Structure, Thermodynamics, and Dynamical Properties of Nucleic Acids, Proteins, and Glass-Forming Liquids

Author: Kristina Streu

STRUCTURE, THERMODYNAMICS, AND DYNAMICAL PROPERTIES OF NUCLEIC ACIDS, PROTEINS, AND GLASS-FORMING LIQUIDS

Kristina Streu

A dissertation

submitted to the Faculty of

the department of Chemistry

in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Boston College

Morrissey College of Arts and Sciences

Graduate School

July 2016

STRUCTURE, THERMODYNAMICS, AND DYNAMICAL PROPERTIES OF NUCLEIC ACIDS, PROTEINS, AND GLASS-FORMING LIQUIDS

Kristina Streu

Advisor: Professor Udayan Mohanty

Abstract:

The stabilization of particular conformations of protein and nucleic acid structure is believed to play an important role in many important biological functions. In chapter one, the α-helical conformation and structural stability of single and double stapled all-hydrocarbon cross-linked p53 peptides when bound and unbound to MDM2 are investigated. Our study provides a comprehensive rationalization of the relationship between peptide stapling strategy, the secondary structural stability, and the binding affinity of p53-MDM2 complex.

In chapter two, we study counterion-mediated collapse of a strongly charged model polyelectrolyte chain by Group-II divalent metal cations using coarse-grained Brownian dynamics simulations. Polyelectrolyte effects govern the association of counterions with the chain. Large ions are less effective in counterion condensation than small ions. However, upon counterion condensation, the reduction of the backbone charge is independent of size of the metal cations. Above a threshold value of Coulomb strength parameter, counterion release entropy drives the formation of counterion-induced compact states.

In chapter three, the nature of surface tension in the random first order theory of supercooled liquid is analyzed within the framework of Landau-Lifshitz fluctuation theory. We show that the surface tension of a droplet satisfies the differential equation $4\pi r^2(\frac{d\sigma}{dr}) + 8\pi r\sigma(r) - Br^{1/2} = 0$, where $B/T = \sqrt{12\pi k_B c_v}$, T is temperature, k_B is Boltzmann constant, and c_v is heat capacity. A consequence is that the slope of the relaxation time at the glass transition temperature, i.e., the fragility index, is expressed as the square of the ratio of heat capacity and configurational entropy of the supercooled liquid.

When backbone extended nucleosides are incorporated into a double helix, a unique helical structure is formed. In chapter four, we find that the predicted stability of modified backbone DNA strands in aqueous solution is in good agreement with experimental melting temperature data. The incorporation of extended backbone nucleosides into a duplex results in elongation of the end-to-end chain distance due to the distortion of the B-DNA conformation at the mutated base-pair insertion. We also find that the modified backbone helical twist is approximately 40 degrees, larger than B-DNA helical twist and closer to the twist angle predicted for D-form DNA.

The folding of RNA tertiary structure has been described as an equilibrium between partially folded I (intermediate) states, and the fully folded native conformation, or N state. RNA is highly sensitive to the ionic environment due to its negative charge, and tertiary structures tend to be strongly stabilized by Mg^{2+}. There is a need for models capable of describing the ion atmosphere surrounding RNA with quantitative accuracy. In chapter 5, we present a generalized Manning condensation model of RNA electrostatics for studying the Mg^{2+}-induced RNA folding of the 58mer ribosomal fragment.

Table of Contents

List of Abbreviations

6'-dA	(2*R*,3*S*,5*R*)-5-(6-amino-9*H*-purin-9-yl)-2-(2-hydroxyethyl)tetrahydrofuran-3-ol
6'-dT	1-((2*R*,4*S*,5*R*)-4-hydroxy-5-(2-hydroxyethyl)tetrahydrofuran-2-yl)-5-methylpyrimidine-2,4(1*H*,3*H*)-dione
A	Adenine
BD	Brownian Dynamics
C	Cytosine
dA	2'-deoxyadenosine
dC	2'-deoxycytidine
dG	2'-deoxyguanosine
dT	2'-deoxythymidine
DH	Debye-Hückel
DNA	β-D-2'-deoxyribonucleic acid
FEP	Free energy perturbation
FP	Filling potential
G	Guanine
M	Molar
M^{2+}	Divalent metal cation
MD	Molecular Dynamics
mM	Millimolar
NaCl	Sodium chloride
NLPB	Non-linear Poisson Boltzmann
PE	Polyelectrolyte
PMF	Potential of mean force
RCM	Ring-closing olefin metathesis
REMD	Replica-exchange molecular dynamics
RFOT	Random first order transition
RG	Renormalization group

RNA	β-D-ribonucleic acid
T	Thymine
T_g	Glass transition temperature
T_m	Melting point
TI	Thermodynamic integration
WHAM	Weighted-histogram analysis method
Å	Angstrom
μ	Micro

Glossary of Terms

Bjerrum length: The length at which the electrostatic interaction between two elementary charges is comparable in magnitude to the thermal energy.

Brownian dynamics: A simulation method that uses the concept of Brownian motion, or random motion, to evolve particles in time. In the equation of motion, the net particle acceleration is comprised of the force due to particle interaction potentials, the friction force, and the random force.

Contact map: The map describes the distance between native contact residue pairs in a protein using a two-dimensional matrix. The *ij* element of the matrix is unity if the distance between residues *i* and *j* are less than a predetermined value, and is otherwise zero.

Counterion condensation: The phenomenon by which counterions remain in close proximity to a charged polyelectrolyte chain, thereby compensating a large percentage of polyelectrolyte charges (i.e. the linear charge density along the chain is reduced below a threshold value). These counterions are said to be "condensed".

Desmond: A software package from Schrödinger, Inc. utilized to carry our robust molecular dynamics simulations of biological systems.

E3 ubiquitin ligase: A protein that binds to specific DNA sequences. This protein controls the flow of genetic information from DNA to mRNA.

End-to-end distance: Considering a vector that points from one end of a polymer to chain to the other end. The magnitude of the vector is the end-to-end distance.

Fragility index: Characterizes the slope of the viscosity (relaxation time, or time required for intermolecular rearrangement) of a material with temperature as the material approaches its glass transition temperature.

Free energy perturbation: A method to calculate the free energy of ligand-protein binding by calculating the bound and unbound components in gas phase and in solvated forms. This method ignores fluctuations of the protein and ligand.

Glass transition temperature, T_g: The range of temperatures over which the short, reversible glass transition from viscous liquid to glassy state, or the reverse from a hard, brittle glassy state to a molten liquid state occurs.

Gō-type formalism: An idea based around the assumption that protein residues that interact in the fully folded structure play a major role in the folding process. The model proposes that energetic contributions of the native interactions act as the sole driving force in the folding process.

Linear charge density: The ratio of the Bjerrum length to the spacing between neighboring charged monomers of the polyelectrolyte chain. For B form DNA in water, the distance between backbone phosphates is ~ 3.4 Å. With two phosphate charges every 3.4 Å, the charge spacing is ~ 1.7 Å, the Bjerrum length is ~ 7 Å; and therefore the linear charge density is 4.2.

MDM2: The protein encoded by the gene of the same name. MDM2 functions as an E3 ubiquitin ligase and inhibits p53 transcriptional activation.

Molecular dynamics: A simulation method that consists of a numerical, step-by-step solution of the classical equations of motion.

Native contact: A contact between amino acid side chains that are not neighbors in the sequence space but are spatially close in the protein crystal structure.

p53: A tumor suppressor protein that is encoded in humans by the *TP53* gene. In multicellular organisms, p53 regulates the cell cycle and functions as a tumor suppressor.

Potential of mean force: The potential that gives the average force over all configurations. It examines how a system's energy changes as a function of a reaction coordinate.

Random first order transition: A spin-glass inspired theory of the glass transition.

Renormalization group analysis: A mathematical analysis that allows systematic investigation of the changes of a physical system as viewed at different distance scales, typically it is implemented when dealing with strongly interacting systems.

Replica-exchange molecular dynamics: A simulation method that generates N identical systems and each system is evolved in time at a different temperature using molecular dynamics simulations. Using Metropolis criterion, configuration exchanges between replicas are made, allowing for configurations once accessible only at high temperatures to be simulated at low temperatures, and vice versa.

Ring-closing olefin metathesis: An olefin metathesis reaction that allows closing of 7-8 member rings.

Spin-glass: A disordered magnet, where the orientation of magnetic poles (the spin on atoms) in 3-D space are not aligned in a regular pattern. The disordered state results in frustrated interactions or distortions of the geometry of atomic bonds that would otherwise be ordered in a regular array in a solid.

Stapled α-helical peptide: A strategy for stabilizing α-helices through an all-hydrocarbon cross-link. One key component of this strategy is α-methylated amino acids with olefinic side-chains of varying length and configured with either *R* or *S* chirality. These unnatural amino acids are incorporated into the peptide at either *i* and *i*+4 or *i* and *i*+7 positions, and fused using olefin metathesis to cross-link one or two turns of the helix.

Supercooled liquid: A glassy state that is broadly defined as liquid which no longer flows. A supercooled state can be obtained by sufficient fast cooling of a liquid to bypass the freezing point T_m to reach T_g, the glass transition temperature.

Thermodynamic integration: A method to calculate the free energy difference between states by defining a thermodynamic path between the states and integrating over ensemble-averaged enthalpy changes along the path.

Umbrella sampling: A technique used to improve conformational sampling of a system by applying a biasing potential to overcome a barrier in the system's energy landscape. The biasing potential is applied along a reaction coordinate to obtain conformations that typically would not be explored in a regular MD simulation, so that one can accurately calculate the PMF along the reaction coordinate.

WaterMap: A software package from Schrödinger, Inc. utilized to calculate thermodynamic properties of solvent exposed binding pocket sites in proteins.

Weighted-histogram analysis method: A method used to unbias the potential applied in umbrella sampling MD simulations to obtain the PMF.

Introduction

This thesis uses theory and computation to address a variety of pertinent biophysical questions important to theorists and experimentalists alike. Some of the questions this work will address are: What are the conformational preferences, interaction profiles, and binding affinities of stapled p53 peptide analogs in complex with the MDM2 receptor? Can one characterize the binding interactions of the stapled peptide constructs to MDM2, and rationalize the experimental binding trends? What are the collapse dynamics effects of different divalent cations in the ion atmosphere around a polyelectrolyte? Are there conformational and energetic consequences of polyelectrolyte collapse under these conditions? Can the surface tension of an entropic droplet be characterized within the framework of Xia-Wolynes random first order transition (RFOT) model? Are phosphate backbone extended DNA oligomers stable and if so, what conformation do they adopt? Can computational modeling of the modified DNA oligomers explain experimental melting temperature trends? Can one design a secondary structure of the 58mer ribosomal RNA fragment to characterize the conformation of the native basin at low Mg^{2+} salt concentration? Does a generalized Manning condensation model capture the ion atmosphere of RNA secondary structure and agree with experimental results?

Below is a summary of the theory and concepts related to the topics investigated in this thesis: (A) probing the origin of structural stability of all hydrocarbon, single and double stapled p53 peptide analogs in complex with MDM2, (B) investigating surface charge density effects on polyelectrolyte folding using Brownian dynamics, (C) the

nature of surface tension in random first order transition model of supercooled state, and (D) conformation and structural stability of nucleic acids.

A. p53 – MDM2 binding interaction

p53 is a protein of interest to many researchers for its multiple roles in anticancer activity.[1,2] The p53 protein can activate pathways for DNA repair when DNA sustains damage, which is an important process of interest to many researchers studying the phenomenon of aging.[3] Also, the cell cycle is halted upon p53 recognition of DNA damage, and p53 can initiate apoptosis if DNA damage is irreparable.[3] The MDM2 protein binds to p53 and negatively regulates p53 activity by transporting it from the nucleus to the cytosol, rendering it inactive.[4] MDM2 also functions as E3 ubiquitin ligase, covalently attaching ubiquitin to p53 and marking it for degradation by the proteasome.[4] To regulate p53 and prevent it from killing healthy cells, the presence of MDM2 is imperative, but in cancerous cells the p53 protein is unable to achieve anticancer function if MDM2 is overexpressed, leading to a common cause of tumor growth.[5,6]

A number of techniques have been applied to restore the native p53 anticancer activity, and this work will focus on inhibiting MDM2 regulation of native p53 by creating stapled p53 peptides. A small α-helical region of the p53 protein binds to MDM2, and a short sequenced peptide representing this p53 region potentially provides a non-genotoxic anticancer agent with a novel mechanism of action.[7,8] One obstacle of using a short p53 peptide segment is that the fragment is less likely to maintain its MDM2 bound conformation while in solution.[9] Enhanced helicity of the p53 peptide

fragment has been shown for peptide analogs with an all-hydrocarbon staple.[10-12] The secondary structure stability of stapled peptide analogs in solution have been compared to the wild type p53 peptide[8], however a detailed computational study on the binding interaction of stapled p53 peptides to MDM2 could optimize the peptide analog most likely to restore native p53 activity.

This work uses replica-exchange molecular dynamics (REMD) simulations, WaterMap calculations, and umbrella sampling molecular dynamics (MD) with potential of mean force (PMF) calculations to characterize the binding of wild type and stapled p53 peptides to the MDM2 protein.[13] WaterMap calculations were performed using the WaterMap module of the Schrodinger Suite 2011.[14] The calculation details and theory have been well established in the literature, and will be thoroughly discussed in Chapter 1.[15,16] Using computer simulations to calculate potential of mean forces along a reaction coordinate has been proven effective to overcome energy barriers using a biasing potential.[17,18] The biased reaction pathway can be unbiased using a weighted-histogram analysis method (WHAM) to obtain the PMF.[19,20] This method is exploited to analyze the p53-MDM2 binding interaction along a chosen reaction coordinate and is discussed in detail in Chapter 1.[13]

B. Polyelectrolyte electrostatics

The ion atmosphere around polyelectrolytes is complex due to the phenomenon of counterion condensation, first realized by Onsager and later worked out by Manning.[1,2] To summarize this phenomenon: the closely placed phosphate charges along a DNA backbone leads to strong electrostatic repulsion. When DNA is immersed in an aqueous

solution containing monovalent or divalent cations from added salts, the thermodynamic properties of the system are electrostatically unstable.[1-4] As cations, also referred to as counterions, from the bulk solution condense onto the backbone of the DNA, the favorable electrostatic interaction leads to a lowering of the free energy of the system.[3]

Each phosphate charge is reduced by a factor 1-θ, where θ is the number of counterions associated with a single phosphate. Manning and Onsager have shown that θ depends on the valence of the counterions, the charge spacing of the polyelectrolyte, and the Bjerrum length: the characteristic distance between a pair of charges at which the coulomb interaction balances thermal fluctuations.[5,6] Manning also found that the percent condensation of monovalent cations on B-form DNA was 76% and that this parameter remained independent of salt concentration.[3]

Brownian dynamics simulations have been used to investigate the thermodynamic properties of counterion condensation in order to characterize the ion atmosphere effects of polyelectrolyte collapse. This technique has been established for polyanion–polycation complexation[7], and we evoke a similar protocol in Chapter 2 for polyanion–M^{2+} complexation using a bead-spring model.[8] The polyelectrolyte behavior of a 120 monomer polyanion is analyzed in the presence of monovalent and divalent ions, with several divalent ion sizes. Specifically, we are interested in the salt dependence of polyelectrolyte chain collapse dynamics by introducing different M^{2+} (Mg^{2+}, Ca^{2+}, Sr^{2+} and Ba^{2+}) into the salt solution. We calculate a free energy and entropy of folding dependence on the divalent cation size and despite using a simplified bead-spring model and Brownian dynamics simulations, the results follow similar trends as experimental measurements for ΔG of folded and unfolded conformations of *Tetrahymena* ribozyme.[9]

C. Structure of a supercooled liquid

In Chapter 3 a detailed background and theoretical derivation on the nature of surface tension in the random first order transition (RFOT) model of a supercooled state is presented. Therefore, this section will serve to outline a few definitions that will clarify the ideas in Chapter 3.

Supercooled, glass-forming liquids are everywhere: from eyeglasses or window panes, to hard plastics like polystyrene or naturally occurring volcanic glass.[1] Even though the glassy chemical state is well-known and characterized physically, researchers still strive to develop theoretical models to describe structural glass-forming liquids.[2-4] A supercooled liquid is a state of matter in which the temperature of a substance has been quickly lowered to bypass the melting point T_m, and reach the glass transition temperature, T_g. The structure of the glass that is achieved through this process is a disordered molecular array that is more representative of the liquid state, even though the material appears hard, brittle, and glass-like. Supercooled liquids are far away from equilibrium, and the driving force toward equilibrium in RFOT theory of supercooled liquids is explored further in Chapter 3.[5] Also, the work in Chapter 3 presents the theoretical result for the fragility index, expressed as the square of the ratio of the heat capacity and configurational entropy.[5]

D. Nucleic acid structure and stability

DNA nucleoside modifications

Modifications to the DNA structure can be utilized to develop pharmaceutical drugs or to create tools that enhance our understanding of the biological roles of nucleic acids.

The modifications can be made to the nucleobase[1], the sugar[2-4], or the phosphate backbone[5,6]. Base modifications have been made to investigate the driving force of helix formation. Duplexes that form in systems lacking hydrogen bonds between base pairs support the claim that base stacking is a primary driving force of helix formation.[7,8] The sugar moiety of the duplex can be modified by changing the ribose ring size, stereochemistry, or by making mutations to the traditional heteroatoms. Cytotoxicity properties are exhibited in DNA with the ribose O4' oxygen substituted with sulfur.[9] Of particular interest to this work, backbone modifications can include elongating or shortening the number of carbons in the backbone, mutating backbone atoms, or changing the entire backbone by changing how nucleosides are linked. Substituting sulfur or a heavy oxygen isotope for a phosphate oxygen generates a stereocenter at the phosphate and allows for stereospecific monitoring of the duplex backbone.[10,11] Peptide nucleic acids have a backbone of pseudo-peptide linked 'nucleosides' instead of a phosphodiester link. These backbone modified oligomers still form duplexes, however they lack the phosphate negative charges, eliminating the charge-charge repulsion present in unmodified DNA.[12,13]

The focus of this work is on DNA backbone extension by inserting a methylene group at either the 3'-carbon (modified adenosine base is referred to as 3'-dA) or 5'-carbon (modified adenosine base is referred to as 6'-dA) position, depicted in Figure 1.

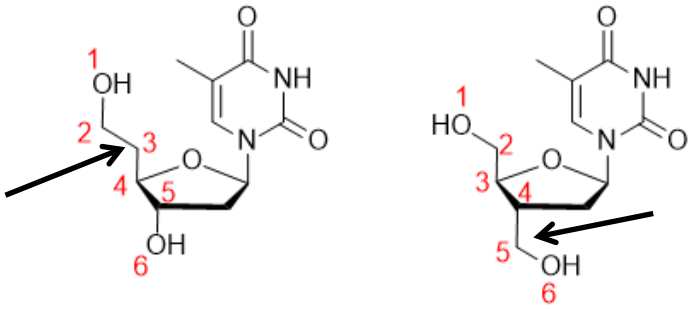


Figure 1: Schematic representation of a modified thymine nucleoside, extended at the 3'-carbon (left) and the 5'-carbon position (right).

The effect on duplex structure and stability due to incorporation of extended nucleosides is investigated using molecular dynamics and the results are presented in Chapter 4. These modified oligomers were synthesized by our collaborators in Professor McLaughlin's lab in the Boston College Chemistry Department. The synthetic routes and experimental procedures are beyond the scope of this work and will be presented in the manuscript in preparation[14], however for comparison to computational results, the experimental melting temperature data is reported in Chapter 4 for the modified DNA oligomers studied.

Generalized Manning model to describe Mg^{2+}-induced RNA folding

Electrostatic models that are capable of describing Mg^{2+}-RNA interactions are needed to match experimental ion atmosphere results and accurately describe the RNA energy landscape.[15-18] Existing models of ion solution electrostatics like Debye-Hückel[19,20] and

Nonlinear Poisson-Boltzmann (NLPB)[21-23] fall short in describing strong Mg^{2+}-RNA interactions, ion-ion correlations, and ion size effects.[24-27]

It is well known that the ionic environment can have a profound impact on RNA and that the presence of Mg^{2+} in the RNA ion atmosphere is typically required to form compact structures.[28] This requirement for RNA folding originates from the highly negative charged RNA phosphate backbone that creates a strong repulsive effect, which is reduced when Mg^{2+} ions condense from the ion atmosphere and interact with the RNA surface.[29,30] As Mg^{2+} concentrations are increased in the ion atmosphere, equilibrium folding of the RNA tertiary structure occurs in two steps: formation of a compact I state intermediate, followed by formation of tertiary contacts that stabilize the N state.[17]

In collaboration with Onuchic's group we have developed a coarse-grained model of RNA that accurately describes the excess ion atmosphere of RNA. The model takes into account the Mg^{2+}-Mg^{2+}, Mg^{2+}-phosphate, phosphate-phosphate correlations by treating the RNA atoms and Mg^{2+} ions explicitly. KCl condensation can vary with the RNA conformation and Mg^{2+} competes with K^+ at the RNA surface. These conditions are considered in our implicit model for the monovalent salt such that the electrostatic interactions, heterogeneity of the phosphates, screening ions, and ion accessibility near the RNA are included. Further details and theory of the model are outlined in a recent publication.[31]

In chapter 5, we describe the 58mer ribosomal fragment system and study its I and N states. Our model can uniquely characterize the ion atmosphere of both states to elucidate the origins of Mg^{2+}-induced folding of RNA tertiary structure.

References

A. p53-MDM2 binding interaction

1. Joerger, A. C.; Fersht, A. R. *Cold Spring Harb. Perspect. Biol.* **2010**, 2, 1-20.

2. Verdine, G. L.; Walensky, L. D. *Clin. Cancer Res.* **2007**, 13, 7264-7270.

3. Huang, F.; Rajagopalan, S.; Settanni, G.; Marsh, R. J.; Armoogum, D. A.; Nicolaou, N.; Bain, A. J.; Lerner, E.; Ying, L.; Fersht, A. R. *Proc. Natl. Acad. Sci. USA* **2009**, 106, 520758/1-520758/9.

4. Chen, J.; Lin, J.; Levine, A. J. *Mol. Med.* **1995**, 1, 142-152.

5. Keleti, J.; Quezado, M. M.; Abaza, M. M.; Raffeld, M.; Tsokos, M. *Am. J. Pathol.* **1996**, 149, 143-151.

6. Taylor, A. C.; Shu, L.; Danks, M. K.; Poquette, C. A.; Shetty, S.; Thayer, M. J.; Houghton, P. J.; Harris, L. C. *Med. Pediatr. Oncol.* **2000**, 35, 96-103.

7. Baek, S.; Kutchukian, P. S.; Verdine, G. L.; Huber, R.; Holak, T. A.; Lee, K. W., Popowicz, G. M. *J. Am. Chem. Soc.* **2012**, 134, 103-106.

8. Guo, Z.; Mohanty, U.; Noehre, J.; Sawyer, T. K.; Sherman, W.; Krilov, G. *Chem. Biol. & Drug Design* **2010**, 75, 348-359.

9. Onuchic, J. N.; Luthey-Schulten, Z.; Wolynes, P. G. *Annu. Rev. Phys. Chem.* **1997**, 48, 545-600.

10. Bernal, F.; Tyler, A. F.; Korsmeyer, S. J.; Walensky, L. D.; Verdine, G. L. *J. Am. Chem. Soc.* **2007**, 129, 2456-2457.

11. Schafmeister, C. E.; Po, J.; Verdine, G. L. *J. Am. Chem. Soc.* **2000**, 122, 5891-5892.

12. Kritzer, J. A.; Lear, J. D.; Hodsdon, M. E.; Schepartz, A. *J. Am. Chem. Soc.* **2004**, 126, 9468-9469.

13. Guo, Z.; Streu, K.; Krilov, G.; Mohanty, U. *Chem. Biol. & Drug Design* **2014**, 83, 631-642.

14. *Schrödinger Suite 2011: WaterMap* version 1.3. New York, NY: Schrödinger, LLC.

15. Abel, R.; Young, T.; Farid, R.; Berne, B. J.; Friesner, R. A. *J. Am. Chem. Soc.* **2008**, 130, 2817-2831.

16. Wang, L.; Berne, B. J.; Friesner, R. A. *Proc. Natl. Acad. Sci. USA* **2011**, 108, 1326-1330.

17. Roux, B. *Comp. Phys. Comm.* **1995**, 91, 275-282.

18. Souaille, M.; Roux, B. *Comp. Phys. Comm.* **2001**, 135, 40-57.

19. Gallicchio, E.; Andrec, M.; Felts, A. K.; Levy, R. M. *J. Phys. Chem. B* **2005**, 109, 6722-6731.

20. Masunov, A.; Lazaridis, T. *J. Am. Chem. Soc.* **2003**, 125, 1722-1730.

B. Polyelectrolyte electrostatics

1. Manning, G. S. *J. Chem. Phys.* **1969**, 51, 924-933.

2. Manning, G. S. *J. Chem. Phys.* **1969**, 51, 3249-3252.

3. Manning, G. S. *Quarterly Review of Biophysics* **1978**, 111, 179-246.

4. Manning, G. S. *Acc. Chem. Res.* **1979**, 12, 443-449.

5. Fenley, M. O.; Manning, G. S.; Olson, W. K. *J. Phys. Chem.* **1992**, 96, 3963-3969.

6. Manning, G. S. *Biophys. Chem.* **2002**, 101-102, 461-473.

7. Ou, Z.; Muthukumar, M. *J. Chem. Phys.* **2006**, 124, 154902/1-154902/11.

8. Streu, K; Mohanty, U. (unpublished).

9. Koculi, E.; Hyeon, C.; Thirumalai, D.; Woodson, S. *J. Am. Chem. Soc.* **2007**, 129, 2676-2682.

C. Surface tension of a supercooled liquid

1. Angell, C. A. *Science* **1995**, 267, 1924-1935.

2. Adam, G.; Gibbs, J. H. *J. Chem. Phys.* **1965**, 43, 139-146.

3. Mohanty, U. *Physica A* **1991**, 177, 345-355.

4. Mohanty, U. *Adv. Chem. Phys.* **2004**, 129, 73-105.

5. Streu, K; Wolynes, P. G.; Mohanty, U. (unpublished).

D. Nucleic acid structure and stability

1. Jacobsen, N.; Bentzen, J.; Meldgaard, M.; Jakobsen, M. H.; Fenger, M.; Kauppinen, S.; Skouv, J. *Nucleic Acids Res.* **2002**, 30, e100/1-e100/7.

2. Hertoghs, K. M. L.; Ellis, J. H.; Catchpole, I. R. *Nucleic Acids Res.* **2003**, 31, 5817-5830.

3. Kumar, N.; Nielsen, K. E.; Maiti, S.; Petersen, M. *J. Am. Chem. Soc.* **2006**, 128, 14-15.

4. Lee, J. A.; Moon, H. R.; Kim, H. O.; Kim, K. R.; Lee, K. M.; Kim, B. T.; Hwang, K. J.; Chun, M. W.; Jacobsen, K. A.; Jeong, L. S. *J. Org. Chem.* **2005**, 70, 5006-5013.

5. Guga, P.; Maciaszek, A.; Stec, W. J. *Org. Lett.* **2005**, 7, 3901-3904.

6. Xu, Q.; Katkevica, D.; Rozners, E. *J. Org. Chem.* **2006**, 71, 5906-5913.

7. Yakovchuk, P.; Protozanova, E.; Frank-Kamenetskii, M. D. *Nucleic Acids Res.* **2006**, 34, 564-574.

8. Schweitzer, B. A.; Kool, E. T. *J. Am. Chem. Soc.* **1995**, 117, 1863-1872.

9. Ober, M.; Muller, H.; Pieck, C.; Gierlich, J.; Carell, T. *J. Am. Chem Soc.* **2005**, 127, 18143-18149.

10. Eckstein, F. *Acc. Chem. Res.* **1979**, 12, 204-210.

11. Knowles, J. R. *Annu. Rev. Biochem.* **1980**, 49, 877-919.

12. Egholm, M.; Buchardt, O.; Christensen, L.; Behrens, C.; Freier, S. M.; Driver, D. A.; Berg, R. H.; Kim, S. K.; Norden, B.; Nielsen, P. E. *Nature* **1993**, 365, 566-568.

13. Wittung, P.; Nielsen, P. E.; Buchardt, O.; Egholm, M.; Norden, B. *Nature* **1994**, 368, 561-563.

14. Streu, K.; Das, R.; McCormick, R.; Mohammad, I.; McLaughlin, L.; Mohanty, U (unpublished).

15. Hennelly, S. P.; Novikova, I. V.; Sanbonmatsu, K. Y. *Nucleic Acids Res.* **2013**, 41, 1922.

16. Leipply, D.; Draper, D. E. *Biochemistry* **2011**, 50, 2790.

17. Grilley, D.; Misra, V.; Caliskan, G.; Draper, D. E. *Biochemistry* **2007**, 46, 10266.

18. Soto, A. M.; Misra, V.; Draper, D. E. *Biochemistry* **2007**, 46, 2973.

19. Hyeon, C.; Thirumalai, D. *Proc. Natl. Acad. Sci. USA* **2005**, 102, 6789.

20. Cho, S. S.; Pincus, D. L.; Thirumalai, D. *Proc. Natl. Acad. Sci. USA* **2009**, 106, 17349.

21. Honig, B.; Nicholls, A. *Science* **1995**, 268, 1144.

22. Chen, S. W.; Honig, B. *J. Phys. Chem. B* **1997**, 101, 9113.

23. Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. *Proc. Natl. Acad. Sci. USA* **2001**, 98, 10037.

24. He, Z.; Chen, S. –J. *J. Chem. Theory and Comput.* **2012**, 8, 2095.

25. Boschitsch, A. H.; Danilov, P. V. *J. Comput. Chem.* **2012**, 33, 1152.

26. Chu, V. B.; Bai, Y.; Lipfert, J.; Herschlag, D.; Doniach, S. *Biophys. J.* **2007**, 93, 3202.

27. Borukhov, I.; Andelman, D. *Phys. Rev. Lett.* **1997**, 79, 435.

28. Draper, D. E. *RNA* **2004**, 10, 335.

29. Heilman-Miller, S. L.; Thirumalai, D.; Woodson, S. A. *J. Mol. Biol.* **2001**, 306, 1157.

30. Wittmer, J.; Johner, A.; Joanny, J. F. *J. Phys. II (France)* **1995**, 5, 635.

31. Hayes, R. L.; Noel, J. K.; Mandic, A.; Whitford, P. C.; Sanbonmatsu, K. Y.; Mohanty, U.; Onuchic, J. N. *Phys. Rev. Lett.* **2015**, 114, 258105/1-258105/6.

Chapter One

Probing the origin of structural stability of all hydrocarbon, single and double stapled p53 peptide analogs in complex with MDM2*

*Reproduced with permission from Guo, Z.; Streu, K.; Krilov, G.; Mohanty, U. "Probing the origin of structural stability of all hydrocarbon, single and double stapled p53 peptide analogs in complex with MDM2" *Chemical Biology & Drug Design* **2014**, 83, 631-642.

1. Introduction

The transcription factor p53 commences arrest of the natural cell cycle in response to DNA damage.[1,2] The level of cellular p53 is governed by its interactions with E3 ubiquitin ligase, MDM2.[1,2] Over-expression of MDM2 in a cell results in the loss of p53 activity.[1.2]

Functional proteins usually fold into distinct conformations in solution. However, short peptide segments of a protein are less susceptible to fold when they are separated from the rest of the macromolecule.[3] These fragments are frequently found to be in disordered states in solution and therefore are not able to maintain proper interactions necessary for binding.[3] Stabilizing the secondary structure of peptide via synthetic cross-links has been recently introduced as a strategy to avoid peptide unfolding. Specifically, stabilization of a 16-residue helical domain in p53 was recently accomplished by introducing unnatural α-methylated amino acids with either *S* or *R* stereochemistry spaced one or two helical turns of a peptide sequence apart, and cross-linked with all-hydrocarbon side-chain tethers of different lengths by ring-closing olefin metathesis.[4-6] Circular dichroism studies indicate that such cross-linking substantially improved α-helicity. In particular, the helical content of a stapled p53-peptide *cis*-sah8, whose sequence is shown in Table 1, was found to be 65%, while only 11% the wild type peptide population was in an α-helical conformation.[4,5] Remarkably, the helically stabilized *cis*-sah8 peptides were shown to slow the growth of cancer cells *in vivo* by activating the p53 mediated apoptotic paths.[4]

Table 1. Sequences of wild type, single and double stapled peptides.

Peptide	Sequence																
wild type	Ac	L	S	Q	E	T	**F**[a]	S	D	L	**W**	K	L	**L**	P	E	N	NH_2
cis-sah3	Ac[b]	L	S	Q	$\mathbf{X_R}$[c]	T	**F**	S	D	L	**W**	$\mathbf{X_S}$	L	**L**	P	E	N	NH_2[d]
cis-sah4	Ac	L	S	Q	E	T	**F**	$\mathbf{X_R}$	D	L	**W**	K	L	**L**	$\mathbf{X_S}$	E	N	NH_2
cis-sah8	Ac	Q	S	Q	Q	T	**F**	$\mathbf{X_R}$	N	L	**W**	R	L	**L**	$\mathbf{X_S}$	Q	N	NH_2
trans-sah8	Ac	Q	S	Q	Q	T	**F**	$\mathbf{X_R}$	N	L	**W**	R	L	**L**	$\mathbf{X_S}$	Q	N	NH_2
cis-dsah8	Ac	Q	S	Q	Q	$\mathbf{X_S}$	**F**	$\mathbf{X_R}$	N	$\mathbf{X_S}$	**W**	R	**L**	**L**	$\mathbf{X_S}$	Q	N	NH_2

a. Dark blue color denotes key hydrophobic residues Phe19, Trp23, and Leu26.
b. Ac denotes N-terminal acetylation.
c. X (red, orange, and green) denotes α-methylated amino-acids with R and S stereochemistry. ($\mathbf{X_S}$,$\mathbf{X_S}$) and ($\mathbf{X_R}$,$\mathbf{X_S}$) denote the respective positions and chiralities of the α-methylated amino-acids anchoring the two linkers and their respective chirality.
d. NH_2 denotes C-terminal primary amide.

In this chapter, we investigate the extent to which the helical propensity of various stapled peptides (Table 1) in solution reflects their conformation preference, interaction profile and binding affinity in complex with the MDM2 receptor. The stapled peptide is obtained by inserting an all-hydrocarbon cross-link anchored by a pair of α-methylated amino acid termini inserted at positions *i* and *i*+7 of an α-helical peptide sequence, such that the cross-link spans two full turns of the helix. In addition, we have also designed a double stapled peptide by introducing a second eight-carbon bridge with *S* chirality amino acids incorporated at positions i and i+4; this cross-link is on the opposite side of the helix of that connecting residues i and i+7. Several modifications of the wild type p53 peptide sequence (Table 1) with the net peptide charge ranging from -2 to +1, which conserve the Phe19, Trp23 and Leu26 triad known to be critical for binding to the MDM2 were explored.[7] These were found to have significant effects on binding affinity and cellular potency. As one of the primary mechanisms believed to drive binding, the various contacts by MDM2 to the hydrophobic face of the peptide are believed to shield the protein-protein interface from the solvent and probably inhibit access to various protein-associated factors.[7] When the p53 peptide is bound to its protein-associated factors, it forms a complex that initiates transcription and transactivation. Since p53 binds to MDM2 using the key residues involved in initiating transactivation, the MDM2 protein prevents the p53 peptide activity; and when p53 is bound to MDM2, transactivation will not occur.

2. Theoretical methods

Simulation system preparation

The sequences of variants of the 16-residue transactivation domain of p53 peptides used in this study are shown in Table 1. The hydrocarbon tether stapled to the peptide at either i, i+7 or i, i+4 residues was incorporated, with the double bond placed in a *cis* or *trans* configuration, using Maestro suite of programs.[8,9] A publicly available crystal structure of *trans*-sah8 bound to MDM2 (PDB ID 3V3B)[10] was used as a template for constructing the initial configurations of other stapled peptide complexes. Protonation states consistent with pH=7 were assigned using the Protein Preparation Wizard.[11-14] The protein chain termini were capped using neutral ACE an NAC caps. Both the free peptide and the complex were placed in an orthorhombic box and solvated with TIP3P water[15] such that the solvent buffer of at least 7 Å was maintained around the protein. An appropriate number of sodium or chloride ions were added to maintain charge neutrality.

Replica-exchange molecular dynamics simulations

Each system was equilibrated at 300 K by a series of constrained minimizations and all atom molecular dynamics simulations using the default equilibration protocol in Desmond.[8] The OPLS-2005 force field and periodic boundary conditions were employed in the simulations.[16,17] Long-range electrostatic interactions were calculated by particle-mesh Ewald method.[18] Short range electrostatic and Van der Waals interactions were truncated at 7.0 Å. To control the pressure, Martina–Tobias–Klein method was used.[19] Constant simulation temperature was maintained by Nose–Hoover thermostats.[20]

Replica-exchange molecular dynamics[21], as implemented in Desmond was used to generate the ensembles of bound and free peptide conformers. In order to maintain the overall positioning of stapled peptide relative to the MDM2 binding groove, a weak harmonic restraining force with a force constant k = 1.0 kcal/mol/Å^2 was applied to both the α-carbon of residue 23 of the peptide and all α-carbons of the protein. Restraining the large-scale relative motions of the protein effectively reduced the number of degrees of freedom, allowing efficient sampling of the binding site interactions with a tractable computational effort. Hence, 64 replicas linearly distributed between 300 K and 600 K were sufficient to provide adequate coverage of the local conformational space. Each replica was evolved in parallel for 20 ns in the NPT ensemble. The equations of motion were integrated using the multistep RESPA integrator[22] with an inner time step of 2.0 fs for bonded interactions and non-bonded interactions within the short range cutoff, and the outer time step of 6.0 fs for other long range interactions. Replica exchanges were attempted every 12 ps, after the first 100 ps, between each pair of nearest neighbor trajectories. After 10 ns period, the configurations were saved in 1 ps interval for the final 10 ns run. In order to assess convergence, ensemble averages of relevant quantities computed over the first and the last half of the production period of the simulation were compared.

Binding free energy calculations

WaterMap analysis

WaterMap calculations were performed using the WaterMap module of the Schrödinger Suite 2011.[23] The structure of the *trans*-sah8 peptide in complex with the MDM2 receptor was obtained from the crystal structure (PDB ID 3V3B).[10] The complex structure was prepared using a standard protocol as implemented in the Protein Preparation Wizard.[24] Hydrogens were added as appropriate, terminals capped, and waters beyond 5Å from the binding site were deleted. The protonation states of ionizable residues were assigned according to their local environment using PROPKA[25-28] such as to optimize the hydrogen bonding network. The structure was then subject to minimization restrained to 0.3 Å RMSD to relax the system, optimize interactions and remove bad contacts.

The binding site for the WaterMap simulation was defined by the position of the co-crystallized stapled peptide. The regions of the protein beyond 10 Å from the binding site were truncated. All interactions were treated using the OPLS-2005 force field.[16,17] The standard protocol, consisting of a series of restrained minimizations and short MD simulations followed by a grand canonical Monte Carlo sampling of the binding site waters, as implemented in WaterMap was applied to equilibrate the solvated system at 298 K. A 5000 ps production MD simulation was performed to collect water statistics with the protein and ligand heavy atoms restrained. The resulting water trajectory was analyzed through clustering to identify hydration sites within 10 Å of the peptide binding sites and compute their thermodynamic properties. The contributions to the binding free

energy due to water displacement were computed from the overlaps of the ligand atoms with the hydration sites according to[29]:

$$\Delta G_{\text{bind}} = \sum_{i,j} E_{\text{rwd}} \left(1 - \frac{|r_i - r_j|}{R_{\text{co}}}\right) \Theta\left(E_j - E_{\text{co}}\right) \Theta\left(R_{\text{co}} - |r_i - r_j|\right)$$
$$-T \sum_{i,j} S_{\text{rwd}} \left(1 - \frac{|r_i - r_j|}{R_{\text{co}}}\right) \Theta\left(S_j^{\text{e}} - S_{\text{co}}\right) \Theta\left(R_{\text{co}} - |r_i - r_j|\right) \quad (1)$$

where the sum runs over *i* ligand atoms and *j* hydration sites; Θ is the Heavyside function, and R_{co}, E_{co}, E_{rwd}, S_{co} and S_{rwd} are empirically determined parameters.[29] The contribution to the binding free energy for filling an evacuated cavity was computed according to the procedure by Wang et al.[30]:

$$\Delta G_{bind} = -kT \ln\left(P_0\right) - 2.36 \quad (2)$$

where $-kT\ln(P_0)$ is the probability of the cavity being unoccupied, as determined from the WaterMap simulation, and 2.36 kcal/mol is the solvation free energy of methane.

Weighted-histogram analysis method (WHAM)

Free energy perturbation (FEP)[31] and thermodynamic integration (TI)[32] represent the most widely used techniques for relative free energy calculations between different thermodynamic states. However, studies showed that obtaining reliable results using the FEP and TI is computationally expensive even for a simple system. Sampling problems show strong influence on the accuracy when a large energy barrier is present. In this case, convergence will be difficult to achieve using regular molecular dynamics.[33,34]

One way to overcome this problem is to combine samples that are obtained under different simulation conditions. In particular we will employ the potential of mean forces

(PMFs) formalism. In a standard canonical simulation, one may not be able to precisely estimate the barrier height since sampling in the high-energy barrier region will not be as sufficient as the low energy minima. However, if umbrella sampling with biasing potentials is used, the system will be confined to small regions of phase space near a particular value of the reaction coordinates of the PMF. This strategy will provide adequate sampling to the entire reduced reaction coordinate of interest. Furthermore, since the biasing potential is known, it can be easily "unbiased" after sufficient sampling.[35]

In our study, the PMFs were obtained by post-processing the data from a series of umbrella sampling MD simulations with a different center of geometry separation distance between the peptide and protein domains at 300K. Weighted-histogram analysis method (WHAM) is used for the post-processing and the detailed procedures are described in the literature.[36,37] We used the program developed by the Grossfield group to perform the WHAM analysis.[38] The umbrella potential was used to restrain the C_α-atom of the central residue of the ligand and receptor chains (Trp23 in p53 and Leu85 in MDM2) with a force constant 2 kcal/mol/Å (Figure 1). The C_α-atom distances between the two domains were set to range from -2 to 24 Å with 1 Å increments. Note that distance 0 Å represents the original crystal structure geometry or initial geometry for those peptide/protein complexes without crystal structures.

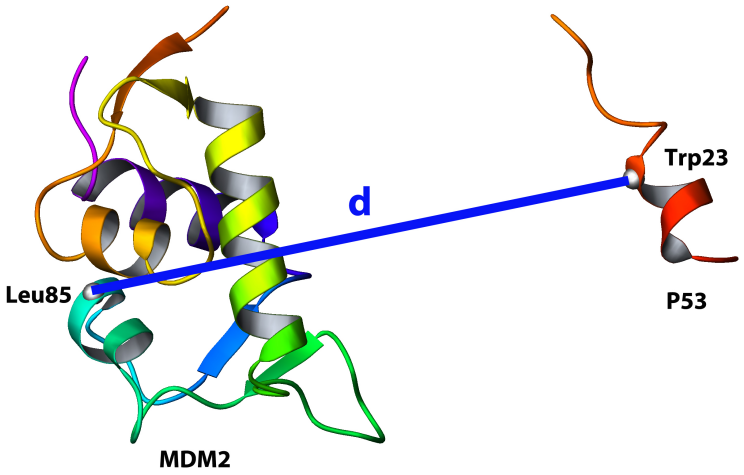


Figure 1. Illustration of the reaction coordinates for the p53 peptide and target protein MDM2.

For each distance, conventional molecular dynamics (MD) simulations were performed using the Desmond package. The OPLS 2005 force field[16,17] was used to model the protein interactions and TIP3P model[15] was used to the water. Particle-mesh Ewald method (PME)[18] was used to calculate long-range electrostatic interactions with grid spacing of 0.8 Å. Van de Waals and short range electrostatic interactions were smoothly truncated at 9.0 Å. Nose-Hoover thermostats[20] were used to maintain the constant simulation temperature and Martina-Tobias-Klein method[19] was used to control the pressure. The equations of motion were integrated using the multistep RESPA integrator with an inner time step of 2.0 fs for bonded interactions and non-bonded interactions with in the short-range cutoff. An outer time step of 6.0 fs was used for non-bonded interactions beyond the cutoff. Periodic boundary conditions were applied. The system was equilibrated with the default protocol provided in Desmond.[39]

After the equilibration, a 2 ns NVT production simulation was performed for each different inter-domain distance at 300K temperature and the simulation system configurations were saved in 4 ps intervals for analysis.

3. Results and Discussion

Peptide stapling and the stability of secondary structural content

Stabilization of secondary structure is believed to play an important role in the binding of p53-derived stapled peptides to the MDM2 receptor. In order to explore this hypothesis, we have computed the conformation population distribution of *cis*-sah3, *cis*-sah4, *cis*-sah8, and *trans*-sah8 that are bound to MDM2. In Figure 2, we show the distribution of the conformational population of the stapled peptides as a function of the helical content (defined as ratio of the number of helical residues to the total number of residues) and the number of $i \rightarrow i+4$ backbone hydrogen bonds indicative of α-helical geometry. Observe that *cis*-sah4, *cis*-sah8 and *trans*-sah8 have multi-peak structures indicating the presence of a fully helical state, several partially folded states, and one unfolded state. Each of the two *cis* stapled peptides shows one significant peak at helicity 35-40% with 1-2 hydrogen bonds. Peptide *cis*-sah4 has two additional strong peaks with helicity at 55% and 25%, each with 2 hydrogen bonds. Meanwhile, *cis*-sah8 has one additional weak peak and one relatively strong peak with helicity at 55% and 25% with 2 hydrogen bonds, respectively. Comparing the conformational distribution of *cis*-sah4 and *cis*-sah8, we found that *cis*-sah4 shows higher α-helical content than *cis*-sah8 when bound to the negative regulator protein MDM2. *cis*-dsah8 exhibits three multipeak structures in the unbound state, a feature also observed for *cis*-sah4 and *cis*-sah8.

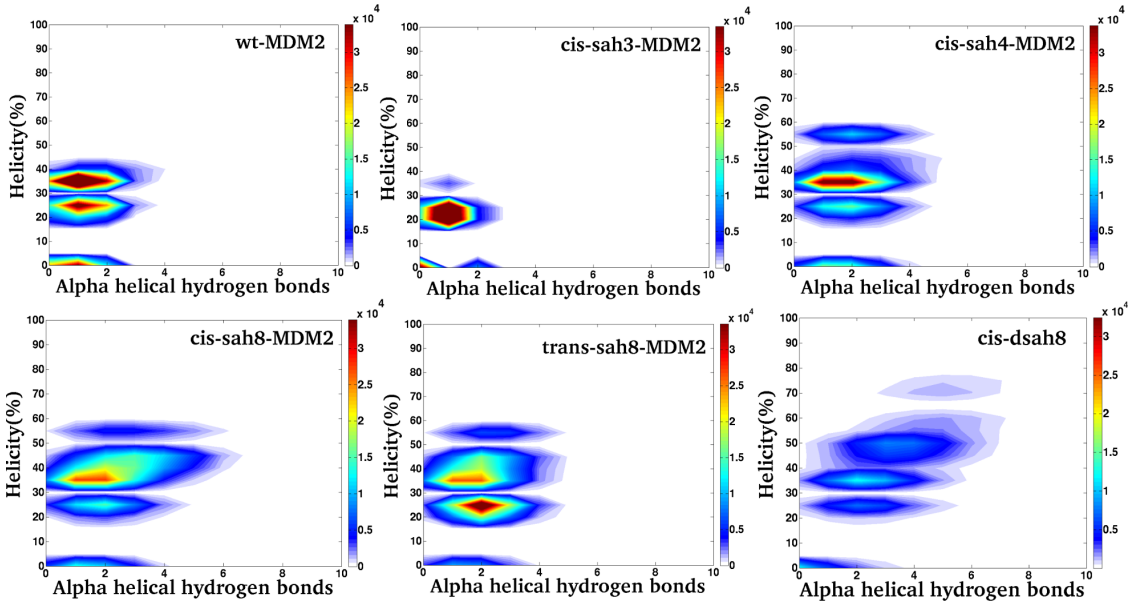


Figure 2. The contour plots of the conformational distribution of the p53 peptide analogs as a function of the number of backbone α-helical hydrogen bonds and the percent α-helical content. More red color indicates higher intensity.

The distribution of (ϕ,ψ) dihedral pairs of the stapled peptides bound to MDM2 is shown in Figure 3. The Ramachandran plot of *cis*-sah4 and *cis*-sah8 exhibits three peaks relative to the wild type. One of the peaks has significant population around the α-helical region. This peak is broad, indicating a distorted helix. There are also two additional peaks located in the extended β-sheet region. The *cis*-dsah8 unbound peptide has one peak in the β-sheet region, in addition to a large population in the α-helical region. We have summarized in Table 2 the percentage of residues of the bound peptides whose dihedral pairs are characteristic of α-helix, i.e., within 30^{o} of (-95^{o},-15^{o}) and (-35^{o},-70^{o}). We find that fraction of the conformational population in the α-helical region is larger for the stapled peptides bound to MDM2 than that of the corresponding free peptides in solution (39). For example, for *trans*-sah8 bound to MDM2, 42.26% of residues are found in the α-helical region, while 36.33% were found to be in the same range for the free peptide. The corresponding values for *trans*-sah4 were 34.75% and 30.45%. In contrast, the fraction of residues with α-helical range dihedrals for the free *cis*-sah8 and *cis*-sah4 was 32% and 40%, respectively[39], while it was found to be 43.80% and 41.57% when bound to MDM2. The α-helical populations for *trans*-sah8 and *cis*-sah8 bound to MDM2 are comparable, while that of free *cis*-dsah8 unbound is the highest. Our results indicate that there are contributions from non α-helical structural motifs to the equilibrium ensemble.[39]

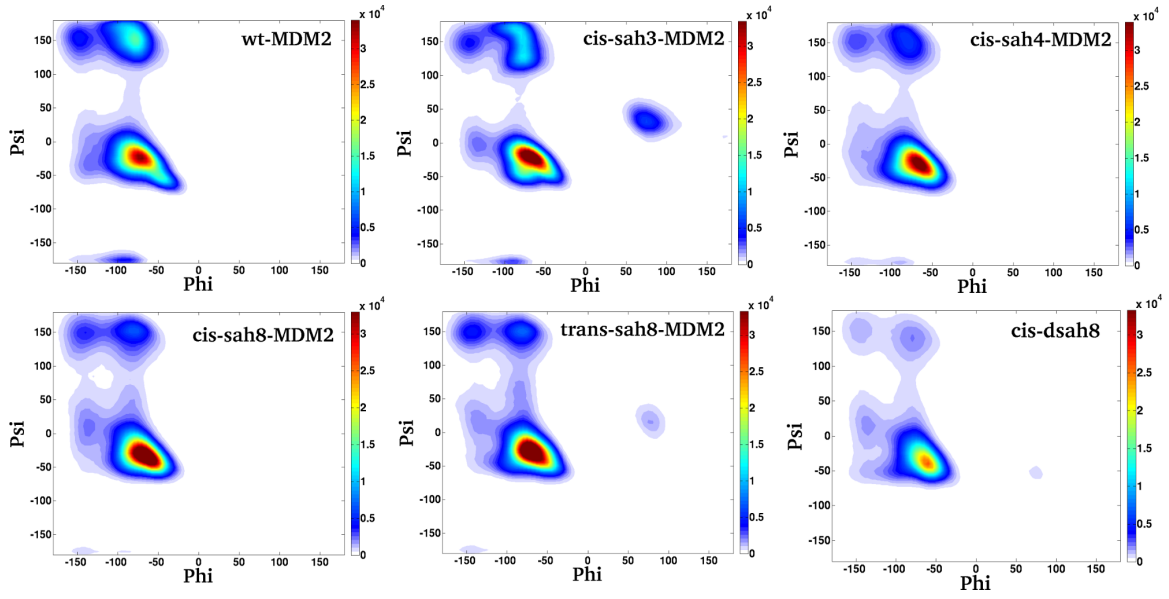


Figure 3. The Ramachandran plot of the distribution of the backbone dihedrals for WT and *cis* and *trans* single and doubly stapled peptides bound to MDM2. More red color indicates higher intensity. Random coil region is significantly higher for the bound wild type and *cis*-sah3 peptides.

Table 2. α-helix dihedrals and backbone hydrogen bonds for single and double stapled peptides.

Peptide	α-helix dihedrals	Backbone H-bonds
wild type	30.19%	3.38
cis-sah3	33.13%	3.54
cis-sah4	41.57%	4.45
cis-sah8	43.80%	4.63
trans-sah8	42.26%	4.64
cis-dsah8	44.99%[a]	5.57[a]

a. The data for *cis*-dsah8 is for unbound state.

The distribution of helical occupancy across the residues of the peptide chains was found to be inhomogeneous (Figure 4). In all cases, the residues spanned by the linker tend to have higher helical occupancy compared to residues outside the linker. It can be observed that, although *trans*-sah8 is cross-linked in the same positions as *cis*-sah8 and *cis*-sah4, it has higher helical content in the region spanned by the linker. Nonetheless, the increase in the stability of the α-helical secondary structure motif imparted by the hydrocarbon cross-link appears to be non-local. In addition, there is an increase in helical content relative to the unstapled wild type peptide persists outside the immediate area spanned by the cross-link, as can be observed for *cis*-sah8, *cis*-sah4, *trans*-sah8, and *cis*-dsah8 as well (Figure 4).

Hence, the stability and propensity to form α-helical conformations depend on the stereochemistry of the hydrocarbon linker, relative position of the cross-link anchor points, and the specific amino acid substitutions.[40,41]

In order to gain further insight into the secondary structure stabilization provided by the cross-linker, the conformational ensembles at several temperatures generated in the course of REMD simulations were analyzed further. To probe this effect, the total helical occupancy as a function of temperature is shown in Figure 4 (bottom). Wild type (WT) and *trans*-sah3 exhibit complete melting of the secondary structure even at relatively low temperatures. In particular, the *trans*-sah3 peptide shows a low occupancy of the α-helical structure for a broad range of temperatures, while doubly stapled *cis*-dsah8 has the highest helical content at room temperature. The helicity of *trans*-sah8 reaches 50% around 340 K and this profile is maintained up to 400 K.

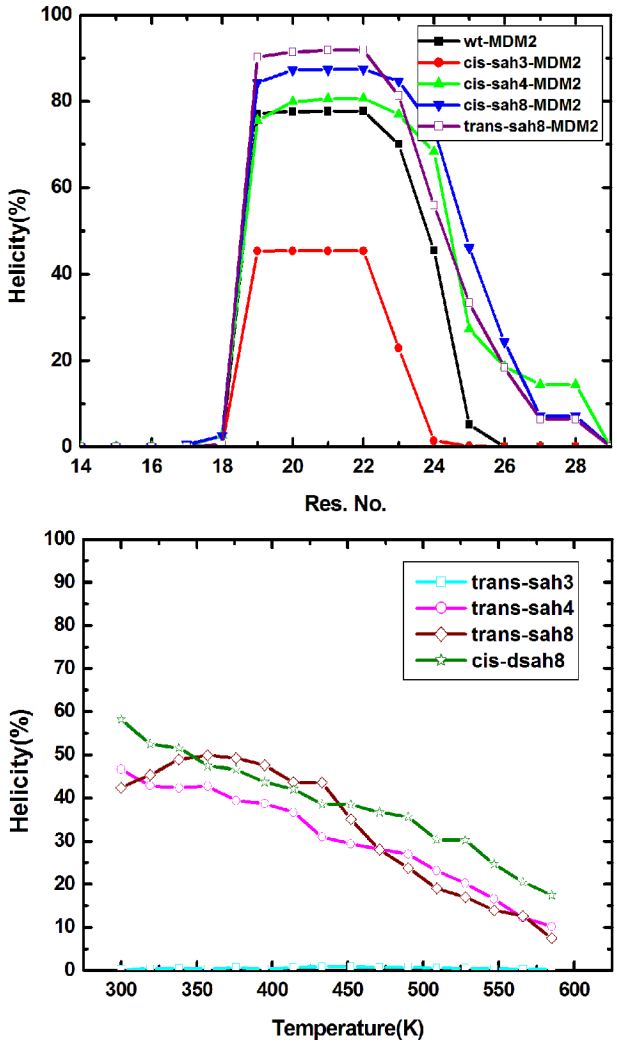


Figure 4. Comparison of percent helical occupancy versus residue number (top) at T = 300 K for WT, *cis*-sah3, *cis*-sah4, *cis*-sah8, and *trans*-sah8 bound to MDM2. The helix melting curves (bottom) for *trans*-sah3, *trans*-sah4, *trans*-sah8, and *cis*-dsah8 in unbound state as a function of temperature.

A similar melting curve is observed for *trans*-sah4 and *cis*-dsah8, however the double stapled peptide consistently maintains a higher helicity over the single stapled peptides. The fastest decay of helical structure is found for *trans*-sah8, indicating that the linker stereochemistry has a significant impact on the degree of secondary structure stabilization, which is consistent with experimental findings. The stereochemistry of the stapled linker is shown to have some effect on the binding affinity of the stapled peptides. The *cis*-sah8 and *trans*-sah8 helicity studies are comparable, however the *cis* linker allows the peptides to maintain slightly higher helical content throughout the simulations. In Figure 5, the double stapled peptide maintains significant helical content, higher than the single stapled peptide, even above 500 K and outside the stapled region. The high content of α-helical structure is just one factor that contributes to the binding affinity of the stapled peptides, however the helical content is not strictly an exclusive evaluation of the binding affinity.

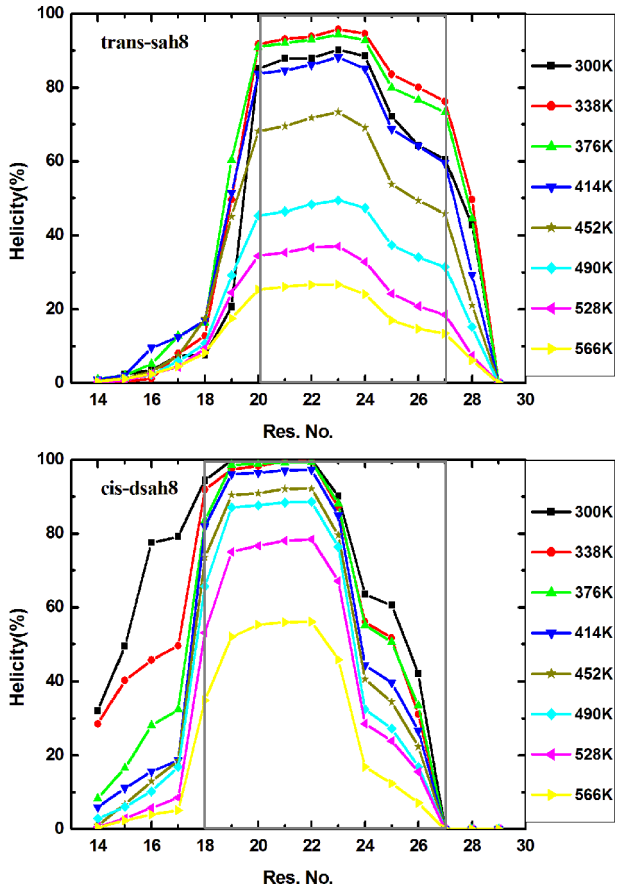


Figure 5. Percentage helicity versus residue number of *trans*-sah8 (top) and *cis*-dsah8 (bottom) at various temperatures.

We next investigated the mechanism through which the binding with MDM2 further stabilizes the receptor-peptide complex. A useful measure of the protein structure stability can be obtained by examining the fraction of native inter-residue contacts Q maintained during the course of the MD simulation, which are captured by the ensemble characteristic P(Q). A native contact is formed when the distance between pairs of C_α atoms of the residues proximal in the crystal structure are within a distance r_c taken to be 6.5 Å. The cutoff value r_c = 6.5 Å for native contacts was chosen based on the occurrence of the first peak in the radial distribution of residues in the interior of proteins.[42] From the crystal structure, we can specify these native contact C_α atom pairs and obtain the total number $N_{crystal}$ of them. Meanwhile, from each MD simulation frame, the distances for the same C_α atom pairs also can be obtained and a number $N_{simulation}$ was used to count the distances smaller than 6.5 Å. A Q value is the ratio between $N_{simulation}$ and $N_{crystal}$. The population distribution of the native contact fraction P(Q) for all the MD simulation frames is shown in Figure 6 (top) for WT and stapled peptides, in which the double bond in the cross-linker is in either the *cis* or *trans* conformation. The ordering of the systems with respect to their Q values is independent to the value of r_c. Specifying a different r_c value will simply result in a shift in the P(Q) distribution for all systems.[43]

To elucidate the contributions of specific peptide residues to Q, we analyzed the contact forming probability per residue, P_i, for the same set of stapled peptides [Figure 6 (bottom)]. The probability, P_i, was computed from the simulation using the relation[43]

$$P_i = \sum_i \sum_j \Delta_{ij} \Theta(x) / t_o \sum_j \Delta_{ij}, \qquad (3)$$

where the contact map of residues i and j is denoted by Δ_{ij}, $\Theta(x)$ is the characteristic function, t_o is the number of simulation frames, and $x = r_c - r_{ij}(t)$.

When the stapled peptide is bound to MDM2, the peak of the distribution is shifted to higher values of Q compared to the wild type complex, indicating an overall stabilization of the complex structure. The time-averaged value (20ns MD simulation for each system) of the fraction of native contacts, $\overline{Q}$, for WT, *trans*-sah8, *cis*-sah8 and *cis*-dsah8 complexed with MDM2 was found to be 0.90±0.04, 0.88±0.05, 0.89±0.04, and 0.93±0.03 respectively. In comparison, the time-averaged value $\overline{Q}$ for the unbound WT and *trans*-sah8 peptides free in solution was 0.79±0.10 and 0.91±0.05, respectively.

The binding free energy ΔG_{bind} between p53 peptides and MDM2 can be decomposed as enthalpical contribution ΔH and entropical contribution $-T\Delta S$. A higher $\overline{Q}$ with smaller error bar can be interpreted as a more stable structure with smaller fluctuation and vice versa. When the wild type p53 peptide binds to MDM2, the time-average value of the native contacts fraction changes from 0.79±0.10 to 0.90±0.04. The structural fluctuation freedom of wild type peptide has been significantly reduced and the entropical changes during the binding process will be a negative ($-T\Delta S$ positive value) contribution to the binding free energy which can be viewed as a enthalpy-entropy compensation situation. Meanwhile, for the binding process of *trans*-sah8 with MDM2, the time-average value of the native contacts fraction $\overline{Q}$ changes is in the opposite direction, indicative of a positive ($-T\Delta S$ negative value) contribution to the binding affinity.

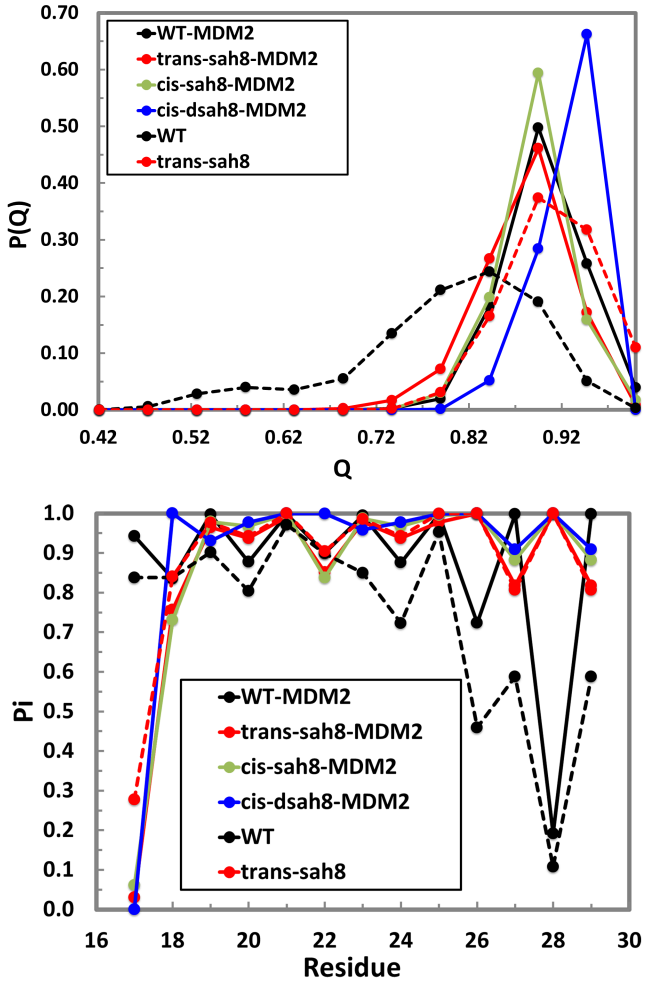


Figure 6. (Top) The distribution P(Q) of the fraction of native contacts Q for the wild type and the stapled peptide MDM2 complexes compared to free peptides in solution. (Bottom) The per residue probability P_i of forming native contacts for the wild type and the stapled peptide MDM2 complexes compared to free peptides in solution.

The probability of forming native contacts of a number of peptide residues is increased when the peptide is stapled, and increased further when the stapled peptides are bound to MDM2. Notably, the stabilization effects of stapling are not local.[39,43]

The crystal structure of wild type p53 bound with MDM2 (PDB ID: 1YCR)[7] shows four native contacts between the three key peptide residues and those lining the MDM2 binding pocket ($Residue_{p53}$, $Residue_{MDM2}$): (Phe19, Gln72), (Phe19, Hie73), (Trp23, Val93), and (Leu26, Hie96). Figure 7 (top) depicts the fraction of the native contacts observed between the three key p53 residues Phe19, Trp23, and Leu26, and MDM2 in the course of the simulation.

A common descriptor of the overall geometry of folded structures such as peptides is given by the distribution of the end-to-end distance. Figure 7 (bottom) shows the distribution of the end-to-end distance between residues 17-29 in WT-MDM2, *trans*-sah8-MDM2, *cis*-sah8-MDM2, *cis*-dsah8-MdM2, and WT and *trans*-sah8 unbound. Observe that the various distributions are localized around two distances. The distributions of the WT peptides, both bound and unbound, are localized around the native value of 23.10 Å obtained from crystal structure of MDM2 bound to p53. The distributions of the stapled peptide analogs, both bound and unbound, are localized around 19 Å, a value significantly lower than the WT end-to-end distance which could be derived from higher α-helical and lower extended secondary structural contents.

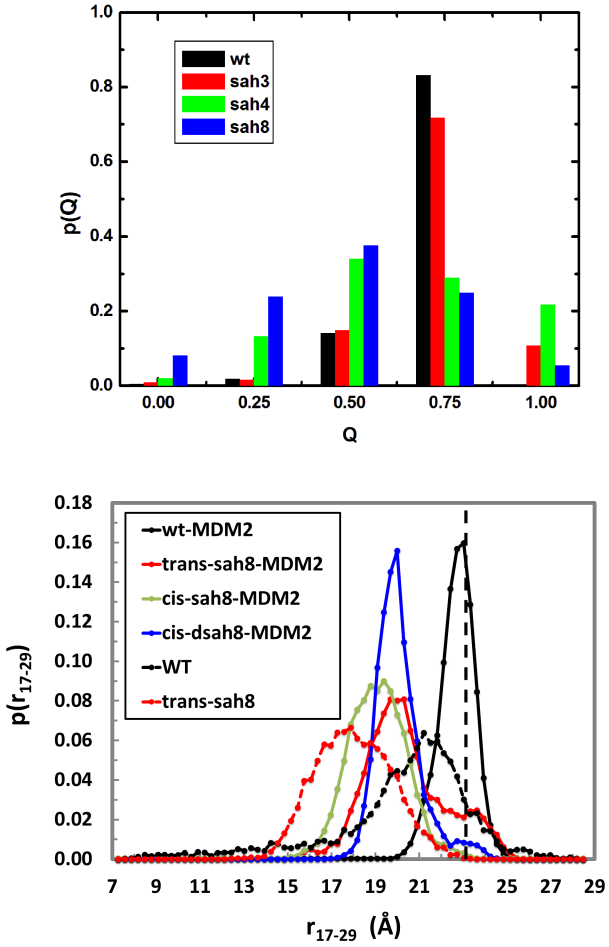


Figure 7. (Top) The distribution P(Q) of the fraction of native contacts Q, for the three key residues of the peptide and MDM2. (Bottom) The end-to-end distance distribution of p53 residues 17-29 in complex with MDM2 and in free solution. The vertical dashed line indicates the value (23.10 Å) of the end-to-end distance obtained from the crystal structure of WT p53-MDM2 (1YCR).

Recently, the rules governing side-chain stereochemistry of the amino acids of an all-hydrocarbon cross-link inserted at *i*, *i*+3 positions into α-helical peptides, and in which the tether bridging 0.83 turns is on one face of the helix, were investigated by Kim *et al*.[44] By exploiting i, i+3 cross-linked systems, these authors designed novel double stapled peptides (44). Specifically, two *i*, *i*+4 cross-links were inserted by ring-closing olefin metathesis on the same face of the helix.[25] The simulations of the double stapled peptide construct explored in this study complements their work and supports the notion that introducing an additional cross-link enhances structural stability of the stapled peptides.

Gō-type formalism[43] has recently been used to study the mechanism through which a single hydrocarbon cross-link leads to improved helical stability of the BH3 peptide[45], excised from the BID protein. The model includes hydrogen bonds along the backbone, interactions between various side chains, and sequence dependent potentials. Quantities such as the distribution of the fraction of native contacts, the average number of native contacts for a given residue, and the distribution of the end-to-end distance were analyzed for the BH3 peptide and the BH3 peptide with a staple.[43]

In agreement with these authors[43], we find that stapling enhances the probability to develop native contacts. We find that the distribution of the end-to-end distance between residues 17-29 is localized around two different distances; specifically the stapled peptides are localized at a smaller end-to-end distance. In contrast, the end-to-end distance between residues 1 and 17 for the BH3 peptide and the stapled BH3 peptide is localized around only the native distance.[43]

To further investigate the relationship between the stability of the secondary structural content and its effect to the binding affinity between p53 peptides with its

target protein MDM2, we have used various calculation techniques to obtain the relative binding affinities for these systems. In the following parts, we will focus our discussion on the results obtained by WaterMap analysis and Potential Mean Forces (PMFs) from WHAM.

Binding affinity of stapled p53 peptides to MDM2

WaterMap analysis

WaterMap simulations yielded a total of 103 hydration sites identifying regions with water density greater than twice that of the bulk. Due to the shallow, solvent exposed nature of the peptide binding cavity, the majority of these sites were found to have thermodynamic properties close to those of bulk water, and are not expected provide a significant contribution to binding. However, several clusters of high-energy (unstable) hydration sites, summarized in Table 3 can be observed.

Table 3. The energies, entropies and free energies (kcal/mol) of the four clusters of high-energy (unstable) hydration sites in the binding pockets of stapled peptides.

Pocket	Hydration Site	ΔH(kcal/mol)	-TΔS(kcal/mol)	ΔG(kcal/mol)
Phe19	28	1.2	1.4	2.6
	29	0.5	1.4	1.9
	64	1.9	0.8	2.7
	66	1.0	0.7	1.8
	41	2.0	1.0	3.0
Trp23	33	0.5	1.2	1.7
Leu26	46	2.9	1.0	3.9
	81	1.2	0.7	2.0
	94	0.5	0.6	1.1
Staple	83	0.2	0.6	0.8
	18	1.9	1.6	3.5
	31	-0.1	1.3	1.2
	72	2.0	1.7	2.7

The first cluster consisting of hydration sites 28, 29, 64, 66, and 41 occupies a broad, shallow hydrophobic pocket lined by Val93, Val75, and Ile61. All of these sites are displaced by Phe19 of the stapled peptide ligands, with the position of the latter largely conserved across the simulations, for a relative contribution of ~12 kcal/mol to binding free energy, obtained by summing the ΔG values in Table 3 of the sites corresponding to the pocket Phe19. The second cluster includes hydration sites 46, 81 and 94, which populate a hydrophobic pocket lined by Ile99, Leu57 and Leu54. All of these sites in turn are displaced by Leu26, which is also found to occupy this pocket for all stapled peptides simulated. In total, displacing the unstable waters from this pocket is estimated to contribute ~7 kcal/mol to binding, obtained by summing the ΔG values in Table 3 of the sites corresponding to the pocket Leu26.

The third cluster, consisting of a single hydration site 33, occupies a deep hydrophobic pocket lined by Val75, Leu57, Ile61, Val93 and Ile99. This site is displaced by Trp23 of the stapled peptides, the residue that occupies this position across all simulations. The displaced hydration site is only marginally unstable, due to favorable hydrogen bonding between the water molecule occupying the site and backbone carbonyl of Leu54, which is functionally replaced by the nitrogen hydrogen bond donor of the Trp side chain. However, the displacement of this site alone is not sufficient to describe the strong affinity Trp exhibits for this pocket. This can be explained through the analysis of cavities (solvent depleted regions) available through the WaterMap interface. The later reveals a large cavity in the back of the binding pocket (outlined in red mesh in Figure 8), which is filled by the phenyl ring of the ligand Trp23 side chain.

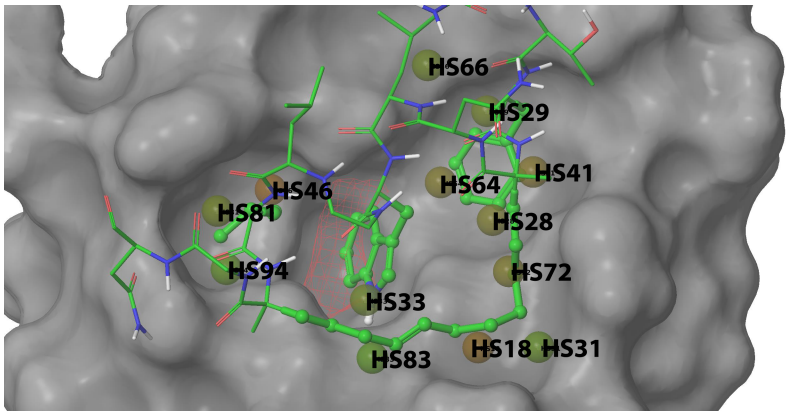


Figure 8. The WaterMap of the MDM2 binding site obtained from the crystal structure of the *trans*-sah8 complex (PDB ID: 3V3B). The three hydrophobic pockets and staple binding ridge of the MDM2 protein are shown as a gray surface, and the ligand is indicated in green. The key ligand side chains and the staple are shown as thick bonds. The hydration sites are shown as colored spheres. The solvent-depleted cavity in the Trp23 binding pocket is shown as red mesh surface.

The filling of solvent-depleted cavities has been shown to provide a large contribution to the binding affinity[46-50], and can be estimated by Eq. 2 to be ~6 kcal/mol.[30] In summary, the three hydrophobic pockets which bind Phe19, Trp23 and Leu26 provide a large positive contribution to the binding free energy for all peptides studied. This is not unexpected, as the former three residues were found to be critical in maintaining the potency of the stapled peptides.

The final cluster of high energy waters comprising hydration sites 83, 18, 31 and 72 occupies a ridge tucked above the three principal hydrophobic binding pockets, formed by the backbone fold of residues 53-58 as well as the side chain of Phe55. These are found to be selectively displaced by the hydrocarbon linker (the 'staple') chain of the peptide ligands. Unlike the principal hydrophobic pockets, which are populated at all times during the peptide simulations, the ridge region is selectively occupied by the staple. In particular, the *trans*-sah8 and *trans*-sah4 peptides displace all four of the hydration sites, whereas the stereochemistry around the double bond of the *cis* ligands precludes *cis*-sah8 and *cis*-sah4 from effectively displacing sites 18 and 31. The different positioning of the staple as well as point mutations in *cis*-sah3 cause the peptide to adapt a different preferred conformation during the simulation with the staple displaced away from the protein surface, and hence unable to displace any of the hydration sites. The contributions to the binding affinity of hydration sites displaced by the staple are summarized in Table 4.

Table 4. The free energy released by displacing the hydration sites populating the staple binding pocket (kcal/mol). The values were obtained from the hydration site free energies via the displacement functional in WaterMap.

Peptide	HS #18(kcal/mol)	HS #31(kcal/mol)	HS #72 (kcal/mol)	HS #83 (kcal/mol)
trans-sah8	-3.5	-0.6	-2.7	-0.8
trans-sah4	-3.4	-0.6	-2.7	-0.8
cis-sah8	-1.8	-0.3	-2.7	-0.8
cis-sah4	-1.8	-0.3	-2.7	-0.8
cis-sah3	0.0	0.0	0.0	0.0

Potential mean forces (PMFs) from WHAM

Calculating the potential of mean force along the peptide–protein binding axis, by varying the distance between the two domains gives insight into the relative binding free energies for the wild type and stapled peptides (Figure 9). Each point along the PMF curve represents a 2ns MD simulation with the peptide-protein distance biased to the corresponding distance along the reaction coordinate. The error bars in Figure 9 at each point reflect the PMF fluctuation at each reaction coordinate position.

Along the reaction coordinate pathway, when the center of geometry distance is zero, it represents the bound state for the p53 peptides and target protein MDM2. The PMFs for all p53 peptide analogs with MDM2 become flat when the inter-domain distances are larger than 14 Å. Therefore, all of the systems should be in the unbound state when the inter-domain distance increased to the final 24 Å. The energy differences between these two states in the PMF plot can be used to assess the relative binding affinity between different ligands. In particular, the larger the difference from the unbound state to the bound state in PMF, the higher the binding affinity of the peptide bound with MDM2

Notably, the binding affinity trend seen in Figure 9 matches the experimental K_d trends for the four peptides that were experimentally tested. Based on the experimental data, the order of the peptides from lowest to highest binding affinity is: *cis*-sah3 (K_d = 1200 nM), WT (K_d = 410 nM), *cis*-sah8 (K_d = 55 nM) and *cis*-sah4 (K_d = 0.92 nM) (4). This is the order of peptides from lowest to highest relative binding affinity that was found from the PMF plot using WHAM analysis: *cis*-sah3 (-17 kcal/mol), WT (-20 kcal/mol), *cis*-sah8 (-24 kcal/mol) and *cis*-sah4 (-34 kcal/mol). The calculated relative binding affinities using PMF are larger than the binding affinities found experimentally

for these peptides. One rationalization is that the PMF analysis does not include the loss of entropy due to binding, as we only sample the peptide approaching the protein locally along one reaction coordinate.

The surface plasmon resonance (SPR) experiments have shown that the processes by which ligands become dissociated from and adsorbed to a target protein take several seconds in general.[51] These actual processes are quite long compared with contemporary MD simulation time scale. Instead of obtaining the actual protein-ligand binding reaction path with expensive computational efforts, our dissociation processes and the relative PMF calculations are based on setting the center of geometry separation as the specific reaction coordinate, which provides a fast and reasonably accurate binding affinity estimate. However, we must note that the profile of PMF depends on the choice of the reaction path. Certainly, there are numerous binding reaction paths during the dissociation process of p53 peptides and MDM2 in reality. With the single center of geometry binding reaction path used in this study, the computational error will be inevitable. To investigate this, a different reaction coordinate pathway was chosen and compared to the reaction path used in Figure 9.

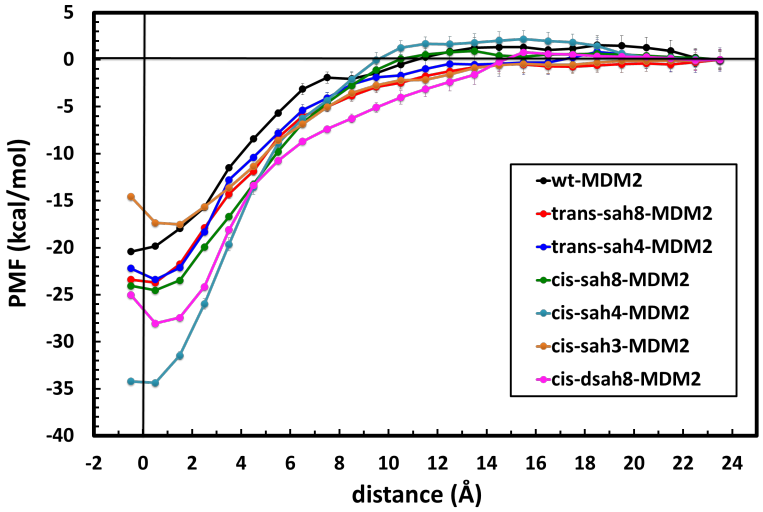


Figure 9. The potential mean forces of the p53/MDM2 binding interaction for the p53 peptides including: the wild type peptide (PDB ID: 1YCR) and the *trans*-sah8 peptide (PDB ID: 3V3B).

The center of geometry of MDM2 and p53 were found to be residue 85 and residue 23 respectively and used as the initial reaction coordinate pathway. As discussed, a different reaction coordinate pathway could result in a different binding affinity. To this end, in Figure10, the PMF along the pathway between residue 85 of MDM2 and residue 19 of p53 for *trans*-sah8 was calculated under the same conditions.

The new reaction coordinate pathway shows a higher binding affinity of the *trans*-sah8 peptide compared to the center of geometry pathway. Multiple dimension reaction coordinates with sufficient sampling will certainly be able to provide more insights for the understanding of protein-ligand association processes. However, this will substantially increase the computational cost.[52] In the study of Yoshifumi *et al*.[53], a filling potential (FP) method based on Taboo search was developed to search for and determine suitable reaction coordinates without setting it *a priori*. With these alternative choices of reaction coordinates, future study will provide more compelling results for the binding affinity of the single and double stapled p53 peptides when bound to MDM2.

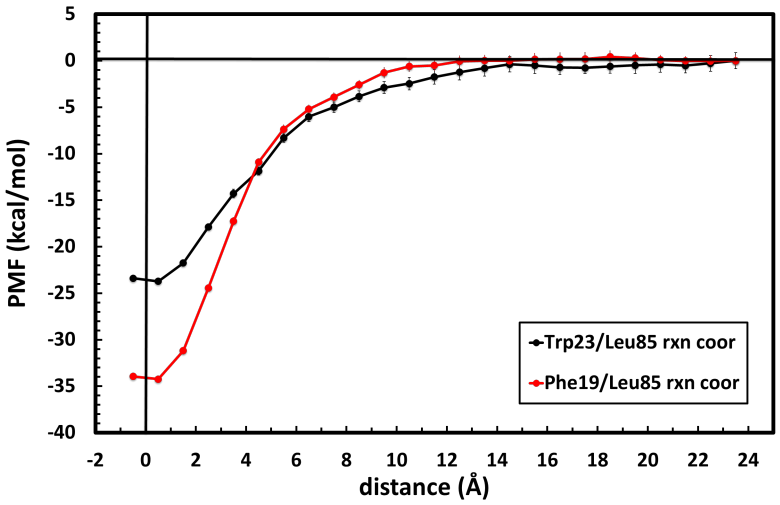


Figure 10. The potential mean forces of the p53/MDM2 binding interaction for the trans-sah8 peptide (PDB ID: 3V3B) along two different reaction coordinates. Trp23/Leu85 is the pathway along the peptide/protein center of geometries (used in Figure 9) while Phe19/Leu85 is the new pathway chosen for comparison.

Summary

In summary, we have investigated α-helical conformation and structural stability of single and double stapled all-hydrocarbon cross-linked p53 peptides free in solution and in complex with the MDM2 receptor. The simulations and data analysis presented in this chapter lead to qualitative understanding of how the variations in the peptide sequence, the stereochemistry of the cross-linker and the length of the tether, affect the relative stability of the secondary structure of stapled peptides. For all peptides studied, hydrophobic pockets which bind Phe19, Trp23 and Leu26 provide a large positive contribution to the binding free energy. We have determined the relative binding free energies and find, in agreement with experimental binding data, the order of the peptides from lowest to highest binding affinity is *cis*-sah3, WT, *cis*-sah8, and *cis*-sah4.

References

1. Joerger, A. C.; Fersht, A. R. *Cold Spring Harb. Perspect. Biol.* **2010**, 2, 1-20.

2. Verdine, G. L.; Walensky, L. D. *Clin. Cancer Res.* **2007**, 13, 7264-7270.

3. Onuchic, J. N.; Luthey-Schulten, Z.; Wolynes, P. G. *Annu. Rev. Phys. Chem.* **1997**, 48, 545-600.

4. Bernal, F.; Tyler, A. F.; Korsmeyer, S. J.; Walensky, L. D.; Verdine, G. L. *J. Am. Chem. Soc.* **2007**, 129, 2456-2457.

5. Schafmeister, C. E.; Po, J.; Verdine, G. L. *J. Am. Chem. Soc.* **2000**, 122, 5891-5892.

6. Kritzer, J. A.; Lear, J. D.; Hodsdon, M. E.; Schepartz, A. *J. Am. Chem. Soc.* **2004**, 126, 9468-9469.

7. Kussie, P. H.; Gorina, S.; Marechal, V.; Elenbaas, B.; Moreau, J.; Levine, A. J.; Pavletich, N. P. *Science,* **1996***,* 274, 948-953.

8. (2009) *Desmond (version 2.2)*. *Desmond (version 2.2)*. New York, NY: Schrödinger, LLC.

9. (2009) *Maestro (version 8)*. *Maestro (version 8)*. New York, NY: Schrödinger, LLC.

10. Baek, S.; Kutchukian, P. S.; Verdine, G. L.; Huber, R.; Holak, T. A.; Lee, K. W., Popowicz, G. M. *J. Am. Chem. Soc.* **2012**, 134, 103-106.

11. (2011) *Schrödinger Suite 2011: Epik* version 2.2. *Schrödinger Suite 2011: Epik version 2.2*. New York, NY: Schrödinger, LLC.

12. (2011) *Impact* version 5.7. *Impact version 5.7*. New York, NY: Schrödinger, LLC.

13. (2011) *Prime* version 2.3. *Prime version 2.3*. New York, NY: Schrödinger, LLC.

14. Madhavi, S. G.; Adzhigirey, M., Day, T.; Annabhimoju, R.; Sherman, W. *J. Comput. Aided Mol. Des.* **2013**, 27, 221-234.

15. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983***,* 79, 926-935.

16. Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996***,* 118, 11225-11236.

17. Kaminski, G. A.; Friesner, R.A.; Tirado-Rives, J.; Jorgensen, W. L. *J. Phys. Chem. B* **2001**, 105, 6474-6487.

18. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, 103, 8577-8593.

19. Martyna, G. J.; Tobias, D. J.; Klein, M. L. *J. Chem. Phys.* **1994**, 101, 4177-4189.

20. Hoover, W. G. *Phys. Rev. A* **1985**, 31, 1695-1697.

21. Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, 314, 141-151.

22. Humphreys, D. D.; Friesner, R. A.; Berne, B. J. *J. Phys. Chem.* **1994**, 98, 6885-6892.

23. (2011) *Schrödinger Suite 2011: WaterMap* version 1.3. *Schrödinger Suite 2011: WaterMap version 1.3*. New York, NY: Schrödinger, LLC.

24. (2011) *Protein Preparation Wizard*. New York, NY: Schrödinger, LLC.

25. Li, H.; Robertson, A. D.; Jensen, J. H. *Proteins: Structure, Function, and Bioinformatics* **2005**, 61, 704-721.

26. Bas, D. C.; Rogers, D. M.; Jensen, J. H. *Proteins: Structure, Function, and Bioinformatics* **2008**, 73, 765-783.

27. Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. *J. Chem. Theory Comput.* **2011**, 7, 525-537.

28. Søndergaard, C. R.; Olsson, M. H. M.; Rostkowski, M.; Jensen, J.H. *J. Chem. Theory Comput.* **2011**, 7, 2284-2295.

29. Abel, R.; Young, T.; Farid, R.; Berne, B. J.; Friesner, R. A. *J. Am. Chem. Soc.* **2008**, 130, 2817-2831.

30. Wang, L.; Berne, B. J.; Friesner, R. A. *Proc. Natl. Acad. Sci. USA* **2011**, 108, 1326-1330.

31. Zwanzig, R. W. *J. Chem. Phys.* **1954**, 22, 1420-1426.

32. Kirkwood, J. G. *J. Chem. Phys.* **1935**, 3, 300-313.

33. Roux, B. *Comp. Phys. Comm.* **1995**, 91, 275-282.

34. Souaille, M.; Roux, B. *Comp. Phys. Comm.* **2001**, 135, 40-57.

35. Gallicchio, E.; Andrec, M.; Felts, A. K.; Levy, R. M. *J. Phys. Chem. B* **2005**, 109, 6722-6731.

36. Masunov, A.; Lazaridis, T. *J. Am. Chem. Soc.* **2003**, 125, 1722-1730.

37. Sun, Y.; Dominy, B. N.; Latour, R. A. *J. Comp. Chem.* **2007**, 28, 1883-1892.

38. Grossfield, A. *http://membrane.urmc.rochester.edu/content/wham*.

39. Guo, Z.; Mohanty, U.; Noehre, J.; Sawyer, T. K.; Sherman, W.; Krilov, G. *Chem. Biol. & Drug Design* **2010**, 75, 348-359.

40. Moretto, A.; Crisma, M.; Kaptein, B.; Broxterman, Q. B.; Toniolo, C. *Biopolymers* **2006**, 84, 553-565.

41. De Poli, M.; Moretto, A.; Crisma, M.; Peggion, C.; Formaggio, F.; Kaptein, B.; Broxterman, Q. B.; Toniolo, C. *Chemistry-A European Journal* **2009**, 15, 8015-8025.

42. Miyazawa, S.; Jernigan, R. L. *Macromolecules* **1985**, 18, 534-552.

43. Hamacher, K.; Hubsch, A.; McCammon, J. A. *J. Chem. Phys.* **2006**, 124, 164907/1-164907/8.

44. Kim, Y. W.; Kutchukian, P. S.; Verdine, G. L. *Org. Lett.* **2010**, 12, 3046-3049.

45. Walensky, L. D.; Kung, A. L.; Escher, I.; Malia, T. J.; Barbuto, S.; Wright, R. D.; et al. *Science* **2004**, 305, 1466-1470.

46. Young, T.; Hua, L.; Huang, X.; Abel, R.; Friesner, R. A.; Berne, B. J. *Bioinformatics* **2010***,* 78, 1856-1869.

47. Malham, R.; Johnstone, S.; Bingham, R. J.; Barratt, E.; Phillips, S. E. V.; Laughton, C. A.; et al. *J. Am. Chem. Soc.* **2005**, 127, 17061-17067.

48. Airenne, T. T.; Kidron, H.; Nymalm, Y.; Nylund, M.; West, G.; Mattjus, P.; et al. *J. Mol. Biol.* **2006**, 355, 224-236.

49. Lario, P. I.; Pfuetzner, R. A.; Frey, E. A.; Creagh, L.; Haynes, C.; Maurelli, A. T.; et al. *EMBO J.* **2005**, 24, 1111-1121.

50. Guo, Z.; Li, B.; Dzubiella, J.; Cheng, L. T.; McCammon, J. A.; Che, J. *J. Chem. Theory Comp.* **2013**, 9, 1778-1787.

51. Giannetti, A. M.; Koch, B. D.; Browner, M. F. *J. Med. Chem.* **2008**, 51, 574-580.

52. Rajamani, R.; Naidoo, K. J.; Gao, J. *J. Comp. Chem.* **2003**, 24, 1775-1781.

53. Fukunishi, Y.; Mikami, Y.; Nakamura, H. *J. Phys. Chem. B* **2003**, 107, 13201-13210.

Chapter 2

Investigating Surface Charge Density Effects on Polyelectrolyte Folding Using Brownian Dynamics*

*Reproduced from Streu, K.; Mohanty, U. "Investigating surface charge density effects on polyelectrolyte folding using Brownian dynamics." *Manuscript in preparation.*

1. Introduction

DNA and RNA are highly charged molecules.[1-3] Folding of RNA into stable tertiary structures is governed by its interactions with magnesium and other divalent and monovalent metal cations. There are varieties of structural and biochemical investigations on the association of divalent and monovalent metal ions with nucleic acids.[1-31] Mg^{2+} can displace other larger metal cations in the ion atmosphere around nucleic acids.[18-20,29,31,32] The basis of how divalent and monovalent metal ions effect DNA and RNA stability, and the characteristics of ion-nucleic interactions are less than well understood.[1-3,7,29,31]

The ionic environment around nucleic acid can be partitioned into diffuse and chelated ions.[31-35] The diffuse ions interact with each other and with the nucleic via long-range electrostatic interactions.[31-35] Both divalent and monovalent ions contribute to the diffuse ion atmosphere. There are several contributions to the free energy when chelated ions interact with nucleic acid. For the chelated ions to make contact with the surface of nucleic acid, it is essential for them to partially dehydrate.[30b,31,34,35] Furthermore, chelated ions can displace ions in the diffuse layer. These energetic influences are counterbalanced by electrostatic interactions between phosphates and the chelated ions.[23,31,34,35]

Models for highly charged polyelectrolyte chains based on Debye-Hückel (DH) electrostatics[36-39] and non-linear Poisson Boltzmann (NLPB),[40-50] treat the ionic environment as a continuum. In these approaches, molecular details such as ion-ion correlations and the discrete effects of the ions are generally ignored. If ion-ion correlations are weak, as they are for monovalent ions, non-linear Poisson Boltzmann provides satisfactory description of the ionic atmosphere of nucleic acids.

A fundamental length scale[32,33] of electrostatic interaction is the Bjerrum length l_B

$$l_B = \frac{e^2}{4\pi\varepsilon_o \varepsilon k_B T}. \quad (1)$$

It is the distance at which the Coulomb energy between a pair of charges in aqueous solution balances thermal energy of the surrounding at temperature T. Another basic scale is the linear charge density or Coulomb strength parameter defined as $\Gamma = l_B / b$ where Γ is the ratio of the Bjerrum length to the average charge spacing b of the phosphate groups. If the Coulomb strength parameter of a polyelectrolyte chain is larger than unity, as it is for RNA and DNA, then monovalent counterions would condense on the nucleic acid to reduce the charge on the phosphate groups. The electrostatic repulsion between the phosphates would be reduced due to the condensed counterions. This reduction of repulsive forces leads to counterion-induced collapse of RNA.[12,13,20a,46-48] Small angle X-ray and small angle neutron scattering experiments on various RNAs indicate that radius of gyration further decreases after counterion condensation.[46-48]

Consider a globular polyelectrolyte chain viewed as a sphere of radius R. The chain has N unit charges e on its surface. The chain is immersed in an electrolyte solution. The salt solution is characterized by an inverse Debye screening length κ. The surface free energy of the sphere can be obtained by using linearized solution of Poisson-Boltzmann equation. The overall free energy has an entropic term that arises since the condensed counterions are brought to close proximity of the chain from the bulk solution. To obtain the equilibrium state, the overall free energy needs to be minimized with respect to θ, the later quantity defined as the number of counterions per unit surface charge of the macroion. In the small sphere approximation, $\kappa R \ll 1$ one obtains[50]

$$N\theta = -(\frac{R}{zl_B})\ln c\,, \qquad (2)$$

where c is the bulk concentration of counterions. When the radius of the sphere is comparable to the Debye screening length, then $\kappa R \approx O(1)$; in this case, one finds that[50]

$$N\theta = -(\frac{2R}{zl_B})(1+\kappa l_B)\ln \kappa l_B\,. \qquad (3)$$

Now consider the case when the ion atmosphere forms a thin layer around the sphere. The radius of the sphere is larger than the Debye screening length. Analysis reveals that counterion condensation sets in when the surface charge density of the sphere exceeds the threshold value characteristic of a planar charged wall in an aqueous solution.[50]

Compared with monovalent ions, multivalent counterions are more effective in neutralizing the charges on the polyelectrolyte chain. Charge fluctuations of the counterions lead to an effective attractive potential, which in the mean-field approximation is $\Delta\Phi(r) \approx (1-z^2)cl_B^3 e^{-\kappa r}/\kappa r$.[39] The strength of the attractive potential increases with valence z of the metal cations. The effective potential can bridge phosphate groups and lead to the reduction of the persistence length.[51] The interplay between counterion condensation and charge fluctuations allows a polyelectrolyte chain to form structures that are compact.[19]

As both valence z and excluded volume interactions between the condensed ions determine the distribution of ions around polyelectrolyte chain, they will also determine the characteristics of ion-induced states of the chain. There are only a handful of experimental studies on cation size dependence of nucleic acid collapse, and even less theoretical studies.[46,47,48,19] Free energy of nucleic acid folding decreases as cation charge

density increases.[46] Larger cations produce more extended RNA structures than smaller metal ions, even when both ions carry the same charge.[48,19] Effects of Group-I monovalent cations on RNA folding and stability indicate that the smallest ions are most stabilizing. There is a stronger dependence of stability on ion size in RNAs with higher charge density and that cellular RNAs have evolved in the presence of K^+ and some RNAs take specific advantage of this ion to achieve specific structures.[49]

In this chapter we investigate the characteristics of ion-induced collapse of a strongly charged polyelectrolyte chain of varying Coulomb strength parameter by carrying out coarse-grained Brownian dynamics with counterions that correspond to Group-II divalent metal cations. The model of the polyelectrolyte chain, the counterions and the coions, and Brownian dynamics are introduced in Section 2. In Section 3, we present our results for the collapse dynamics of the polyelectrolyte as change in its radius of gyration with time for Group-II metal cations at low salt conditions where the association of the counterions with the chain is governed by polyelectrolyte effects. We determine the free energy and the entropy changes of the collapsed state as a function of the Coulomb energy parameter of the chain, as well as the charge density ζ of the counterion, the later defined as ze/V, where V is the volume of the metal cation and ze is the charge. For the collapsed chain conformation, we study the divalent ion and the monomer radial distribution function, the adsorbed ions around the chain in the radial direction, and the integrated radial charge as a function of the hydrodynamic radius of the Group-II metal counterions.

2. Model and Setup

A bead spring model for a polyelectrolyte chain consists of N=120 spherical monomers or beads. Two consecutive monomers are at an equilibrium distance a. Each monomer with a van der Waals radius of $a/2$ carries a charge of -1e. N metal counterions (+1e) represented as spheres are added to neutralize the system.

The total potential energy U is the sum of all bonded interactions and non-bonded interactions, that includes electrostatic and excluded volume effects[52-54]

$$U = U_{bond} + U_{LJ} + U_c . \tag{4}$$

Bond stretching potential U_{bond} in the chain describes the interactions between two consecutive beads i and $i+1$ which are connected by harmonic spring with a potential[52-54]

$$U_{bond} = \frac{k_a}{2}(r_{i,i+1} - a)^2 . \tag{5}$$

k_a is the spring constant and $r_{i,i+1}$ is the distance between two consecutive beads. The effective short range repulsions between beads i and j is modeled by the repulsive part of the Lennard-Jones potential[52-54]

$$U_{LJ} = \varepsilon_{LJ}[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - 2\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6} + 1], \tag{6}$$

where r_{ij} is the distance between charged species i and j, ε_{LJ} is the interaction strength, the radii $\sigma_{ij} = \sigma_i + \sigma_j$ depend on the type of interaction. For counterion-monomer interaction, σ_i is given by the hydrodynamic radius of divalent ions in Table 1, while

$\sigma_j = a/2$. Finally, electrostatic interactions between pair of ions or monomers are described by the Coulomb potential[52-54]

$$U_c = \frac{z_i z_j e^2}{4\pi\varepsilon_o \varepsilon r_{ij}}, \tag{7}$$

where z_i is the valency of the ith ion or bead, e is the elementary charge, ε_o is the permittivity of vacuum, ε is the dielectric constant of the solvent, and r_{ij} is the distance between charged species i and j.

The equations of motion for the ith particle is governed by Brownian dynamics[54,46,19]

$$\vec{r}_i(t+\Delta t) = \vec{r}_i(t) - \frac{D\Delta t}{k_B T}\vec{\nabla}_{r_i} U + \sqrt{6D\Delta t}\,\vec{f}_i(t). \tag{8}$$

Here, $\vec{r}_i$ is the coordinate of the ith particle, ζ is the friction coefficient, $\vec{\nabla}_{r_i} U$ is the gradient of the total potential energy U, k_B is the Boltzmann constant, T is the temperature, D is the diffusion coefficient, $\vec{f}_i(t)$ is Gaussian random noise and satisfies the fluctuation-dissipation theorem, $\left\langle \vec{f}(t)\cdot\vec{f}(t')\right\rangle = \delta_{ij}\delta(t-t')$.

The polyelectrolyte chain consists of $N = 120$ monomers. The equilibrium distance between successive monomers is $a = 0.6$ nm. The van der Waals radius of the monomer is taken to be 0.3 nm. The spring constant k_a is set to 120 kJ mol^{-1} nm^{-2}. To generate an initial configuration and to achieve charge neutrality of the system, the polyelectrolyte along with N metal monovalent metal cations (+1e) were randomly distributed within the simulation box. The counterions are spheres of radius a. The simulation box is a cube with side of length 30 nm. For a typical simulation, the procedure begins with a

Brownian dynamics simulation to equilibrate the polyelectrolyte chain along with the monovalent counterions. Then, m =10 divalent counterions (+2e) along with 2m coions (-1e) are added to the system to investigate ion-mediated collapse dynamics of the polyelectrolyte chain.

During the initial equilibrium stage with only monovalent counterions, the chain never folds to a compact state. Equilibration led to charge neutralization due to condensed monovalent metal cations onto the chain. The Bjerrum length l_B in water sets the energy scale. The precise value of l_B will not change the relationship between how the energy of the collapsed state varies with ion size. We initiate rapid counterion-induced collapse of the chain, for each divalent metal cation, by varying the Coulomb interactions relative to thermal fluctuations, i.e., the Bjerrum length, such that the Coulomb strength parameter $\Gamma = l_B / a$ is between 0.1 and 10.

The hydrodynamic radii R_{hyd} of Group-II metal cations (+2e) are listed in Table 1.[19] We assume the co-ions (-1e) are of the same radius as the monovalent counterions. To minimize statistical errors, four independent trajectories were generated for each Group-II divalent metal cation. The total duration of each simulation is $3x10^8$ time steps. The simulation time is long enough to produce a collapse state of the chain that resembles a globular structure.

In Brownian dynamics, periodic boundary conditions were employed and the integration time step Δt was 0.36 ps. Electrostatics of the system is computed using Ewald summation procedure. The diffusion coefficients D_m of the monomers are assigned to be 1.0 x 10^{-3} nm^2/ps, while that of the ions are obtained from the Stokes-Einstein relation, $D_i = (a/2R_{hyd}) D_m$.

Table 1: Hydrodynamic radius of Group-II metal divalent cations from reference 19.

Divalent Cation, X	Radius (nm)
Mg^{2+}	.207
Ca^{2+}	.233
Sr^{2+}	.266
Ba^{2+}	.300

3. Results and Discussion

Using the radius of gyration R_g of the chain, chain folding is quantitatively monitored throughout the simulation. The radius of gyration was determined by the relation $R_g = \Sigma_{i<j} r_{ij}^2 / N$, where r_{ij} is the distance between monomers i and j and N is the number of monomers. The initial rise is due to electrostatic repulsion between the monomer beads. This rise is followed by decrease that lead to formation of compact globular structure at long times (Figure 1). Smaller ions tend to be more effective in driving the chain collapse than larger ions. The compactness of the globule as measured by radius of gyration increases slightly from Mg^{2+} to Ba^{2+} as shown in Figure 2.

The chain always folds into a compact globule, no matter the size of the divalent metal cation, so long as the dielectric constant of the solvent is less than 90. However, for dielectric larger than this value, the polyelectrolyte chain in the presence of Sr^{2+} and Ba^{2+} (lower surface charge density ζ) do not exhibit significant folding over the course of the simulation. In contrast, although the chain simulated with Mg^{2+} and Ca^{2+} (higher surface charge density ζ) does fold, the Mg^{2+} simulation (highest surface charge density ζ) folds faster.

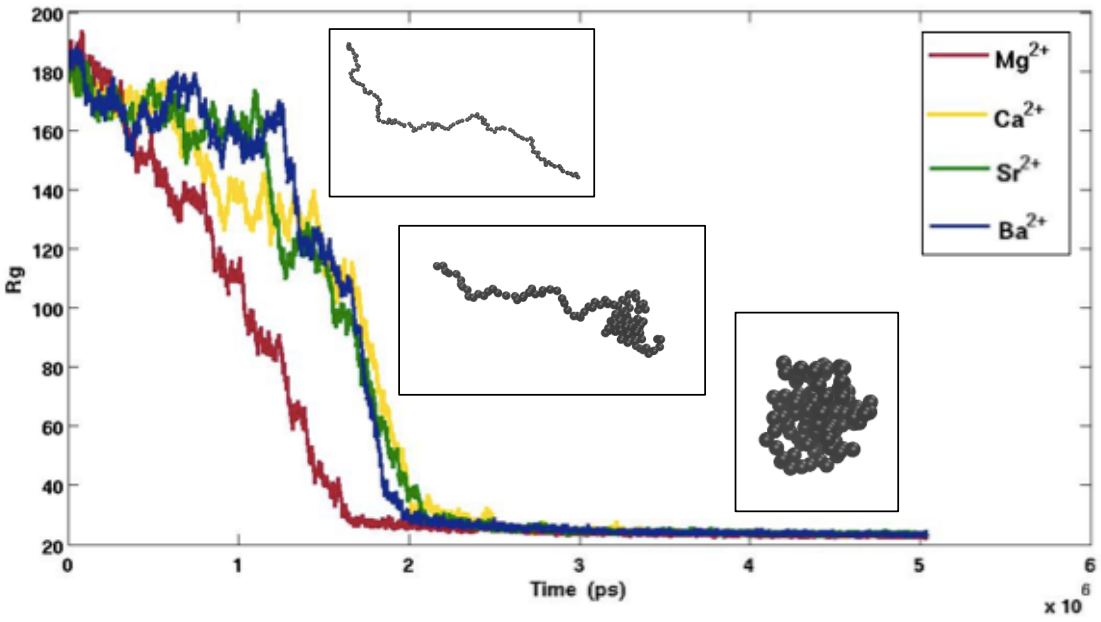


Figure 1: The radius of gyration of the polyelectrolyte chain throughout the production simulation, for the four different divalent cations at Coulomb strength parameter Γ = 1.19 (dielectric constant = 80).

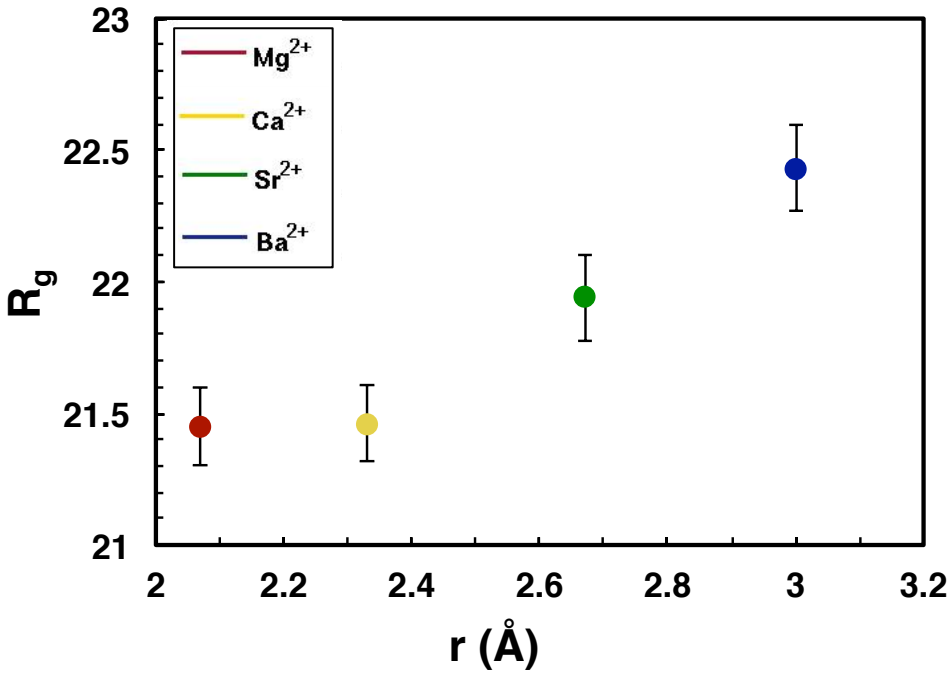


Figure 2: Size-dependence of polyelectrolyte chain chain on Group-II divalent metal cations in the collapsed state.

The internal energy of the simulated system consists of three parts: bond stretching energy, Lennard-Jones excluded volume interaction, and Coulomb interaction. Only the Coulomb energy shows noticeable change in going from the extended chain to folded chain conformation. This is due to the bead-spring model that allows the polyelectrolyte chain to alter their conformations to assist close pairing without substantial distortion of bonds. Therefore, the enthalpy of polyelectrolyte chain collapse can be approximated by the Coulomb energy difference ΔE between that extended and the collapsed state. Figure 3 shows $\Delta E / T$ at different Γ for simulations varying the Group-II divalent metal cation size.

The attractive energy ΔE between the chain and divalent metal cations is small and negative for values of Coulomb strength parameter $\Gamma < 0.2$. The Coulomb energy change is most negative around $\Gamma = 1.5$, after which it gets less negative, and at $\Gamma > 2.5$, starts to turn positive. This positive Coulomb energy change is noticeable for the larger divalent metal cations. At higher values of Coulomb strength parameter, one expects most of the counterions to be absorbed and this should lead to leveling off of the Coulomb energy change.

Observe that the negative Coulomb energy change ΔE reaches a minimum around $\Gamma = 1.5$ (Figure 3). At $\Gamma < 1.5$ a small fraction of the total counterions adsorb on the chain. As a result, most of polyelectrolyte charges are unneutralized. Net Coulomb energy gain occurs when the divalent cations condense to further neutralize the charges on the chain. With further increase in Γ, the favorable Coulomb energy change is diminished; it is energetically less than favorable to displace interactions between counterions and monomers with monomer-monomer interactions.

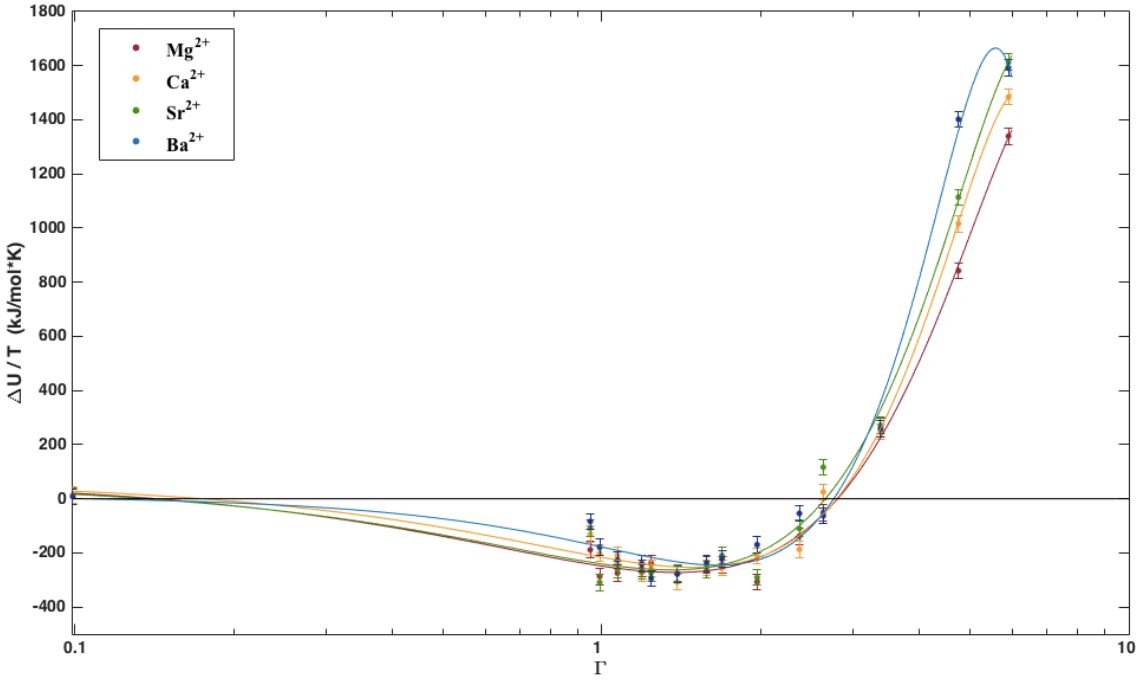


Figure 3: Net Coulombic energy gain due to counterion-induced collapse of polyelectrolyte chain by Group-II divalent metal cations is plotted as a function of Coulomb strength parameter. Chain length is N =120. The results were averaged over four independent simulations to reduce statistical errors.

The free-energy change is obtained by the relation[53]

$$\frac{\Delta F}{T} = -\int_{T_{ref}}^{T} \frac{\Delta E}{T'^2} dT' \tag{9}$$

where T is the temperature of interest and T_{ref} is the reference temperature, which in this study is taken to be when linear charge density of the chain is the smallest ($\Gamma = 0.1$). The entropy change is then calculated by subtracting the free energy from the enthalpy of chain collapse, $T\Delta S = \Delta E - T\Delta S$. To carry out the integration in Eq. 6, $\Delta E / T$ was fitted by a smooth fifth order polynomial function of Coulomb strength parameter Γ, and then the integration performed in Γ space.[52]

Figures 4 and 5 compare the free energy ΔF and the entropy ΔS changes as a function of the Coulomb strength parameter for Group-II divalent metal cations. Free energy of chain collapse is a concave-up function of Γ (Figure 4). It closely follows the decreasing negative values of $\Delta E / T$ for $\Gamma < 1.5$, and at $\Gamma = 1.5$ the entropy contribution continually lowers free energy of collapse. This effect is ion size dependent, as the higher surface charge density ions have a more pronounced lowering in the free energy. For $\Gamma \geq 3$, the magnitude of the negative free energy change of chain collapse starts to reduce due to the fact that Coulomb enthalpy now becomes increasingly positive. Again this effect is ion size dependent, as the lower surface charge density ions have a more positive Coulomb enthalpy curve.

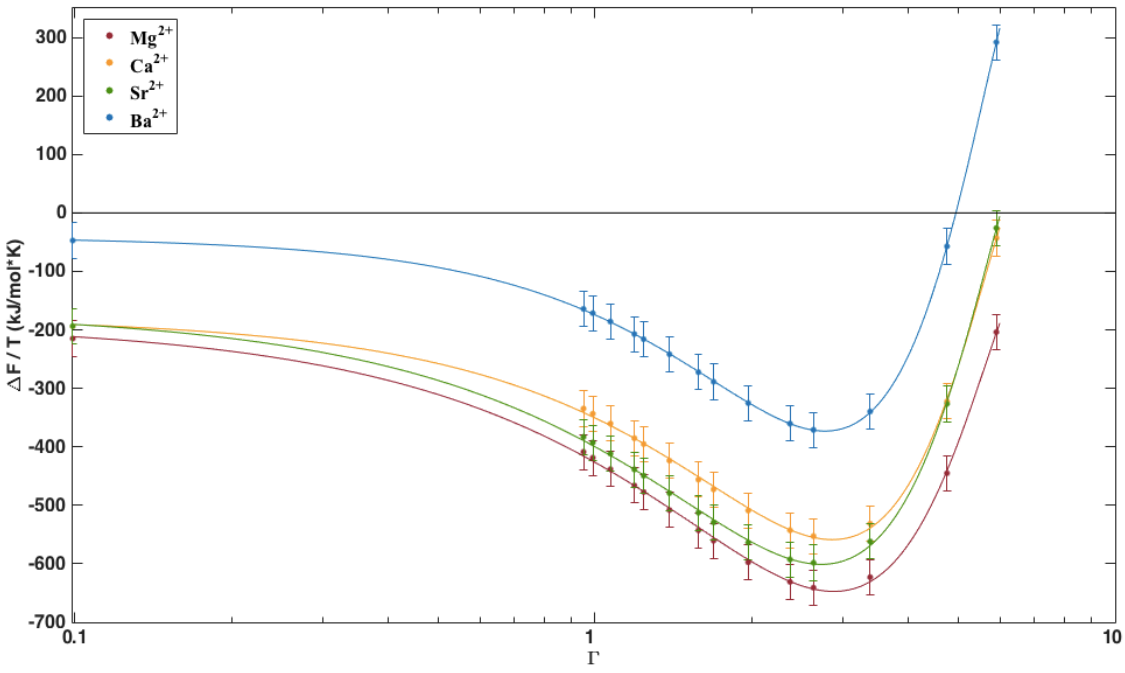


Figure 4: Net free energy gain due to counterion-induced collapse of polyelectrolyte chain by Group-II divalent metal cations is plotted as a function of Coulomb strength parameter. Chain length is N =120. The results were averaged over four independent simulations to reduce statistical errors.

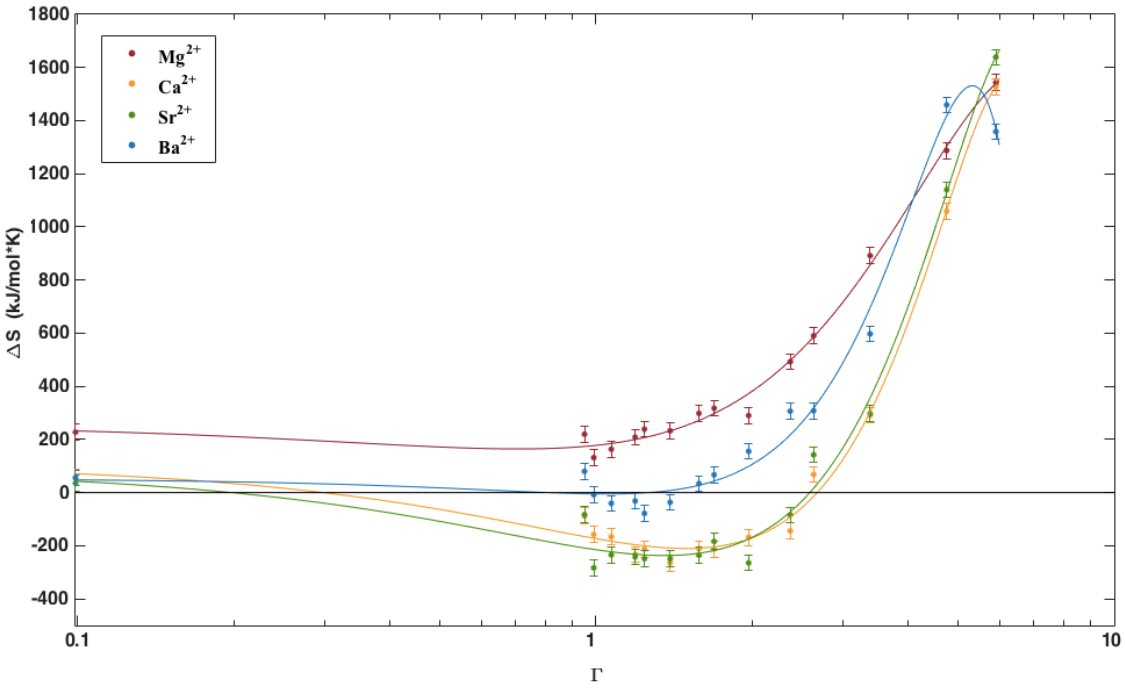


Figure 5: Net entropy change due to counterion-induced collapse of polyelectrolyte chain by Group-II divalent metal cations is plotted as a function of Coulomb strength parameter. Chain length is N = 120. The results were averaged over four independent simulations to reduce statistical errors.

In the collapsed chain conformation, the configurational, translational, and rotational entropies are reduced. The chain collapse is also accompanied by a release of adsorbed counterions that increases the entropy of the system. The entropy of chain collapse (Figure 5) is negligible for $\Gamma < 1.0$, but quickly increases with increasing Coulomb strength parameter. For $\Gamma > 1.0$, and for metal cations with high surface charge density, such as Mg^{2+} and Ca^{2+}, the counterion release entropy is the dominant contribution to the entropy change due to polyelectrolyte collapse.

We can define a threshold Coulomb strength parameter Γ^* where the magnitudes of the ΔE and $T\Delta S$ balance each other. Γ^* depends in a non-linear fashion on the surface charge density of the Group-II metal cations. Below $\Gamma^* < 2.5$ chain collapse is driven by the negative Coulomb energy change derived from electrostatic attraction between the charged monomers and the divalent cations, while the counterion release entropy plays a secondary role. Above $\Gamma^* > 2.5$, the counterion release entropy contributes significantly and it contributes to lowering of the free energy until $\Gamma^* > 2.5$, especially at high values of Coulomb strength parameter Γ where the Coulomb energy change is positive.

The characteristics of the free energy and the entropy change versus the counterion surface charge density (Figure 6) of the counterions for a fixed value of the Coulomb strength parameter allow for comparison with experimental data on RNAs. Woodson *et al*. found that the change in free energy of folding of various RNAs decrease with increasing surface charge density.[19] While polyamines were used experimentally to investigate surface charge density dependence effects on folding, the trend in free energy is similar to the free energy trend found in this study.

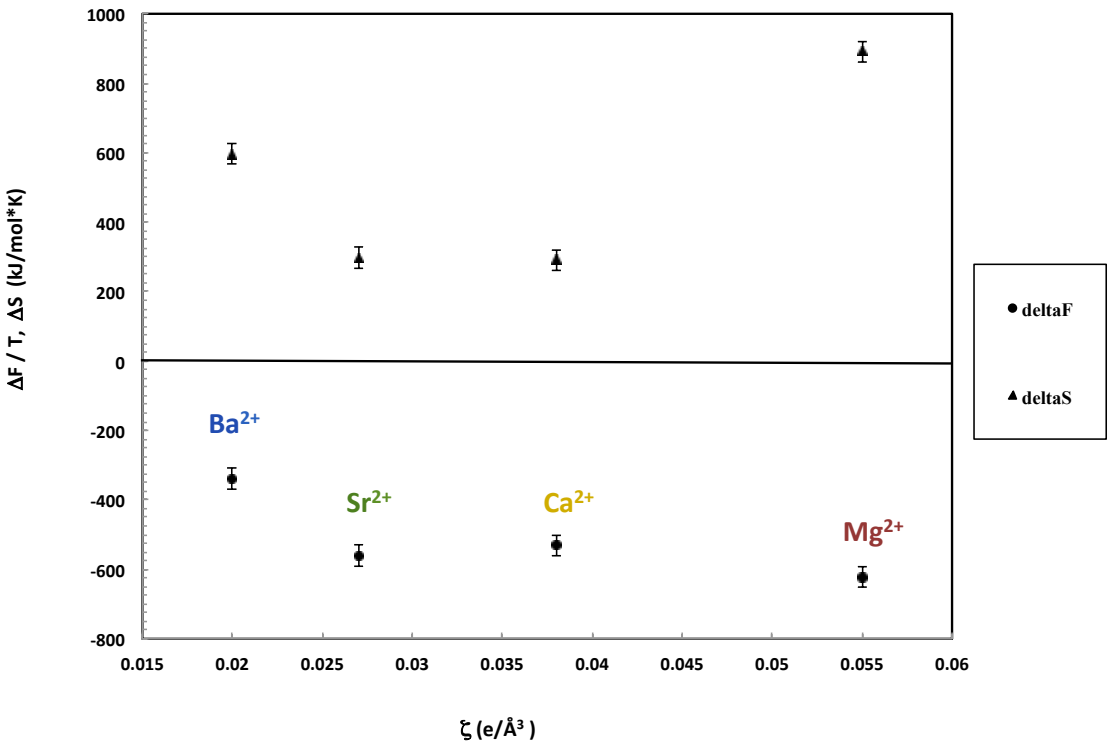


Figure 6: Divalent cation surface charge density dependence of free energy and entropy change due to counterion-induced collapse of polyelectrolyte chain Coulomb strength parameter $\Gamma = 3.38$ (dielectric constant = 28).

The surface charge density dependence of the entropy change on folding has not yet been investigated experimentally. However, simulations predict that for Coulomb strength parameter less than 2, the change in entropy does not exhibit much variation with surface charge density. The change in entropy of folding increases as the Coulomb strength parameter increases. This trend is in agreement with investigation of entropy of folding of a polyelectrolyte chain by coarse-grained Langevin simulations.[52] In contrast, the change in internal energy of the collapsed state varies slightly in the same range of linear charge density of Group-II metal cations investigated (Figure 7).

The origins of the stability between the chain and counterion charge density is further revealed in the radial distribution function $g(r)$ between the divalent metal cations and the charged monomers. The ion-monomer radial distribution function $g(r)$ is calculated as the averaged density function of the divalent cations around a polyelectrolyte chain monomer dividing by the mean density of the divalent cations in the simulated system. The results for the four divalent metal cations are calculated as an average of four independent simulations with a statistical error of ± 0.001 and are shown in Figure 8. The distance of closest approach of the divalent metal cations to the chain monomers increase with increase of hydrodynamic radii. As the divalent cation size increases, the various peaks shift to the right. This reflects excluded volume effects. The first peak occurs at a distance that is approximately the sum of the radius of the monomer (3Å) and the radius of the Group-II divalent metal cations (Table 1). The sharp first and second peaks in $g(r)$ indicate noticeable correlation between the chain and divalent metal cations regardless of the size.

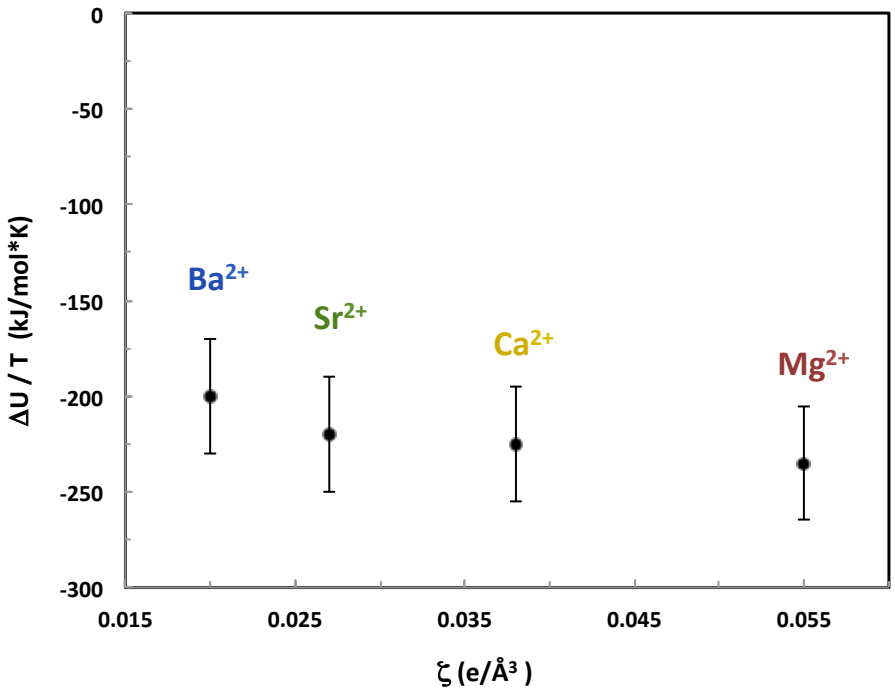


Figure 7: Divalent cation surface charge density dependence of internal energy change due to counterion-induced collapse of polyelectrolyte chain for Coulomb strength parameter $\Gamma = 3.38$ (dielectric constant = 28).

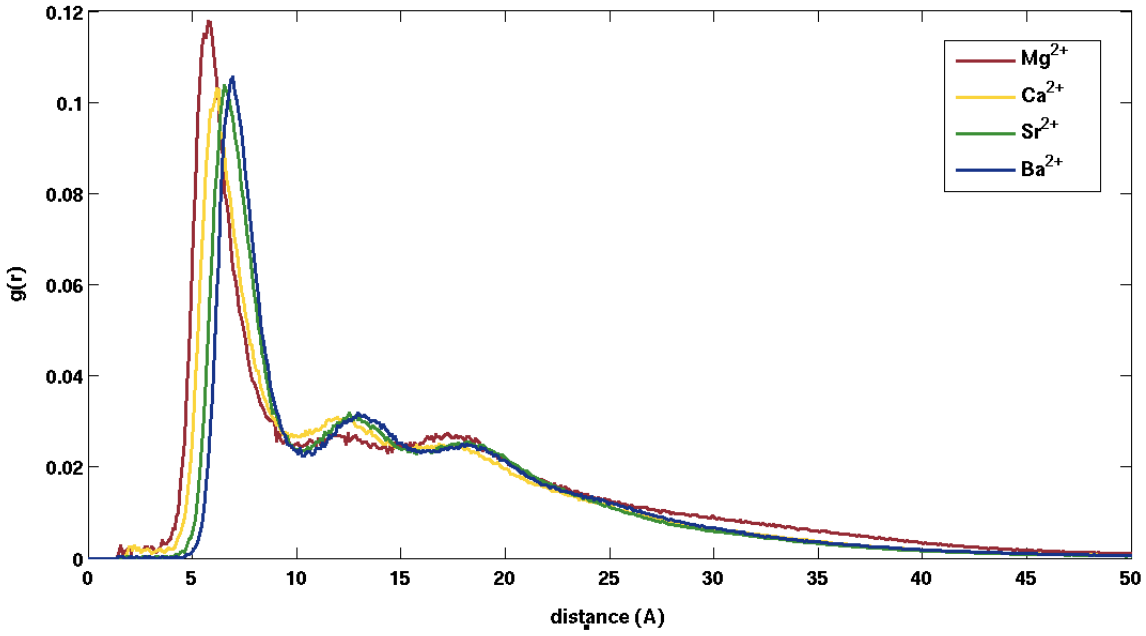


Figure 8: Radial distribution function g(r) between the polyelectrolyte chain monomers and the Group-II divalent metal cations at Coulomb strength parameter Γ = 4.75 (dielectric constant = 20). Each curve is averaged over four independent simulations to minimize the statistical errors.

We have determined the ion distribution around the chain in the radial direction. A sphere of radius r_t is drawn that is centered at each bead. The union of the spheres encases a tube of this radius around the polyelectrolyte chain.[53] For each type of ion, the average number of ions within the tube was calculated as function of tube radius (Figure 9) and averaged over four independent simulations with a statistical error of ± 0.5. For $r_t > 7$, the average number $N_t(r_t)$ of divalent counterions is constant. Nearly all divalent counterions condense on the chain in the low salt region. The other Group-II divalent ions have similar ion distributions around the polyelectrolyte chain in the radial direction.

Upon condensation of the divalent cations, the backbone charge reduction is independent of the size of the ions. However, there are subtle trends in the ion distribution due to surface charge density of the counterions. The Mg^{2+} ions are found at a closer radius to the chain, and are spread further out around the chain tube; the other divalent ions follow: Ca^{2+}, then Sr^{2+}, and then Ba^{2+}. This effect is partially due to ion size; Mg^{2+} ions can penetrate closer to the polyelectrolyte chain, while Ba^{2+} are limited to radii 6 angstroms and further.

The integrated charge distribution $Q_t(r_t)$ of the chain is obtained by summing the charges of each type in the tube region and averaging it over four independent simulations (statistical error is ± 1.0). The size dependence of divalent metal cations on integrated charge distribution is shown in Figure 10. The smaller Group-II metal cations neutralize the chain at smaller radius; this is more than just a cation excluded volume effect. For large radius, the divalent cations neutralize the chain equally, i.e. there is no size effect.

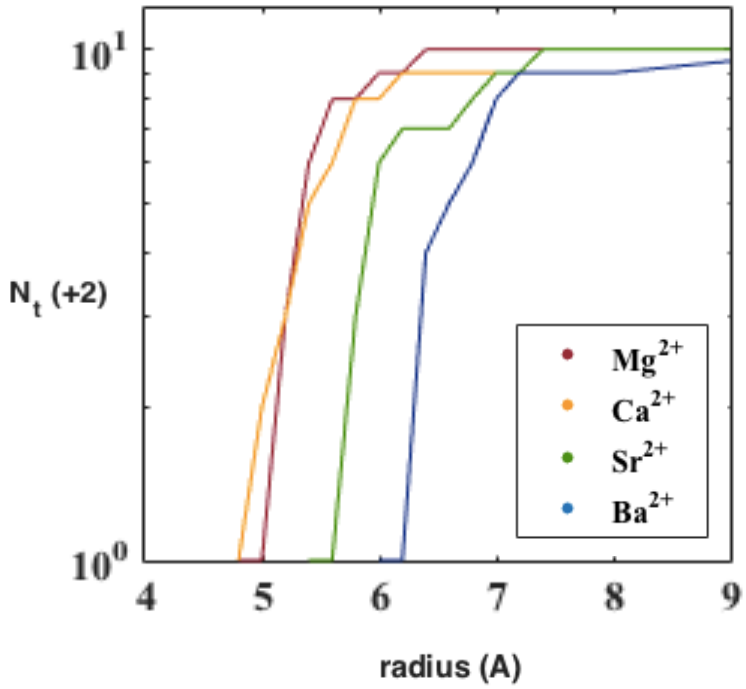


Figure 9: Ion distribution around the chain in the radial direction. Average number of divalent cations, $N_t(r_t)$ inside a tube-like region is plotted as function of tube radius at Coulomb strength parameter $\Gamma = 4.75$ (dielectric constant = 20). Each curve is averaged over four independent simulations to minimize the statistical errors.

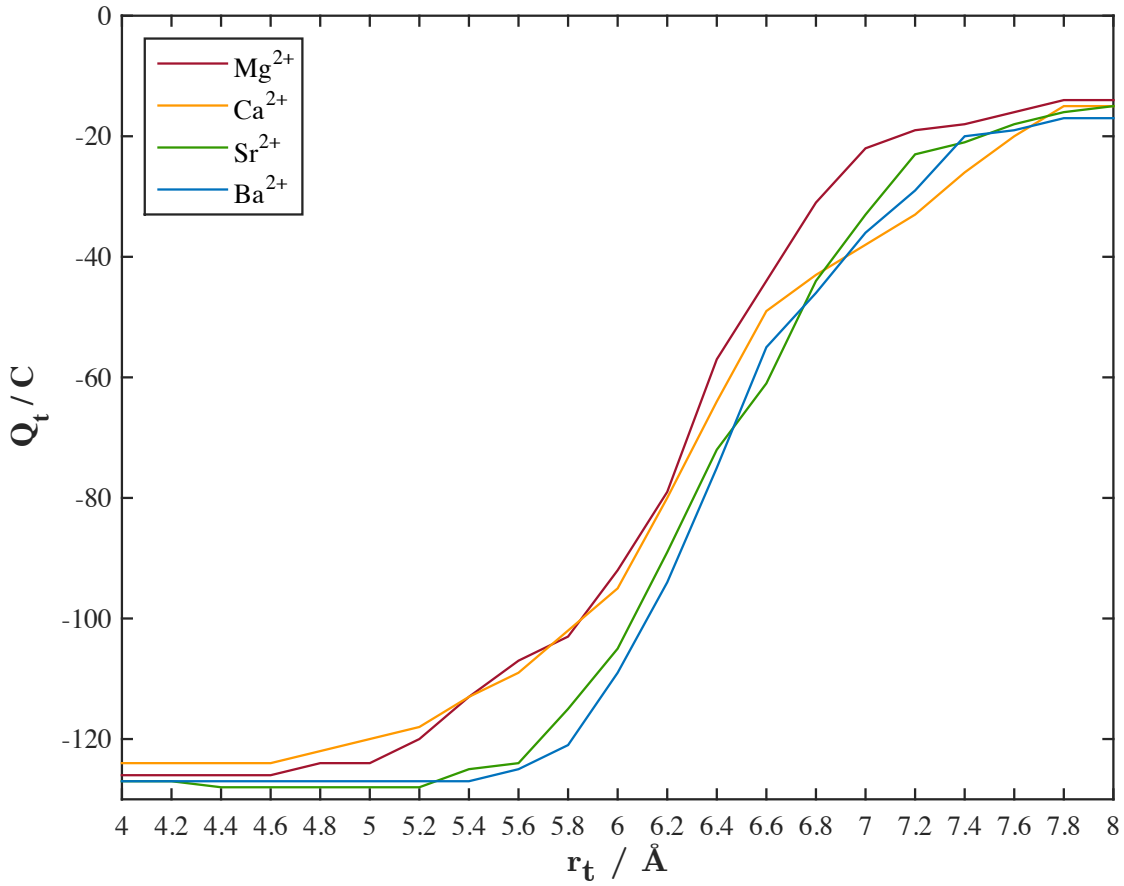


Figure 10: Integrated charge distribution around the chain in the radial direction. The integrated charge distribution $Q_t(r_t)$ in the coiled conformation versus radial distance from the chain backbone with different Group-II divalent metal cations at Coulomb strength parameter Γ = 4.75 (dielectric constant = 20). Each curve is averaged over four independent simulations to minimize the statistical errors.

As shown by Manning,[32] counterion condensation occurs when Γ is larger than unity. Hence, it would be fruitful to look at how the integrated charge distribution in the chain along the radial direction varies with Coulomb strength parameter (Figure 11). Observed that there is a less negative overall charge on the chain as the Coulomb strength parameter increases. This can be rationalized as follows. The Coulomb strength parameter is inversely proportional to the dielectric constant; therefore, as the dielectric constant decreases, more ions condense onto the polyelectrolyte chain to neutralize the overall negative charge.

In summary, coarse-grained Brownian dynamics simulations indicate that larger Group-II metal cations are less effective in counterion condensation than the smaller ions in the same column of the periodic table. The reduction of the backbone charge upon counterion condensation is independent of size of the divalent metal cations. The characteristics of counterion-induced collapse are governed by the charge density of the Group-II metal cations. The change in entropy as a result of counterion-induced collapse increases with increasing counterion charge density for Coulomb strength parameter larger than unity. Above a threshold value of Coulomb strength parameter, counterion release entropy drives the formation of counterion-induced compact states.

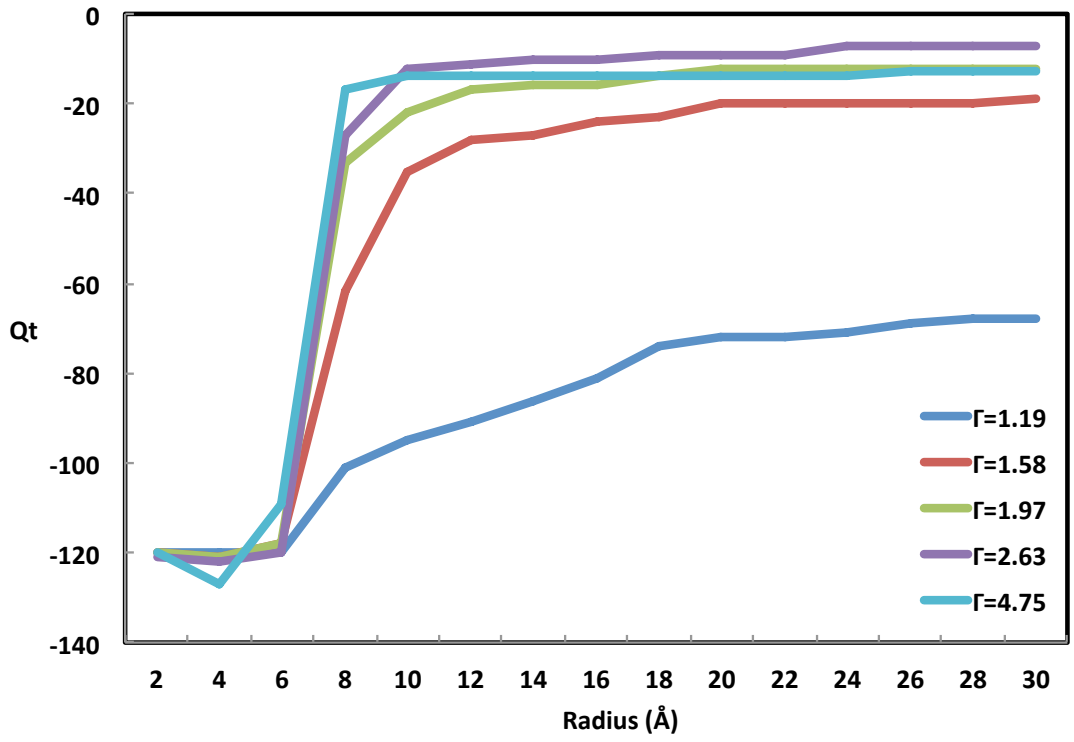


Figure 11: The integrated charge in the coiled conformation versus radial distance from the chain backbone with Mg^{2+} divalent cations at various Coulomb strength parameter: Γ = 1.19, 1.58, 1.97, 2.63, 4.75. Each curve is averaged over four independent simulations to minimize the statistical errors.

References

1. Woodson, S. A. *Curr. Opin. Chem. Biol*. **2005**, 9, 104-109.

2. Pyle, A. M. *J. Biol. Inorg. Chem*. **2002**, 7, 679-690.

3. Braunlin, W. H. *Adv. Biophys. Chem*. **1995**, 5, 89-139.

4. Cech, T. R.; Bass, B. L. *Annu. Rev. Biochem*. **1986**, 55, 599-630.

5. Laing, L. G.; Gluick, T. C.; Draper, D. E. *J. Mol. Biol.* **1994**, 237, 577-587.

6. Dahm, S. C.; Uhlenbeck, O. C. *Biochemistry* **1991**, 30, 9464-9469.

7. Auffinger, P.; Grover, N.; Westhof, E. *The Royal Society of Chemistry* **2011**, 9, 1-36.

8. Bai, Y.; Greenfeld, K. J.; Travers, V. B.; Chu, V. B.; Lipfert, J.; Doniach, S.; Herschlag, D. *J. Am. Chem. Soc*. **2007,** 129, 14981-14988.

9. Bujalowski, W.; Graeser, E.; McLaughlin, L. W.; Proschke, D. *Biochemistry* **1986**, 25, 6365-6371.

10. Schimmel, P. R.; Redfield, A. G. *Annu. Rev. Biophys. Bioeng*. **1980**, 9, 181-221.

11. Romer, R.; Hach, R. *Eur. J. Biochem*. **1975**, 55, 271-284.

12. Cole, P. E.; Yang, S. K.; Crothers, D. M. *Biochemistry* **1972**, 4358-4480.

13. Stein, A.; Crothers, D. M. *Biochemistry* **1976**, 15, 160-167.

14. Krakauer, H.; Sturtevant, J. M. *Biochemistry* **1974**, 12, 2579-2589.

15. Yarus, M. *FASEB J*. **1993**, 7, 31-39.

16. Grilley, D.; Soto, A. M.; Draper, D. E. *Proc. Natl. Acad. USA* **2006**, 103, 14003-14008.

17. Leipply, D.; Draper, D. E. *Biochemistry* **2010**, 49, 1843-1853.

18. Leipply, D.; Lambert, D.; Draper, D. E. *Methods Enzymol.* **2009**, 469, 427-457.

19. Koculi, E. C.; Hyeon, C.; Thirumalai, D.; Woodson, S. A. *J. Am. Chem. Soc*. **2007**, 129, 2676-2682.

20. Heilman-Miller, S. L.; Pan, J.; Thirumalai, D.; Woodson, S. A. *J. Mol. Biol.* **2001**, 309, 57-68.

21. Murthy, V. L.; Rose, G. D. *Biochemistry* **2000,** 39, 14365-14370.

22. Baird, N. J.; Westhof, E.; Qin, H.; Pan, T.; Sosnick, T. R. *J. Mol. Biol*. **2005**, 352, 712-722.

23. Soto, A. M.; Misra, M.; Draper, D. E. *Biochemistry* **2007**, 46, 2973-2983.

24. Blanchard, S. C.; Gonzalez, Jr. R. L.; Kim, H. D.; Chu, S.; Puglisi, J. D. *Nature Structural Biology* **2004**, 11, 1008-1014.

25. Lee, T. H.; Blanchard, S. C.; Kim, H. D.; Puglisi, J. D.; Chu, S. *Proc. Natl. Acad. Sci. USA* **2007**, 104, 13661-13665.

26. Klein, D. J.; Moore, P. J.; Steitz, T. A. *RNA* **2004**, 10, 1366-1379.

27. Selmer, M.; Dunham, C. M.; Murphy, F. V. IV.; Weixlbaumer, A.; Petry, S.; Kelley, A. C.; Ramakrishnan, V. *Science* **2006**, 313, 1935-1942.

28. Kankia, B. I.; Marky, L. A. *J. Phys. Chem. B* **1999**, 103, 8759-8767.

29. Lipfert, J.; Doniach, S.; Das, R.; Herschlag, D. *Annual Review of Biochemistry* **2014**, 83, 813-841.

30. Guo, Z; Gibson, M.; Sanyasi, S.; Chu, S.; Mohanty, U. *Proc. Natl. Acad. USA* **2011**, 108, 3947-3951.

31. Draper, D. E. *RNA* **2004**, 10, 335-343.

32. Manning, G. S. *Quart. Review of Biophysics* **1978**, 11, 179-246.

33. Manning, G. S.; Mohanty, U. *Physica A* **1997**, 247, 196-204.

34. Hayes, R. L.; Noel, J. K.; Mohanty, U.; Whitford, P.; Hennelly, S.; Onuchic, J. N.; Sanbonmatsu, K. *J. Am. Chem. Soc*. **2012**, 134, 12043-12053.

35. Hayes, R. L.; Noel, J. K.; Whitford, P. C.; Mohanty, U.; Sanbonmatsu, K. Y.; Onuchic, J. N. *Biophys. J.* **2014**, 106, 1508-1519.

36. Barrat, J. L.; Joaany, J. F. *Adv. Chem. Phys*. **1996**, 94, 1-66.

37. Kjellander, R. *J. Phys. Chem*. **1995**, 99, 10392-10407.

38. Song, S.; Xiao, T. *J. Chem. Phys*. **2011**, 135, 104104-104114.

39. (a) Moriera, A. G.; Netz, R. R. *Eur. Phys. J. D* **2000**, 81, 145-149. (b) Netz, R. R. Orland, H. *Eur. Phys. J. E* **2000**, 1, 203-214.

40. Fuoss, R. M.; Katchalsky, A; Lifson, S. *Proc. Natl. Acad. USA* **1951**, 37, 579-589.

41. Mandel, M. *J. Phys. Chem*. **1992**, 96, 3934-3942.

42. Katchalsky, A. *Pure Appl. Chem*. **1971**, 26, 327-372.

43. Forsman, J. *J. Phys. Chem. B* **2004**, 108, 9236-9245.

44. Netz, R. R.; Orland, H. *Europhys. Lett*. **1999**, 45, 726-732.

45. Attard, P. *J. Phys. Chem*. **1995**, 99, 14174-14181.

46. Koculi, E.; Lee, N.; Thirumalai, D. and Woodson, S. A. *J. Mol. Biol*. **2004**, 341, 27-36.

47. Moghaddam, S.; Caliskan, G.; Chauhan, S.; Hyeon, C.; Briber, R. M.; Thirumalai, D.; Woodson, S. A. *J. Mol. Biol.* **2009**, 393, 753-764.

48. Heilman-Miller, S. L.; Das, J.; Thirumalai, D.; Woodson, S. A. *J. Mol. Biol.* **2001**, 309, 57-68.

49. Lambert, D.; Leipply, D.; Shiman, R.; Draper, D. E. *J. Mol. Biol*. **2009**, 390, 791-804.

50. Manning, G. S. *J. Phys. Chem. B* **2007**, 111, 8554-8559.

51. Wittmer, J.; Johner, A.; Joanny, J. F. *J. Phys. II France* **1995**, 5, 635-654.

52. Ou, Z; Muthukumar, M. *J. Chem. Phys*. **2006**, 124, 154902/1-154902/11.

53. Wei, Y. F.; Haiso, P. Y. arXiv:0912.4070v [con-mat.soft] pages 1-28.

54. Geyer, T. *Biophysics* **2011**, 4:7, 1-20.

Chapter Three

On the nature of surface tension in random first order transition model of supercooled state*

*Reproduced from Streu, K.; Wolynes, P. G.; Mohanty, U. "On the nature of surface tension in random first order transition model of supercooled state." *Manuscript in preparation.*

1. Introduction

There has been a renewed interest in trying to unravel the supercooled state of matter and the characteristics of structural glass transition. A supercooled liquid is far away from equilibrium. Yet, under appropriate conditions of temperature and pressure, one can maintain it be a liquid and measure various equilibrium and dynamical properties. These experiments indicate various universal features of the relaxation dynamics in a supercooled liquid.[1-6]

Theoretical models to describe structural glass-forming liquids are based on dynamical perspective called mode coupling,[7-9] thermodynamic approaches,[10-23] and mixture of thermodynamic and dynamical points of view.[24-31] Correlated dynamics in structural glass formers has been formulated based on the underlying structures of the trajectory space in terms of dynamic heterogeneity and excitation lines.[32,33] The partition function of the trajectory space reveals a singularity that is characteristic of first-order phase change between non-ergodic and ergodic states.[32,33]

A model of supercooled polymer melts based on activated barrier hopping ideas has been developed and extended to non-equilibrium glassy state.[34,35] This model assumes that locally molecular motions in supercooled polymer melts resemble a solid. Freed and coworkers have developed lattice models to describe glass formation in polymer fluids.[14,36] Specifically, these models are capable of describing thermodynamic and fragility of polymer glasses with variations of cohesive energy, monomer structure, and flexibility of backbone and side groups.[36]

There is a deep correspondence between liquids below the freezing point and a class of mean-field spin glasses.[37,38] The later system exhibits both a dynamic transition

temperature where system gets trapped in one of the metastable states whose free-energy is larger than the liquid state, and a thermodynamic transition with no latent heat to a glassy state at a lower temperature.[37,38] Between these two temperatures, the difference in free energy between the metastable states and the liquid is equal to the complexity, i.e., one can depict it to be the configurational entropy of the liquid.[37,38]

In the random first order theory (RFOT) of supercooled liquids,[39-47] the driving force towards equilibrium that occurs by escape from a local metastable configuration region is dictated by competition between the configuration entropy of the other states to which it would escape to, and the surface energy of creating a liquid-like droplet. Near the dynamic crossover temperature, random first order theory and model coupling are essentially equivalent.[45,46] The theory predicts that fragility index which is proportional to the kinetic fragility is inversely related to change of heat capacity per bead of the substance.[39,40,46] In the theory, dynamic heterogeneity length scales like the critical radius of the entropic drop, is found to be inversely proportional to two thirds power of the deviation of temperature from the Kauzmann temperature.[46] The fragility index[18,19] is also inversely proportional to Narayanaswamy-Moynihan-Tool nonlinear parameter, a quantity that is a measure of how far a structural glass is away from equilibrium. Various dynamical characteristics of glasses at low temperatures have been elucidated by random first order theory (RFOT).[41,42,46]

Extension of random first order transition theory for finite dimensional systems is based on fusion of kinetic and thermodynamic views, namely that those particles in large clusters are considered to be frozen as a result of large activation energies, while particles in small clusters are thermodynamically frozen in states that are of low energies.[30,31,45] In

this methodology, a characteristic length scale appears beyond which mean-field picture is not valid and supports RFOT concept of a mosaic of states.[30,31,45]

In this work we study the characteristics of the surface tension of an entropic droplet within the framework of Xia-Wolynes random first order model for the description of supercooled liquids.[40,41,46] In section 2, the probability of a fluctuation in a supercooled liquid is expressed in terms of intensive variables that allow us to deduce the minimum work required to carry out reversible changes in a small part of the supercooled liquid. In random first order model, the wetting of a droplet by a specific density wave by an outside layer from another density wave, act to lower the surface free energy. This is accounted for by the surface tension of the droplet that depends on the radius of the drop. A differential equation for the surface tension is derived in section 3. In the last section, we compare some of the predictions of the model with random first order theory of supercooled liquids.

2. Fluctuation theory

The probability of fluctuation within the framework of Landau-Lifshitz fluctuation theory[48] is $P \sim \exp(-W_{min}/k_B T)$, where k_B is the Boltzmann's constant, T is the absolute temperature, and W_{min} is the minimum work required to carry out reversible changes in a small part of the supercooled liquid. At fixed temperature and pressure P, one can express the minimum work $W_{min} = \Delta E - T\Delta S + P\Delta V$ in terms of energy change ΔE, entropy change ΔS, and volume change ΔV as a result of fluctuation.[48]

It is fruitful to express the minimum work in terms of *intensive variables*, $\tilde{e} = E/V$, $\tilde{s} = S/V$, and $n = N/V$, where N is the number of particles on the system. Then,

$\Delta E - T\Delta S + P\Delta V - \mu\Delta N$ is equal to $V(\Delta\tilde{e} - T\Delta\tilde{s} - \mu\Delta n)$ since $\tilde{e} - T\tilde{s} + P - \mu n$ vanishes, where μ and N denotes the chemical potential of the system. Now expand $\Delta\tilde{e} - T\Delta\tilde{s} + -\mu\Delta n$ $\Delta\tilde{e} - T\Delta\tilde{s} + -\mu\Delta n$ to second order in deviation from equilibrium

$$\frac{1}{2}(\frac{\partial^2\tilde{e}}{\partial\tilde{s}^2})_n(\Delta\tilde{s})^2 + \frac{1}{2}(\frac{\partial^2\tilde{e}}{\partial n^2})_{\tilde{s}}(\Delta n)^2 + 2(\frac{\partial^2\tilde{e}}{\partial n\partial\tilde{s}})_n(\Delta\tilde{s})(\Delta n), \quad (1)$$

where we have made use of the identities $(\frac{\partial\tilde{e}}{\partial\tilde{s}})_n = -T$ and $(\frac{\partial\tilde{e}}{\partial n})_{\tilde{s}} = -\mu$. Next, we observe that Equation 1 can be re-expressed as

$$1/2((\Delta\tilde{s})\Delta(\frac{\partial\tilde{e}}{\partial\tilde{s}})_n + (\Delta n)\Delta(\frac{\partial\tilde{e}}{\partial n})_{\tilde{s}})$$

$$= 1/2((\Delta\tilde{s})(\Delta T) + (\Delta n)(\Delta\mu)). \quad (2)$$

The probability of fluctuation in terms of intensive variables is $P \sim \exp(-V/2k_BT[(\Delta\tilde{s})(\Delta T) + (\Delta n)(\Delta\mu)])$. On expanding ΔT with n held fixed, we observe that the probability of fluctuations scale as $P \sim \exp(-(V/2k_BT)(\Delta\tilde{s})^2/2\tilde{c}_v)$, where $\tilde{c}_v$ is the heat capacity per unit volume. Hence, we obtain the desired expression for fluctuations in entropy $\sqrt{k_B\tilde{c}_v/V}$. We will make use of this result in the next section.

3. Surface tension of entropic drops in random first order transition theory

In random first order transition model of supercooled liquid, the driving force towards equilibrium occurs by escape from a local metastable configuration region, and is dictated by competition between the configuration entropy of the other states to which it would escape to, and the surface energy of creating a liquid-like droplet.[39,40]

The droplet free energy as a function of the radius r of the droplet is given by

$$F(r) = 4\pi r^2 \sigma(r) - \frac{4\pi r^3}{3} T s_c \,, \tag{3}$$

where s_c is the configurational entropy density and T is the absolute temperature. The interface between the various aperiodic minima is complex, and the wetting of droplet by a specific density wave by an outside layer from another density wave, act to lower the surface free energy. This is accounted for by the surface tension $\sigma(r)$ of the droplet that depends on the radius of the drop. The maximum of the free energy satisfies

$$\frac{dF(r)}{dr} = -4\pi r^2 T s_c + 4\pi r^2 (\frac{d\sigma}{dr}) + 8\pi r \sigma(r) = 0 \,. \tag{4}$$

Let us add and subtract a term, $Br^{1/2}$, on the left hand side of Equation 4

$$-4\pi r^2 T s_c + Br^{1/2} + 4\pi r^2 (\frac{d\sigma}{dr}) + 8\pi r \sigma(r) - Br^{1/2} = 0 \,, \tag{5}$$

where B is function of temperature. B is determined as follows.

Let us assume that the sum of the first two terms in Equation 5 vanishes. This condition leads to a relation between configuration entropy density and B

$$s_c = \frac{B}{4\pi T} r^{-3/2} \,. \tag{6}$$

Due to fluctuations in the driving force, i.e., configurational entropy, the cooperative rearranging regions or mosaic like structures, fluctuate in size. However, we showed in section 2 that fluctuations in the configurational entropy of a droplet are $\{\sqrt{(3/4\pi)k_B c_v}\} r^{-3/2}$. On equating this expression with Equation 6, we obtain an explicit expression for B

$$B/T = \sqrt{12\pi k_B c_v} \,. \tag{7}$$

In other words, we have thus shown that if B is given by Eq. (7), then the sum of the first

two terms on the left hand side of Equation 5 vanishes.

The sum of the remaining terms in Equation 5 vanishes

$$4\pi r^2 (\frac{d\sigma}{dr}) + 8\pi r\sigma(r) - Br^{1/2} = 0 . \quad (8)$$

Equation 8 describes the variation of the surface tension of an entropic drop with size r. In the next section, the solution of Equation 8 for the surface tension is compared with that obtained from renormalization group analysis[39,40] of random field Ising magnet by Xia and Wolynes.

4. Discussion

The solution to the differential equation for the surface tension is of the form

$$\sigma(r) = \sigma_o (r / r_o)^{\alpha} , \quad (9)$$

where σ_o is the surface tension in the absence of wetting. The boundary condition on the entropic droplet is that $r \to \infty$, $\sigma \to 0$, and that the short-range value of surface tension is set to σ_o at $r = r_o$.[39,40] On substituting Equation 9 in Equation 8, and making use of the boundary conditions, we find that $\alpha = 1/2$, and that B is expressible in terms of σ_o and r_o

$$B = 6\pi\sigma_o r_o^{1/2} . \quad (10)$$

On equating Equations 7 and 10, we obtain an explicit expression for the temperature dependence of the product $\sigma_o r_o^{1/2}$

$$\sigma_o r_o^{1/2} = T\sqrt{k_B c_v / 3\pi} . \quad (11)$$

We next compare this prediction with that based on renormalization group (RG) analysis of random first order transition entropic picture of supercooled liquids.[39,40] The wetting

point of the droplet by a density wave by an outside layer from a different density wave, leads via analogy with random field Ising magnet,[30,39] to a differential renormalization group formulation for the surface tension $\sigma(r)$

$$\sigma^{1/3} d\sigma = -\alpha \frac{dr}{r^{5/3}}, \tag{12}$$

where $\alpha = 3x4^{-4/3}(T\sqrt{k_B \tilde{c}_v})^{4/3}$ and k_B is the Boltzmann's constant. $\Delta\tilde{c}_p$ is the jump in heat capacity per unit volume. Integrating the above equation between r_o and r leads to

$$\int_{\sigma_o}^{\sigma} \sigma^{1/3}\, d\sigma = -\alpha \int_{r_o}^{r} \frac{dr}{r^{5/3}}. \tag{13a}$$

$$[(\frac{\sigma}{\sigma_o})^{4/3} - 1] = \frac{2\alpha}{(\sigma_o^2 r_o)^{2/3}}[(\frac{r_o}{r})^{2/3} - 1]. \tag{13b}$$

The boundary condition on the drop is $r \to \infty$, $\sigma \to 0$. This leads to a relation between α and σ_o, namely $2\alpha = (\sigma_o^2 r_o)^{2/3}$; the later relation is re-expressed as

$$\sigma_o r_o^{1/2} = (\frac{6^{3/4}}{4})T\sqrt{k_B \tilde{c}_v}. \tag{14}$$

Observe that the temperature dependence of the product $\sigma_o r_o^{1/2}$ is the same is that predicted by Equation 11; the two expressions differ by a numerical factor.

The activation barrier $\Delta F^{\dagger}$ in the random first order theory is[30,39,46]

$$\Delta F^{\dagger} = \frac{3\pi\sigma_o^2 r_o}{Ts_c}. \tag{15}$$

This expression is further simplified using Equation 11 to yield $\Delta F^{\dagger}/k_B T = \frac{c_v}{s_c}$. The slope of the relaxation time at the glass transition temperature, i.e., the fragility index[18,19]

is

$$m = \{\frac{d \ln \tau}{d(T_g / T)}\}_{T=T_g} \approx \frac{1}{\ln(10)}(\frac{c_v}{s_c})^2 . \qquad (16)$$

In deriving Equation 16, we have ignored the temperature variation of the heat capacity near the glass transition temperature.

Several points are in order. First, observe that fragility index m is expressed as the ratio of heat capacity and configurational entropy of the supercooled liquid.[13,15] Second, no parameters appear on the right hand side of Equation 10; both heat capacity and configurational entropy at the glass transition temperature can be either obtained or estimated from experimental data. Third, the fragility of the liquid as predicted by random first order transition model is $(\frac{27}{16})\pi(\frac{nk_B}{\Delta c_p})\ln^2(\frac{\alpha_L r_o^2}{\pi e})$, where n is the density of particles, r_o is the mean lattice spacing, $\alpha_L^{-1/2}$ is the root mean square ms displacement around an aperiodic minimum. If the Lindemann ratio, $\alpha_L^{-1/2} / r_o$, is taken to be 0.1, then the liquid fragility expressed in terms of the gas constant R and the change in heat capacity per mole, is $32R / \Delta c_p$.[40,41,46] Third, the near universality of $\sigma_o / nr_o k_B T$ in random first order transition model is based on universality of the Lindemann ratio. From Equation 11, the universality of σ_o at the laboratory glass transition temperature is only approximate for glass forming liquids.

Finally, we can generalize our results to entropic droplets in arbitrary spatial dimensions d. The droplet free energy as a function of the radius r of the drop is $F(r) = S_d r^{\theta}\sigma(r) - \Omega_d r^d T s_c$, where Ω_d and S_d denotes the volume and surface area of a

unit sphere in dimensions, and θ is an exponent not necessarily equal to d-1. We minimize the free energy with respect to the radius of the drop, and add and subtract the term Br^{ω} to the minimized equation, where ω is an exponent. Then, following the same strategy discussed in section 3, we find that the differential equation for the surface tension of the entropic droplet is $\frac{d\sigma(r)}{dr} + \frac{\theta\sigma(r)}{r} - \frac{B}{S_d r^{\theta-\omega}} = 0$. The boundary conditions on the drop are as before, namely, $r \to \infty$, $\sigma \to 0$, and the short-range value of surface tension is σ_o at $r = r_o$. The solution of the differential equation for the surface tension of the drop is $\sigma(r) = \sigma_o(\frac{r_o}{r})^{\alpha}$, where the exponent $\alpha = \theta - \omega - 1$. Furthermore, we find that $\omega = d - 1$, $B/T = d(k_B c_v \Omega_d)^{1/2}$, the temperature dependence of the product $\sigma_o r_o^{\alpha}$ is expressible in a form analogous to Equation 14, and that the slope of the relaxation time at the glass transition temperature is given by Equation 16.

In summary, we have analyzed the nature of surface tension in the random first order theory of supercooled liquid within the framework of Landau-Lifshitz fluctuation theory. A differential equation for the surface tension of the drop is derived. The characteristics of the surface tension are in agreement with that based on random first order model of glass forming liquids. The fragility index is given by the square of the ratio of heat capacity and configurational entropy of the supercooled liquid.

References

1. Ngai, K. L. *J. Non-Crys. Solids* **2000**, 275, 7-51.

2. Angell, C. A. *Science* **1995**, 267 1924-1933.

3. Mohanty, U. Advances in Chemical Physics, Vol. 89, Chapter 2, pages 89-158, ed. Prigogine, I.; Rice, S. A. (John Wiley & Sons, Inc., **1994**).

4. Angell, C. A.; Ngai, K. L.; McKenna, G. B.; Martin, S. W. *J. Appl. Phys*. **2000**, 88, 3113-3157.

5. Eidger, M. D. *Annu. Rev. Phys. Chem*. **2000**, 51, 99-128.

6. Richert, R. *J. Phys. Condens. Matter* **2002**, 14, R703-738.

7. Götze, W. *J. Phys. Condens. Matter* **1999**, 11, A1-45.

8. Biroli, G. Bouchard, J. P.; Miyazaki, K.; Reichman, D. R. *Phys. Rev. Lett*. **2006**, 97, 195701/1-195701/4.

9. Bhattacharyya, S. M.; Bagchi, B.; Wolynes, P. G. *Proc. Natl. Acad. Sci. USA* **2008**, 105, 10677-16082.

10. Adam, G.; Gibbs, J. H. *J. Chem. Phys*. **1965**, 43, 139-146.

11. Gibbs, J. H.; DiMarzio, E. A. *J. Chem. Phys*. **1958**, 28, 373-383.

12. Mohanty, U. *Physica A* **1991**, 177, 345-355.

13. Mohanty, U.; Craig, N.; Fourkas, J. T. *Phys. Rev. E* **2001**, 64, 10501-10502.

14. Ducowitz, J.; Freed, K. F.; Douglas, J. F. *J. Chem. Phys*. **2006**, 124, 064901-064915.

15. Mohanty, U.; Craig, N.; Fourkas, J. T. *J. Chem. Phys*. **2001**, 114, 10577-10578.

16. Stillinger, F. *J. Chem. Phys*. **1988**, 88, 7818-7825.

17. Angell, C. A. *J. Res. Natl. Inst. Stand. Technol*. **1997**, 102, 171-185.

18. Richert, R.; Angell, C. A. *J. Chem. Phys*. **1998**, 108, 9016-9026.

19. Mohanty, U. *Adv. Chem.Phys*. **2004**, 129, 73-105.

20. Hall, R. W.; Wolynes, P. G. *J. Phys. Chem*. *B* **2008**, 112, 301-312.

21. Ito, K.; Moynahan, C. T.; Angell, C. A. *Nature* **1999**, 398, 492-495.

22. Mezard, M.; Parisi, G. *J. Chem. Phys*. **1999**, 111, 1076-1095.

23. Wang, S.; Wolynes, P. G. *J. Chem. Phys*. **2013**, 138, 12A521/1-12A52/10.

24. Sarmakar, S.; Dasgupta, C.; Sastry, S. *Proc. Natl. Acad. Sci. USA* **2009**, 106, 3675-3679.

25. Mallamace, F.; Branca, C.; Corsaro, C.; Leone, N.; Spooren, J.; Chen, S. H.; Stanley, H. E. *Proc. Natl. Acad. Sci*. *USA* **2010**, 107, 22457-22462.

26. Casalini, R.; Mohanty, U.; Roland, C. M. *J. Chem. Phys*. **2006**, 125, 014505-014513.

27. Casalini, R.; Capaccioli, S.; Roland, C. M. *J. Phys. Chem. B* **2006**, 110, 11491-11495.

28. Capaccioli, S.; Ruocco, G.; Zamponi, F. *J. Phys. Chem. B* **2008**, 112, 10652-10658.

29. Giovambattista, N.; Buldyrev, S. V.; Starr, F.; Stanley, H. E. *Phys. Rev. Lett*. **2003**, 90, 085506/1-085506/4.

30. Bouchard, J. P.; Birol, G. arXiv:cond-mat/0406317v1, pages 1-8.

31. Biroli, G.; Bouchard, J. P.; Cavagna, A.; Grigera, T. S.; Verricchio, P. arXiv:cond-mat/0805.4427v1, pages 1-15.

32. Chandler, D.; Garrahan, J. P. *Ann. Rev. Phys. Chem*. **2010**, 61, 191-217.

33. Garrahan, J. P.; Chandler, D. *Phys. Rev. Lett*. **2002**, 89, 035704/1-035704/4.

34. Schweizer, K. S. *J. Chem. Phys*. **2005**, 123, 244501/1-244501/13.

35. Chen, K.; Schweizer, K. S. *J. Chem. Phys*. **2007**, 126, 014904/1-014904/10.

36. Freed, K. F. *Acc. Chem. Res*. **2011**, 44, 194-203.

37. Kirkpartick, T. R.; Thirumalai, D.; Wolynes. P. G. *Phys. Rev. A* **1989**, 40, 1045-1054.

38. Kirkpartick, T. R.; Wolynes, P. G. *Phys. Rev. A* **1987**, 35, 3072-3080.

39. Xia, X.; Wolynes, P. G. *Proc. Natl. Acad. Sci. USA* **2000**, 97, 2990-2994.

40. Xia, X.; Wolynes, P. G. *Phys. Rev. Lett*. **2001**, 86, 5526.

41. Lubchenko, V.; Wolynes, P. G. *Proc. Natl. Acad. Sci. USA* **2003**, 100, 1515.

42. Lubchenko, V.; Wolynes, P. G. *J. Chem. Phys*. **2003**, 119, 9088-9105.

43. Stevenson, J. D.; Schmalian, J.; Wolynes, P. G. *Nat. Phys*. **2006**, 2, 268-274.

44. Stevenson, J. D.; Wolynes, P. G. *J. Phys. Chem. B*. **2005**, 109, 15093-15097.

45. Biroli, G.; Bouchard, J. P. arXiv:0912.2542v1, 1-53.

46. Lubchenko, V.; Wolynes, P. G. *Annu. Rev. Phys. Chem.* **2007**, 58, 235-266.

47. Rabociy, P.; Lubchenko, V. *J. Phys. Chem. B* **2012**, 116, 5729-5737.

48. Landau, L. D.; Lifshitz, E. M. *Statistical Mechanics*; Pergamon Press: Elmsford, NY, **1980**.

Chapter Four

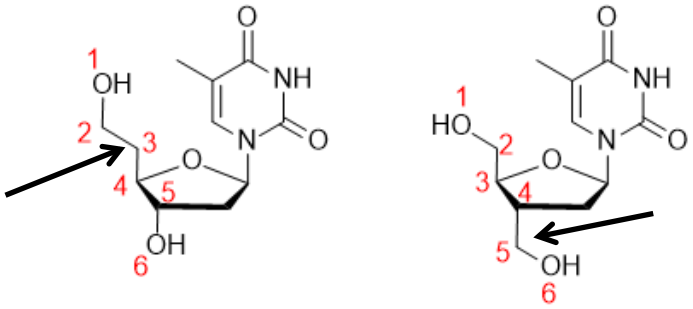
Conformation and structural stability of modified DNA oligomers*

*Reproduced from Streu, K; Das, R.; McCormick, R; Mohammad, I.; McLaughlin, L.; Mohanty, U. "Conformation and structural stability of modified DNA oligomers." *Manuscript in preparation.*

1. Introduction

Modifying the phosphodiester backbone of a DNA polymer can alter how tightly the DNA double helix is wound as well as the stability of the molecule. There are previous studies that include both increasing and decreasing the complexity of a standard phosphodiester nucleic acid linkage. Many of these modification strategies have yielded novel nucleic acid structures that exhibit useful properties. The synthesis and structure characterization of modified DNA backbone nucleosides is complex and often unsuccessful. Some examples of successful modifications are the creation of a phosphorothioate linkage between nucleosides, which provide a synthetic route for incorporating sulfur atoms into the DNA backbone.[1] More recently, researchers have created phosphorothioester linked nucleic acids with a resistance to phosphodiester bond cleavage and a decrease in duplex melting temperature, which is indicative of a decrease in duplex stability.[2] McLaughlin and coworkers have previously achieved the synthesis of shortened and simplified backbone nucleic acid oligomers.[3-8] One of the shortened backbone structures is a propagation of threose nucleic acids (TNA) in which the 3' methylene group is deleted on each nucleoside, and the hydroxyl group is shifted adjacent to the base on the sugar. The other simplified backbone structure is a propagation of glycol nucleic acids (GNA) in which the TNA structure is further modified by deleting the oxygen and adjacent carbon in the ribose ring, creating more flexibility in the nucleic acid linker.[3-8] Both of these bonding patterns form a truncated phosphodiester backbone that is simplistic in nature and may be representative of precursors to the evolved nucleic acid structure we know today.[9]

In this study, we investigate elongated nucleic acid linkers and the modification influences on duplex structural stability. The 5' to 3' sequence of the DNA duplex is as follows: GGCGAATTCCGG, with a complementary base-paired strand. For each modified duplex, four extra carbons are inserted into each duplex, two per each single strand that is base-paired together. These carbon insertions occur at both adenine (A) residues or at both thymine (T) residues in the sequence. If the carbons are inserted on A residues in one chain, carbons are inserted on the corresponding base-paired T residues in the complimentary chain, and vice versa. The added carbon modifications were strategically placed directly across from each other in the double helix to match any conformational change that may occur. Figure 1 shows the position of the carbon insertion into the backbone chain, in either the 3' or the 6' position of the nucleoside. In Table 1 we report the combinations of modifications that were applied to DNA dodecamers, as well as the notation used for each. Finally, the T_m melting data were taken, and MD simulations performed, at two different salt (NaCl) concentrations: 0.1 M and 1.0 M.



3'-dT **6'-dT**

Figure 1: Schematic representation of a modified thymine nucleoside, extended at the 3'-carbon (left) and the 5'-carbon (right).

Table 1: Sequences of the unmodified and extended backbone modified DNA dodecamers.

DNA Duplex		Nucleotide Sequence # and Identity												
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	
Unmodified	**5'**	G	G	C	G	A	A	T	T	C	C	G	G	**3'**
	3'	C	C	G	C	T	T	A	A	G	G	C	C	**5'**
$GA_62\_CT_32$	**5'**	G	G	C	G	A_6	A_6	T	T	C	C	G	G	**3'**
	3'	C	C	G	C	T_3	T_3	A	A	G	G	C	C	**5'**
$CA_62\_GT_32$	**5'**	G	G	C	G	A	A	T_3	T_3	C	C	G	G	**3'**
	3'	C	C	G	C	T	T	A_6	A_6	G	G	C	C	**5'**

[a]Subscript of 3 denotes that an extra carbon was added to the 3' site of the nucleoside.

[b]Subscript of 6 denotes that an extra carbon was added to the 6' site of the nucleoside.

[c]The 2 in the 12mer nomenclature denotes that the nucleoside modification was made twice in that sequence.

2. Methods

Simulation System Preparation

The sequences of the DNA duplexes used in the study are reported in Table 1. The starting DNA conformation for each system was the canonical B-DNA double helix generated using the Maestro suite of programs.[10,11] Extended backbone structures were generated by inserting carbons in the designated locations in Table 1 and depicted in Figure 1. The modified and unmodified structures were placed in a 10 Å cubic box with periodic boundary conditions, and solvated with TIP3P[12] water using the Desmond System Builder.[10] An appropriate number of sodium ions were added to maintain charge neutrality, and each system was solvated with an additional NaCl concentration of 0.10 M and 1.0 M, consistent with experimental salt concentrations.

Molecular Dynamics Simulations

Molecular dynamics (MD) simulations were performed for each DNA sequence in Table 1 using Desmond. Each system was equilibrated at 300 K by a series of minimizations and all atom molecular dynamics simulations using the default equilibration protocol in Desmond.[10] The OPLS-2005 force field and periodic boundary conditions were employed in the simulations.[13,14] Long-range electrostatic interactions were calculated by particle mesh Ewald method.[15] Short-range electrostatic and Van der Waals interactions were truncated at 7.0 Å. To control the pressure, Martyna-Tobias-Klein method was used.[16] Constant simulation temperature was maintained by Nose-Hoover thermostats.[17] Simulations were performed using the NPT ensemble class, with a

constant pressure of 1 atm and temperature of 300 K. The length of each simulation was 30 ns with energies saved in 1-ps intervals and trajectories saved in 4-ps intervals.

3. Results and Discussion

Helix Stability

Results of helix stability were evaluated by monitoring base pair hydrogen bonds maintained throughout the simulation. Each hydrogen bond distance in the duplex is calculated for each trajectory output throughout the length of the simulation (32 hydrogen bonds in total for each 12mer duplex) and Figure 2 compares all three sequences at high salt concentration. The shorter the hydrogen bond, the more stable the helix. Notably, the extended backbone duplexes remain as stable as, or more stable than, the unmodified 12mer duplex.

The stability of each 12mer duplex system was compared in the low and high salt concentrations that were used in melting temperature analysis. The unmodified and backbone extended 12mer duplexes exhibited more stability in 1.0 M salt over 0.10 M salt concentration, as reflected by the hydrogen bond distances of the base-pairs. Overall, more hydrogen bonds were shorter and maintained bonded throughout the simulation in 1.0 M salt for all duplexes. This is graphed in Figure 3 for a central base-pair hydrogen bond in the $CA_3$2-$GT_3$2 backbone extended duplex.

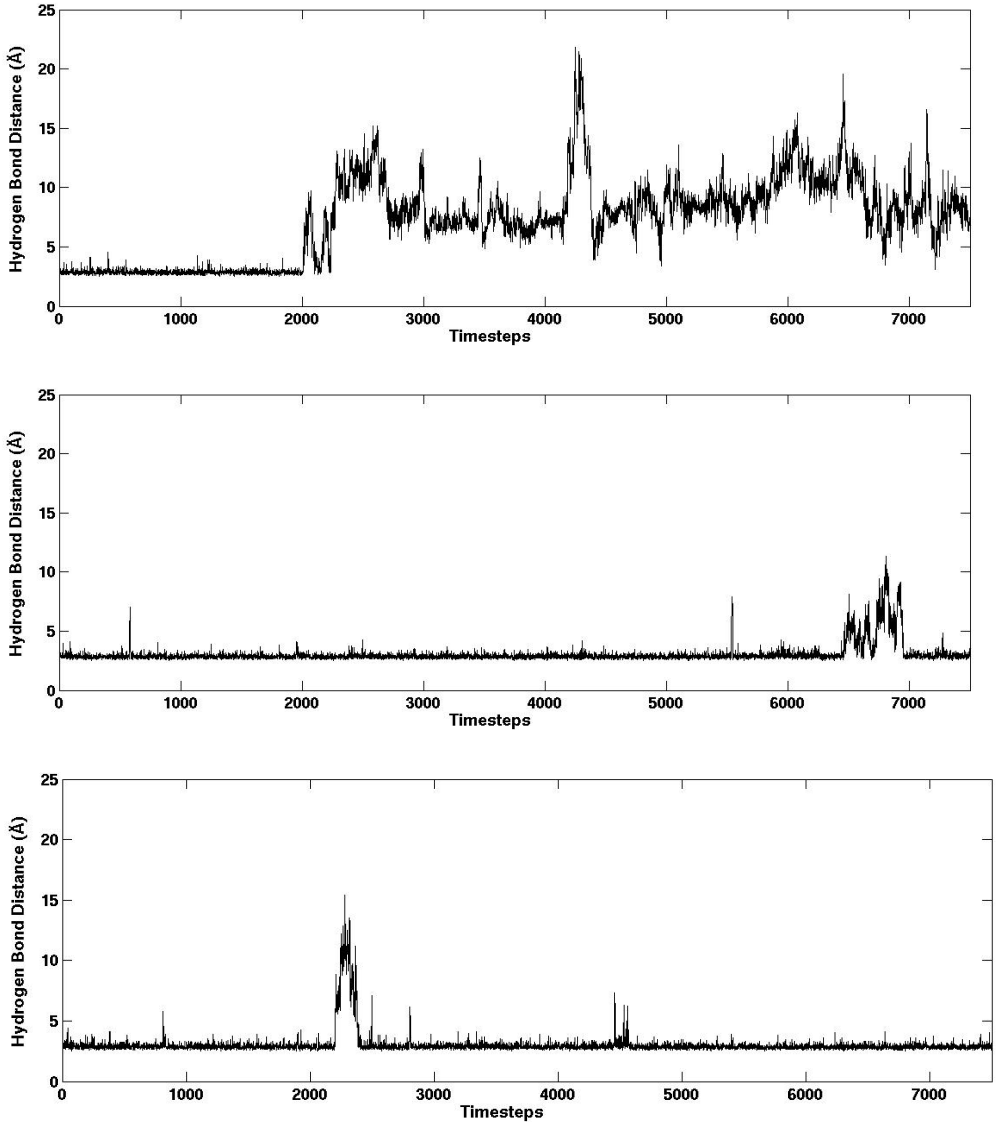


Figure 2: Hydrogen bond length of a terminal DNA duplex base-pair, at the 1 position in the duplex sequence (Unmodified – top, $CA_6$2-$GT_3$2 – middle, $GA_6$2-$CT_3$2 – bottom) throughout 30-ns MD simulation in 0.1 M salt concentration.

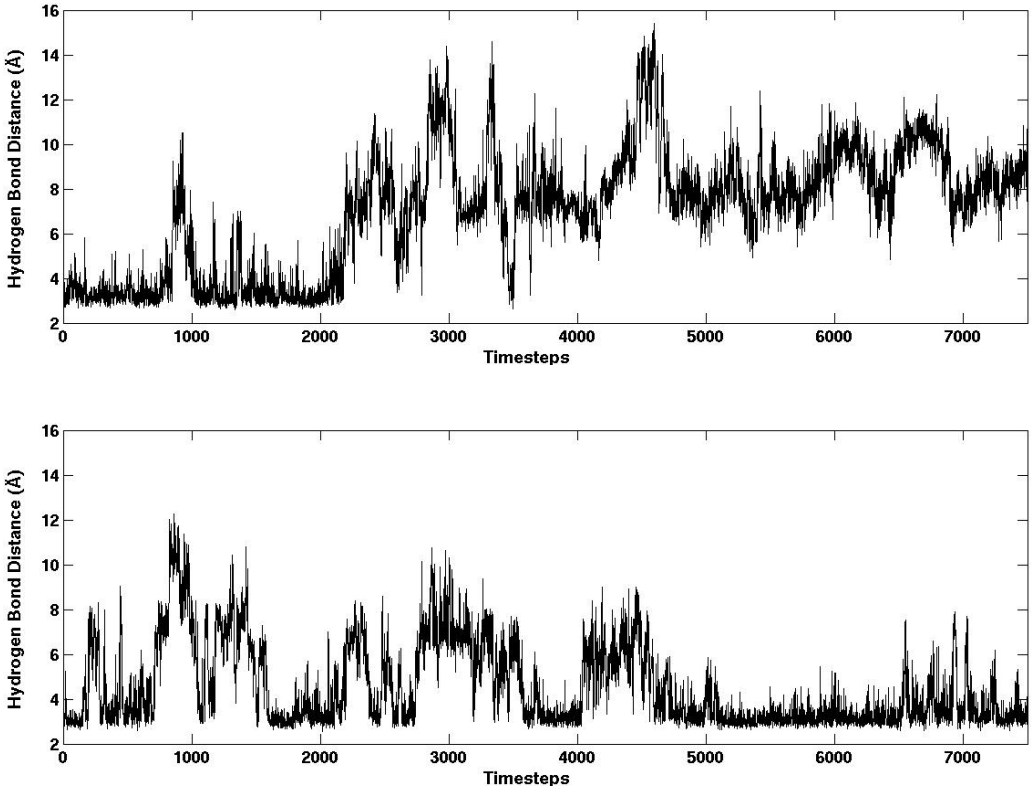


Figure 3: Hydrogen bond length of the 6 position in the dodecamer sequence, a central base-pair, for the modified $CA_6$2-$GT_3$2 DNA duplex in 0.10 M salt (top) to 1.0 M salt (bottom).

Backbone and Helical Conformation

The conformational characteristics of the DNA duplexes are first analyzed by the phosphate distances between duplex base-pairs (helix width) and phosphate distances end-to-end (helix length). These values are reported in Table 2 and Figure 4 depicts the duplex conformations. In Figure 4, the conformations of the modified duplexes appear distorted compared to the unmodified duplex. In general, the results in Table 2 can be summarized as the duplexes form more elongated and slimmer conformations at higher NaCl concentrations. The elongated, skinny DNA structure is characteristic of a more stable, strongly hydrogen-bonded, duplex. One of the modified duplexes, $GA_6$2_$CT_3$2, has an even longer helix length, and smaller helix width compared to the unmodified duplex at the same NaCl concentration. And the other modified duplex, $CA_6$2_$GT_3$2, has a conformation very comparable to the unmodified duplex at the same salt concentration. Therefore, extended backbone duplexes have slightly different conformations compared to the unmodified duplexes, but we find that the modified DNA oligomers retain helix stability.

Table 2: Helix end-to-end and helix width distances measured between phosphates of the DNA backbone for all 12mer duplexes studied.

	Unmodified 0.10 M	Unmodified 1.0 M	$GA_62\_CT_32$ 0.10 M	$GA_62\_CT_32$ 1.0 M	$CA_62\_GT_32$ 0.10 M	$CA_62\_GT_32$ 1.0 M
End-to-end distance (Å)	46.84	53.16	48.15	44.34	44.09	54.08
DNA width (Å)	17.21	14.81	16.99	16.14	17.03	15.77

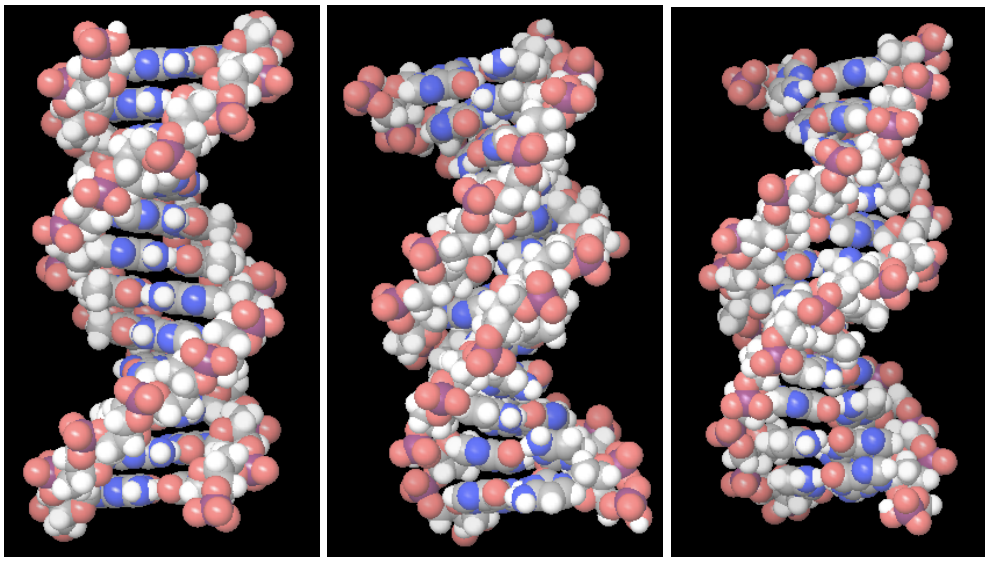


Figure 4: Conformations of the three DNA duplexes: Unmodified (left), $CA_6$2-$GT_3$2 (center), $GA_6$2-$CT_3$2 (right).

Conformation of the backbone extended DNA duplexes is also analyzed by helical twist angle. We utilize 3DNA, a software package for the analysis of three-dimensional nucleic acid structures, to calculate the helical twist at each base-pair of the duplex.[18] The twist angles are plotted in Figure 5 to compare the modified regions of the backbone extended duplexes with the unmodified duplex. The sharp increase in twist angle for the backbone extended duplexes at base-pair 'A6T7' shows a shift from the expected B-DNA twist angle of 36^{o} up to over 40^{o} which is the reported D-form DNA helical twist.[19]

T_m Experimental Melting Data Comparison

The T_m melting data is reported in Table 3 for the DNA duplexes, and there are some consistencies and inconsistencies with the structural stability results from the molecular dynamics simulations. Each of the duplexes studied have a higher T_m for the higher NaCl concentration, indicative of highly stable duplex formation and this experimental T_m trend is in agreement with the computational results. Both modified duplexes show lower T_m values when compared to the unmodified duplex, which suggest that the extended backbone modifications destabilize the helix. These results are inconsistent with the MD simulation results and can be rationalized as follows. The added carbon to extend the backbone of the modified duplexes causes a disruption in the favorable base-stacking interactions of the sequential base-pairs along the duplex chain. The loss of strong base stacking interactions between adjacent base-pairs causes instability in the DNA duplex and results in lower T_m values. The loss of duplex stability due to disruption of base stacking interactions could not be quantified using MD simulations alone. Instead, the computational results reflect increased or similar stability for modified the duplex

backbone, with shorter base-pair hydrogen bonds and more hydrogen bonds maintained throughout the MD simulation.

In summary, while not all of the duplex conformation and structural stability computational results are consistent with the experimental data, the inconsistencies reveal the origin of the extended backbone duplex instability. We find that backbone extended DNA duplexes have a similar or more elongated chain with more stable hydrogen bonding between base-pairs, distorted base-pair stacking interactions, and a helical twist angle at the modified site that is consistent with D-DNA, while the unmodified oligomers are B-DNA duplexes.

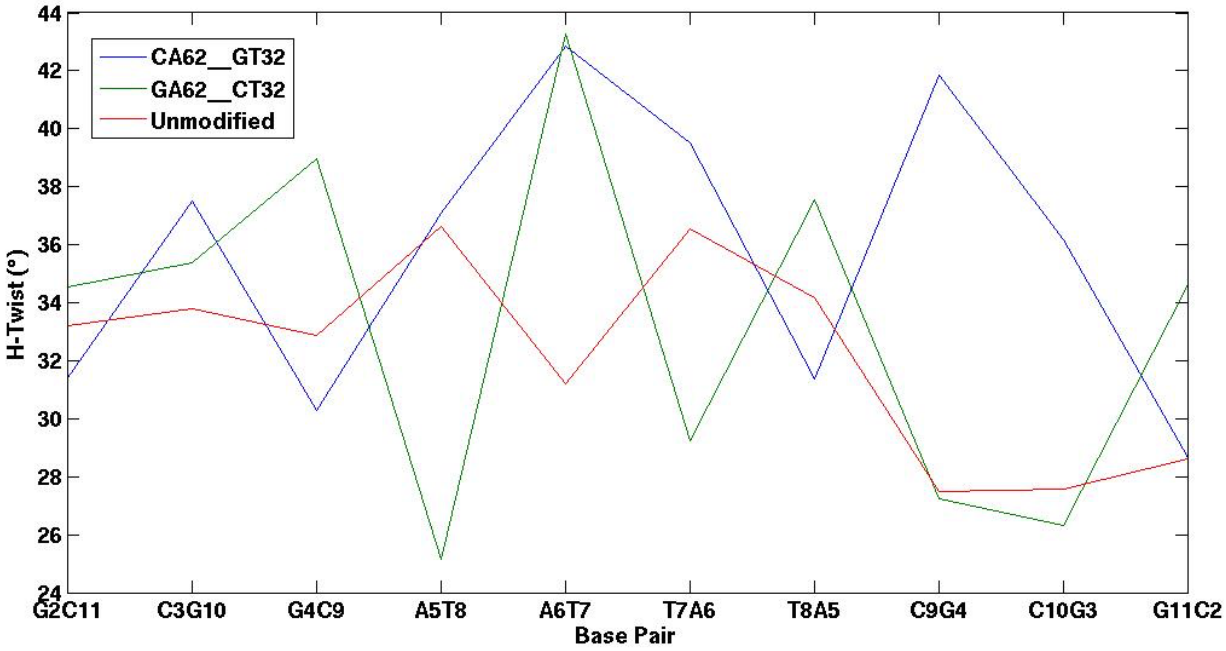


Figure 5: Calculated helical twist angles of DNA duplex base-pairs (Unmodified – red, CA_62-GT_32 – blue, GA_62-CT_32 – green) using 3DNA software.[18]

Table 3: Melting temperature, T_m, values for DNA duplexes at 0.1 M and 1 M salt concentration.

DNA Duplex	NaCl (M)	T_m (^oC)
Unmodified	0.10	53
	1.0	59
$GA_6$2_$CT_3$2	0.10	42
	1.0	49
$CA_6$2_$GT_3$2	0.10	42
	1.0	48

References

1. Cook, A. F. *J. Am. Chem. Soc.* **1970**, 92, 190-195.

2. Xu, Y.; Kool, E. T. *Nucleic Acids Research* **1998**, 26, 3159-3164.

3. Chen, J. J.; Tsai, C. H.; Cai, X; Horhota, A. T.; McLaughlin, L. W.; Szostak, J. W. *PLOS One* **2009**, 4.

4. Horhota, A. T.; Zou, K. Y.; Ichida, J. K.; Yu, B.; McLaughlin, L. W.; Szostak, J. W. *J. Am. Chem. Soc.* **2005**, 127, 7427-7434.

5. Horhota, A. T.; Szostak, J. W.; McLaughlin, L. W. *Org. Lett.* **2006**, 8, 5345-5347.

6. Ichida, J. K.; Horhota, A.; Zou, K. Y.; McLaughlin, L. W.; Szostak, J. W. *Nucleic Acids Research* **2005**, 33, 5219-5225.

7. Ichida, J. K.; Horhota, A.; Yu, B.; McLaughlin, L. W.; Szostak, J. W. *J. Am. Chem. Soc.* **2005**, 127, 2802-2803.

8. Zou, K. Y.; Horhota, A.; Yu, B.; Szostak, J. W.; McLaughlin, L. W. Org. Lett. 2005, 7, 1485-1487.

9. Chaput, J. C.; Ichida, J. K.; Szostak, J. W. *J. Am. Chem. Soc.* **2003**, 125, 856-857.

10. (2009) Desmond (version 2.2). New York, NY: Schrödinger, LLC.

11. (2009) Maestro (version 8). New York, NY: Schrödinger, LLC.

12. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, 79, 926-935.

13. Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, 118, 11225-11236.

14. Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. J. *Phys. Chem. B* **2001**, 105, 6474-6487.

15. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, 103, 8577-8593.

16. Martyna, G. J.; Tobias, D. J.; Klein, M. L. *J. Chem. Phys.* **1994**, 101, 4177-4189.

17. Hoover, W. G. *Phys. Rev. A* **1985**, 31, 1695-1697.

18. Lu, X. L.; Olson, W. K. *Nucleic Acids Res.* **2003**, 31, 5108-5121.

19. Saenger, W. *Principles of Nucleic Acid Structure*, Springer-Verlag: New York, 1984.

Supplemental Data

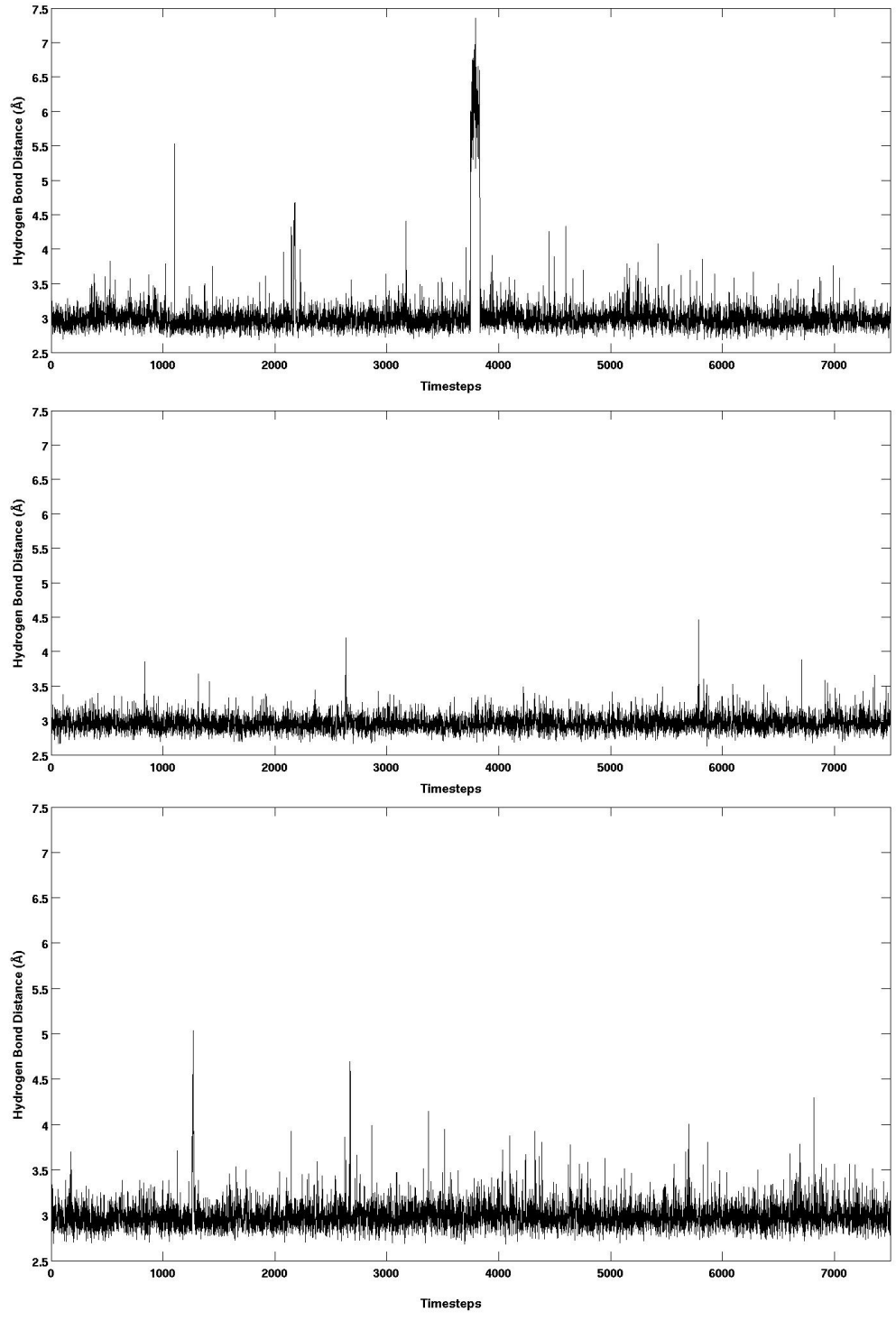


Figure S1: Hydrogen bond length of a DNA duplex base-pair at the 3 position in the duplex sequence (Unmodified – top, $CA_6$2-$GT_3$2 – middle, $GA_6$2-$CT_3$2 – bottom) throughout 30-ns MD simulation in 1.0 M salt concentration.

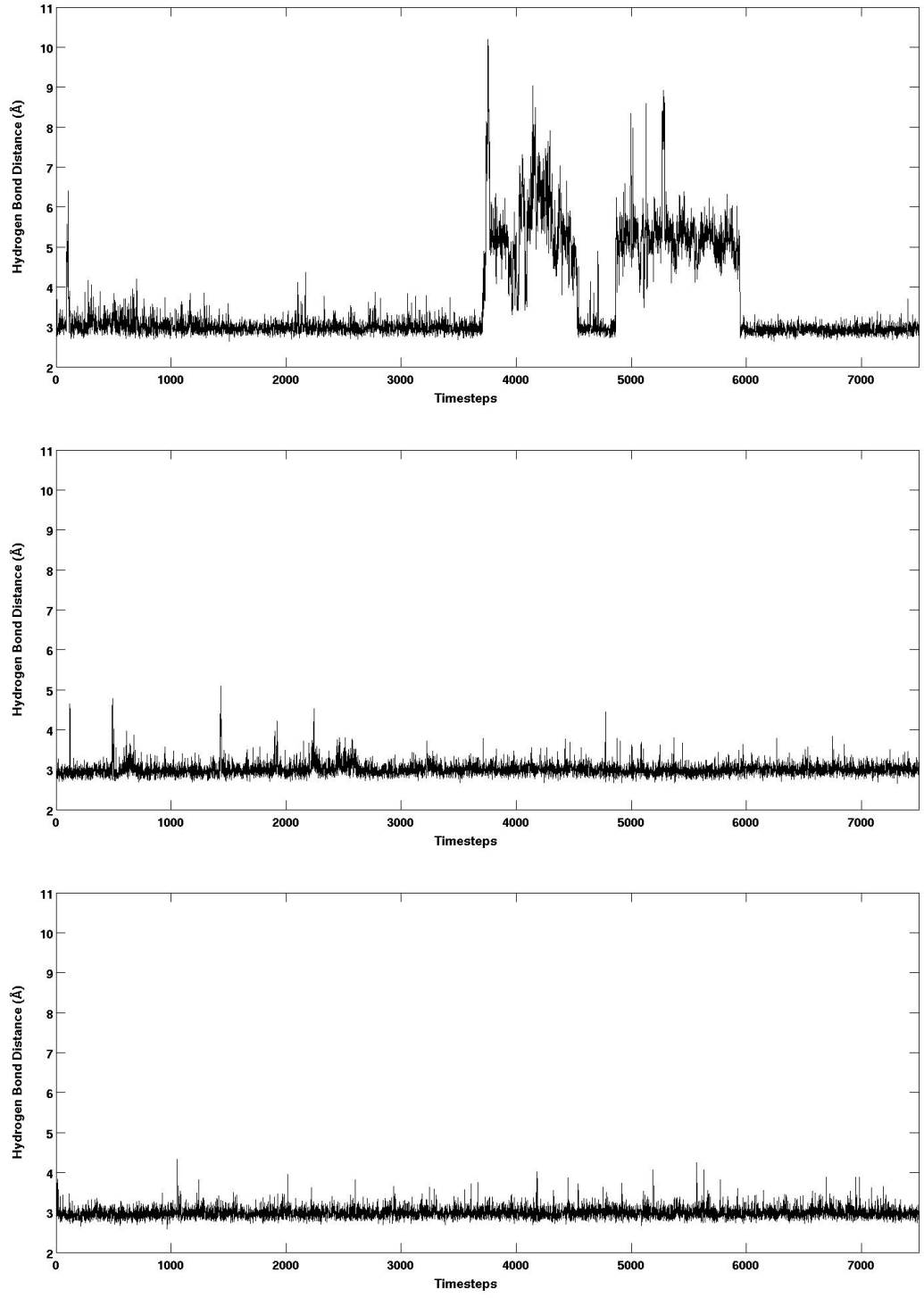


Figure S2: Hydrogen bond length of a DNA duplex base-pair at the 9 position in the duplex sequence (Unmodified – top, $CA_62\text{-}GT_32$ – middle, $GA_62\text{-}CT_32$ – bottom) throughout 30-ns MD simulation in 1.0 M salt concentration.

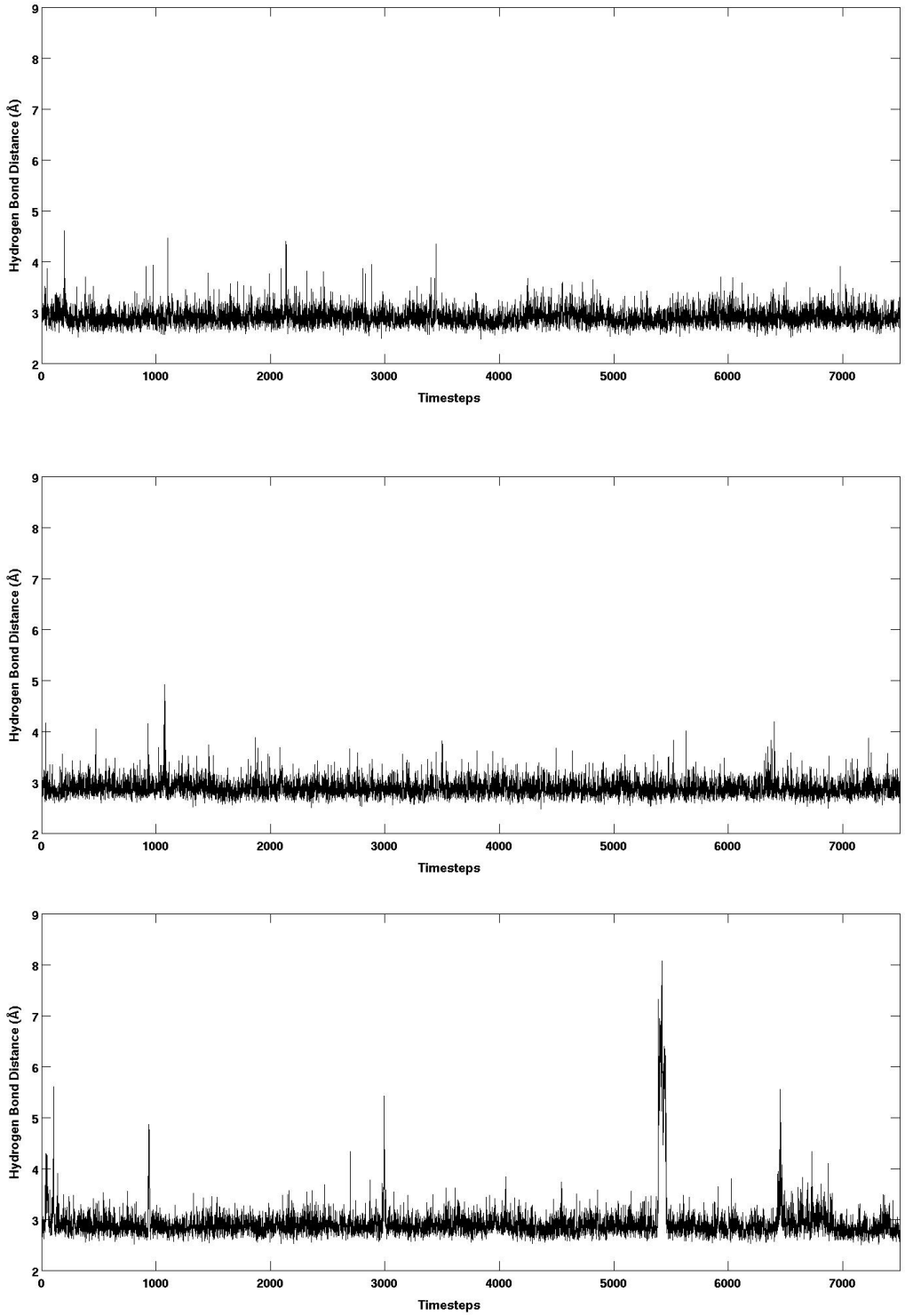


Figure S3: Hydrogen bond length of a DNA duplex base-pair at the 4 position in the duplex sequence (Unmodified – top, CA_62-GT_32 – middle, GA_62-CT_32 – bottom) throughout 30-ns MD simulation in 1.0 M salt concentration.

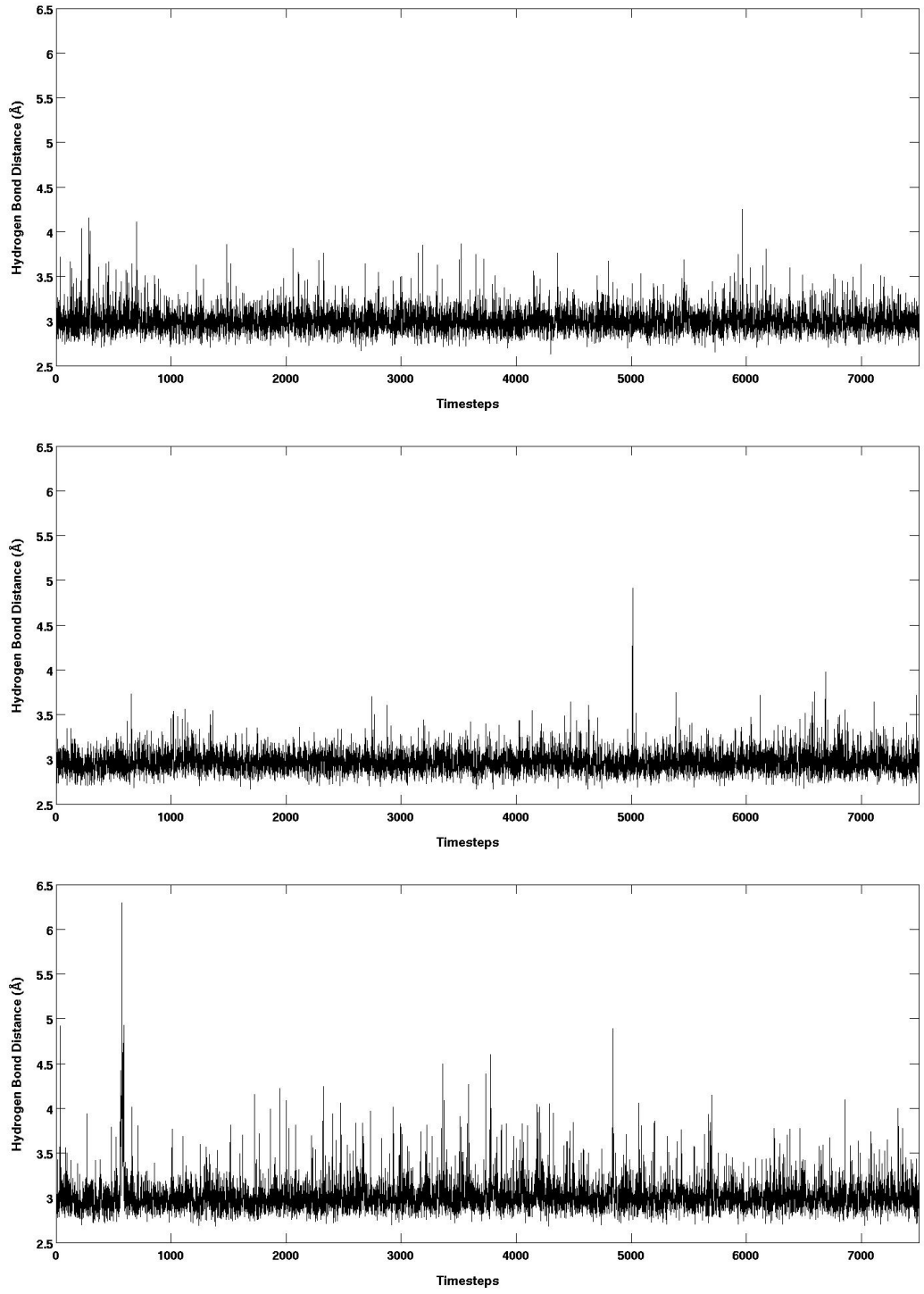


Figure S4: Hydrogen bond length of a DNA duplex base-pair at the 9 position in the duplex sequence (Unmodified – top, $CA_62\text{-}GT_32$ – middle, $GA_62\text{-}CT_32$ – bottom) throughout 30-ns MD simulation in 0.1 M salt concentration.

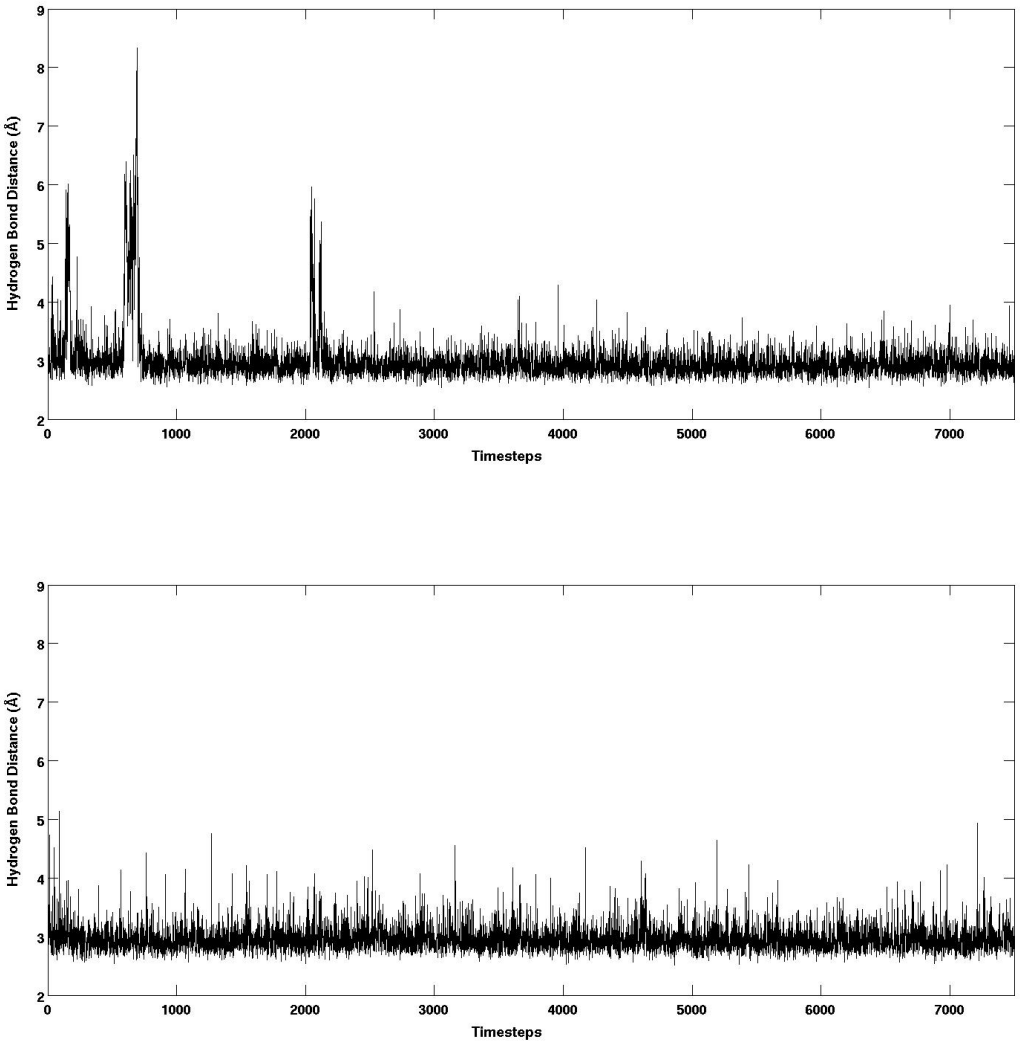


Figure S5: Hydrogen bond length of the DNA base-pair at the 10 position in the dodecamer sequence for the unmodified DNA duplex in 0.10 M salt (top) to 1.0 M salt (bottom).

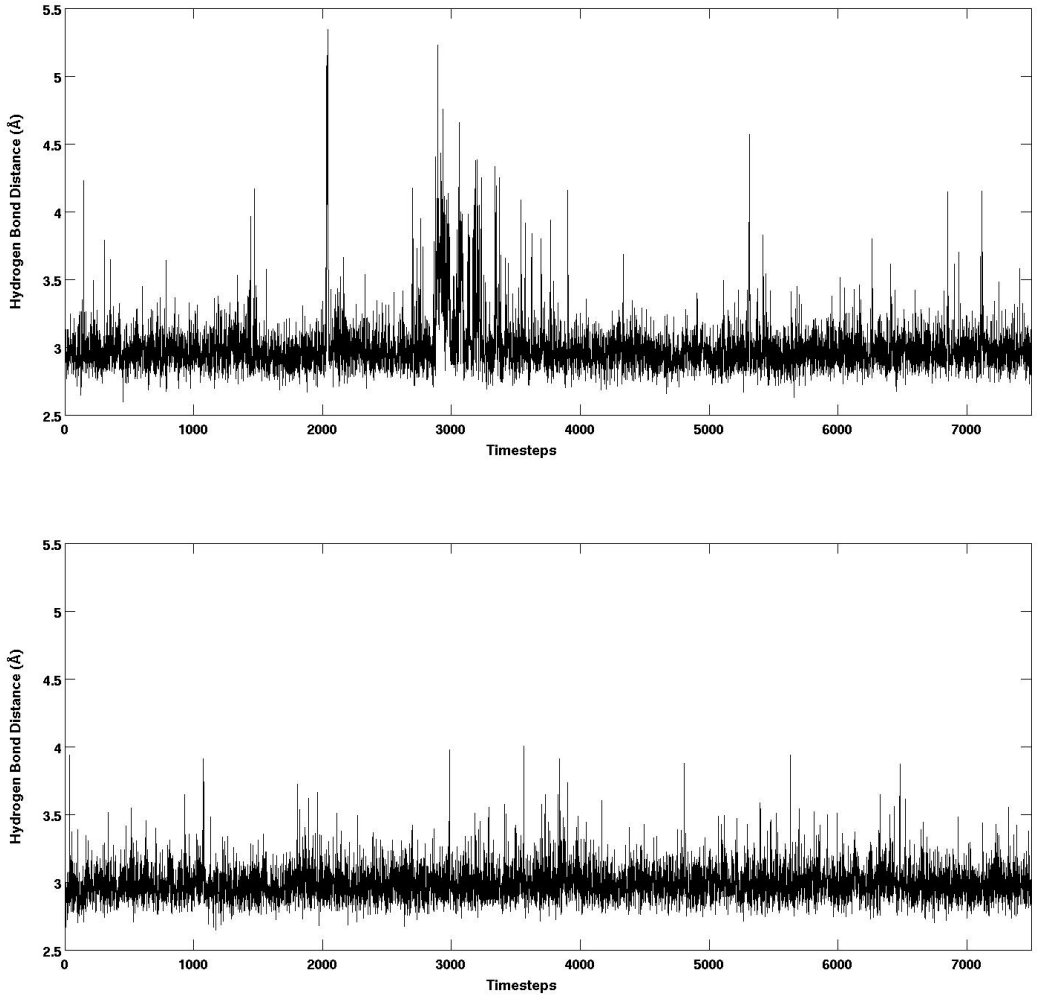


Figure S6: Hydrogen bond length of the DNA base-pair at the 4 position in the dodecamer sequence for the modified $CA_6$2-$GT_3$2 DNA duplex in 0.10 M salt (top) to 1.0 M salt (bottom).

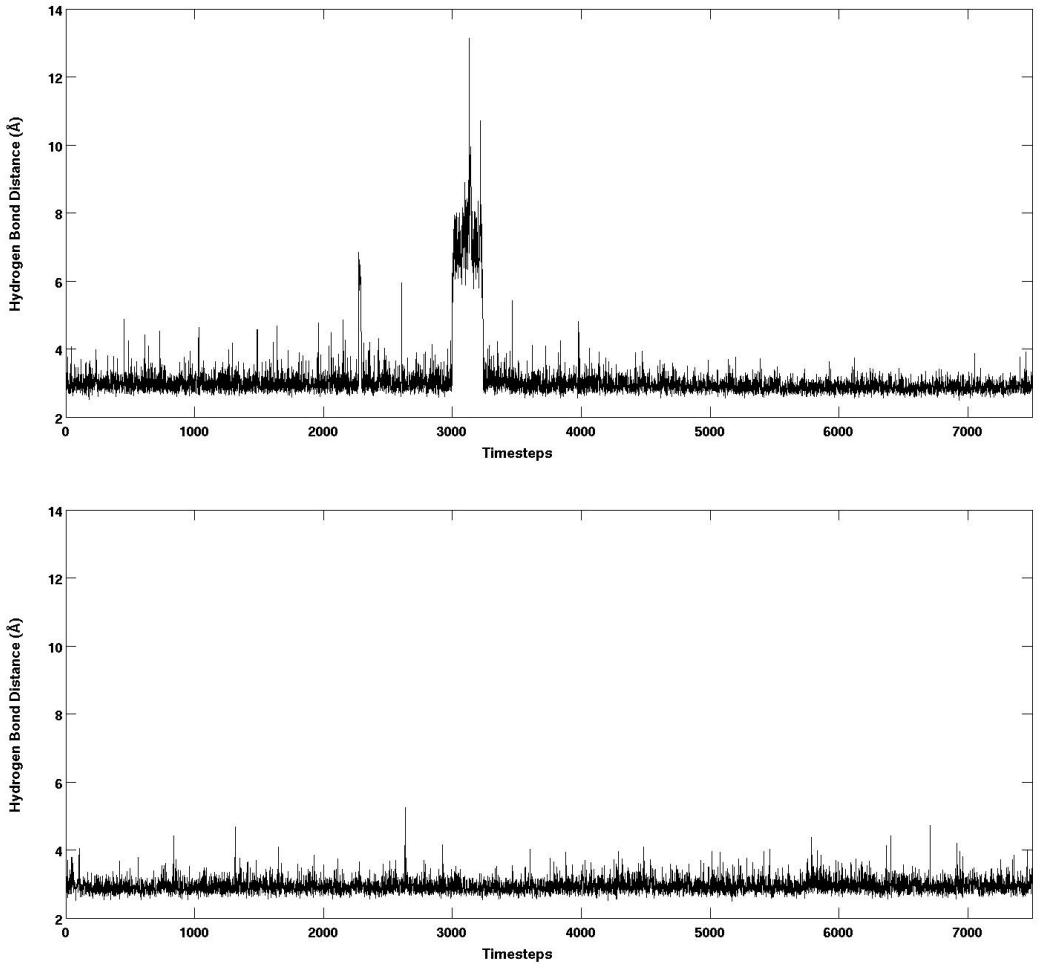


Figure S7: Hydrogen bond length of the DNA base-pair at the 3 position in the dodecamer sequence for the modified $GA_6$2-$CT_3$2 DNA duplex in 0.10 M salt (top) to 1.0 M salt (bottom).

Chapter Five

Designing the secondary structure and predicting the Mg^{2+}-induced folding free energy of the 58mer ribosomal fragment

1. Introduction

RNA molecules generally adopt a set of partially folded conformations containing only secondary structure when Mg^{2+} is absent from the ion atmosphere. We call these RNA structures the intermediate or I state, which consist of a potential broad distribution of many different conformations with similar energies.[1] Upon addition of Mg^{2+} the native tertiary structure, N state, is strongly stabilized relative to the I state. This preference to form tertiary contacts and fold into the native structure implies that the N state interacts more favorably with the Mg^{2+} compared to the I state.[2] While the focus of many investigations on this difference in RNA-ion interaction free energies has been centered around strong interactions of ions with folded RNAs[3-5], some studies have found strong interactions of Mg^{2+} with partially unfolded RNAs.[6-8] Therefore we should consider how to accurately describe the I state RNA ion atmosphere when investigating the effects of Mg^{2+} on RNA stability.

Draper and co-workers directly address the important consideration of the I state ion atmosphere in their studies on the 58-nucleotide fragment of rRNA.[1] In this work, an I state secondary structure is designed by making a mutation that disrupts an important tertiary interaction, shown in Figure 1. The tertiary structure, or N state of the 58mer, also referred to as U1061A (Figure 1, left) is stabilized by the noncanonical hydrogen bonding between bases and a number of tertiary interactions.[9]

The three-helix junction I state, or A1088U (Figure 1, right) has the potential to adopt extended, I_E, or compact, I_C, conformations. Upon reaching a sufficient concentration of Mg^{2+}, c_{2+}, the compact I state conformation folds to the N state. The folding process of I_E ➔ I_C ➔ N is depicted in Figure 2.

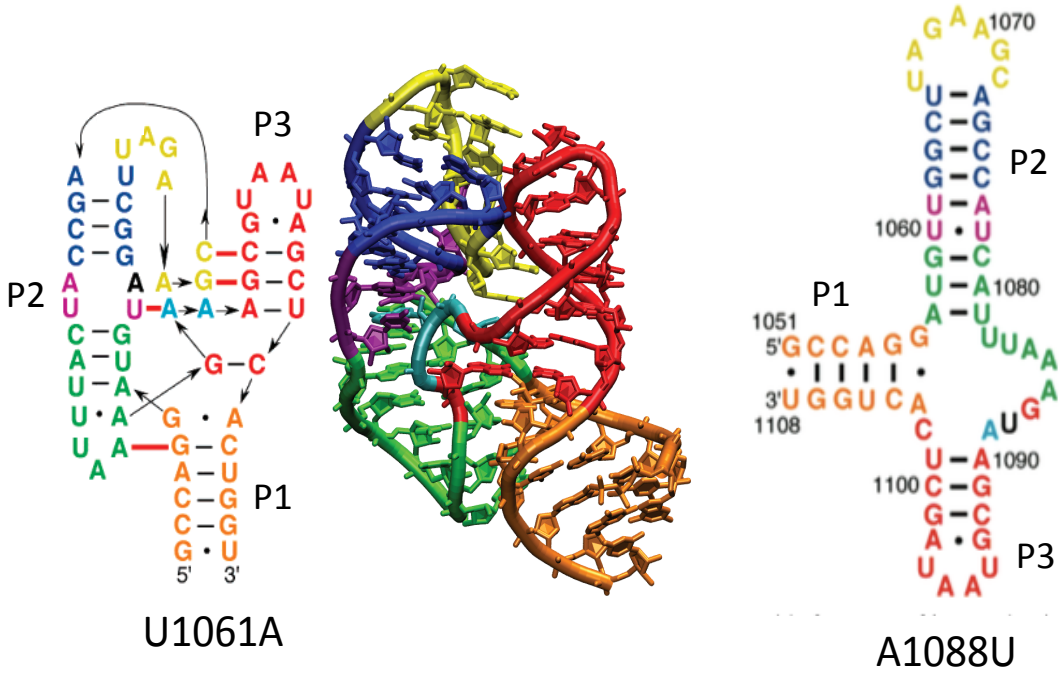


Figure 1: Variants of the 58-nucleotide fragment of large subunit ribosomal RNA from *E. coli* (nucleotides 1051-1108) used in this study. Mutations from the *E. coli* sequence are colored black. Left: U1061A RNA drawn to indicate base pairs (black horizontal lines) and tertiary base-base hydrogen bonding (red bars) in the folded RNA.[1,9] Center: 3-D conformation of the U1061A tertiary structure.[9] Right: A1088U RNA, in which the A1088 - U1060 tertiary interaction has been disrupted, drawn with the secondary structure used in modeling of I state conformations.[1]

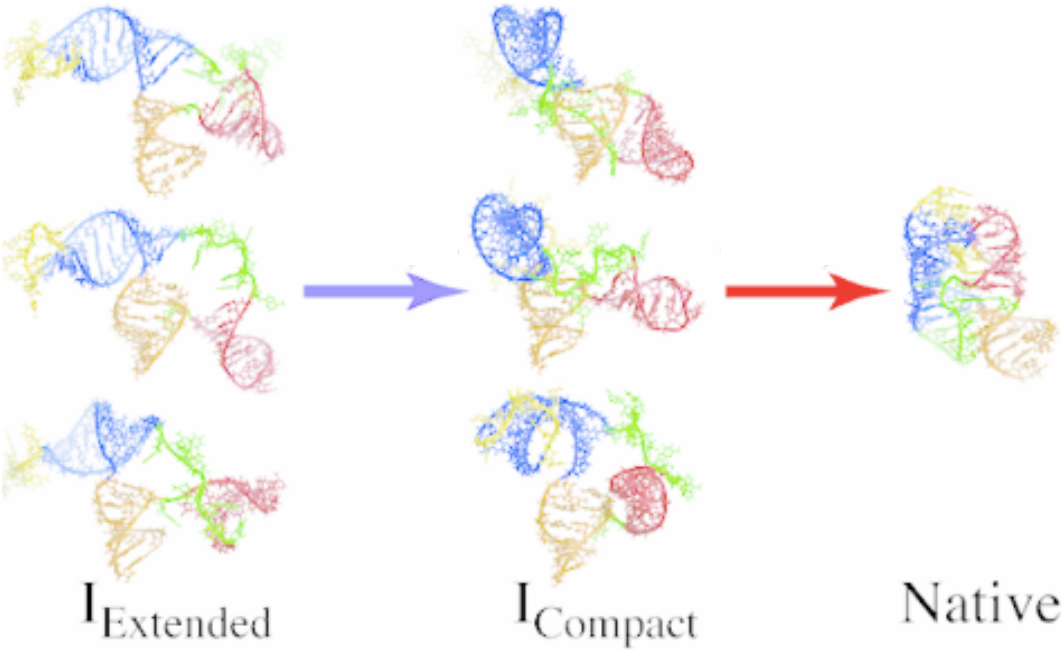


Figure 2: Schematic of possible extended I state conformations which adopt I state compact conformations that then fold into the N state tertiary structure.[1]

The free energy associated with going from the I-state RNA in absence of Mg^{2+} to the N state in the presence of Mg^{2+} is defined as $\Delta\Delta G_{Mg2+}$. A thermodynamic cycle can be constructed to calculate $\Delta\Delta G_{Mg2+}$ two different ways for the 58mer fragment from experimental results and is shown in Figure 3. Using the free energies shown in Figure 3, $\Delta\Delta G_{Mg2+}$ is defined as[7]:

$$\Delta\Delta G_{Mg^{2+}} = \Delta G^{o}_{obs,Mg^{2+}} - \Delta G^{o}_{obs,0} = \Delta G_{N-Mg^{2+}} - \Delta G_{I-Mg^{2+}} \quad (1)$$

The free energies reported in Figure 3 are determined experimentally using a titration method. The Mg^{2+} interaction with the RNA can be quantified as Mg^{2+} is titrated into a solution by making use of a fluorescent dye and integrating the fluorescence titration curve.[7]

More specifically, the fluorescent chelator dye measures a thermodynamic parameter, Γ_{2+}, known as the number of excess ions that associate with an RNA.[1,7-8,10] The excess ion atmosphere, Γ_{2+} can also be calculated from simulations.[11,12] The Mg^{2+}-RNA interaction free energy, ΔG_{Mg2+} for a particular 58-mer I or N state, is directly related to the excess Mg^{2+}, integrated as a function of the Mg^{2+} concentration.[7]

$$\Delta G_{Mg^{2+}} = -k_B T \int_0^{c_{2+}} \Gamma_{2+} d\ln c'_{2+} \quad (2)$$

In this chapter, our goal is to correctly predict Γ_{2+} from simulations for the 58mer RNA secondary structure and tertiary structure, over a range of Mg^{2+} concentrations. If this is achieved, one can calculate ΔG_{Mg2+} for the N state and I state by integrating Γ_{2+} over the range of Mg^{2+} concentrations, and determine a predicted $\Delta\Delta G_{Mg2+}$ to compare to experiment by subtracting the I state ΔG_{Mg2+} from the N state ΔG_{Mg2+}.

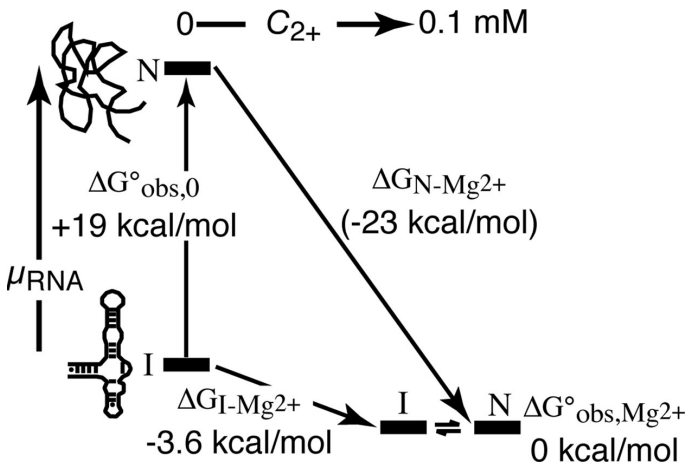


Figure 3: Thermodynamic cycle and energy level diagram for Mg^{2+}-induced folding of a 58-mer rRNA fragment. The cycle separates the folding reaction (vertical arrows) from Mg^{2+} association (sloping horizontal arrows), where I is the partially folded state of the RNA and N represents the native structure. Three of the free energies are derived from experiment; the one in parentheses is calculated from the other three.[7]

There has yet to be an accurate theoretical predication for the free energy of RNA folding, $\Delta\Delta G_{Mg2+}$, for the 58mer ribosomal fragment reported in the literature because of the lack of conformational information for the I states. This serves as the motivation for our work presented in this chapter, where we design a secondary structure I state and utilize our generalized Manning condensation model to predict $\Delta\Delta G_{Mg2+}$.

2. Methods

Secondary structure design

The P1 (Figure 1, orange nucleic acids) and P3 (Figure 1, red nucleic acids) domains of the 58mer RNA have the same base-pairing and conformation in both the secondary and tertiary structures. To model the secondary structure P1 and P3 domains, we make use of the crystal structure information for the tertiary structure (PDB ID: 1HC8).[9] We designed the P2 domain of the secondary structure using a nucleic acid builder for modeling RNA structure.[14-16] Figure 4 shows the conformation and base-paired interactions of the P2 domain that we designed for the secondary structure. The P2 structure file generated from the nucleic acid builder was then stitched together with the P1 and P3 structure information from the tertiary crystal structure to yield the secondary structure I state. The result of the design is a single stranded helical RNA that corresponds to a topology file with all of the secondary structure contacts.

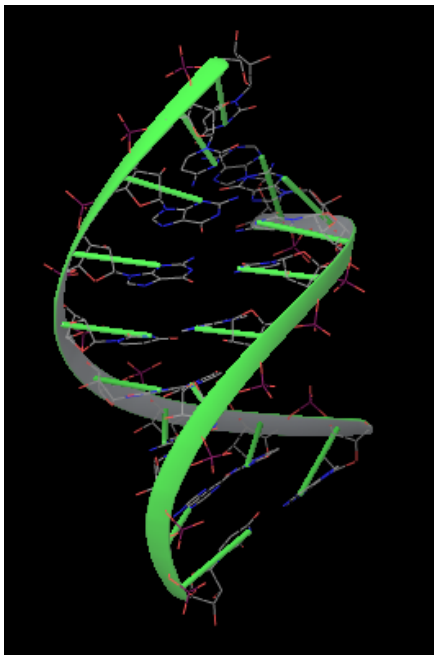


Figure 4: Design of the P2 domain for secondary structure using the Nucleic Acid Builder.[14-16]

Theoretical model and simulation

We make use of a coarse-grained all heavy atom structure-based model of RNA to capture native basin fluctuations, where the theoretical base in the energy landscape is designed to be the crystal structure (see Appendix for details). Mg^{2+} is treated explicitly in the model to account for ion-ion correlations, while KCl condensation is described implicitly by the generalized Manning model. The model takes the folded RNA to be a compact and irregular structure with varying counterion condensation on each phosphate. This varying condensation can differ from one phosphate to the next, and can be dynamic with fluctuations from one trajectory to the next. The model addresses the inherent electrostatic heterogeneity of the phosphates by accounting for these counterion condensation differences.[17] Near the RNA surface there are ion inaccessible volumes due to the excluded volume of the RNA. Our model accounts for this inaccessible volume by preventing implicit screening ions from condensing and occupying this space.[17]

3. Results and Discussion

Figure 5 shows the experimental and theoretical results for the 58mer tertiary structure. The model is able to reproduce Γ_{2+} at several KCl concentrations. At lower Mg^{2+} concentrations an inflection point appears in the experimental data and below this Mg^{2+} concentration the model no longer predicts accurate Γ_{2+} for the 58mer N state. We propose that at low concentrations of Mg^{2+} the N state tertiary interactions are not stabilized and the 58mer adopts the I state conformation. Therefore, the Γ_{2+} predicted by our model at low Mg^{2+} concentration will be inaccurate as the 58mer is in a different native basin than the one designed around the N state.

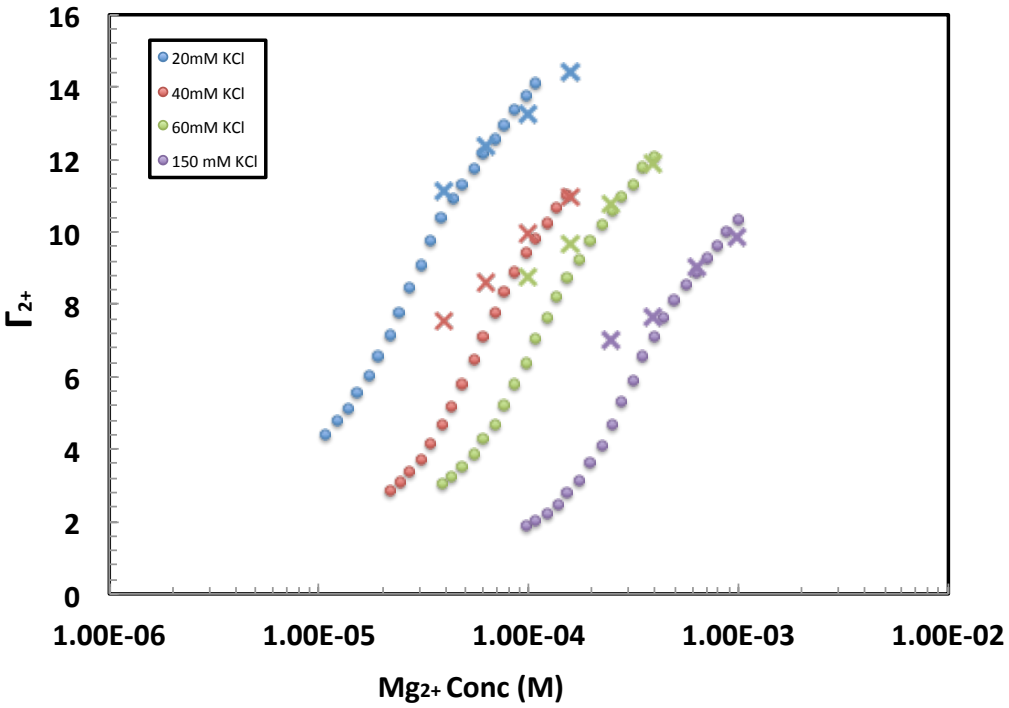


Figure 5: The generalized Manning condensation model captures Mg^{2+} over a range of concentrations for Mg^{2+} at four different KCl concentrations. The experimental results[1] are plotted as dots and simulation results are plotted as x.

Our design of the I state enables us to predict 58mer fragment excess ion atmosphere at lower Mg^{2+} concentrations using the model for a new native basin. The results for Γ_{2+} of the I state over a range of Mg^{2+} concentrations at 60 mM KCl are shown in Figure 6. The N state Γ_{2+} prediction is included for comparison. The I state experimental Γ_{2+} values are ~1 Mg^{2+} ion higher than the Γ_{2+} for the 58mer ribosomal fragment predicted by our model.

Our reasoning for the lower predicted Γ_{2+} is that we begin the simulation in the extended helix conformation, and much of the simulation time is spent folding into the secondary structure conformation. This is depicted in Figure 7 with the extended single strand helix starting structure on the left. On the right is one of the folded I state conformations that we see towards the end of the simulation. We propose that starting a new production simulation using the folded 58mer secondary structure will result in predicted Γ_{2+} values closer to those reported from experiment. This work is ongoing in our lab, and we are also studying ion chelation effects on structure stabilization. We hope that by more accurately modeling the excess ion atmosphere of the I state, we can predict the free energy of folding for the 58mer ribosomal fragment.

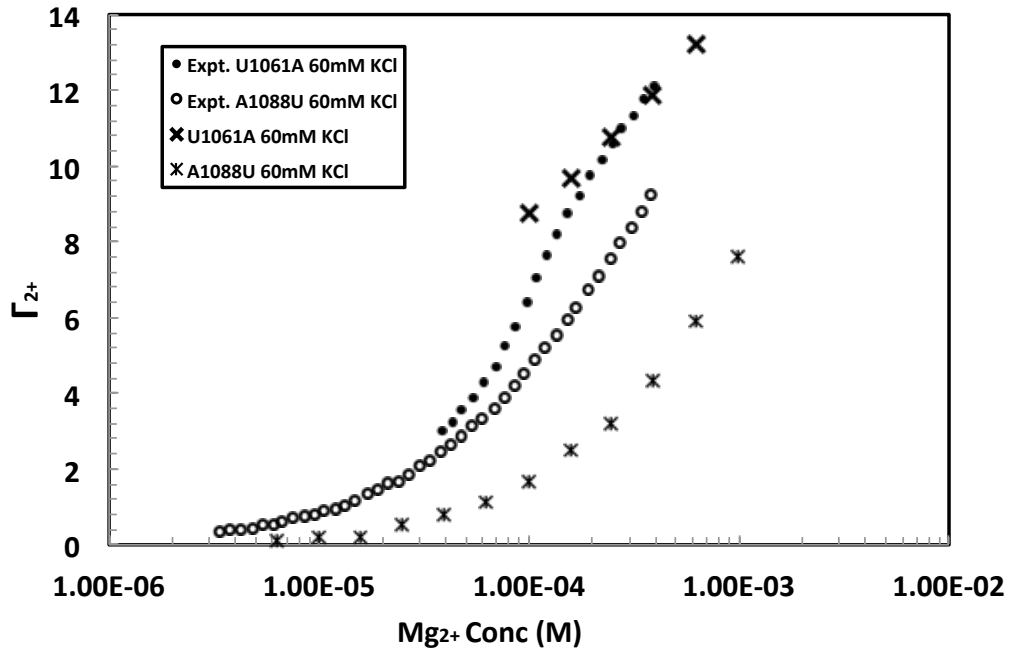


Figure 6: The generalized Manning condensation model predicts 58mer excess ion atmosphere over a range of Mg^{2+} concentrations at 60 mM KC. The experimental results[1] are plotted as open dots (I state, A1088U) and closed dots (N state, U1061A) and simulation results are plotted as * (I state) and x (N state).

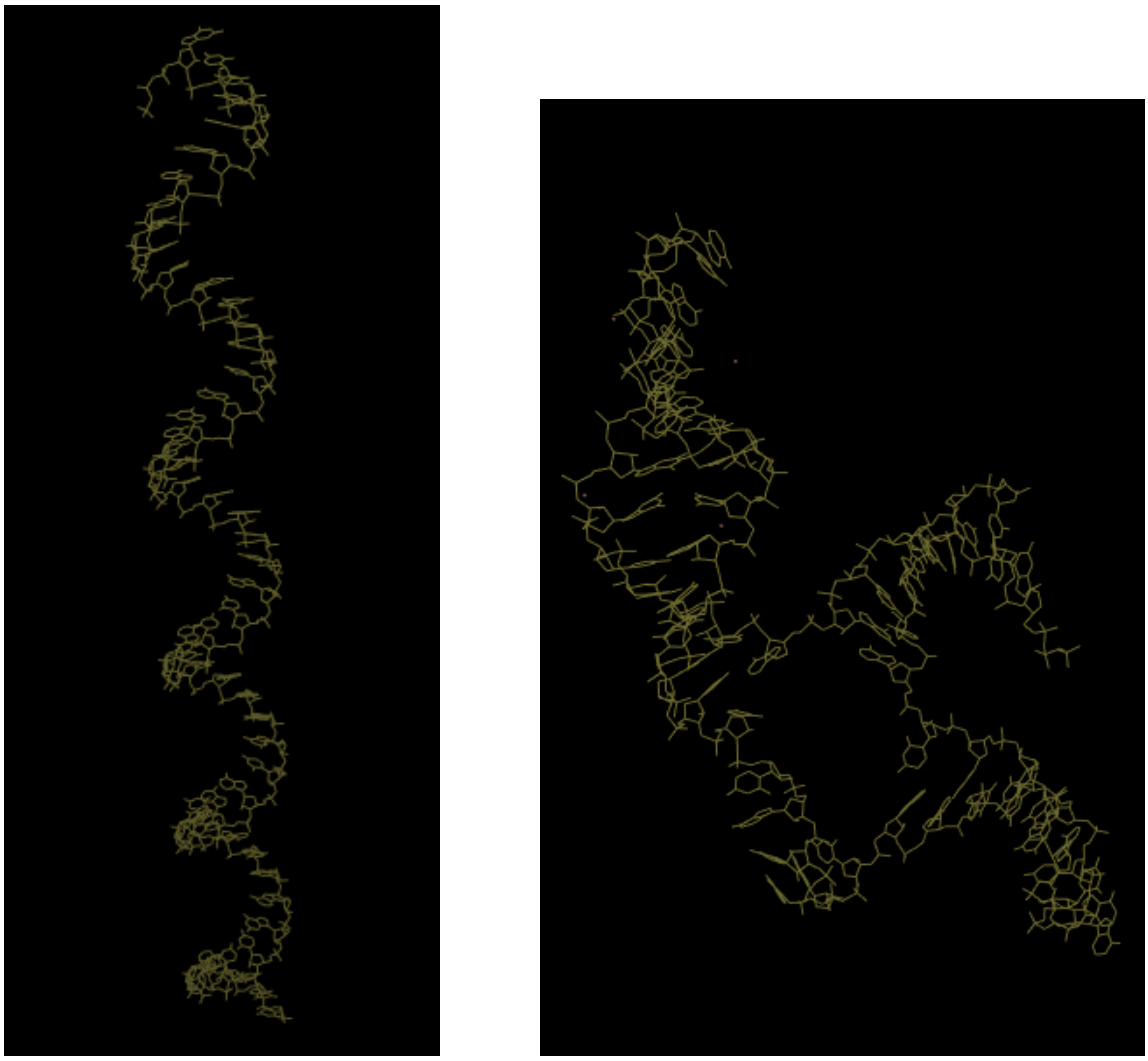


Figure 7: Secondary structure begins as a single RNA helix (left) and folds into the designed 3-helix junction I state conformation (right).

References

1. Grilley, D.; Misra, V.; Caliskan, G.; Draper, D. E. *Biochem*. **2007**, 46, 10266-10278.

2. Grilley, D.; Soto, A. M.; Draper, D. E. *Methods Enzymol*. **2009**, 455, 71-94.

3. Misra, V. K.; Draper, D. E. *Proc. Natl. Acad. Sci. USA* **2001**, 98, 12456-12461.

4. van Buren, B. N.; Hermann, T.; Wijmenga, S. S.; Westhof, E. *Nucleic Acids Res*. **2002**, 30, 507-514.

5. Petrov, A. S.; Lamm, G.; Pack, G. R. *Biopolymers* **2005**, 77, 137-154.

6. Misra, V. K.; Draper, D. E. *J. Mol. Biol.* **2002**, 317, 507-521.

7. Grilley, D.; Soto, A. M.; Draper, D. E. *Proc. Natl. Acad. Sci. USA* **2006**, 103, 14003-14008.

8. Soto, A. M.; Misra, V.; Draper, D. E. *Biochemistry*, **2007**, 46, 2973-2983.

9. Conn, G. L.; Draper, D. E.; Lattman, E. E.; Gittis, A. G. *Science* **1999**, 284, 1171-1174.

10. Leipply, D.; Draper, D. E. *Biochemistry* **2011**, 46, 10266.

11. Hayes, R. L.; Noel, J. K.; Whitford, P. C.; Mohanty, U.; Sanbonmatsu, K. Y.; Onuchuc, J. N. *Biophys. J*. **2014**, 106, 1508.

12. Hayes, R. L.; Noel, J. K.; Mohanty, U.; Whitford, P. C.; Hennelly, S. P,; Onuchic, J. N.; Sanbonmatsu, K. Y. *J. Am. Chem. Soc.* **2012**, 134, 12043.

13. Conn, G. L.; Grittis, A. G.; Lattman, E. E.; Misra, V. K.; Draper, D. E. *J. Mol. Biol.* **2002**, 318, 963-973.

14. Stroud, J. http://structure.usc.edu/make-na/index.html

15. Arnott, S.; Hukins, D. W. L.; Dover, S. D.; Fuller, W.; Hodgson, A. R. *J. Mol. Biol.* **1973**, 81, 107-122.

16. Fuller, W.; Wilkins, M. H. F.; Wilson, H. R.; Hamilton, L. D.; Arnott, S. *J. Mol. Biol.* **1965**, 12, 60-76.

17. Hayes, R. L.; Noel, J. K.; Mandic, A.; Whitford, P. C.; Sanbonmatsu, K. Y.; Mohanty, U.; Onuchic, J. N. *Phys. Rev. Lett.* **2015**, 114, 258105/1-258105/6.

Appendix

Structure-based model (SBM)

We use an all-atom SBM, which explicitly includes all heavy atoms and Mg^{2+} ions to properly represent Mg^{2+}-induced phosphate attraction, Mg^{2+}-Mg^{2+} correlation, and the dense outer-sphere Mg^{2+} population. The all-atom SBM potential, which is designed to consider fluctuations around the native basin, is given by[11]:

$$
\begin{aligned}
V_{\mathrm{SBM}} = & \sum_{i}^{\text{bonds}} \frac{\epsilon_r}{2}(r_i - r_{0i})^2 + \sum_{i}^{\text{angles}} \frac{\epsilon_\theta}{2}(\theta_i - \theta_{0i})^2 \\
& + \sum_{i}^{\substack{\text{impropers} \\ \text{or planar}}} \frac{\epsilon_\chi}{2}(\chi_i - \chi_{0i})^2 \\
& + \sum_{i}^{\substack{\text{proper} \\ \text{dihedrals}}} \epsilon_\phi F_{\mathrm{D}}(\phi_i - \phi_{0i}) \\
& + \sum_{ij}^{\text{contacts}} \epsilon_{\mathrm{C}} \left(\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right) \\
& + \sum_{ij}^{\substack{\text{RNA-RNA} \\ \text{non-contacts}}} \epsilon_{\mathrm{NC}} \left(\frac{\sigma_{\mathrm{NC}}}{r_{ij}} \right)^{12},
\end{aligned}
\tag{1}
$$

where

$$
F_{\mathrm{D}}(\phi) = (1 - \cos\phi) + \frac{1}{2}(1 - \cos 3\phi).
\tag{2}
$$

Geometric parameters (r_{0i}, θ_{0i}, χ_{0i}, ϕ_{0i}, and σ_{0i}) are set by their values in a crystal structure, so that the crystal structure is the global minimum in energy.[11] Energetic parameters of the SBM have been calibrated and are presented in reference 11. The final term of the SBM potential controls the RNA-RNA excluded volume and is given by $\sigma_{NC} = \sigma_{RNARNA} =$ 1.7 Å.[11]

The excluded volume of the Mg^{2+} ions is controlled by $V_{\text{Mg-size}}$, which is given by

$$V_{\text{Mg-Size}} = \sum_{ij}^{\text{Mg-RNA}} \epsilon_{\text{MgRNA}} \left(\frac{\sigma_{\text{MgRNA}}}{r_{ij}} \right)^{12} + \sum_{ij}^{\text{Mg-Mg}} \epsilon_{\text{MgMg}} \left(\frac{\sigma_{\text{MgMg}}}{r_{ij}} \right)^{12}. \quad (3)$$

Extensive all-atom Amber MD simulations were done to determine the excluded volume parameters: $\sigma_{\text{MgRNA}} = 3.4$ Å and $\sigma_{\text{MgMg}} = 5.6$ Å.[11]

Electrostatic calculations

The K^+ and Cl^- implicit model is implemented by making use of Manning counterion condensation theory and dividing the ion distributions into two populations: screening ions and Manning condensed ions. The screening ion density is given by a linearized Poisson-Boltzmann distribution. The condensed ion density distribution goes beyond classical Manning condensation and is generalized to account for compact and irregular structures of RNA.[17] In Debye-Hückel (DH) theory, screening ions of species *s* have a local density

$$n_{\text{DH},s}(\vec{r}) = c_s \left(1 - \frac{z_s \Phi_0(\vec{r})}{k_B T} \right) \quad (4)$$

that varies linearly with the electrostatic potential Φ_0, where c_s and z_s are the concentration and charge of ionic species *s*. Debye-Hückel electrostatics are used in modeling ion distributions far from the RNA, and Manning counterion condensation theory captures the deviations in DH electrostatics near RNA. In our model we consider the screening ion density distribution and the density of Manning condensed ions as a

sum of two normalized Gaussian distributions $P(r,\sigma)$ centered on the position of each RNA phosphate[17]:

$$n_{\mu,s}(\vec{r}) = \sum_i \mu_{is} P(|\vec{r} - \vec{r}_i|, \sigma_\mu), \tag{5}$$

$$n_{\eta,s}(\vec{r}) = \sum_i \eta_{is} P(|\vec{r} - \vec{r}_i|, \sigma_\eta), \tag{6}$$

to treat condensation on each phosphate individually. The total density of ions is then $n_{\mathrm{DH},s} + n_{\mu,s} + n_{\eta,s}$.[17] The mixing Gaussian controls mixing free energy and the size $\sigma_\mu = 0.7$ nm is set to the Bjerrum length. The hole Gaussian enforces the ion accessibility by offsetting any ions too close to the RNA and the size $\sigma_\eta = 0.34$ nm is set approximately to the closest approach of a hydrated ion to RNA. A detailed description of the calibration of these parameters is presented in the supplemental information for reference 17. The Manning condensed ions of species s at a charged atom i can then be calculated as $\theta_{is} = \mu_{is} + \eta_{is}$.

The interaction between two point charges in the model is[17]

$$\phi(r_{ij}, 0) = k_B T \frac{l_B}{r_{ij}} \exp(-\kappa r_{ij}), \tag{7}$$

while the interaction between a point and a Gaussian is[17]

$$\phi(r_{ij}, \sigma) = k_B T \sum_{a=\pm 1} \frac{-1}{2} \frac{l_B}{a r_{ij}} \exp\left(\frac{1}{2}\kappa^2\sigma^2 + \kappa a r_{ij}\right) \times \left[1 - \mathrm{erf}\left(\frac{a r_{ij} + \kappa\sigma^2}{\sigma\sqrt{2}}\right)\right], \tag{8}$$

and the interaction between two Gaussians is $\phi\left(r_{ij}, \sqrt{\sigma_m^2 + \sigma_n^2}\right)$.[17]

Therefore, the total electrostatic free energy G_E can be given in terms of the above Debye-Hückel interactions[17]

$$G_E = \frac{1}{2}\sum_{ij}\sum_{mn} q_{m,i}q_{n,j}\phi(r_{ij}, \sqrt{{\sigma_m}^2 + {\sigma_n}^2}), \tag{9}$$

and the electrostatic potential Φ is[17]

$$\Phi_m(\vec{r}) = \sum_{j}\sum_{n} q_{n,j}\phi(\vec{r} - \vec{r}_j, \sqrt{{\sigma_m}^2 + {\sigma_n}^2}), \tag{10}$$

where m and n runs over the three labels $\{0, \mu, \eta\}$, denoting points, mixing Gaussians, and hole Gaussians respectively, and i and j runs over all charged atoms. This model therefore allows for condensed ions that are modeled explicitly to interact with individual phosphates.

The local ion density around an RNA phosphate $n_{\mathrm{Mix},is}$ can be approximated by averaging $n_{\mathrm{DH},s}$ over the mixing Gaussian and adding $n_{\mu,s}$[17]:

$$n_{\mathrm{Mix},is} = c_s\left(1 - \frac{z_s\Phi_\mu(\vec{r}_i)}{k_BT}\right) + n_{\mu,s}(\vec{r}_i). \tag{11}$$

The effective volume a Gaussian occupies can be estimated as the inverse of the local Gaussian density[17]

$$V_{m,i} = 1/\sum_{j} P(r_{ij}, \sigma_m), \tag{12}$$

where the sum on j runs over all phosphates. Our model makes use of Equation 11 and 12 to reformulate the classical Manning counterion condensation mixing free energy to include screening ions by expressing it in terms of the local ion density and condensation volume. Thus, the mixing free energy is approximately[17]

$$G_{\mathrm{Mix}} = \sum_i \sum_s k_B T n_{\mathrm{Mix},is}(V_{\mu,i} - V_{\eta,i}) \ln (n_{\mathrm{Mix},is}/ec_s). \tag{13}$$

The potential ensures that $n_{\mathrm{Mix},is} \geq 0$, to avoid any negative concentration of screening ions.[17]

The electrostatic free energy G_{ES} of the screening ions is included in the model because the mixing free energy of the screening ions is included, and G_{ES} is given by[17]

$$G_{ES} = \frac{1}{2}\sum_i \sum_s z_s c_s \left(1 - \frac{z_s \Phi_\mu(\vec{r}_i)}{k_B T}\right)(V_{\mu,i} - V_{\eta,i})\Phi_\mu(\vec{r}_i). \tag{14}$$

The concentration of each ionic species $n_{\mathrm{Hole},is}$ within the excluded volume of the polyelectrolyte particle i is[17]

$$n_{\mathrm{Hole},is} = c_s \left(1 - \frac{z_s \Phi_\eta(\vec{r}_i)}{k_B T}\right) + n_{\mu,s}(\vec{r}_i) + n_{\eta,s}(\vec{r}_i), \tag{15}$$

where the screening ions have been averaged over the hole Gaussian. To enforce ion accessibility near the RNA, the ions found to be within the excluded volume of the RNA are removed by $n_{\mathrm{Hole},is} = 0$. A strong harmonic restraint[17]

$$G_{\mathrm{Hole}} = \frac{1}{2} k_{\mathrm{Hole}} \sum_i \sum_s {n_{\mathrm{Hole},is}}^2 \tag{16}$$

is added to the model potential to keep η within 0.01 ions of the correct value. To maintain stability, μ_{is} and η_{is} are weakly harmonically restrained by a term G_{Rest}. Combining equations 4-16 yields our model potential: $G_E + G_{\mathrm{Mix}} + G_{ES} + G_{\mathrm{Hole}} + G_{\mathrm{Rest}}$.[17] The four implicit condensation variables for each phosphate (μ_{i+}, μ_{i-}, η_{i+}, and η_{i-}) are treated as coordinates, and together with explicit Mg^{2+} and RNA positions, they evolve with Langevin dynamics on this potential.[17]

Acknowledgements

We thank J. Gong and M. Gibson for discussions and work performed during the initial stages of the work presented in Chapter 1. We thank and acknowledge Brian Stamm, Alex Friedman, Alex Zhuang, and Dr. Z. Guo for their help in the early stages of the work presented in Chapter 2. Thank you to Irwin Oppenheim for discussions on Chapter 3. Finally, thank you to Ailun Wang for collaborative work on Chapter 5. And we thank Ryan L. Hayes and José Onuchic for discussions on design of 58mer secondary structure and on structure based model for RNA.