol /K for the process intermediate \rightarrow final state. It is important to recognize that experimental results for ΔC_p are extremely difficult to determine due to the limited temperature range between the intermediate \rightarrow final state transition temperature and the solvent boiling point; hence, any comparison between experimental and computed values for ΔC_p are not reliable.

1.3.4 Root -Mean- Square -Deviations of Different models of Quadruplexes

In order to assess the comparative conformational stability of different folding isomers of human telomere quadruplexes, the (RMSD) were determined for each simulation. The calculation of RMSD is described in Appendix II. The reference frame for the RMSD calculations was the final frame. The results for these analyses are shown in Table IV. Also, Collie *et al.*⁷⁰ used variations in RMSD between the telomeric RNA and DNA quadruplexes during MD simulations as a measure of comparative structural stability. They argued that the structures with lower RMSD exhibit greater stability.

As seen in Table IV, the unfolded structure showed the greatest RMSD as expected. Hybrid structures 1 and 2 showed the lowest RMSD and the parallel structure showed intermediate values. The parallel state, being intermediate between the unfolded and hybrid states, is consistent with our total energy and ΔC_p analysis above. This suggests that the most stable structures are the h1 and h2 isomers with the parallel isomer being the intermediate state. Between these two most stable isomers, we find that h2 has slightly greater stability. Our analysis of the transition from initial state \rightarrow intermediate \rightarrow final state as described above indicates that the most likely initial state is h1 in contrast with the RMSD analysis. However, if we consider the standard deviation of these two states, the RMSD difference is not significant

Table IV
Average RMSD per residue for different models of human telomere quadruplexes

Structure	RMSD Å	Standard deviation
Unfolded State (uf)	3.53	0.88
1kfl(p)	2.14	0.42
2gku (h1)	1.87	0.38
2j pz (h2)	1.46	0.18

1.4 Concluding Remarks

The total energy, SASA, (ΔC_p), and RMSD results of the investigated human quadruplex models consistently and independently show that the equilibrium low temperature state of the human telomere quadruplex (TTAGGG)₄ in K⁺ solution can not be the all-parallel structure and that the ground state must be a hybrid structure, most likely the h1 state.

Our molecular dynamic stimulations suggest that the mixed parallel and antiparallel hybrid states are the most stable and probably the most favorable conformation. The parallel conformation is the next favorable structure in K^+ solution.

Previously reported results of thermal denaturation indicate that an intermediate of indeterminate structure appears prior to unfolding to the denatured random coil state.⁶⁷ Our results indicate that this intermediate state is probably the all-parallel state and it has sufficient stability to occur as a defined intermediate.

The identity of the ground (low temperature) state is identified as the h1 hybrid conformation, which can be used in rational drug design to target this particular structure in the development of anticancer therapeutics.

CHAPTER II

Molecular Dynamics Simulations of the Bacteriorhodopsin Membrane Protein

2.1. Introduction

The cell membrane performs a great variety of functions including separation of intra- and extra-cellular components, protection,¹ and perhaps, most importantly, provides a matrix in which membrane proteins can function. Membrane proteins in eukaryotic cells provide control for the passage of material into and out of the cell and also provide a key regulatory mechanism for cell function. In every cell, the membrane incorporates associated proteins that function as sensors, allowing the cell to efficiently adapt to environmental changes.

All biological membranes have a common structure that is an extremely thin coat of lipid and protein molecules. The lipid layer is typically a bilayer composed of amphoteric lipid molecules that have the polar “head group” facing outward and the nonpolar region in the membrane interior. This arrangement causes the solvent molecules to be excluded from the membrane and can not pass freely from one side to the other.

Membrane proteins are not covalently bound to the lipid molecules and therefore can translate in two dimensions within the membrane. Membrane proteins in the lipid bilayer perform specific functions such as transferring molecules or participating in ATP production. Various membrane proteins can help to bind the membrane to the cytoskeleton and some of them act as receptors and participate in chemical signaling. The lipid and protein design of the interior and exterior sides of the cell vary from each other and this distinguishes the diverse operations on the two sides of the membrane.

2.1.1 Membrane Proteins

The quantity and type of proteins in a membrane are extremely dependent on the type of the cell, For example, in the myelin membrane, which forms an electrical

insulating layer for nerve cell axons, only 25% of the membrane mass is protein, but in some membranes %75 is comprised of protein. Most membranes have approximately 50% by weight protein composition. Membrane proteins are commonly glycosylated with oligosaccharide chains bounded to them.⁷¹

2.1.2 Membrane proteins and lipid bilayer

There are three basic types of membrane proteins depending on their association with the membrane. Some membrane proteins are located completely in the cytoplasm and are linked to the bilayer solely by one or a few covalently bonded fatty acid chains or different types of lipid chains named prenyl groups. These are referred to as peripheral membrane proteins and are associated with only one side of the membrane (intra- or extra-cellular). Integral membrane proteins are embedded into the lipid membrane. Some integral membrane proteins, because of their amphoteric nature extend to both sides of cell membrane and are referred to as “trans-membrane”. Their hydrophobic sections exist primarily within the membrane and interact with the hydrophobic ends of the lipid bilayers in the core of the bilayers.⁷¹ Their hydrophilic sections are exposed to the aqueous solvent on both sides of the membrane. The hydrophobicity of a number of membrane proteins is increased by covalent addition of a fatty acid chain. Figure 12 illustrates different kinds of membrane proteins and an example of G coupled receptor protein is provided in Figure 13.

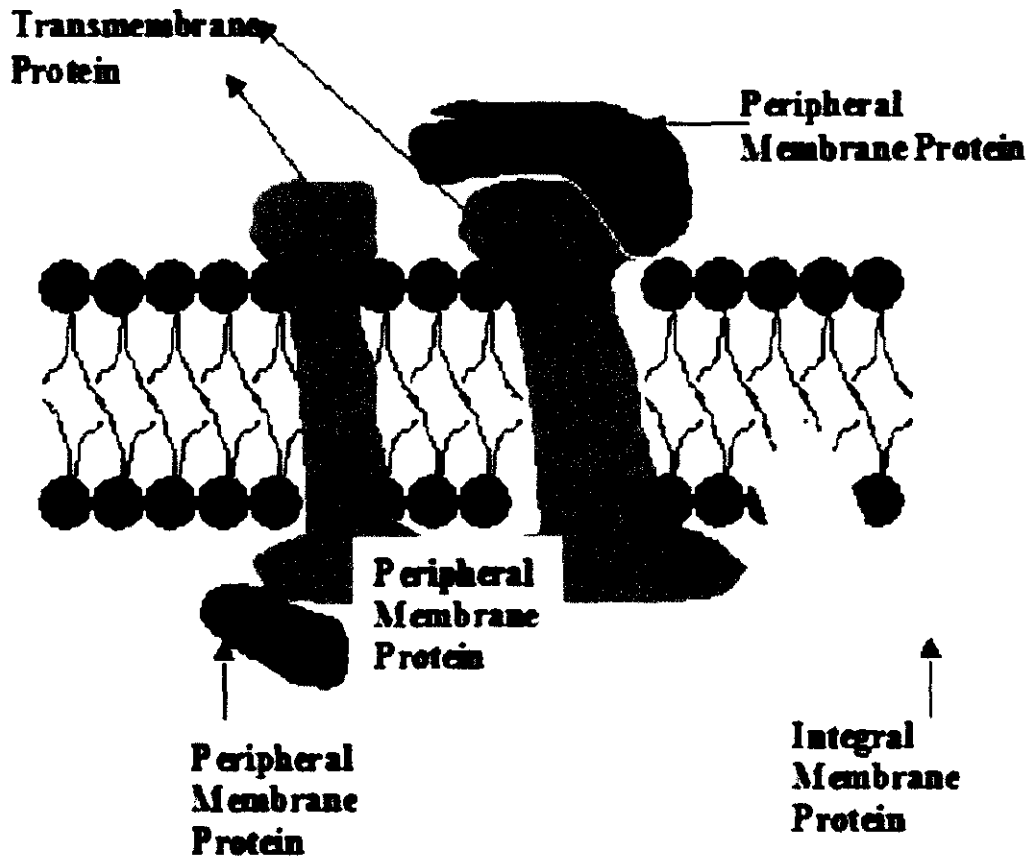


Figure 12: Different forms of membrane proteins

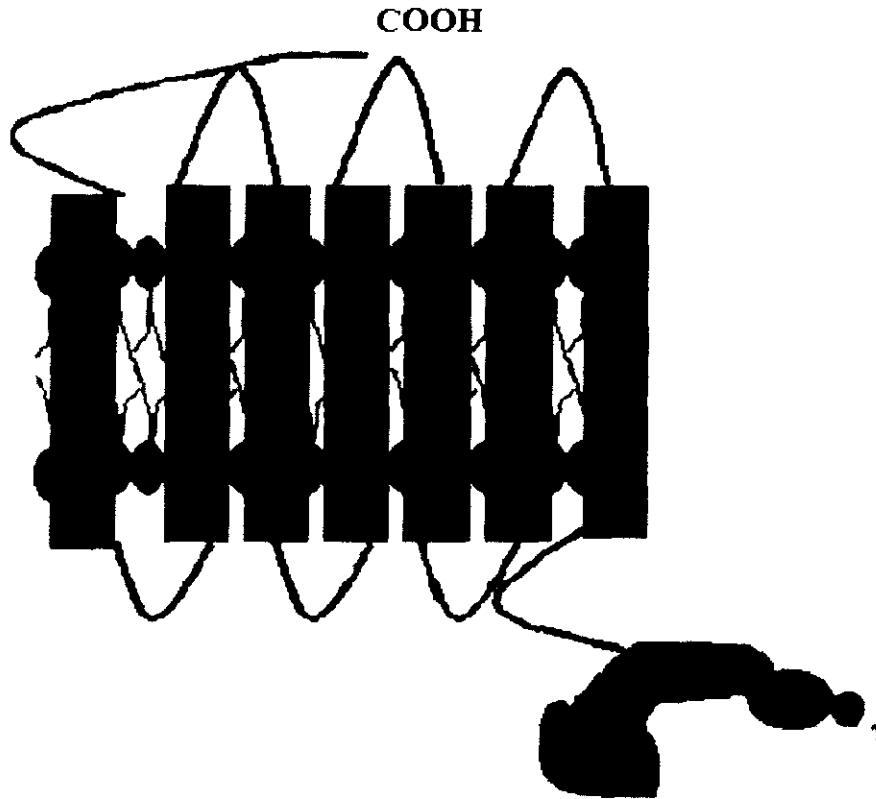


Figure 13: G coupled receptor protein

2.1.3 Alpha Helical Arrangement of Trans-membrane Proteins

Trans-membrane proteins have specific orientations in the membrane.¹ This asymmetrical structure is determined by the role the protein has in the membrane. The secondary structure of the protein is commonly composed of α helices in which the helices are oriented approximately normal to the membrane plane and form a kind of interior tube which excludes both solvent and lipids.⁷¹ In single-pass trans-membrane proteins, a single α helix polypeptide passes through the membrane, whereas in multi-pass trans-membrane proteins, several (commonly seven) α helices transect the membrane.

Since transmembrane proteins are exceedingly difficult to crystallize, only a small number of them have been studied completely by (XRD). The three-dimensional

structures of the majority of them have not elucidated.¹ However, the primary sequence of membrane proteins were determined using DNA cloning and sequencing methods.⁷²

2.1.4 Bacteriorhopsin

Bacteriorhodopsin (bR) is considered to be the ideal model for investigating transmembrane proteins. First of all, several XRD structures have been published.⁷³⁻⁷⁵ Second, it has been long studied to better understand its photophysical properties, therefore a great deal is already known about it. The *Halobacterium salinarium*, which is found in salt mines and salt lakes and represents a particularly rich source of bR since this organism expresses the protein at extremely high rates. Bacterioopsin is the product of the *bop* gene in the *Halobacterium salinarium* genome and has 248 amino acids with a molecular weight of 26 KDa.^{76,77} Its tertiary structure is composed of seven alpha helices (named A to G) joined through loops on both sides of the membrane. Retinal is enclosed by the helices in the center of the protein. Further Retinal is covalently bonded to the protein through a protonated Schiff base with Lys216. Three molecules of bR are clustered into trimers which are arranged into a hexagonal 2-dimensional lattice.⁷⁸

In the dark-adapted ground state of bR, the conformation of the retinal molecule is in equilibrium between the all trans and 13-cis forms with a ratio of 1:2 respectively.⁷⁹ In the dark-adapted state the absorbance maximum of is 558 nm. Upon illumination, the dark-adapted bR state is activated to produce the light adapted form, with an absorbance maximum of 568 nm. Retinal is surrounded by the hydrophobic and aromatic residues of the protein, which undergo a cascade of events that result in proton transfer. Among the nearest neighbor residues to the retinal are aromatic amino acids like Trp (86, 138, 182, 189) and Tyr (57, 83, 185). These residues are bulky and appear to be essential to

prevent the retinal from adopting static conformations that would prevent the conformational changes needed in the photocycle.^{73,80-83} Some polar and hydrophilic residues (Thr90, Asp212, Asp85) are located in the retinal binding pocket as well. Hydrophobic residues appear to have a structural function while more hydrophilic residues are believed to play a functional role in the photocycle. Figure 14 shows bR in dark adapted ground state (pdb 1C3W).



Figure 14: Structure of Bacteriorhodopsin in the ground state

Residues Thr90 or Pro186 have no connection with the proton pathway but can be considered as accommodating conformational variations in the retinal. Asp85, Asp212, and Arg82 are not part of the retinal binding pocket, but are able to provide a counter-ion charge, which protects the positive charge of the Schiff Base. A number of water molecules are associated with these counterions.⁸⁴⁻⁸⁸

Following light stimulation, the retinal changes its conformation from all trans in the dark-adapted state to a cis isomer, leading to a cascade of structural modifications in the protein interior. The result is that charge transfer produces proton flow from the cytoplasm inner part to the external side of the protein.⁸⁹ These protons are subsequently consumed by ATP synthase to generate ATP.⁸⁹

In an anaerobic atmosphere, *H salinarium* discontinues growth and generates a purple membrane. This type of membrane is seen in patches in the membrane and takes up approximately 50% of the membrane plane as shown in Figure 15. It is called purple membrane because of its purple color.⁸⁹ bR accounts for up to 75% of the mass of the purple membrane, with the remainder composed of lipids such as squalene, glycolipids sulfate, and phosphatidyl glycerolphosphates.⁹⁰



Figure 15: Purple membrane of Bacteriorhodopsin

2.1.4.1 Major differences between helices of Bacteriorhodopsin

As described above, bR is a transmembrane protein with the majority of the secondary structure in the α helical conformation. All the important residues, such as Asp85, Asp96, and Arg82, which are engaged in proton transfer, are found in Helix C. Trp86, Arg82, Leu93, and Tyr83 are residues, which help to develop the retinal binding pocket and are also located in Helix C. The retinal molecule is a Schiff-base bound to Lys216 in Helix G.

Helices G, F, B and C demonstrate a number of dissimilarities from each other.⁹¹ Helix G possesses Lys216, which is attached to retinal. Helices B, C and F have a proline residue near the middle of the helix. In substitution defect studies, shows the proline in these was helices substituted with another residue but did not result in structural changes in the helix.⁹²⁻⁹³ Hence, the functional role of these prolines is unknown. But proline breaks the α helical secondary structure of proteins.

2.1.4.2 Bacteriorhodopsin CT

Helix G has a 23-residue peptide sequence following the final residue in the helix, which extends into the cytoplasm (Figure 16). Since this sequence is on the carboxyl terminus, it is referred from here on as the (CT). To date, no X-ray studies of bR have indicated a defined structure for the CT. Hence, there can be no structural insight into the possible functional or structural role it might play. Only a few studies using mutagenic analysis of specific amino acids in the CT.⁹⁴⁻⁹⁷ The results of these studies are discussed in section 1.2.3 below.

-N-SRAIFGEAEAPEPSGADGAAATS-C-

Figure 16: Amino acids of C terminus of Bacteriorhodopsin. Non-polar residues are indicated in red, acidic residues are indicated in underlined blue, and basic residues are indicated in blue italics.

2.1.4.3 bR Photocycle

The process that initiates with the absorption of a photon and ends with the transfer of a proton across the membrane is referred to as the photocycle. Upon illumination at 568 nm, a photon is absorbed by retinal, initially in the all trans state.⁹⁸ After excitation, retinal undergoes structural isomerization to the cis isomer, which causes a proton to transfer from the Schiff base NH^+ to a neighboring Asp residue. The cascade of events leading to the proton transfer across the membrane are described below.

Starting from the ground state of bR, the cycle comprises six intermediate steps labeled J, K, L, M, N and O. Some crystallographic structures for these photo intermediates were reported.^{74,99-104} Some studies, applying diverse techniques for crystallizing bR,¹⁰⁵⁻¹⁰⁸ show inconsistencies for the structures of the same intermediate.¹⁰⁹ However, the important characteristics¹¹⁰ of the intermediate states are shown here.

bR: This is the resting state, which is called bR. Retinal is in the all trans conformation and the Schiff base N atom is protonated in this phase.¹¹¹⁻¹¹²

J: Retinal remains in the all trans form and the Schiff base is protonated. A reformation of the electronic density of retinal among C=C and C-C is observed. Some evidence for the presence of torsion of the polyenic chain was reported.¹¹³

K: The Schiff base remains protonated and retinal isomerizes from all trans to 13-cis.¹¹³⁻¹¹⁴

L: Water molecules located between Asp96 and the Schiff base undergo rearrangement subsequent to retinal isomerization, which modifies the local environment around ASP96.¹¹⁵ These movements result in relaxation in the helices such that the new conformation of the retinal is stabilized.⁹⁹

M: M intermediate has two photo intermediates, known as M1 and M2. The main aspect of this intermediate is the Schiff base deprotonation.¹¹⁶⁻¹¹⁷ The proton dissociated from the Schiff base moves to Asp85, which becomes protonated.

M1: The Schiff base deprotonates and the proton relocates to Asp85.¹¹⁸ This move causes discharge of a proton to the extracellular area through a residue or a complex known as X or proton release group of undetermined structure.

M2: The helices F and G change in a way that the cytoplasmic half of the protein expands to an open channel to accommodate the entrance of water molecules.¹¹⁹

N: In this step, the retinal is reprotonated from Asp 96 and flexes again. The movement of helices F and G opens a narrow channel through which Asp 96 is reprotonated.

O: Asp96 becomes protonated and the retinal switches to its all trans structure.

The source for Asp96 reprotonation is not clear. Some hypotheses consider an antenna complex, which can attract the proton from the bulk, and then transfer it to Asp96. During O, Asp85 stays protonated, later in the O → bR step, this residue deprotonates and the proton release group protonates.¹²⁰

All these actions are part of the photocycle and must be considered in order to find out the course of the proton through the protein. However, the crystal structures obtained recently allow a better understanding of the proton route and the structure of some intermediates. These intermediate structures were prepared at low temperatures and in non-physiological condition for the protein. However, some ideas can be inferred concerning the details of each intermediate in the photocycle. The photocycle of bR is shown in Figure 17.

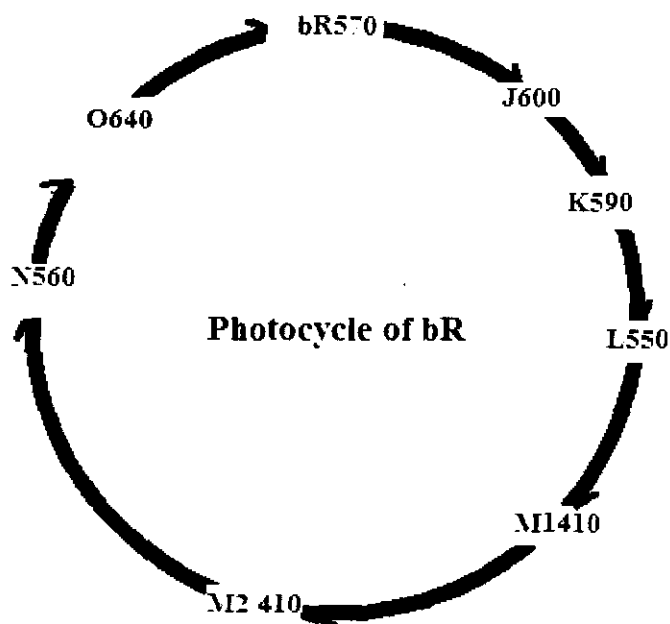


Figure 17: Photocycle of bR. The letter designations for each state are the same as in the text description. The numbers indicate the absorption wavelength maximum (in nm) for each state. bR570 is the dark-adapted rest state investigated in this project.

2.1.4.4 Experimental Studies of the CT of bR

The mechanism for transferring protons into the cytoplasmic channel in bR was investigated using both structural analysis⁸⁶ and site-directed proton pulsing.¹²¹ Both of these methods indicate that some component of the loop sections at the cytoplasmic side of the channel draws protons from the solution (a “proton antenna”). Interestingly, the bR CT contains five amino acids with carboxylate groups.^{73,76} Therefore, the bR CT would appear to be an interesting candidate for taking part in proton sequestration. Proteolytic deletion and spectroscopic analyses suggest that the CT is divided into domains with diverse properties.¹²²

Turner *et al.* studied the role of different domains of the bR CT by using truncation.¹²³ Two truncations were done on the CT of bR: one of the truncations detached 12 amino acids from the distal region and another deleted 10 amino acids from the proximal part of the CT. Removal of the CT amino acids in the proximal section reduced the rate of proton uptake, while removal of the distal amino acids had no outcome on proton uptake. However, both truncations had a significant effect on the stability of bR.

In Turner *et al.* experiments,¹²³ proton uptake at the end of the M stage (M → N) were found to be correlated with full truncation of the CT. They suggested that their results represented additional support for the idea that proton recruitment at the cytoplasmic surface of bR is tied to M state decay which was previously observed by Lanyi and Balashov.^{124,125} The role of long-range electrostatic forces in bR is poorly understood, at present.^{126,127} The CT residues that contain amine side chains ($-\text{NH}_3^+$) can repel protons and the carboxylate-containing side chains could attract protons or make ion pairs with basic side chains, lipid head groups, or solvent cations. Either one of these

phenomena, or a mixture of both could function to adjust the pKa's of acceptor proton groups positioned in or near the cytoplasmic proton channel. Some mutagenic studies suggest that Arg227, Gly231, and Glu234, which are located in membrane proximal part of the CT, may be involved in proton uptake.^{94, 96, 97}

Balashov *et al.* has suggested that some or perhaps all of the five carboxyl terminal side chains in the CT which may act as a proton antenna.¹²⁴ As described above, studies showed that deprotonation of the Schiff base and release of a proton to the extracellular part occurs at the end of the M step.¹²⁸⁻¹²⁹ It was also observed that mutation of Arg227, a carboxyl-terminal amino acid, caused a reduction in the rate of M state decay.⁹⁷ Some studies suggested that Glu234 may also be involved in this process.⁹⁶ The finding that Arg227 and Glu234 have roles in the proton pumping mechanism is consistent with Turner *et al.* analysis that the proximal part of the CT might be related to the reprotonation mechanism.

2.2 Objectives of Our Research

The above mentioned experimental studies have shown that removal of or truncation of the CT leads to a dramatic loss of stability and function of bR. Although the primary sequence of the CT and the primary, secondary, and tertiary structure of the transmembrane portion of bR have been resolved, the secondary and tertiary structure of the CT is currently unknown. Hence, the structure one of the most important portions of the molecule remains unknown. It should be emphasized that the lack of XRD data simply indicates that no long range periodic lattice is observed for the CT. This does not necessarily indicate that the CT does not adopt a particular folding structure, but that any such structure is not periodic in crystalline form under which XRD data must be collected.

In the current study, MD is used to explore the equilibrium structure of the CT to help establish 1) a likely secondary structure, at least capable of determining intra-CT electrostatic or other short-range interactions, 2) the possible interactions and associations between the CT and the transmembrane portion of bR, 3) any possible interactions with the lipid membrane, and 4) the likely solvent associations that could lead to proton sequestration. In particular, we will investigate by MD simulations the equilibrium structure of the CT of bR by itself in solvent, in proximity to bR, and bound to bR. Also, the nature of any association of the CT with either the bR loop structure or lipid membranes is explored to help establish structural associations that could lead to the observed enhanced stability. The use of complementary computer simulations in studying these systems was shown to be well-validated and robust.¹³⁰

2.3 Methodology for simulation of the CT of bR

The reader is directed to the Appendix I for background details of MD simulation techniques. In this section the simulation details specific to the CT of bR will be described.

The starting structure for creating the initial pdb and psf input structure and topology files was built from known primary structure of the CT. In order to find the equilibrium structure of the CT, pdb and psf input files were prepared which contained approximately 12,182 explicit water molecules and placed in different salt concentrations from 1 – 4 M [NaCl] as shown in Figure18. Due to the inherent dilution effects of investigating small systems (even ones with 10 – 15 thousand solvent molecules), it turns out to be easier to prepare higher concentration salt solutions than dilute ones. Moreover,

H. salinarium archaea grow in 4 M [NaCl], so high salt concentration systems are consistent with typical physiological conditions.



Figure 18: CT of bR

In all cases, periodic boundary conditions were employed and the Ewald Particle Mesh method was used for long-range electrostatic potentials. All timesteps were set to 1.0 fs. The annealing processes were run under canonical ensemble conditions (constant NVT). Each production run was performed under isothermal, isobaric ensemble conditions (NPT). The simulations were run as follows:

- 1) Equilibration at $T = 310$ K for 500 ps (500,000 steps).
- 2) Annealing with rigid bonds at $T = 600$ K for 2 ps followed by 2 ps at $T = 550$ K, 2 ps at 500 K, 2 ps at 450 K, 2 ps at 400 K, 2 ps at 350 K, finally 2 ps at 310 K

3) Production runs were performed using a short 5 ps run at 310 K with rigid bonds followed by 10 ns (total) at 310 K without bond constraints.

All simulations were run on parallel high performance computer servers in parallel using the NAMD¹³¹ and CHARMM¹³² forcefields. Typical compute timings were 2 ns per day of wall clock time.

2.3.1 Methodology for Simulation of Different Models of bR

The input models for all simulations of bR were created using the most complete available structures published in the Protein Data Bank (PDB)¹¹⁹. As discussed above, no existing structure contains the CT, however, the primary structure is known for the entire 248 residue sequence. The study included investigations of four system types: 1) the bR molecule without CT in a membrane (Figure 19), 2) the CT alone in solvent, 3) bR in a membrane with the CT close but not bound to bR (Figure 20), and 4) bR with the CT bound (Figure 21). Input structure files for the CT structure and lipid membrane were generated starting from the primary sequence. All other components of the studied systems were obtained from published structures.



Figure 19: bR without carboxyl terminus



Figure 20: bR with CT in proximity



Figure21: bR with CT bound

2.3.2 Building a Structural Model of Bacteriorhodopsin

We used the 1C3W.pdb structure from the PDB database which contains the atom coordinates of bR in the ground state.¹¹⁹ The published structure contained some water molecules and some remnants of cardiolipin lipids, which are components of the matrix used to crystallize the bR molecule. These waters and lipid molecules were first removed

prior to generating the solvated psf and pdb files for the simulation. All psf and pdb files were generated using the VMD package as described in Appendix I.

A 50:50 mixture of dimyristoylphosphatidylcholine (DMPC) and dihexadecylphosphatidylcholine (DHPC) was used for generating the membrane of bR. This bilayer mixture mimics the physiological membrane into which bR is reconstituted prior to experimental investigations. This system was solvated and ions added to make a 1 M of NaCl. The bilayer was first equilibrated for 1 ns at 310 K under fixed bond length constraints. Following this, the system was thermally annealed up to 600 K in 1 ns steps and quenched to 310 K under constant NVT (canonical) ensemble conditions. Finally, the bilayer system was equilibrated at 310K without constraints for 10 ns at isobaric, isothermal (NPT) ensemble conditions.

The bR protein was added to the membrane by removing sufficient lipids from the equilibrated DMPC/DHPC bilayer and re-equilibrating this system similar to described above for the bilayer. A tcl script (lipwat.tcl) was used for this purpose. Systems with a salt concentration of 4 M NaCl were prepared.

The systems containing bR/membrane and the CT (bound or unbound) were generated by adding the equilibrated CT structure from above into the bR/membrane structures with the proximal (N) terminus of the CT either near or bound to the C terminus of bR.

2.3.3 Simulation of Bacteriorhodopsin

All simulations are done using the CHARMM 27 force fields¹³³ and the NAMD simulation package¹³¹ with additional force fields for the lysine-retinal residue from Nina, *et al.*¹³⁴ Each structure was equilibrated using simulated annealing under NVT (to

conserve water density) conditions followed by NPT (1 atm, 310 K) production runs sufficiently long to stabilize the energies, volume, pressure, and temperature. All NPT production runs used the Langevin temperature and pressure coupling algorithm¹³⁵ with 1 fs time steps and unconstrained bond lengths, angles, and dihedrals and continued for 10ns.

We ran a “minimization” run, which places the system in the nearest local energy minimum. Energy minimization then proceeded by an equilibration with the protein constrained, to allow the bR to accommodate to the system environment. We measured the stability of the protein via calculation of the RMSD (Appendix II) of the resulting trajectories using VMD.⁶⁸

2.4 Results and Discussion

The goal of this study was to probe the interaction of the CT with the lipid head groups and exposed bR loop structure. The first part of the analysis was aimed at determining the presence of any intra-molecular interactions within the CT itself. In subsequent studies we investigated the interaction of the CT in proximity with and bound to the bR molecule. These results are discussed below.

2.4.1 CT in solvent

In this section we investigate the properties of the CT in solvent without the presence of the bR. These investigations include an analysis of the structure, the presence of H-bonding within the CT, and salt concentration dependence.

2.4.1.1 Structure analysis

The structure of the CT was first analyzed for the presence of any α or β secondary structure. Shown in Figure 22, is a trajectory snapshot of the CT in 1 and 4 M NaCl.

Although the structure in 4M NaCl appears to be more compacted than in 1M salt, the results of the following analyses indicate that our simulations do not reveal significant structural or dynamical dependence on NaCl concentration.

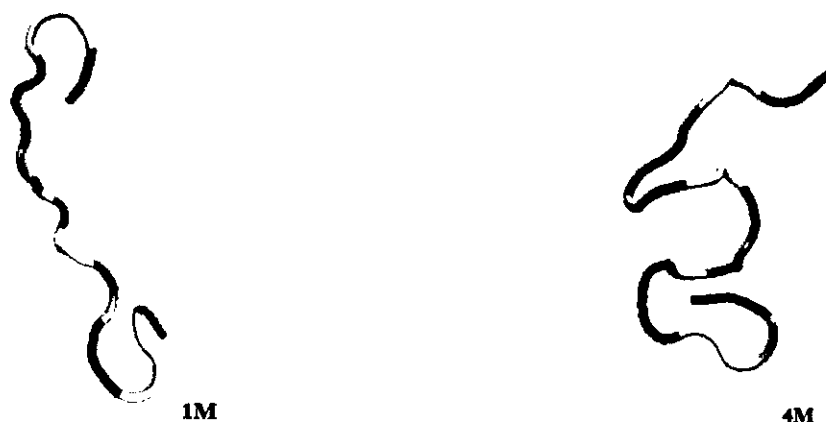


Figure 22: Snapshot of trajectories of CT of bR in 1 and 4M salt concentration in 10 ns simulation

The value of the angles Psi and Phi, as described in Appendix II.iv describe how these can be used to predict the secondary structure of the protein.

Figure 23 shows a Ramachandran plot for this equilibrium structure. It is evident that no clear evidence of α or β secondary structure is present. This is consistent with the XRD results that do not show any repeating pattern as well. Our results do indicate the presence of static coil structure near the ends of the CT. This is shown most clearly by the RMSD data described later.

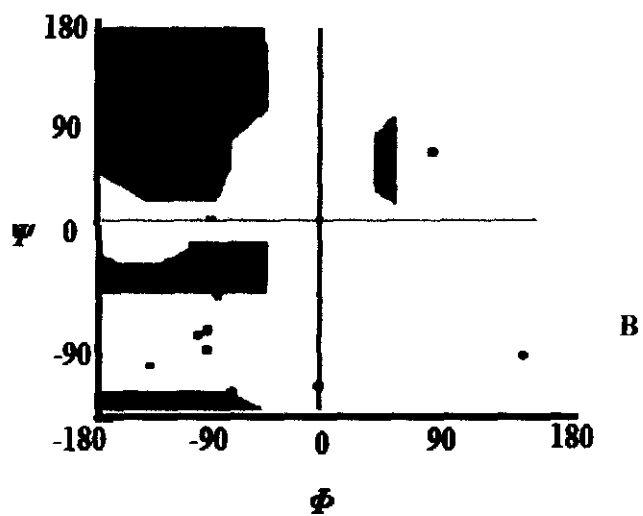
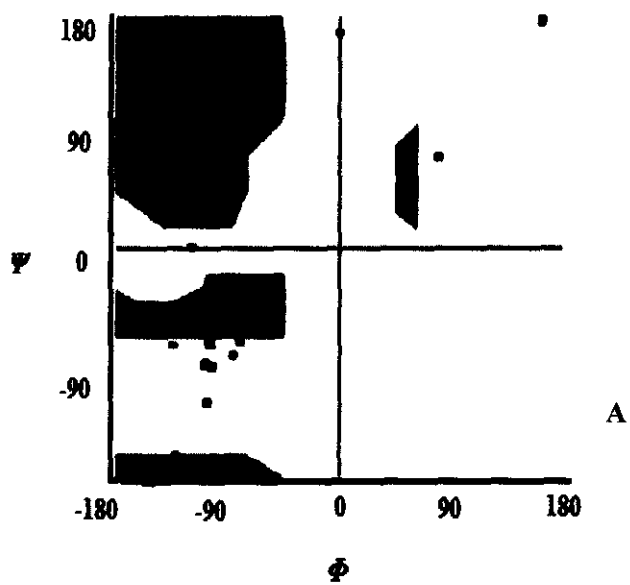


Figure 23: Ramachandran Plot of CT of bR in 1M(A) and 4M(B) salt concentration in 10 ns simulation

2.4.1.2 H-bonding

Shown in Figure 24 is a trajectory snapshot of the CT, indicating the H-bonding structure. As we see for CT in 4M salt concentrations more hydrogen bonding between residues are seen, which might be a reason that 4M salt concentration is a better environment for CT and these hydrogen bonds might help their stability.

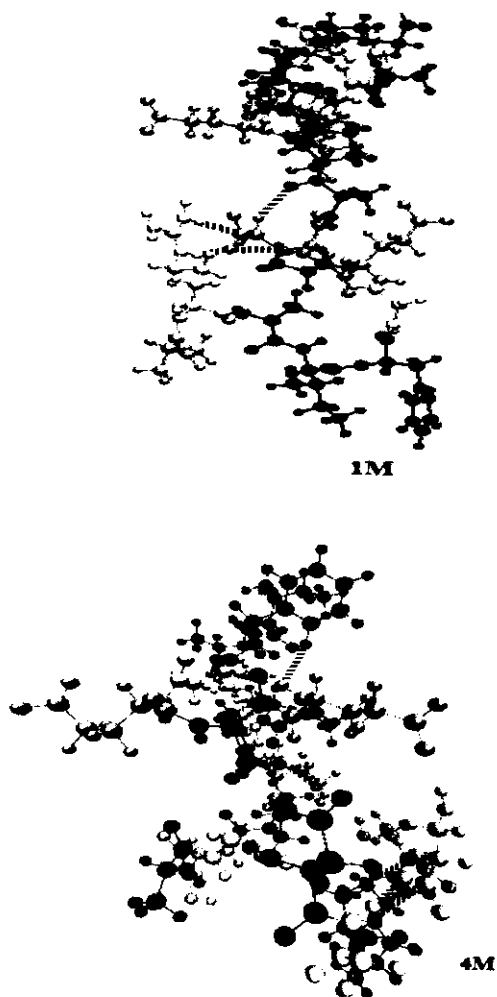


Figure 24: Hydrogen Bonding of CT in 1 and 4M salt concentration in 10ns simulation

2.4.1.3 RMSD and Diffusion constant analysis

As described in Appendix II, the RMSD can be used to help establish the stability of the solute. Shown in Figure 25 is the RMSD as a function of residue number for both concentrations of NaCl.

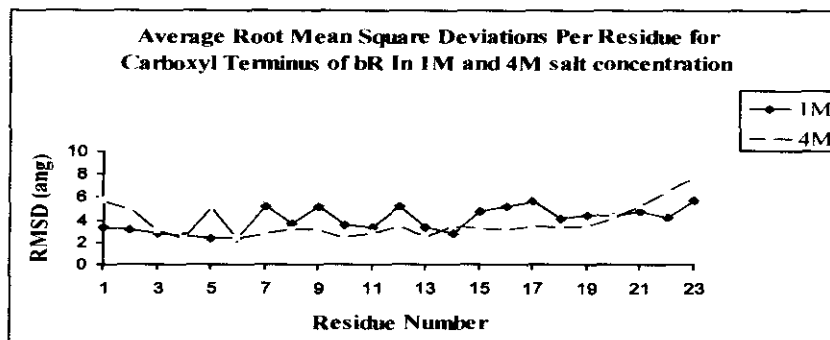


Figure 25: Average RMSD per residue for the CT of bR

We calculated the diffusion constant for the CT of bR in 1 M and 4 M NaCl. For 1 M NaCl, the diffusion was found to be $1.093 \times 10^{-10} \text{ m}^2 \text{ s}^{-1}$ and in 4 M NaCl, it is $0.6433 \times 10^{-10} \text{ m}^2 \text{ s}^{-1}$. The average diffusion constant is smaller in 4M than in 1M salt although the reasons for this slight difference, if important, are not evident.

2.4.1.4 SASA analysis

The average SASA of the CT over 10 ns are shown in Figure 26. The average SASA diminishes from 2753.4 \AA^2 to 24761.6 \AA^2 for the CT in 4 M NaCl possibly indicating that the hydrophobic residues of the CT are less accessible to the solvent. The reduced SASA in 4M NaCl is consistent with the slightly compacted structure of the CT compared with 1M NaCl, as observed in Figure 22. Comparisons of the SASA for the free CT and when in proximity to bR will be discussed below.

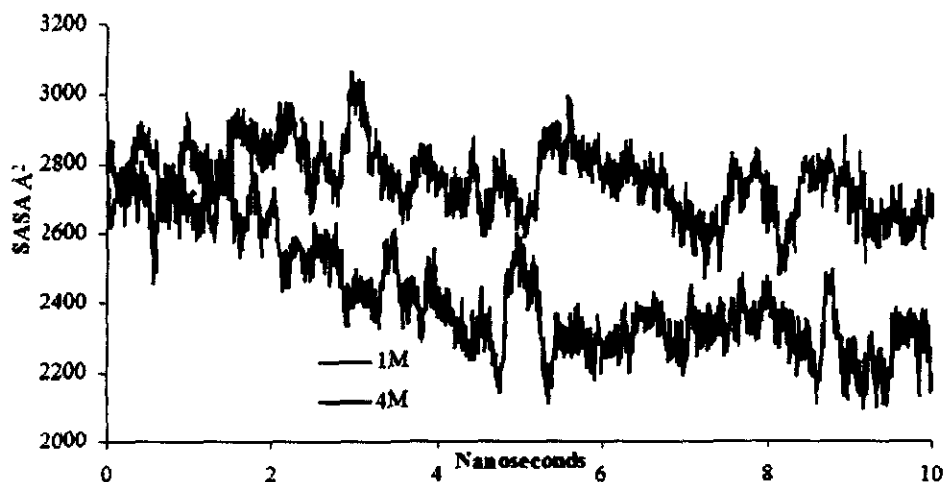


Figure 26: SASA of the CT of bR in different salt concentrations

2.4.2 Analysis of bR without CT

We investigate the properties of the bR in equilibrium in the absence of the CT in order to compare with the results in the following sections for bR with CT attached or in proximity. These investigations include an analysis of the structure, H-bonding, salt bridges, electrostatic energies, and SASA analysis.

2.4.2.1 Structure analysis

Shown in Figure 27, is a Ramachandran Plot (explained in Appendix II) of bR without the CT. As expected, the plot shows mostly α helical structure which is consistent with the known secondary structure of bR.¹¹⁹ The non- α components are primarily the loop structure, which has no defined secondary structure. In addition, there are nine proline residues which are also outside the α region.

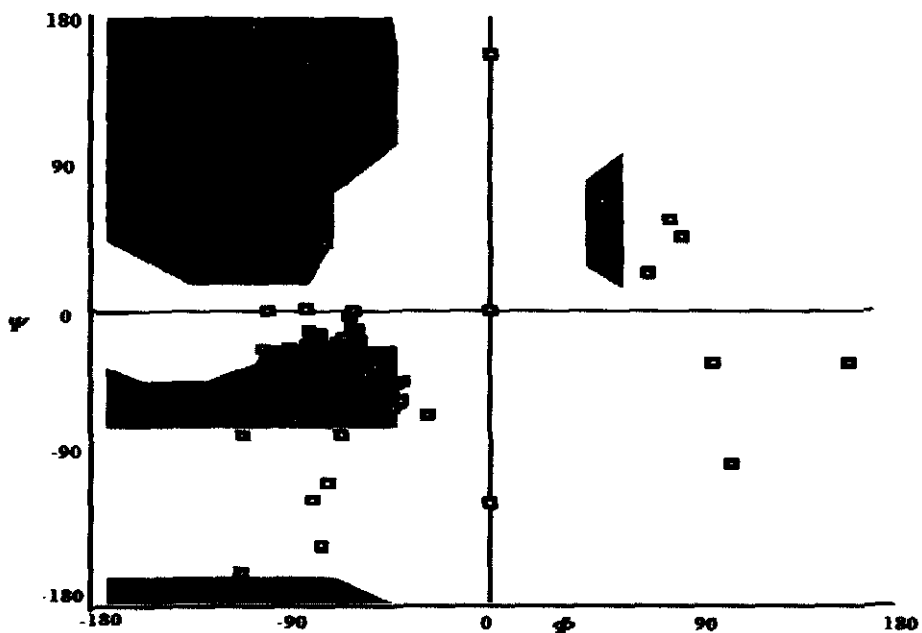


Figure 27: Ramachandran Plot of bR without CT

2.4.2.2 Electrostatic Energies

The electrostatic energies for the bR system only include interactions between the bR protein and the lipids. Figure 28 shows the overall electrostatic interaction energy between bR and the lipid membrane. The trajectory average is -276.83 kcal/mol. This result will be compared with results for bR with CT in the next sections.

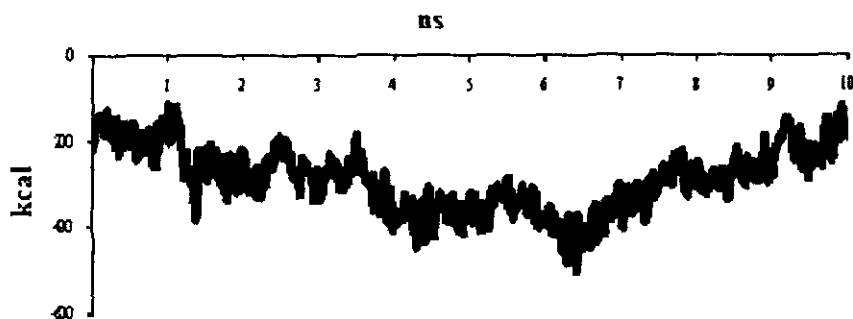


Figure 28: Electrostatic energy between membrane lipids and bR without the CT

2.4.2.3 Hydrogen Bonding

Our simulations revealed the presence of 219 internal H-bonds in the bR molecule. The number of H-bonds external to bR between the molecule and the lipid was found to be 26.

2.4.2.4 Salt bridges

We determined the number of and type of salt bridges internal to the bR molecule which is listed below.

GLU9-ARG7
GLU204-ARG82
ASP212-LYS216
ASP85-LYS216
GLU194-ARG134
GLU194-ARG82
ASP38-ARG164

2.4.2.5 SASA analysis

The SASA is calculated for bR without the CT model and the results are shown in Figure 29. The average total SASA for the trajectory are distributed into polar and non-polar as 6511.94 Å² and 5048.13 Å² respectively.

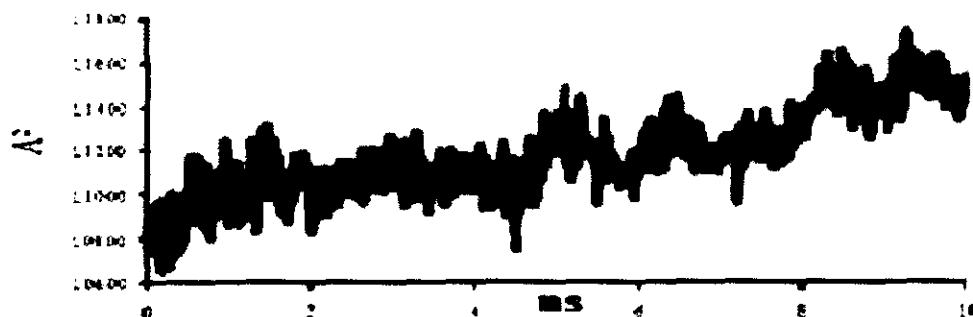


Figure 29: SASA of bR without the CT

2.4.3 Analysis of bR with the CT in Proximity (bR (CT))

The results above provide a benchmark from which comparisons with the CT near or bound to bR are made. We investigate the properties of the CT in equilibrium, in

proximity to but not bound to bR. These investigations include an analysis of the CT structure, the presence of H-bonding within the CT and between the CT and bR and the presence of salt bridges. We also report results of the electrostatic energy of interaction between the membrane lipids and bR, SASA analysis, and heat capacity in 4M salt concentrations. Since *H. Salanarium* typically grows in 4 M NaCl, we chose this salt concentration for our studies.

2.4.3.1 Structure analysis

It is apparent that after 10 ns that the CT appears to be stabilized near bR and that some interactions are likely to be present between the CT and bR. To establish the presence of any new secondary structural motifs, we searched for likely α or β forms using a Ramachandran plot as shown in Figures 30-32. The Ramachandran plot for bR shows that it is mostly α helical as expected. The plot for the CT does not indicate any new alpha or beta structure also as expected.

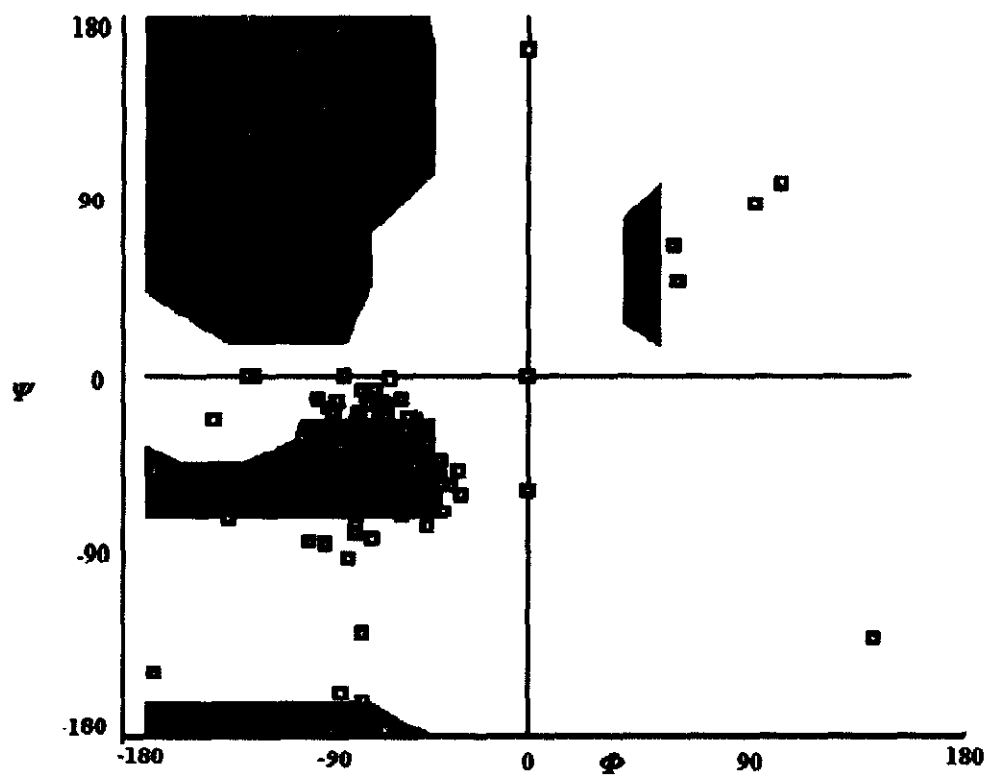


Figure 30: Ramachandran Plot of bR and the CT in proximity

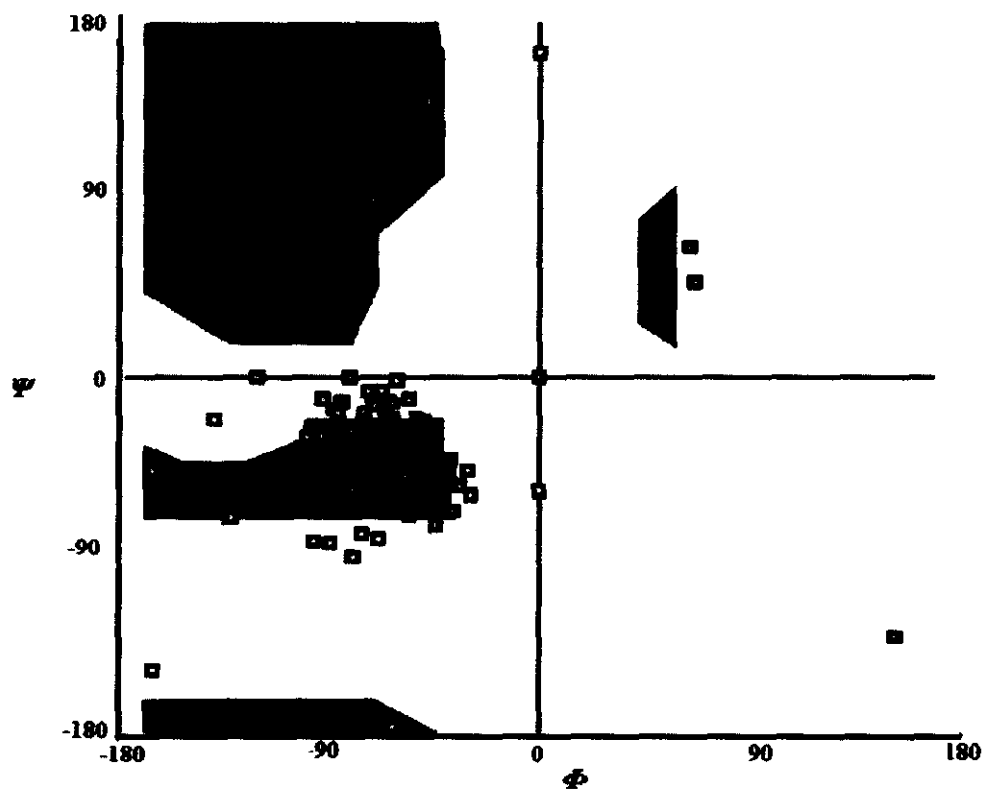


Figure 31: Ramachandran plot of bR without the CT in bR and the CT in Proximity



Figure 32: Ramachandran plot of the CT in bR and the CT in Proximity

2.4.3.2 Electrostatic Energies

In high concentration protein systems such as bR, prior reports suggest that significant lipid-protein electrostatic interactions occur in purple membrane (PM).¹³⁶ in order to help elucidate the effects of electrostatic lipid-protein interactions, we calculated the electrostatic energies between bR without the CT [bR], with the CT in proximity, and with the CT bound [bR (CT)]. In this section, we report the electrostatic interactions between bR and the lipid in the absence of and presence of unbound CT. Figure 33 shows a plot of the electrostatic energy interactions between bR and lipid in bR and the CT in proximity and bR without the CT models. It must be noted that the total electrostatic energy of the bR (CT) would be expected to more negative simply because they have more amino acids present to contribute to this energy. Moreover, it is clear that the bR system has significantly greater fluctuations over time than the bR (CT) system and so the trajectory averages are not a reliable metric for comparison. As described below for the bR-CT system (with the CT attached), the fact that the bR system

undergoes greater fluctuations is an indication that the presence of the CT may help promote enhanced stability over time. It is apparent that the CT induces some effect (either bound or unbound) that reduces fluctuations in the electrostatic interactions that could lead to stronger interactions of bR with the lipids membrane.

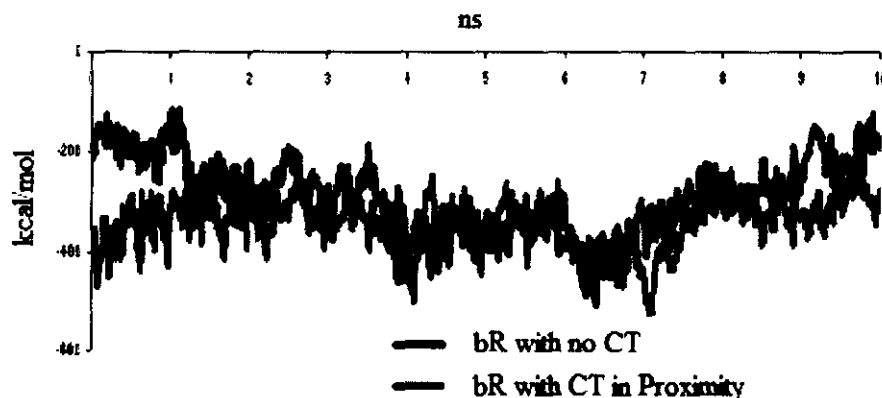


Figure 33: Electrostatic energies between membrane lipids and bR with CT in proximity and bR without the CT

2.4.3.3 Hydrogen Bonding

Previous studies have shown that there exist direct correlations between the number and type of H-bonds for various thermophilic and mesophilic proteins.¹³⁷ In our studies we investigate the change in H-bond content in bR with the CT present and compare with the CT absent. Some of these H-Bonds are between bR and the CT. In order to quantify the effect of the CT presence, we report the change in H-bonds in bR and between bR and the CT in Table V. The number of H-bonds external to bR between the molecule and the lipid was found to be 33.

Table V
of H-bonds between bR and the CT in proximity

Model	# H-bonds
bR in bR and CT in Proximity	211
CT in bR and CT in Proximity	11
bR and CT In Proximity (total)	220
Surplus	+2

2.4.3.4 Salt Bridges

Proteins from thermophiles and hyperthermophiles exhibit more, and frequently networked, salt bridges than proteins from their mesophilic counterparts.¹³⁸ Increasing the thermal stability of proteins by optimization of charge charge interactions is a good example for an evolutionary solution utilizing physical factors.

We calculated the number of salt bridges between bR with the CT in proximity. The results are shown in Table VI. It is evident that the number of salt bridges within bR significantly increases in the presence of the CT, even though none of these new bridges occur between the CT and the bR protein. We also note that one salt bridge appears between GLU237 and ARG227 in the CT and that no salt bridge was found for the CT alone. Since the occurrence of salt bridges indicates enhanced stability, it is apparent that the presence of the CT increases the overall stability of bR.

Table VI
Comparison of number, location, and type of salt bridges in bR alone and with the CT in proximity

System	Salt bridges
bR alone	GLU9-ARG7 GLU204-ARG82 ASP212-LYS216 ASP85-LYS216 GLU194-ARG134 GLU194-ARG82 ASP38-ARG164
bR with CT (proximity)	GLU204-ARG82 GLU237-ARG227 (within the CT) ASP212-LYS216 ASP38-LYS41 GLU166-ARG164 GLU9-ARG7 ASP102-ARG164 ASP85-LYS216 GLU194-ARG82 GLU194-ARG134 ASP102-LYS41 ASP38-ARG164

2.4.3.5 SASA analysis

A folded protein normally has its hydrophobic amino acid side chains shielded from the polar solvent. In this case, the intracellular loop regions of bR would be expected to be partially shielded from the solvent by the CT. Shown in Figure 34 is the SASA plot for bR and for bR with the CT in proximity and bR with no CT. The average SASA for bR without CT is 1156.14 Å².

These plots were made using a 1.4 Å probe sphere. It is clear that the CT shields bR and, as described further below, the SASA for the CT is also reduced. Hence, it is clear that the loop region of bR is shielded from the solvent.

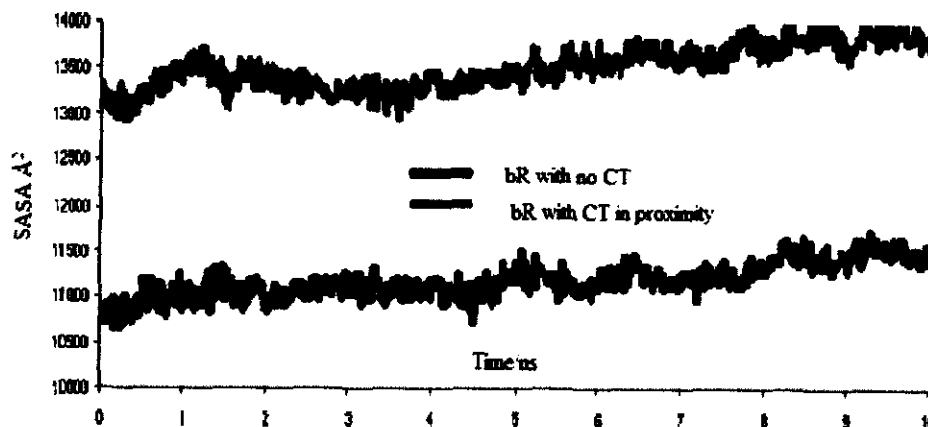


Figure 34: Average SASA of bR in the CT with proximity and bR with no CT

2.4.3.6 Heat Capacity

From the calculated SASA values, we distinguish the hydrophilic and hydrophobic SASA for bR alone and for bR with the CT in proximity. Then, by using the coefficients for the polar and nonpolar SASA parts of the proteins (0.32 and 0.14, respectively) reported by Spolar *et al.*,⁶⁹ and using the equation for calculating ΔC_p (see Appendix II), we report $\Delta C_p = +1.66$ cal/mol K. The results are shown in Table VII. Clearly the small value for ΔC_p indicates only small thermodynamic energy change for the solvent when the CT is not attached to the bR molecule. This is consistent with our result showing that the SASA value for free CT and that for bR alone are essentially the same as that for the CT (bR) system.

Table VII
Average SASA for hydrophobic and hydrophilic residues and calculated ΔC_p
during 10ns of simulations in 4M NaCl for bR without CT and bR with the CT in
proximity

Molecule	SASA_{np} (Å²)	SASA_p (Å²)	ΔC_p (cal/mol K)
bR without CT	6511.94	5048.13	
CT (in solvent)	1017.15	1383.07	
bR with CT close	7527.86	6416.52	
			+1.66

2.4.4 Analysis of bR with bound CT (bR-CT)

The final system investigated was the membrane bound bR molecule with the CT covalently bound. These investigations include an analysis of the structure, the presence of H-bonding and salt bridges, SASA analysis, and heat capacity of this model in 4M salt concentration.

2.4.4.1 Structure analysis

It is obvious that after 10 ns the CT appears to be stabilized by the presence of interactions between the CT and bR. To establish the presence of any new secondary structural motifs, we searched for likely α or β structural forms, as expected, using a Ramachandran plot. The Ramachandran plot for bR shows that it is mostly α helical as expected (Figures 35-37). The plot for the CT does not indicate any new α or β structure, also as expected. Our results indicate that the CT secondary structure is essentially “random” in the free form when near or attached to the bR molecule. Since XRD structure refinements do not clearly show the presence of the CT, our results are consistent with this, albeit negative conformation.

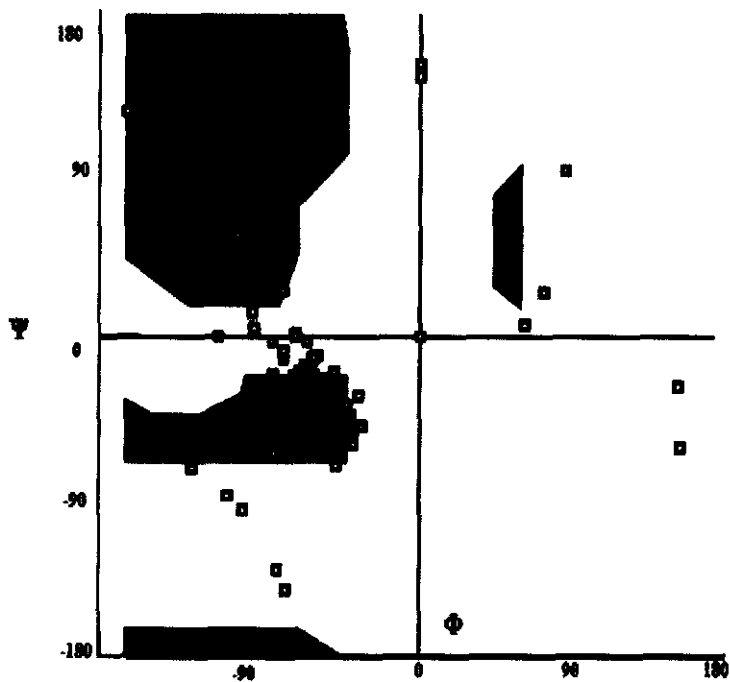


Figure 35: Ramachandran Plot of bR and CT bound

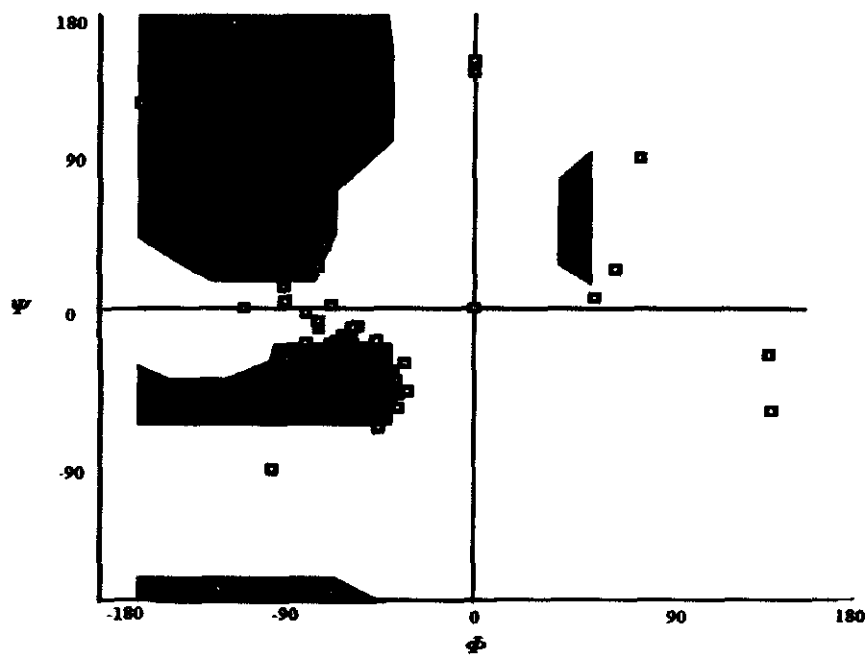


Figure 36: Ramachandran Plot of bR without CT in bR and CT bound Model

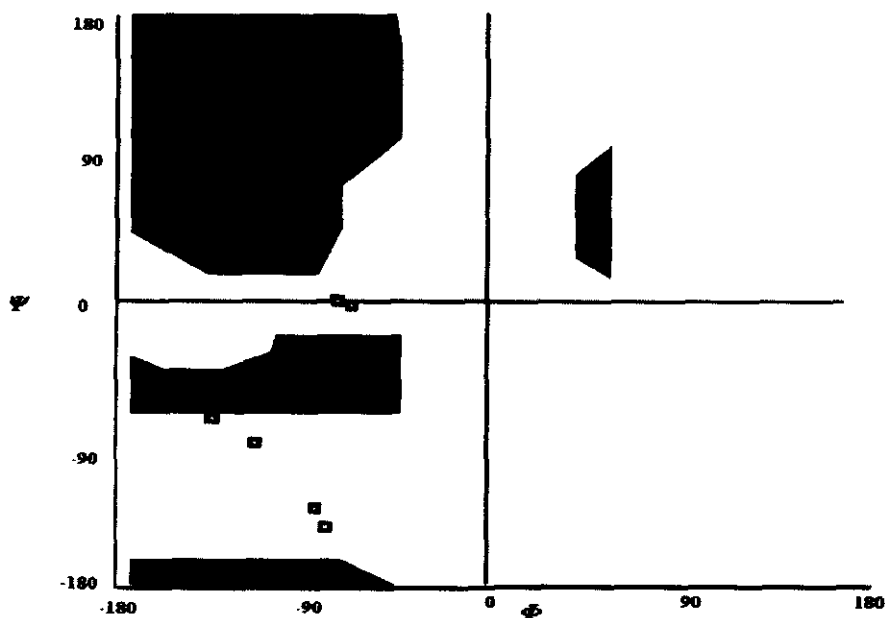


Figure 37: Ramachandran Plot of CT in bR and the CT bound model

2.4.4.2 Electrostatic Interactions

As previously discussed in 2.4.3.2, a significant lipid-protein interaction may occur between bR-CT and the lipid membrane.¹³⁶ In order to further elucidate the effects of the presence of the CT on electrostatic lipid-protein interactions, we calculated the electrostatic energies between bR-CT and the lipid membrane. In this section we compare the electrostatic interactions between bR-CT with the lipid and bR with the lipid. Figure 38 shows a plot of this comparison. As previously discussed in section 2.4.3.2 for the bR and bR(CT) systems, the trajectories showed a high degree of fluctuation during the run, particularly for the bR without CT (in blue). This may indicate that the presence of the CT reduces structural fluctuations over time which may lead to an overall increase in stability. A direct comparison of the electrostatic energy indicates that the bR-CT system was generally more negative (indicating greater stability) compared with the bR

alone. However, since there was greater fluctuation in the bR compared to the bR-CT system, a simple time average of the electrostatic energy is not a reliable metric. It is apparent, however, that the CT induces some structural modifications that lead to reduced fluctuations over time. This appears to be the case whether the CT is absent or is present but not bound to the bR molecule. Again, even though the reason for this remains uncertain, this could be a contributing factor in the experimental finding by Turner *et al.*¹²³ that the CT helps increase overall stability of the protein.

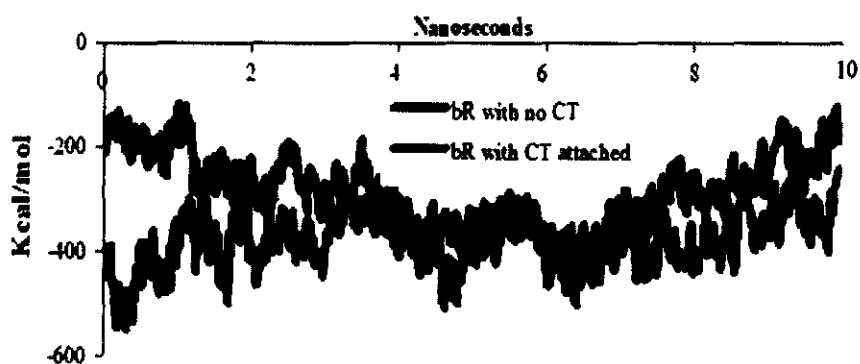


Figure 38: Electrostatic energies between membrane lipids and bR with CT attached

2.4.4.3 Hydrogen Bonding

Previous studies have shown that there exist direct correlations between the number and type of H-bonds for various thermophilic and mesophilic proteins and protein stability.¹³⁷ In our studies, we investigated the change in H-bond content in bR-CT with bR. These values are collected in Table VIII. We find that the number of H-bonds in the entire CT-bR complex contains 17 additional H-bonds compared with the CT or bR portions separately. These 17 additional H-bonds all occur between the CT and

the bR. Hence, we observe that there is some additional stabilization afforded by the presence of the CT in the form of H-bonds.

Table VIII
Number of H-bonds in studied systems

Model	# H-bonds
bR in bR-CT	263
CT in bR-CT	26
bR-CT (total)	272
Surplus	+17

The number of H-bonds external to bR between the protein and the lipid was found to be 40. This is greater than either of the systems of bR alone (26) or bR(CT) (33).

2.4.4.4 Salt Bridges

Thermophile and hyperthermophile proteins frequently exhibit more salt bridges than mesophilic proteins, as we mentioned in section 2.4.3.4.¹³⁸ We calculated the number of salt bridges between bR and CT in bR-CT. The results are shown in Table IX. It is evident that the number of salt bridges within bR significantly increases in the presence of the CT and three of these new bridges (indicated in red) occurred between the CT and the bR protein. Since the occurrence of salt bridges promotes enhanced stability, it is apparent that the presence of the CT increases the overall stability of bR through the formation of salt bridges as well.

Table IX
Salt bridges between CT and bR

System	Salt bridges
bR alone	GLU9-ARG7 GLU204-ARG82 ASP212-LYS216 ASP85-LYS216 GLU194-ARG134 GLU194-ARG82 ASP38-ARG164
bR-CT	ASP96-LYS41 GLU166-ARG164 GLU237-ARG225 GLU204-ARG82 ASP212-LYS216 GLU9-ARG7 ASP38-LYS41 ASP242-ARG225 ASP85-LYS216 GLU194-ARG134 ASP36-ARG164 GLU237-ARG227 ASP38-ARG164

2.4.4.5 SASA analysis

In this section we describe comparisons of SASA measurements (see Appendix II) between bR-CT and bR. Figure 39 shows the SASA trajectory for bR-CT and bR. The average SASA for bR is 111,56.1 Å² and for bR-CT it is 13,616.1 Å². From section 2.4.1.4, the SASA for the CT alone was 2,476.6 Å²; hence, the solvation surface of the CT when it is attached to bR is essentially the same. This indicates that even though we find clear evidence of H-bond and electrostatic interactions between the CT and bR, the

space between the CT and the bR loop region is still solvated to roughly the same extent as the CT alone in solvent.

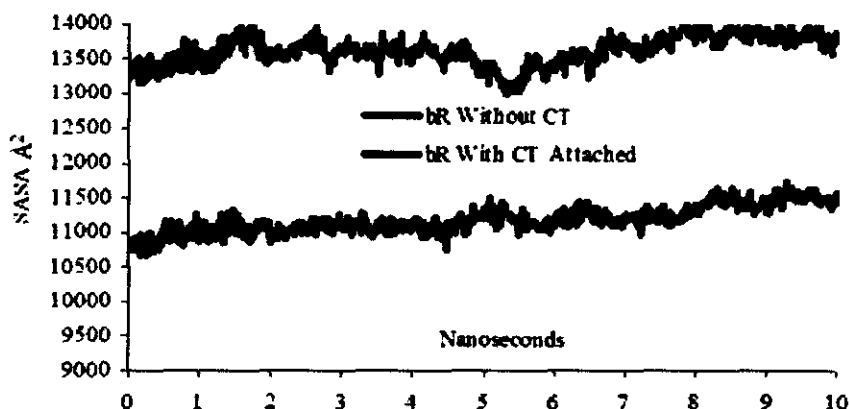


Figure 39: Average SASA of bR in the presence and absence of CT

2.4.4.6 Heat Capacity

We distinguished the SASA for the hydrophilic and hydrophobic residues of bR and CT for the bR-CT system. Using 0.32 and 0.14 for the polar and nonpolar residues respectively (Appendix II), we obtain values for ΔC_p as shown in Table X.

Table X
Average SASA for hydrophobic and hydrophilic residues and calculated ΔC_p for bR and bR-CT

Molecule	SASA _{np} (Å ²)	SASA _p (Å ²)	ΔC_p (cal/mol/ K)
bR without CT	6511.94	5048.13	
CT (in solvent)	1017.15	1383.07	
bR with CT bound	7759.63	6297.57	
			+55.06

These results indicate that, compared to bR (CT), the attached CT-bR system has a larger (measurable) ΔC_p , indicating that there exists a greater degree of association (as

indicated by the exclusion of solvent between the CT and bR) between the CT and bR when the CT is bound.

2.5 Concluding Remarks

Experimental results¹²³ suggested that the CT plays a crucial role in the overall stability and function of bR.

- The solvent accessible surface area and the root-mean-square-displacement analyses indicated that the bR-CT is more stable in 4 M NaCl than 1M NaCl, which is consistent with experimental results.
- The sum of the SASA for the CT alone and for bR alone is approximately equivalent to the SASA of bR-CT bound. This suggests that any interaction resulting from H-bonding or other electrostatic forces does not require the exclusion of solvent from the interfacial region between bR and the CT.
- Heat capacity analysis indicated that ΔC_p was more positive for bR with the CT bound. This indicates that bR is stabilized by the presence of the CT.
- Electrostatic (salt bridges, H-bonding) analyses indicate significant interaction between the CT and bR.
- The degree of H-bonding between the transmembrane part of bR and the lipid bilayer was found to increase as a function of the presence of the CT, even though no particular H-bonding was noticed between the CT and the lipids.

These results demonstrate that the CT does help stabilize the bR molecule in general and causes greater interactions between the lipid membrane and the transmembrane portion of bR. The experimental results, which indicate that truncation or elimination of the CT leads to loss of stability, are confirmed. The likely causes of this enhanced stability are manifest in several interactions that appear to be directly correlated with the presence of the CT.

Appendix I

I. Molecular Dynamics Simulations

Molecular dynamics (MD) simulations¹³⁹ continue to contribute a considerable share in molecular biology. MD programs extensively used are CHARMM,¹⁴⁰ GROMOS,¹⁴¹ and AMBER,¹⁴² which are capable of modeling systems consisting of up to 10^5 or more atoms. An essential approach of reducing the computational demand is via parallel computer architectures. Numerous MD programs, such as X-PLOR and CHARMM,¹³² have been customized to allow them to exploit the power of parallel computers. These programs have undergone significant revision from the original serial code in order to allow parallel computational methods to be employed. In contrast, the program NAMD¹³¹ was specifically designed for distributed memory parallel computers and so has become the application of choice for most modern simulations.

The program makes use of a spatial decomposition scheme to split the region in a manner that put forward the peak scalability. Self-regulating threads of control are consumed to offer extendable load-balancing capacities at the same time sustaining a non-problematical, even decomposition design. Message-driven scheduling can be put into operation to put together the execution of these threads of control in a manner that weakens the power of communication latency. These regulations lead to an elevated performance between large amounts of processors. The goal of MD simulations is to record the evolution of an ensemble subject to a set of thermodynamic constraints using well-validated equations of motion.¹⁴³ The ensemble may be any set of particles, which can be defined in terms of their fundamental properties. At basis, the state of the system would be defined in terms of a few fundamental properties including mass (m), position ($\mathbf{r}(x, y, z)$), velocity ($\mathbf{v}(\mathbf{r})$) and forces acting on them ($\mathbf{F}(m \cdot \dot{\mathbf{v}})$) yielding nine variables plus mass for each particle. The standard notation of bold for vectors and dots for time

derivatives are used throughout. Once the forces are computed, the Newtonian equations of motion are used to determine a displacement vector that would relocate each particle at $\mathbf{r}(t)$ to a new location at $\mathbf{r}(t + dt)$ in a given length of time, dt (timestep). The process is repeated until the ensemble evolves to an equilibrium state and a sufficiently long trajectory is generated to allow property calculations from it. Starting from an initial ensemble configuration, the system follows a path toward least total energy.¹⁴³ The integrated Newton's equations of motion, as applied to MD simulations take the form of the following relations:

$$\begin{aligned}\mathbf{r}(t + dt) &= \mathbf{r}(t) + \mathbf{v}(t)dt + \frac{1}{2}\mathbf{a}(t)dt^2 \\ \mathbf{v}(t + dt) &= \mathbf{v}(t) + \mathbf{a}(t)dt \\ \mathbf{a}(t) &= \frac{1}{m}\mathbf{F}(t) \\ \mathbf{F}(t) &= -\frac{dV}{dt}\end{aligned}$$

To determine a new location at $\mathbf{r}(t + dt)$, one needs to determine the force on the particle from the fourth equation based on the time derivative of a predetermined potential V . This force is then used to calculate the acceleration (\mathbf{a}), the velocity (\mathbf{v}) and location $\mathbf{r}(t + dt)$. In the initial step, no velocities are set and the usual approach is to set the initial velocities using a stochastic algorithm (based on a random number).

To determine the resulting forces on the ensemble components, several levels of approximation are typically used as dictated by the complexity of the ensemble and the properties sought. For small systems ($10^1 - 10^2$ atoms), quantum mechanical (QM) approximations can be used, particularly if electronic properties are needed. In this case, the forces would be computed from solution of the many-body Schrödinger equation. The ensembles investigated in this work encompass on the order of $10^3 - 10^5$ particles,

which require the use of classical models and parameterized forcefields. The advantage of classical MD methods over QM methods are that long simulation times of several millions of time steps can be accomplished over a reasonable timeframe (from days to weeks).

The timesteps typically used in MD simulations are of the order of 1 to 2 femtoseconds. Such a short timestep helps ensure that the typical vibrational periods ($10^{-12} - 10^{-14}$ s) are well resolved. The simulation (wall clock) time can be estimated from the following scenario: There are 9 variables that must be handled for each atom (3 position, 3 velocity, and 3 force). For 10^4 atoms this gives about 10^5 variables that must be processed by the computer. In addition, the intermolecular (two-body) interactions are also processed for each time step. These interactions may amount to 10^3 or so. Typical processor speeds are 10^9 floating point operations per second (flops). Hence, to process one timestep it takes approximately 10^5 operations or approximately 10^{-4} seconds. For a nanosecond (10^9 timesteps) it takes 10^5 seconds or approximately one day of wall clock time per nanosecond of simulation time. Using advanced algorithms incorporated into NAMD, typical processing speeds for the simulations reported in this work were approximately 2 days/ns.

Classical MD parameterized forcefields have been developed to model the most important interactions among the particles in the ensemble, including long-range electrostatic (Coulomb) forces, short-range dispersive forces (van der Waals) between molecular particles, and intra-molecular forces (2-body bond length, 3-body angle, and 4-body dihedral and improper angles). These forcefields are superior to hard-sphere elastic interaction potentials, which cannot include soft interactions that significantly contribute

interaction potentials, which cannot include soft interactions that significantly contribute to the thermodynamic properties of the ensemble. In the following section, the details of the inter- and intra-molecular forcefields are described.

I.A Potential Energy Functions

i. Intra-molecular Forcefield

Potential energy functions correspond to all bonded and non-bonded interactions of the system through the use of parameterized forcefields. These forcefields are generally produced using quantum mechanical methods, which are used to determine the forces and charges of all atoms in a particular molecule. The potentials are well validated against published experimental results of several properties. Among the available forcefield potentials are the AMBER,¹⁴² CHARMM,^{133, 144, 145} and GROMACS¹⁴⁶ force fields.

In our study we used CHARMM potential function and we added the forcefield for retinal moiety of bR, which was optimized to work with CHARMM forcefield.

ii. CHARMM Potential Energy Function

The form of the potential energy function, which is used in CHARMM for intra- and intermolecular interactions of atoms, is shown below:

$$\begin{aligned}
 V = & \sum_{\text{bonds}} \kappa_b (r - r_0)^2 + \sum_{\text{angles}} \kappa_\theta (\theta - \theta_0)^2 + \sum_{\text{dihed}} \kappa_\phi [1 + \cos(n\phi - \delta)] \\
 & + \sum \kappa_\omega (\omega - \omega_0)^2 + \sum \kappa_u (u - u_0)^2 \\
 & + \sum_{\text{nonbonded}} \epsilon \left[\left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{4 \epsilon_0 r_{ij}}
 \end{aligned}$$

moved. The second term in the equation accounts for the bond angles where k_θ the angle force is constant and $\theta - \theta_0$ is the angle from equilibrium between 3 bonded atoms.

The third term is for the dihedral (torsion angles) where k_ϕ is the dihedral force constant, n is the multiplicity of the function, ϕ is the dihedral angle and δ is the phase shift. The fourth term accounts for the improper angles, i.e., out of plane bending, where k_ω is the force constant and $\omega - \omega_0$ is the out of plane angle. The Urey-Bradley component (cross-term accounting for angle bending using 1-3 nonbonded interactions) comprises the fifth term, where k_u is the respective force constant and u is the distance between the 1 and 3 atoms in the harmonic potential.

Nonbonded interactions between pairs of atoms i, j are represented by the 6th and 7th terms. By definition, the nonbonded forces are only applied to atom pairs separated by at least three bonds. The van der Waals non-bonded energy is calculated with a standard 12-6 Leonard-Jones potential and the electrostatic energy with a Coulombic potential. In the Leonard-Jones potential above, σ is the separation distance in the function where the potential goes to zero and ϵ is the energy minimum at the bottom of the potential well.

The electrostatic interactions are handled by the 7th term, which is essentially Coulomb's law. The only fitting parameter in this term is the ionic charge assigned to each atom. These charges are generally determined from quantum mechanical calculations. Since the Coulombic term is very long range, it is generally necessary to include large interaction distances in its calculation. This can result in a significant slowdown in the calculation speed and, therefore, alternative methods have been developed.

The most common method to improve electrostatic computational speed is the Ewald method.¹⁴⁷ This is a technique for calculating the interaction energies of periodic systems. This summation is a unique case of the Poisson summation formula, substituting the summation of interaction energies in real space with an equivalent summation in Fourier space. The benefit of this method is the fast convergence of the Fourier-space summation relative to its real-space analogue particularly when the real-space interactions are long-range. Because electrostatic energies include both short- and long-range interactions, it is greatly beneficial to separate the interaction potential into a short-range fraction summed in real space and a long-range part summed in Fourier space.

I.B Ensemble Types

An ensemble is a collection of all possible thermodynamic systems, which have diverse microscopic states but the same macroscopic or thermodynamic state. The properties of an ensemble are subject to specific constraints as listed below:¹⁴³

- **Microcanonical ensemble (NVE):** The thermodynamic state can be described by a fixed number of atoms (N), volume (V) and energy (E).
- **Canonical Ensemble (NVT):** The thermodynamic state is defined with a fixed number of atoms (N), volume (V), and temperature (T).
- **Isobaric-Isothermal Ensemble (NPT):** This ensemble is characterized by a fixed number of atoms (N), pressure (P), and temperature (T).
- **Grand canonical Ensemble (μVT):** The thermodynamic state for this ensemble is distinguished by a fixed chemical potential (μ), volume (V), and temperature (T).

These ensembles and the corresponding statistical mechanical partition function are listed below:

Table XXI
Different kinds of ensembles

Ensemble	Constraint	$kT \cdot \ln(q)$
μ -Canonical	N, V, E	ST
Canonical	N, V, T	-A
Isobaric-isothermal	N, P, T	-G
Grand	μ , V, T	PV

I.C Controlling Pressure and Temperature in Ensembles

In order to control the temperature and pressure in canonical, grand, and isobaric-isothermal ensembles, several algorithms have been developed, including Berendsen,¹⁴⁸ Nose-Hoover,¹⁴⁹ and , Langevin.¹³⁵ Since the temperature is a function of the atom velocities:

$$E = \frac{3}{2} k_B T = \frac{1}{2} m v^2$$

$$T = \frac{1}{3} \frac{m v^2}{k_B}$$

A common method for controlling temperature is to scale the velocities (up or down) in succeeding timesteps to maintain the temperature setpoint. Such modifications, on the other hand, can trigger the system to perform in a non -Newtonian behavior. So the Berendsen technique is able to diminish the rate of scaling by pairing to an outside temperature source. The Nose- Hoover scaling temperature applies a thermal reservoir

and friction expressions to the equations of motion. For pressure control, the box vectors and atom coordinates are scaled at every time step.

I.D Periodic Boundaries

Periodic boundary conditions make it possible for a simulation to be completed by a reasonably small number of particles using a procedure in which the particles are exposed to forces as if they were in a bulk fluid. Periodic boundary conditions are employed using a box of particles, which are replicated in all directions to give a periodic array. Increasing or decreasing integral multiples of the lattice vectors easily calculates the coordinates of the particles in the image boxes. If a particle leaves the box during the simulation, it is substituted with an image particle that comes in from the opposite side, as illustrated in Figure I.1. Hence, the number of particles inside the central box is conserved.

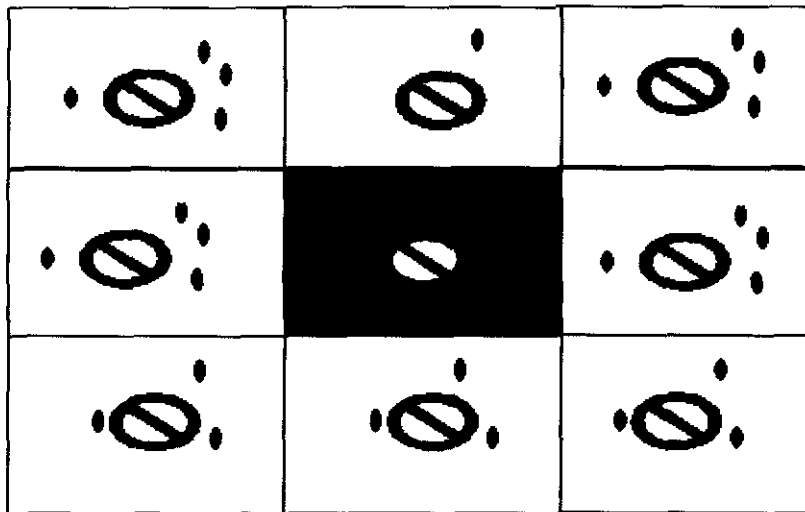


Figure I.1: periodic boundary conditions

I.E General Scheme for Running Molecular Dynamics Simulations

i. Initiation

The initial state of the system is described by a set of position coordinates for each atom in the system. The initial state is constructed from a published structure for the molecular solute (protein or nucleic acid) from which the rest of the systems (lipids, solvent, ions) is constructed. The PDB database¹⁵⁰ contains the necessary initial state file and describes the 3-D starting positions of the biomolecules. It is also possible to make a theoretical structure by homology modeling and use that as the starting point. The initial structure should be as close to the desired structure as possible. In some cases (e.g., the CT), there is no initial structure but one can be built from the primary sequence provided the expected tertiary structure is not overly complex. The coordinate structure for the system is contained in a file formatted in the Protein Database structure (pdb).

Next, one has to create a molecular topology of the molecule, which includes bond, angle, and dihedral, and charge information. In the NAMD/VMD^{68,131} application, this information is contained in a formatted Protein Structure file (psf) that can be generated using VMD. The pdb and psf files are used to add solvent and ions to the system to create the simulation input files used by NAMD.

ii. Equilibration

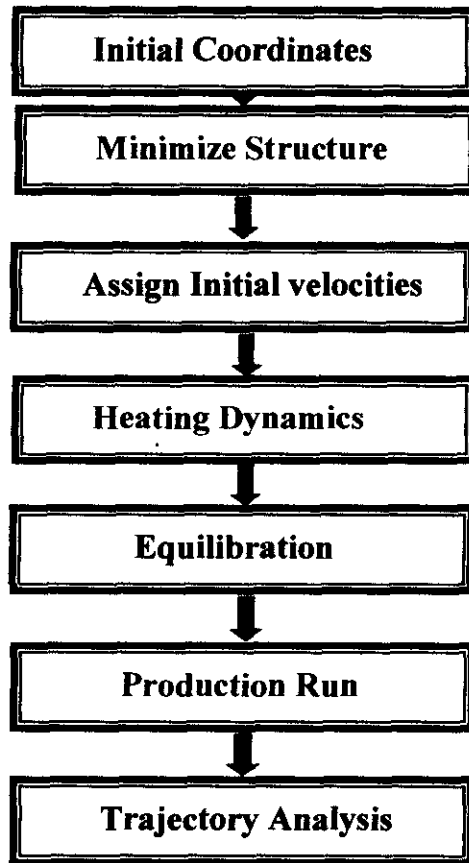
Since the initial state of the system is often well away from equilibrium, it is essential to carefully bring the system into thermal, barometric, and energetic equilibrium through a series of equilibration steps. The general method is to relax the intermolecular positions first while holding the internal degrees of freedom (bond lengths, angles, etc.) fixed. In systems that contain solute secondary structures that were not originally obtained from structural methods (e.g., the CT and membrane lipids), it is advisable to

conduct thermal annealing of the system. This is accomplished by running simulations at elevated temperatures under fixed volume (NVT) conditions so as to remove initial state artifacts that could bias the final state. Once the system equilibrates at elevated temperature, it is thermally quenched in a series of steps down to the desired simulation temperature (typically 300-310K).

The equilibration phase of the simulation includes a run at fixed internal coordinates for the molecular solute and solvent followed by relaxation of the internal coordinates, which results in a fully equilibrated system ready for the production runs that yield the final results.

iii. Production phase

The simulation proceeds in the production period for a time that is determined by the type of analysis needed. This can be from several hundred picoseconds to tens of nanoseconds during the production phase until the thermodynamic parameters are stabilized and sufficient trajectory states are recorded to allow reliable trajectory averages to be obtained. The general scheme for a simulation is described in the following flowchart.



Flowchart 1: General flowchart for designing molecular dynamics simulations

iv. Standard units

When NAMD is in use, most important facts on the progress of the simulation can be demonstrated in a regular output on the console but this output might be written to a file (log file). The units in all output files report values in the following units:

Energy:	kcal/mol
Time	fs
Length	Å
Volume	Å ³
Pressure	bar

Appendix II

II. Analysis Methods

Various software packages other than NAMD were used to evaluate simulation results including:

- VMD⁶⁸ which is a molecular graphics program for visualizing trajectories and analyzing results
- 2-D plotting program xmgrace, Matlab, Mathematica, or Excel for plotting data.

A. Simulation properties

i. Root-mean-square-displacement (RMSD)

The RMSD distinguishes the extent that a particular particle translates from a defined reference point in the simulation system. The RMSD is calculated according to:

$$RMSD = \sqrt{\frac{\sum_{t_0}^{t_f} (r(t) - r(t_0))^2}{N}} = \sqrt{\langle \Delta r^2 \rangle}$$

Where $r(t_0)$ is the reference position, $r(t)$ is the location of a particle in timestep, t , and N is the simulation step.

ii. Diffusion Constant

The diffusion constant D is a measure of the rate of translation for a particle in the system and is calculated according to:

$$D = \lim_{t \rightarrow \infty} \frac{\langle r^2(t) \rangle}{6t} = \frac{RMSD^2}{6t}$$

Where t is the time of the current timestep.

iii. Interaction energies

The interaction energy between different solute molecules and between solute and solvent can be calculated over the course of the trajectories. This is accomplished by

selecting the type or class of molecules to be considered for the calculation and are easily selected using the VMD package.

iv. Ramachandran plots

Protein secondary structure can be predicted from two angles in the peptide backbone. The two angles are shown below.

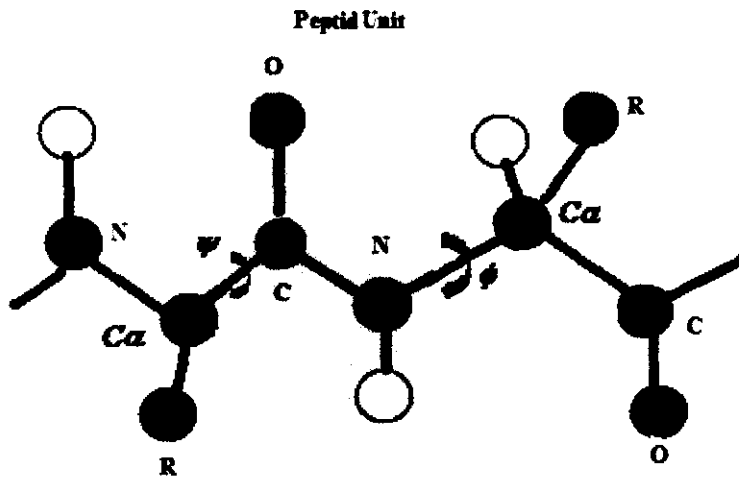


Figure II.1: Alpha and psi angle in protein

It has been shown that particular combinations of these angles leads to particular secondary structure (e.g., beta-sheet, alpha-helix) and a plot of the angle Ψ as a function of Φ , referred to as a Ramachandran plot¹⁵¹, can be used to determine the propensity for a particular type of secondary structure. An example of a Ramachandran plot is shown in Figure II.1. The regions of the plot, which suggest the various types of secondary structure, are indicated. In the example plot, a much large proportion of the protein is seen to be in the α -helix region as would be expected for this protein (bR).

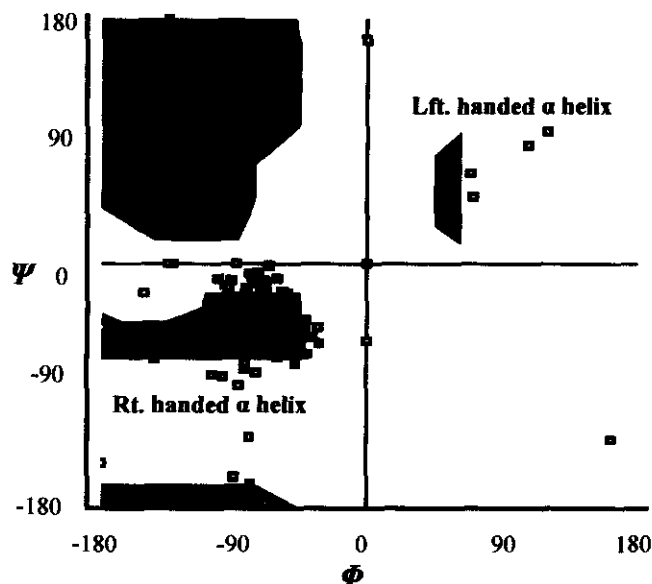


Figure II.2: A Ramachandran Plot of a protein

v. H-bonds

Formation of H-bonds between donors and acceptors is defined by two parameters: the distance between the donor and acceptor, and 2) the angle defined by D—H—A. In our measurements, the distance cutoff was set at 3.0 Å and the maximum angle was set to 20°.

vi. Solvent accessible surface area (SASA)

The solvent accessible surface area is defined as the surface area that would be exposed to the solvent. In general, the SASA is measured by assuming that the solvent molecules are spheres with a specific radius. These spheres are “rolled” over the solute surface such that the minimum separation between the solvent and solute atoms is the sum of the solvent radius and the van der Waals radius of the solute atoms. The measurement is described in Figure II.

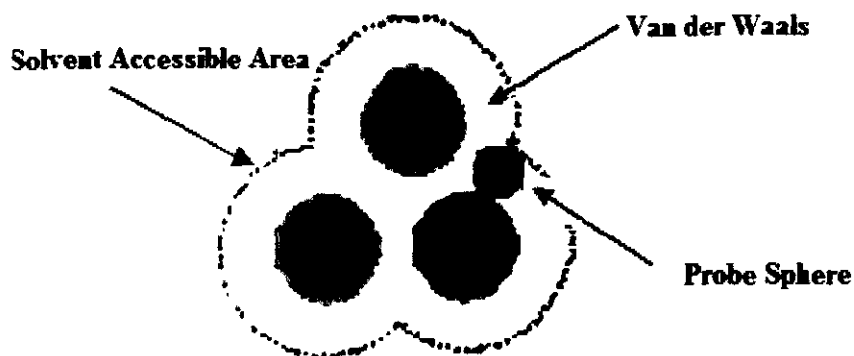


Figure II.3 Calculating Solvent Accessible Surface Area

vii. Heat capacity

Calculation of the heat capacity at constant pressure (C_p) can be used to directly compare with experimental DSC and/or ITC results. In general, a straightforward but difficult method to accomplish this is to use the trajectory energy fluctuations to determine the C_p directly. From a trajectory, one can determine the trajectory average energy \overline{H} and the energy of each step H_i to determine the heat capacity.¹⁴³

$$C_p = \frac{\sum_i (H_i - \overline{H})^2 / n}{k_B T^2} = \frac{\overline{(H - \overline{H})^2}}{k_B T^2} = \frac{\overline{H^2} - \overline{H}^2}{k_B T^2}$$

Although this method will work, it turns out that in order to get a reliable value for C_p , the trajectories must be inordinately long (at least hundreds of ns). Fortunately, an alternative method was developed that does not require extremely large trajectory data sets and is particularly useful for native \rightarrow unfolded phase transitions in biomolecules.¹⁵² In this method one must recognize the physical differences between the native folded and unfolded states. In the native (folded) state most (all) nonpolar residues are buried inside the protein so the solvent is only exposed to polar residues. Since in the isothermal-isobaric ensemble framework:

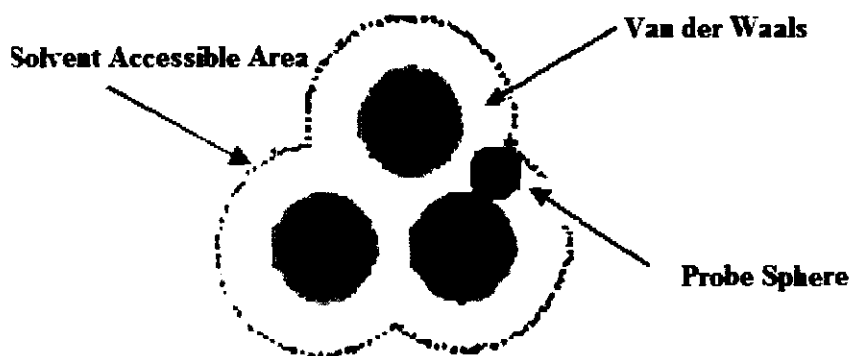


Figure II.3 Calculating Solvent Accessible Surface Area

vii. Heat capacity

Calculation of the heat capacity at constant pressure (C_p) can be used to directly compare with experimental DSC and/or ITC results. In general, a straightforward but difficult method to accomplish this is to use the trajectory energy fluctuations to determine the C_p directly. From a trajectory, one can determine the trajectory average energy \overline{H} and the energy of each step H_i to determine the heat capacity:¹⁴³

$$C_p = \frac{\sum_i (H_i - \overline{H})^2 / n}{k_B T^2} = \frac{\overline{(H - \overline{H})^2}}{k_B T^2} = \frac{\overline{H^2} - \overline{H}^2}{k_B T^2}$$

Although this method will work, it turns out that in order to get a reliable value for C_p , the trajectories must be inordinately long (at least hundreds of ns). Fortunately, an alternative method was developed that does not require extremely large trajectory data sets and is particularly useful for native \rightarrow unfolded phase transitions in biomolecules.¹⁵² In this method one must recognize the physical differences between the native folded and unfolded states. In the native (folded) state most (all) nonpolar residues are buried inside the protein so the solvent is only exposed to polar residues. Since in the isothermal-isobaric ensemble framework:

$$S = k_B \ln \Delta + k_B T \left(\frac{\partial \ln \Delta}{\partial T} \right)_{N,p} \quad \text{and} \quad C_p = \left(\frac{\partial S}{\partial \ln T} \right)_{N,p} \quad \text{we obtain:}$$

$$\begin{aligned} C_p &= k_B \frac{\partial \ln \Delta}{\partial \ln T} + k_B \frac{\partial \ln \Delta}{\partial T} \cdot \frac{\partial T}{\partial \ln T} + k_B T \frac{\partial \left(\frac{\partial \ln \Delta}{\partial T} \right)}{\partial \ln T} \\ &= k_B \frac{\partial \ln \Delta}{\partial \ln T} + k_B T \frac{\partial \ln \Delta}{\partial T} + k_B T^2 \frac{\partial^2 \ln \Delta}{\partial T^2} \\ &= k_B \frac{\partial \ln \Delta}{\partial \ln T} + k_B \frac{\partial \ln \Delta}{\partial \ln T} + k_B T^2 \frac{\partial^2 \ln \Delta}{\partial T^2} \\ &= 2k_B T \frac{\partial \ln \Delta}{\partial T} + k_B T^2 \frac{\partial^2 \ln \Delta}{\partial T^2} \end{aligned}$$

which relates C_p to the partition function Δ . The value of Δ varies with the population of occupied quantum states, including translational, vibrational, rotational, and electronic states. Of these, the vibrational states make the largest contribution to the change in C_p since the other modes do not change (comparatively) as much during the transition. There are two primary changes that occur during the unfolding process. 1) The vibrational states of the folded molecule in the native state are coupled and form somewhat higher frequency modes. Upon unfolding to the denatured state, these modes become less coupled and shift to lower frequencies transferring their energy to other bond vibrational modes. The increase in population of the higher frequency bond vibration modes leads to a larger Δ and, consequently, a larger C_p compared to the native state. 2) Upon unfolding, residues that were previously shielded from the solvent become exposed to the polar solvent environment. Solvent molecules must form structured solvation shells around these nonpolar residues leading to fewer available solvent states and thus lower entropy. Importantly, the structure of these solvation shells decreases significantly with increasing temperature leading to a large positive $C_p = \frac{\partial S}{\partial \ln T}$, for the “upper” unfolded state relative to the “lower” folded state.¹⁵³⁻¹⁵⁶

This formulism can be used to estimate the change in C_p upon transition from the native to unfolded state. The method requires information on the change in the SASA of the polar and nonpolar residues in the two states:^{153,154}

$$\Delta C_p = B_{np} \Delta A_{np} - B_p \Delta A_p$$

In this equation, ΔA_{np} and ΔA_p are the changes in SASA for the nonpolar and polar residues, respectively, for the melting transition. B_{np} and B_p are fitting parameters taken from experimental values for the type of biomolecules under consideration. Typical values are $B_{np} = 0.32$ and $B_p = 0.14$.⁶⁹ More data has been collected for proteins than nucleic acids and so the method has been mostly applied to the former.

viii. Salt bridges

Electrostatic interactions provide a standard method to determine the strength of interactions among charged groups. Their ionization state is dependent on pH and on the particular local environment, which includes interactions between the partial charges of polar but non-ionizable groups (dipoles) and formally charged groups. Generally, charged groups often are exposed to the solvent. In many cases, these negative and positive charged groups are sufficiently near each other to develop ionic pairs or salt bridges. It must be considered that charge-charge attractive or repulsive forces could be significant even if the geometrical centers of the functional groups are divided by spaces in excess of 4 Å which is the distance commonly applied to determine the presence of a salt bridge. Long-range electrostatic forces are significant, not only for the overall stability of protein secondary and tertiary structure, but also are important for function.

REFERENCES

1. Nelson, D.; Cox, M., *Lehninger: Principles of Biochemistry*. WH Freeman: **2005**.
2. Watson, J. D.; Crick, F. A structure for deoxyribose nucleic acid. *Nature*. **1953**, 171, 737-8.
3. Gellert, M.; Lipsett, M. N.; Davies, D. R. Helix Formation by Guanylic Acid. *P.N.A.S.* **1962**, 48, (12), 2013-18.
4. Simonsson, T.; Sjöback, R. G-quadruplex DNA structures--variations on a theme. *Biol. Chem.* **2001**, 382, (4), 621-8.
5. Ou, T.m.; Lu, Y.j.; Tan, J.h.; Huang, Z.S.; Wong, K.Y.; Gu, L.q. G-Quadruplexes: Targets in Anticancer Drug Design. *ChemMedChem*. **2008**, 3, (5), 690-713.
6. Huppert, J. L.; Balasubramanian, S. G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res.* **2007**, 35, 406-13.
7. Blackburn, E. H. Switching and Signaling at the Telomere. *Cell*. **2001**, 106, (6), 661-73.
8. Parkinson, G. N.; Lee, M. P.; Neidle, S. Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature*. **2002**, 417, (6891), 876-80.
9. Redon, S.; Bombard, S.; Elizondo-Riojas, M. A.; Chottard, J.-C. Platinum cross-linking of adenines and guanines on the quadruplex structures of the AG₃(T₂AG₃)₃ and (T₂AG₃)₄ human telomere sequences in Na⁺ and K⁺ solutions. *Nucleic Acids Res.* **2003**, 31, 1605-13.
10. He, Y.; Neumann, R. D.; Panyutin, I. G. Intramolecular quadruplex conformation of human telomeric DNA assessed with ¹²⁵I-radioprobng. *Nucleic Acids Res.* **2004**, 32, 5359.
11. Rujan, I. N.; Meleney, J. C.; Bolton, P. H. Vertebrate telomere repeat DNAs favor external loop propeller quadruplex structures in the presence of high concentrations of potassium. *Nucleic Acids Res.* **2005**, 33, 2022-31.
12. Xu, Y.; Noguchi, Y.; Sugiyama, H. The new models of the human telomere d[AGGG(TTAGGG)₃] in K⁺ solution. *Bioorganic & Medicinal Chemistry* **2006**, 14, (16), 5584-91.

13. Ambrus, A.; Chen, D.; Dai, J.; Bialis, T.; Jones, R. A.; Yang, D., Human telomeric sequence forms a hybrid-type intramolecular G-quadruplex structure with mixed parallel/antiparallel strands in potassium solution. *Nucleic Acids Res.* **2006**, *34*, (9), 2723-35.
14. Luu, K. N.; Phan, A. T.; Kuryavyi, V.; Lacroix, L.; Patel, D. J. Structure of the Human Telomere in K⁺ Solution: An Intramolecular (3 + 1) G-Quadruplex Scaffold. *J. Am. Chem. Soc.* **2006**, *128*, (30), 9963-70.
15. Xue, Y.; Kan, Z.y.; Wang, Q.; Yao, Y.; Liu, J.; Hao, Y.-h.; Tan, Z. Human Telomeric DNA Forms Parallel-Stranded Intramolecular G-Quadruplex in K⁺ Solution under Molecular Crowding Condition. *J. Am. Chem. Soc.* **2007**, *129*, (36), 11185-91.
16. Stewart, S. A.; Weinberg, R. A. Telomeres: Cancer to Human Aging. *Ann. Rev. Cell Dev. Biol.* **2006**, *22*, 531-57.
17. Chai, W.; Shay, J. W.; Wright, W. E. Human Telomeres Maintain Their Overhang Length at Senescence. *Mol. Cell. Biol.* **2005**, *25*, 2158-68.
18. Feldser, D. M.; Greider, C. W. Short Telomeres Limit Tumor Progression In Vivo by Inducing Senescence. *Cancer Cell.* **2007**, *11*, (5), 461-69.
19. Sedivy, J. M. Telomeres Limit Cancer Growth by Inducing Senescence: Long-Sought In Vivo Evidence Obtained. *Cancer Cell.* **2007**, *11*, (5), 389-91.
20. Cosme-Blanco, W.; Shen, M.F.; Lazar, A. J. F.; Pathak, S.; Lozano, G.; Multani, A. S.; Chang, S., Telomere dysfunction suppresses spontaneous tumorigenesis in vivo by initiating p53-dependent cellular senescence. *EMBO Rep.* **2007**, *8*, (5), 497-503.
21. Hahn, W. C. Telomere and Telomerase Dynamics in Human Cells. *Curr. Molec. Med.* **2005**, *5*, (2), 227-31.
22. Shay, J. W.; Bacchetti, S. A survey of telomerase activity in human cancer. *Eur. J. Cancer.* **1997**, *33*, (6), 787-91.
23. Cuesta, J.; Read, M. A.; Neidle, S. The Design of G-quadruplex Ligands as Telomerase Inhibitors. *Mini Rev. Med. Chem.* **2003**, *3*, (1), 11-21.
24. Guittat, L.; Alberti, P.; Gomez, D.; De Cian, A.; Pennarun, G.; Lemarteleur, T.; Belmokhtar, C.; Paterski, R.; Morjani, H.; Trentesaux, C.; Mandine, E.; Boussin, F.; Mailliet, P.; Lacroix, L.; Riou, J.-F.; Mergny, J.-L. Targeting human telomerase for cancer therapeutics. *Cytotechnology.* **2004**, *45*, (1-2), 75-90.

25. Kerwin, S. M. G-Quadruplex DNA as a Target for Drug Design .*Curr. Pharm. Design.* **2000**, 6, (4), 441.
26. Pendino, F.; Tarkanyi, I.; Dudognon, C.; Hillion, J.; Lanotte, M.; Aradi, J.; Ségal-Bendirdjian, E. Telomeres and Telomerase: Pharmacological Targets for New Anticancer Strategies? *Curr. Cancer Drug Targets.* **2006**, 6, (2), 147-80.
27. Zhou, J. M.; Zhu, X. F.; Lu, Y. J.; Deng, R.; Huang, Z. S.; Mei, Y. P.; Wang, Y.; Huang, W. L.; Liu, Z. C.; Gu, L. Q.; Zeng, Y. X. Senescence and telomere shortening induced by novel potent G-quadruplex interactive agents, quindoline derivatives, in human cancer cell lines. *Oncogene.* **2006**, 25, (4), 503-11.
28. Pennarun, G.; Granotier, C.; Gauthier, L. R.; Gomez, D.; Hoffschir, F.; Mandine, E.; Riou, J. F.; Mergny, J. L.; Mailliet, P.; Boussin, F. D. Apoptosis related to telomere instability and cell cycle alterations in human glioma cells treated by new highly selective G-quadruplex ligands. *Oncogene.* **2005**, 24, (18), 2917-28.
29. Riou, J. F.; Guittat, L.; Mailliet, P.; Laoui, A.; Renou, E.; Petitgenet, O.; Mégnin-Chanet, F.; Hélène, C.; Mergny, J. L. Cell senescence and telomere shortening induced by a new series of specific G-quadruplex DNA ligands. *Proc. Nat. Acad. Sci.* **2002**, 99, (5), 2672-77.
30. Tauchi, T.; Shin-ya, K.; Sashida, G.; Sumi, M.; Nakajima, A.; Shimamoto, T.; Ohyashiki, J. H.; Ohyashiki, K. Activity of a novel G-quadruplex-interactive telomerase inhibitor, telomestatin (SOT-095) against human leukemia cells: involvement of ATM-dependent DNA damage response pathways. *Oncogene.* **2003**, 22, (34), 5338-47.
31. Incles, C. M.; Schultes, C. M.; Kempfski, H.; Koehler, H.; Kelland, L. R.; Neidle, S. A G-quadruplex telomere targeting agent produces p16-associated senescence and chromosomal fusions in human prostate cancer cells. *Molec. Cancer Therap.* **2004**, 3, (10), 1201-6.
32. Gomez, D.; O'Donohue, M. F.; Wenner, T.; Douarre, C.; Macadré, J.; Koebel, P.; Giraud-Panis, M. J.; Kaplan, H.; Kolkes, A.; Shin-ya, K. S.; Riou, J. F. The G-quadruplex Ligand Telomestatin Inhibits POT1 Binding to Telomeric Sequences In vitro and Induces GFP-POT1 Dissociation from Telomeres in Human Cells. *Cancer Res.* **2006**, 66, (14), 6903-07.
33. Brassart, B.; Gomez, D.; De Cian, A.; Paterski, R.; Montagnac, A.; Qui, K. H.; Temime-Smaali, N.; Trentesaux, C.; Mergny, J. L.; Gueritte, F.; Riou, J. F. A New Steroid Derivative Stabilizes G-Quadruplexes and Induces Telomere Uncapping in Human Tumor Cells .*Mol. Pharmacol.* **2007**, 72, (3), 631-40.
34. Gomez, D.; Wenner, T.; Brassart, B.; Douarre, C.; O'Donohue, M. F.; El Khoury, V.; Shin-ya, K.; Morjani, H.; Trentesaux, C.; Riou, J. F., Telomestatin-induced Telomere Uncapping Is Modulated by POT1 through G-overhang Extension in HT1080 Human Tumor Cells .*J. Biol. Chem.* **2006**, 281, (50), 38721-29.

35. Tahara, H.; Shin-ya, K.; Seimiya, H.; Yamada, H.; Tsuruo, T.; Ide, T. G-Quadruplex stabilization by telomestatin induces TRF2 protein dissociation from telomeres and anaphase bridge formation accompanied by loss of the 3' telomeric overhang in cancer cells. *Oncogene*. **2005**, *25*, (13), 1955-66.
36. Huppert, J. L.; Balasubramanian, S. Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.* **2005**, *33*, (9), 2908-16.
37. Todd, A. K.; Johnston, M.; Neidle, S. Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res.* **2005**, *33*, (9), 2901-07.
38. Burge, S.; Parkinson, G. N.; Hazel, P.; Todd, A. K.; Neidle, S. Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res.* **2006**, *34*, (19), 5402-15.
39. Paces, J.; Pavlíček, A.; Paces, V. HERVd: database of human endogenous retroviruses. *Nucleic Acids Res.* **2002**, *30*, (1), 205-6.
40. Rawal, P.; Kummarasetti, V. B. R.; Ravindran, J.; Kumar, N.; Halder, K.; Sharma, R.; Mukerji, M.; Das, S. K.; Chowdhury, S. Genome-wide prediction of G4 DNA as regulatory motifs: Role in Escherichia coli global regulation. *Genome Res.* **2006**, *16*, (5), 644-55.
41. Kostadinov, R.; Malhotra, N.; Viotti, M.; Shine, R.; D'Antonio, L.; Bagga, P. GRSDB: a database of quadruplex forming G-rich sequences in alternatively processed mammalian pre-mRNA sequences. *Nucleic Acids Res.* **2006**, *34*, D119-24.
42. Zanotti, K. J.; Lackey, P. E.; Evans, G. L.; Mihailescu, M. R. Thermodynamics of the Fragile X Mental Retardation Protein RGG Box Interactions with G Quartet Forming RNA. *Biochemistry*. **2006**, *45*, (27).
43. Eddy, J.; Maizels, N. Gene function correlates with potential for G4 DNA formation in the human genome. *Nucleic Acids Res.* **2006**, *34*, (14), 3887-96.
44. Raghavan, S. C.; Swanson, P. C.; Wu, X.; Hsieh, C. L.; Lieber, M. R. A non-B-DNA structure at the Bcl-2 major breakpoint region is cleaved by the RAG complex. *Nature* **2004**, *428*, 88-93.
45. Bonaglia, M. C.; Giorda, R.; Mani, E.; Aceti, G.; Anderlid, B.-M.; Baroncini, A.; Pramparo, T.; Zuffardi, O. Identification of a recurrent breakpoint within the SHANK3 gene in the 22q13.3 deletion syndrome. *J. Med. Genet.* **2006**, *43*, (10), 822-28.
46. Liu, W.; Sun, D.; Hurley, L. H. Binding of G-quadruplex-interactive agents to distinct G-quadruplexes induces different biological effects in MiaPaCa cells. *Nucleosides, nucleotides & nucleic acids*. **2005**, *24*, (10-12), 1801-15.

47. Shafer, R. H.; Smirnov, I. Biological aspects of DNA/RNA quadruplexes. *Biopolymers* **2001**, *56*, (3), 209-27.
48. Siddiqui-Jain, A.; Grand, C. L.; Bearss, D. J.; Hurley, L. H. Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proc. Nat. Acad. Sci.* **2002**, *93*, (18), 11593-98.
49. Phan, A. T.; Modi, Y. S.; Patel, D. J. Propeller-Type Parallel-Stranded G-Quadruplexes in the Human c-myc Promoter. *J. Am. Chem. Soc.* **2004**, *126*, (28), 8710-16.
50. Fernando, H.; Reszka, A. P.; Huppert, J.; Ladame, S.; Rankin, S.; Venkitaraman, A. R.; Neidle, S.; Balasubramanian, S. A Conserved Quadruplex Motif Located in a Transcription Activation Site of the Human c-kit Oncogene. *Biochemistry.* **2006**, *45*, (25), 7854-60.
51. Rankin, S.; Reszka, A. P.; Huppert, J.; Zloh, M.; Parkinson, G. N.; Todd, A. K.; Ladame, S.; Balasubramanian, S.; Neidle, S. Putative DNA Quadruplex Formation within the Human c-kit Oncogene. *J. Am. Chem. Soc.* **2005**, *127*, (30), 10584-89.
52. Sun, D.; Guo, K.; Rusche, J. J.; Hurley, L. H. Facilitation of a structural transition in the polypurine/polypyrimidine tract within the proximal promoter region of the human VEGF gene by the presence of potassium and G-quadruplex-interactive agents. *Nucleic Acids Res.* **2005**, *33*, (18), 6070-80.
53. De Armond, R.; Wood, S.; Sun, D.; Hurley, L. H.; Ebbinghaus, S. W. Evidence for the Presence of a Guanine Quadruplex Forming Region within a Polypurine Tract of the Hypoxia Inducible Factor 1 α Promoter. *Biochemistry.* **2005**, *44*, (49), 16341-50.
54. Dai, J.; Dexheimer, T. S.; Chen, D.; Carver, M.; Ambrus, A.; Jones, R. A.; Yang, D. An Intramolecular G-Quadruplex Structure with Mixed Parallel/Antiparallel G-Strands Formed in the Human BCL-2 Promoter Region in Solution. *J. Am. Chem. Soc.* **2006**, *128*, (4), 1096-98.
55. Dexheimer, T. S.; Sun, D.; Hurley, L. H. Deconvoluting the Structural and Drug-Recognition Complexity of the G-Quadruplex-Forming Region Upstream of the bcl-2 P1 Promoter. *J. Am. Chem. Soc.* **2006**, *128*, (16), 5404-15.
56. Cogoi, S.; Xodo, L. E. G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription. *Nucleic Acids Res.* **2006**, *34*, (9), 2536-49.
57. Xu, Y.; Sugiyama, H. Formation of the G-quadruplex and I-motif structures in retinoblastoma susceptibility genes (Rb). *Nucleic Acids Res.* **2006**, *34*, (3), 949-54.

58. Phan, A. T.; Kuryavyi, V.; Gaw, H. Y.; Patel, D. J. Small-molecule interaction with a five-guanine-tract G-quadruplex structure from the human MYC promoter. *Nature Chem. Biol.* **2005**, 1, (3), 167-73.
59. Wang, Y.; Patel, D. J. Solution structure of the human telomeric repeat d[AG3(T2AG3)3] G-tetraplex. *Structure.* **1993**, 1, (4), 263-82.
60. Wheelhouse, R. T.; Jennings, S. A.; Phillips, V. A.; Pletsas, D.; Murphy, P. M.; Garbett, N. C.; Chaires, J. B.; Jenkins, T. C. Design, Synthesis, and Evaluation of Novel Biarylpyrimidines: A New Class of Ligand for Unusual Nucleic Acid Structures. *J. Med. Chem.* **2006**, 49, (17), 5187-98.
61. Cevec, M.; Plavec, J. Role of Loop Residues and Cations on the Formation and Stability of Dimeric DNA G-Quadruplexes. *Biochemistry.* **2005**, 44, (46), 15238-46.
62. Hazel, P.; Huppert, J.; Balasubramanian, S.; Neidle, S. Loop-Length-Dependent Folding of G-Quadruplexes. *Journal of the American Chemical Society.* **2004**, 126, (50), 16405-16415.
63. Rachwal, P. A.; Brown, T.; Fox, K. R. Sequence effects of single base loops in intramolecular quadruplex DNA. *FEBS Lett.* **2007**, 581, (8), 1657-60.
64. Rachwal, P. A.; Findlow, I. S.; Werner, J. M.; Brown, T.; Fox, K. R. Intramolecular DNA quadruplexes with different arrangements of short and long loops. *Nucleic Acids Res.* **2007**, 35, (12), 4214-22.
65. Risitano, A.; Fox, K. R. Influence of loop size on the stability of intramolecular DNA quadruplexes. *Nucleic Acids Res.* **2004**, 32, (8), 2598-2606.
66. Vglask, V.; Bauer, L.; Tlukov, K. Structural Features of Intra- and Intermolecular G-Quadruplexes Derived from Telomeric Repeats. *Biochemistry.* **2010**, 49, (10), 2110-2120.
67. Antonacci, C.; Chaires, J. B.; Sheardy, R. D. Biophysical Characterization of the Human Telomeric (TTAGGG)₄ Repeat in a Potassium Solution. *Biochemistry.* **2007**, 46, 4654-60.
68. Humphrey, W.; Dalke, A.; Schulten, K. VMD - Visual Molecular Dynamics. *J. Molec Graphics.* **1996**, 14, (1), 33-38.
69. Spolar, R. S.; Ha, J.H.; Record, T. Hydrophobic effect in protein folding and other noncovalent processes involving proteins. *Proc. Natl. Acad. Sci. USA* **1989**, 86, 8382-8385.
70. Collie, G. W.; Haider, S. M.; Neidle, S.; Parkinson, G. N. A crystallographic and modelling study of a human telomeric RNA (TERRA) quadruplex. *Nucleic Acids Research.* **2010**, 1-12.

71. Lodish, H.; Berk, A.; Zipursky, L.; Matsudaira, P.; Baltimore, D.; Darnell, J. *Molecular Cell Biology, 4th edition*. W. H. Freeman: New York, 2000.
72. Goswami, P. P.; Chaudhuri, P.; Tiwari, V.; Parihar, N. S.; Harbola, P. C. Cloning and Sequencing of a 16 kDa Outer Membrane Protein Gene of *Pasteurella multocida* P52. *Veterinary Research Communications*. **2004**, *28*, (1), 17-25.
73. Grigorieff, N.; Ceska, T. A.; Downing, K. H.; Baldwin, J. M.; Henderson, R. Electron-crystallographic Refinement of the Structure of Bacteriorhodopsin. *J. Mol. Biol.* **1996**, *259*, (3), 393-421.
74. Kamikubo, H.; Kataoka, M.; Váró, G.; Oka, T.; Tokunaga, F.; Needleman, R.; Lanyi, J. K. Structure of the N intermediate of bacteriorhodopsin revealed by x-ray diffraction. *Proc. Nat. Acad. Sci.* **1996**, *93*, (4), 1386-90.
75. Heymann, J. B.; Müller, D. J.; Landau, E. M.; Rosenbusch, J. P.; Pebay-Peyroula, E.; Büldt, G.; Engel, A. Charting the Surfaces of the Purple Membrane. *Journal of Structural Biology*. **1999**, *128*, (3), 243-249.
76. Dunn, R.; McCoy, J.; Simsek, M.; Majumdar, A.; Chang, S. H.; Rajbhandary, U. L.; Khorana, H. G. The bacteriorhodopsin gene. *Proc. Nat. Acad. Sci.* **1981**, *78*, (11), 6744-48.
77. Ovchinnikov, Y. A.; Abdulaev, N. G.; Feigina, M. Y.; Kiselev, A. V.; Lobanov, N. A. The structural basis of the functioning of bacteriorhodopsin: an overview *FEBS Lett.* **1979**, *100*, (2), 219-24.
78. Blaurock, A. E.; Stoeckenius, W. Structure of the purple membrane. *Nature New Biol.* **1971**, *233*, (39), 152-5.
79. Scherrer, P.; Mathew, M. K.; Sperling, W.; Stoeckenius, W. Retinal isomer ratio in dark-adapted purple membrane and bacteriorhodopsin monomers. *Biochemistry*. **1989**, *28*, (2), 829-34.
80. Hashimoto, S.; Obata, K.; Takeuchi, H.; Needleman, R.; Lanyi, J. K. Ultraviolet Resonance Raman Spectra of Trp-182 and Trp-189 in Bacteriorhodopsin: Novel Information on the Structure of Trp-182 and Its Steric Interaction with Retinal. *Biochemistry*. **1997**, *36*, (39), 11583-90.
81. Hashimoto, S.; Sasaki, M.; Takeuchi, H.; Needleman, R.; Lanyi, J. K. Changes in Hydrogen Bonding and Environment of Tryptophan Residues on Helix F of Bacteriorhodopsin during the Photocycle: A Time-Resolved Ultraviolet Resonance Raman Study. *Biochemistry*. **2002**, *41*, (20), 6495-503.
82. Hatanaka, M.; Kashima, R.; Kandori, H.; Friedman, N.; Sheves, M.; Needleman, R.; Lanyi, J. K.; Maeda, A. Trp86 → Phe Replacement in Bacteriorhodopsin Affects a Water Molecule near Asp85 and Light Adaptation. *Biochemistry*. **1997**, *36*, (18), 5493-98.

83. Petkova, A. T.; Hatanaka, M.; Jaroniec, C. P.; Hu, J. G.; Belenky, M.; Verhoeven, M.; Lugtenburg, J.; Griffin, R. G.; Herzfeld, J. Tryptophan Interactions in Bacteriorhodopsin: A Heteronuclear Solid-State NMR Study. *Biochemistry* **2002**, *41*, (7), 2429-37.
84. Edman, K.; Nollert, P.; Royant, A.; Belrhali, H.; Pebay-Peyroula, E.; Hajdu, J.; Neutze, R.; Landau, E. M. High-resolution X-ray structure of an early intermediate in the bacteriorhodopsin photocycle. *Nature*. **1999**, *401*, 822-26.
85. Grudinin, S.; Büldt, G.; Gordeliy, V.; Baumgaertnerz, A. Water Molecules and Hydrogen-Bonded Networks in Bacteriorhodopsin—Molecular Dynamics Simulation of the Ground State and the M-Intermediate. *Biophys. J.* **2005**, *88*, 3252-61.
86. Kimura, Y.; Vassilyev, D. G.; Miyazawa, A.; Kidera, A.; Matsushima, M.; Mitsuoka, K.; Murata, K.; Hirai, T.; Fujiyoshi, Y. Surface of bacteriorhodopsin revealed by high-resolution electron crystallography. *Nature*. **1997**, *389*, (6647), 206-11.
87. Luecke, H.; Richter, H. T.; Lanyi, J. K. Proton Transfer Pathways in Bacteriorhodopsin at 2.3 Angstrom Resolution. *Science* **1998**, *280*, (5371), 1934-37.
88. Mitsuoka, K.; Hirai, T.; Murata, K.; Miyazawa, A.; Kidera, A.; Kimura, Y.; Fujiyoshi, Y. The structure of bacteriorhodopsin at 3.0 Å resolution based on electron crystallography: implication of the charge distribution. *J. Mol. Biol.* **1999**, *286*, (3), 861-82.
89. Oesterhelt, D.; Hess, B. Reversible Photolysis of the Purple Complex in the Purple Membrane of Halobacterium halobium. *Eur. J. Biochem.* **1973**, *37*, (2), 316-26.
90. Henderson, R. The structure of the purple membrane from Halobacterium halobium: analysis of the X-ray diffraction pattern. *J. Mol. Biol.* **1975**, *93*, (2), 123-38.
91. Subramaniam, S.; Lindahl, M.; Bullough, P.; Faruqi, A. R.; Tittor, J.; Oesterhelt, D.; Brown, L.; Lanyi, J.; Henderson, R. Protein conformational changes in the bacteriorhodopsin photocycle. *J. Mol. Biol.* **1999**, *287*, (1), 145-61.
92. Faham, S.; Yang, D.; Bare, E.; Yohannan, S.; Whitelegge, J. P.; Bowie, J. U. Side-chain Contributions to Membrane Protein Structure and Stability. *J. Mol. Biol.* **2004**, *335*, (1), 297-305.
93. Yohannan, S.; Faham, S.; Yang, D.; Whitelegge, J. P.; Bowie, J. U. The evolution of transmembrane helix kinks and the structural diversity of G protein-coupled receptors. *Proc. Nat. Acad. Sci.* **2004**, *101*, (4), 959-63.

94. Alexiev, U.; Mollaaghababa, R.; Khorana, H. G.; Heyn, M. P. Evidence for Long Range Allosteric Interactions between the Extracellular and Cytoplasmic Parts of Bacteriorhodopsin from the Mutant R82A and Its Second Site mutant R82. *J. Biol. Chem.* **2000**, *275*, (18), 13431-40.
95. Brown, L. S. Reconciling crystallography and mutagenesis: a synthetic approach to the creation of a comprehensive model for proton pumping by bacteriorhodopsin. *Biophys. Biochim. Acta.* **2000**, *1460*, (1), 49-59.
96. Checover, S.; Marantz, Y.; Nachliel, E.; Gutman, M.; Pfeiffer, M.; Tittor, J.; Oesterhelt, D.; Dencher, N. A., Dynamics of the Proton Transfer Reaction on the Cytoplasmic Surface of Bacteriorhodopsin. *Biochemistry.* **2001**, *40*, (14), 4281-92.
97. Stern, L. J.; Khorana, H. G. Structure-function studies on bacteriorhodopsin. X. Individual substitutions of arginine residues by glutamine affect chromophore formation, photocycle, and proton translocation. *J. Biol. Chem.* **1989**, *264*, (24), 14202-08.
98. Lanyi, J. K. Proton transfers in the bacteriorhodopsin photocycle. *Biochimica et Biophysica Acta.* **2006**, *1757*, 1012-1018.
99. Edman, K.; Royant, A.; Larsson, G.; Jacobson, F.; Taylor, T.; van der Spoel, D.; Landau, E. M.; Pebay-Peyroula, E.; Neutze, R. Deformation of Helix C in the Low Temperature L-intermediate of Bacteriorhodopsin. *J. Biol. Chem.* **2004**, *279*, (3), 2147-58.
100. Facciotti, M. T.; Cheung, V. S.; Nguyen, D.; Rouhani, S.; Glaeser, R. M. Crystal Structure of the Bromide-Bound D85S Mutant of Bacteriorhodopsin: Principles of Ion Pumping. *Biophys. J.* **2003**, *85*, (1), 451-58.
101. Facciotti, M. T.; Rouhani, S.; Glaeser, R. M. Crystal structures of bR(D85S) favor a model of bacteriorhodopsin as a hydroxyl-ion pump. *FEBS Lett.* **2004**, *564*, (3), 301-06.
102. Lanyi, J. K. X-Ray Crystallography of Bacteriorhodopsin and Its Photointermediates: Insights into the Mechanism of Proton Transport. *Biochem. (Mosc)* **2001**, *66*, (11), 1192-96.
103. Lanyi, J. K. X-ray diffraction of bacteriorhodopsin photocycle intermediates (Review). *Molec. Membranr Biol.* **2004**, *21*, (3), 143-50.
104. Lanyi, J. K.; Schobert, B. Mechanism of Proton Transport in Bacteriorhodopsin from Crystallographic Structures of the K, L, M1, M2, and M2' Intermediates of the Photocycle. *J. Mol. Biol.* **2003**, *328*, (2), 439-50.
105. Faham, S.; Bowie, J. U. Bicelle crystallization: a new method for crystallizing membrane proteins yields a monomeric bacteriorhodopsin structure. *J. Mol. Biol.* **2002**, *316*, (1), 1-6.

106. Gouaux, E. It's not just a phase: crystallization and X-ray structure determination of bacteriorhodopsin in lipidic cubic phases. *Structure*. **1998**, 6, (1), 5-10.
107. Rummel, G.; Hardmeyer, A.; Widmer, C.; Chiu, M. L.; Nollert, P.; Locher, K. P.; Pedruzzi, I.; Landau, E. M.; Rosenbusch, J. P. Lipidic Cubic Phases: New Matrices for the Three-Dimensional Crystallization of Membrane Proteins. *J. Struct. Biol.* **1998**, 212, (2), 82-91.
108. Takeda, K.; Sato, H.; Hino, T.; Kono, M.; Fukuda, K.; Sakurai, I.; Okada, T.; Kouyama, T. A novel three-dimensional crystal of bacteriorhodopsin obtained by successive fusion of the vesicular assemblies. *J. Mol. Biol.* **1989**, 283, (2), 463-74.
109. Lanyi, J. K. What is the real crystallographic structure of the L photointermediate of bacteriorhodopsin? *Biochim. Biophys. Acta*. **2004**, 1658, (1-2), 14-22.
110. Haupts, U.; Tittor, J.; Oesterhelt, D. Closing In on Bacteriorhodopsin: Progress in Understanding the Molecule. *Ann. Rev. Biophys. Biomol. Struct.* **1999**, 28, 367-99.
111. Earnest, T. N.; Roepe, P.; Braiman, M. S.; Gillespie, J.; Rothschild, K. J. Orientation of the bacteriorhodopsin chromophore probed by polarized Fourier transform infrared difference spectroscopy. *Biochemistry*. **1986**, 25, (24), 7793-98.
112. Siebert, F.; Mantele, W., Investigation of the Primary Photochemistry of Bacteriorhodopsin by Low-Temperature Fourier-Transform Infrared Spectroscopy. *Eur. J. Biochem.* **1983**, 130, (3), 565-73.
113. Atkinson, G. H.; Ujj, L.; Zhou, Y. Vibrational Spectrum of the J-625 Intermediate in the Room Temperature Bacteriorhodopsin Photocycle. *J. Phys. Chem. A* **2000**, 104, (18), 4130-39.
114. Atkinson, G. H.; Blanchard, D.; Lemaire, H.; Brack, T. L.; Hayashi, H. Picosecond time-resolved fluorescence spectroscopy of K-590 in the bacteriorhodopsin photocycle. *Biophys. J.* **1989**, 55, (2), 263-74.
115. Maeda, A.; Sasaki, J.; Ohkita, Y. J.; Simpson, M.; Herzfeld, J. Tryptophan perturbation in the L intermediate of bacteriorhodopsin: Fourier transform infrared analysis with indole-nitrogen-15 shift. *Biochemistry*. **1992**, 31, (50), 12543-45.
116. Braiman, M. S.; Ahl, P. L.; Rothschild, K. J. Millisecond Fourier-transform infrared difference spectra of bacteriorhodopsin's M412 photoproduct. *Proc. Nat. Acad. Sci.* **1987**, 84, (14), 5221-25.
117. Rothschild, K. J.; Clark, N. A. Polarized infrared spectroscopy of oriented purple membrane. *Biophys. J.* **1979**, 25, (3), 473-87.
118. Braiman, M. S.; Bousché, O.; Rothschild, K. J. Protein dynamics in the bacteriorhodopsin photocycle: submillisecond Fourier transform infrared spectra

- of the L, M, and N photointermediates. *Proc. Nat. Acad. Sci.* **1991**, *88*, (6), 2388-92.
119. Luecke, H. S., B.; Richter, T.; Cartailier, P.; Lanyi, J. Structure of bacteriorhodopsin at 1.55 Å resolution. *J. Mol. Biol.* **1999**, *291*, (4), 899-911.
 120. Zscherp, C.; Schlesinger, R.; Tittor, J.; Oesterhelt, D.; Heberle, J. In situ determination of transient pKa changes of internal amino acids of bacteriorhodopsin by using time-resolved attenuated total reflection Fourier-transform infrared spectroscopy. *Proc. Nat. Acad. Sci.* **1999**, *96*, (10), 5498-503.
 121. Checover, S.; Nachliel, E.; Dencher, N. A.; Gutman, M. Mechanism of Proton Entry into the Cytoplasmic Section of the Proton-Conducting Channel of Bacteriorhodopsin. *Biochemistry.* **1997**, *36*, (45), 13919-28.
 122. Renthal, R. D., N.; Tuley, J.; Horowitz, P. Constraints on the flexibility of bacteriorhodopsin's carboxyl-terminal tail at the purple membrane surface. *Biochemistry.* **1983**, *22*, (1), 5-12.
 123. Turner, G. J. C., S.; Pohren, L.; Hines, Kirk G.; Correia, J.; Mitchell, D. The Bacteriorhodopsin Carboxyl-Terminus Contributes to Proton Recruitment and Protein Stability. *Biochemistry.* **2009**, *48*, 1112-1122.
 124. Balashov, S. P. Imasheva, E. S.; Govindjee, R.; Ebrey, T. G. Titration of aspartate-85 in bacteriorhodopsin: what it says about chromophore isomerization and proton release. *Biophys. J.* **1996**, *70*, (1), 473-81.
 125. Lanyi, J. K. Understanding structure and function in the light-driven proton pump bacteriorhodopsin. *J. Struct. Biol.* **1998**, *124*, 164-78.
 126. Balashov, S. P. Protonation reactions and their coupling in bacteriorhodopsin. *Biochim. Biophys. Acta.* **2000**, *1460*, (1), 75-94.
 127. Balashov, S. P. Lu, M.; Imasheva, E. S.; Govindjee, R.; Ebrey, T. G.; Othersen, B. 3rd.; Chen, Y.; Crouch, R. K.; Menick, D. R. The proton release group of bacteriorhodopsin controls the rate of the final step of its photocycle at low pH. *Biochemistry.* **1999**, *38*, (7), 2026-39.
 128. De Groot, H. J.; Smith, S. O.; Courtin, J.; van den Berg, E.; Winkel, C.; Lugtenburg, J.; Griffin, R. G.; Herzfeld. Solid-state ¹³C and ¹⁵N NMR study of the low pH forms of bacteriorhodopsin. *Biochemistry.* **1990**, *29*, (29), 6873-6883.
 129. Oesterhel, D. The structure and mechanism of the family of retinal proteins from halophilic archaea. *Curr. Opin. Struct. Biol.* **1998**, *8*, (4), 489-500.
 130. Tanizaki, S.; Feig, M. Molecular dynamics simulations of large integral membrane proteins with an implicit membrane model. *J. phy. Chem. B* **2006**, *110*, 548-556.

131. Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comp. Chem.* **2005**, *26*, (16), 1781-1802.
132. Brooks, B. R.; Hodoscek, M. Parallelization of CHARMM for MIMD Machines. *Chemical Design Automation News (CDA News)*. **1992**, *7*, (12), 16-22.
133. MacKerell, A. J.; Banavali, N.; Foloppe, N. Development and current status of the CHARMM force field for nucleic acids. *Biopolymers*. **2000-2001**, *56*, (4), 257-65.
134. Nina, M.; Roux, B.; Smith, J. C. Functional interactions in bacteriorhodopsin: a theoretical analysis of retinal hydrogen bonding with water. *Biophys J.* **1995**, *68*, (1), 25-39.
135. Grest, G. S.; Kremer, K. Molecular-dynamics simulation for polymers in the presence of a heat bath. *Phys. Rev. A.* **1986**, *33*, (5), 3628-3631.
136. Pomerleau, V.; Harvey-G, E.; Boucher, F. Lipid-protein interactions in the purple membrane: structural specificity within the hydrophobic domain. *Biochimica et Biophysica Acta*. **1995**, *1234*, 221-224.
137. Vogt, Gerhard.; Argos, P. Protein thermal stability: hydrogen bonds or internal packing? *Folding and Design*. **1997**, *2*, S40-S46.
138. Honing, B.; Hubbel, L. Stability of "salt bridges" in membrane proteins. *Proc. Natl. Acad. Sci. USA*. **1984**, *81*, 5412-5416.
139. McCammon, J. A.; Harvey, S. C. *Dynamics of Proteins and Nucleic Acids*. Cambridge University Press: 1987.
140. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D. ; States, D. J.; Swaminathan, S.; Karplus, M. A program for macromolecular energy, minimization, and dynamics calculations. *J. Comp. Chem.* **1983**, *4*, (2), 187-212.
141. van Gunsteren, W. F.; Berendsen, H. J. C. *GROMOS Manual*, Lab. of Phys. Chem., Univ. of Groningen: 1987.
142. Weiner, P. K.; Kollman, P. A. AMBER: Assisted Model Building with Energy Refinement. A General Program for Modeling Molecules and Their Interactions. *J. Comp. Chem.* **1981**, *2*, 287-303.
143. Leach, A. *Molecular Modeling Principles and Applications*. 2001.
144. MacKerell, J. D.; Bashford, M.; Bellott, R. L.; Dunbrack Jr, J. D.; Evanseck, M. J.; Field, S. ; Fischer, J.; Gao, H.; Guo, S.; Ha, D.; Joseph-McCarthy, L.; Kuchnir, K. ; Kuczera, F. T. K. ; Lau, C. ; Mattos, S.; Michnick, T. ; Ngo, D. T.; Nguyen, B.; Prodhom, W. E.; Reiher, III, B. ; Schlenkrich, Roux, M. ; Smith, J. C. ; Stote, R. ; Straub, Watanabe, J.; Wiórkiewicz-Kuczera, M.; Yin, D. ; Karplus, M. All-atom

- empirical potential for molecular modeling and dynamics studies of protein. *J. Phys. Chem B.* **1998**, 102, 3586-3616.
145. MacKerell Jr, A. D.; Wiorkiewicz Kuczera, K, J.; Karplus, M . An all-atom empirical energy function for the simulation of nucleic acids. *J. Am. Chem. Soc.* **1995**, 117, (48), 11946-11975.
 146. Van Der Spoel, D.; Lindhal, E.; Hess, B.; Groenhof, G.; Mark, A.; Berendsen, H. GROMACS: Fast, flexible and free. *Journal of Computational Chemistry.* **2005**, 26, (16), 1701-1718.
 147. Ewald, P. Die Berechnung optischer und elektrostatischer Gitterpotentiale (Evaluation of optical and electrostatic lattice potentials). *Ann. Phys.* **1921**, 64, 253-287.
 148. Berendsen, H. J. C.; Postoma, J.P.M.; Van Gunsteren, W.F.; Dinola, A.; Haak, J.R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, 81, (8), 3684-90.
 149. Evans, D. J.; Holian, B. The Nose-Hoover thermostat. *J. Chem. Phys.* **1985**, 83, (8), 4069-4074.
 150. Berman, H.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.; Bourne, P. The Protein Data Bank. *Nucleic Acids Research.* **2000**, 28, (1), 235-242.
 151. Ramachandran, G. N.; Ramakrishnan, C.; Sasisekharan, V. Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.* **1963**, 7, 95-99.
 152. Shortle, D. The denatured state (the other half of the folding equation) and its role in protein stability. *FASEBJ* **1996**, 10, 27-34.
 153. Lazaridis, T.; Karplus, M., Heat Capacity and Compactness in Denatured Proteins. *Biophys. Chem.* **1999**, 78, 207-17.
 154. Prabhu, N. V.; Sharp, K.A. Heat Capacity in Proteins. *Ann. Rev. Phys. Chem.* **2005**, 56, 521-48.
 155. Kauzmann, W. Some factors in the interpretation of protein denaturation. *Adv. Protein Sci.* **1959**, 14, 1-63.
 156. Myers, J. K.; Pace, C.N.; Scholtz, J.M. Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Sci.* **1995**, 4, (10), 2138-2148.