

Copyright
by
Sara Yuengge Cheng
2018

The Dissertation Committee for Sara Yuengee Cheng certifies that this
is the approved version of the following dissertation:

Computational and Experimental Studies of Biomolecules

Committee:

Ernst-Ludwig Florin, Supervisor

Pengyu Ren, Co-Supervisor

Vernita Gordon

Michael Marder

Rick Russell

Computational and Experimental Studies of Biomolecules

by

Sara Yuengge Cheng

Dissertation

Presented to the Faculty of the Graduate School of
The University of Texas at Austin
in Partial Fulfillment
of the Requirements
for the degree of
Doctor of Philosophy

The University of Texas at Austin

August, 2018

For my parents

Acknowledgements

I would like to thank my supervisor, Dr. Ernst-Ludwig Florin. Dr. Florin encouraged me to explore my own ideas and to work on interdisciplinary research. I learned how to perform single-molecule and biochemical experiments. As a result, I gained a deep appreciation for the challenges of experimental research. I am very thankful for the many hours Dr. Florin spent working with me on my scientific writing and research presentations. I will always remember to start from the end and to tell a story with my research.

I would like to thank my co-supervisor, Dr. Pengyu Ren. Dr. Ren for welcoming me to his biomolecular engineering group and allowed me to work on a diverse set of projects. I appreciate Dr. Ren's willingness to work through research problems with me, and I am grateful for his guidance and mentorship throughout my graduate career. I will always remember to see the big picture and to engage in research that pushes the boundaries of computational biophysics.

I want to thank the members of my dissertation committee from the physics department, who have provided me with support and advice throughout my graduate career. I am grateful for Dr. Michael Marder's openness to discuss my ideas on research and teaching. I am grateful to Dr. Vernita Gordon for introducing me to experimental biophysics research and her encouragement throughout my time at UT.

I would like to thank my research collaborators: Dr. Rick Russell, Dr. Philip Yangyuoru, Dr. Tim Yeh, and Dr. Judy Obliosca. Dr. Russell kindly allowed me to work in his research lab, and he is a great source of knowledge for all things RNA. Dr. Yangyuoru worked tirelessly with me at the bench to prepare single-molecule samples and perform biochemical experiments. I am very grateful for his mentorship and his friendship. I am grateful to have worked with Dr. Tim Yeh, and I appreciate his advice on career paths. Dr. Judy Obliosca met with me weekly, for almost a year, to share her experimental results with me and to collaborate with me on my computational work. I admire her work ethic, and I greatly enjoyed working with her. I would also like to thank Aaron Foote for his time and patience, working with me to generate single-molecule pulling data.

I would like to thank the students and the professors at the Center for Nonlinear Dynamics (CNLD). I want to thank Dr. Harry Swinney for his leadership and vision for the CNLD. I would like to acknowledge my fellow graduate students at the CNLD, who are some of the friendliest people I have ever met. I have enjoyed working with members Dr. Florin's lab, and I am grateful for their support and friendship. I would also like to thank the lab members of the Dr. Ren's lab who are amongst the sharpest people I have ever met. My friends, I will miss seeing you all every day. I cherish our food adventures and coffee breaks.

Finally, I would like to thank my family. My parents have been my greatest supporters during my PhD studies. They instilled upon me the importance of education, and they have always encouraged me to pursue my dreams. I am grateful for their love and support, and I dedicate this thesis to them.

Computational and Experimental Studies of Biomolecules

Publication No. _____

Sara Yuengge Cheng, Ph.D.
The University of Texas at Austin, 2018

Supervisors: Ernst-Ludwig Florin, Pengyu Ren

Integrating experiments and computational modeling is critical for understanding the structure and dynamics of biomolecules. Beyond providing validation for experimental results, computational modeling, that incorporates accurate physical models and enhanced sampling methods, can provide insight into the mechanisms underlying experimental observations. I will present four projects where experiments and computational modeling were used together, to understand mechanisms underlying the structure and dynamics of biomolecules. The first project involves using enhanced sampling to improve the efficiency of calculating the hydration free energies of small molecules using a polarizable force field. These predictions are compared with a conventional free energy method, and excellent agreement is found between the methods. The second project involves using atomic molecular dynamics simulations to determine the molecular mechanism underlying the ability of nanosensor to detect point-mutations in a DNA sequence. By analyzing the nearest-neighbor hydrogen bonding profile, from simulations of the nanosensor, a molecular mechanism was proposed to explain the experimental data. The third project involves the incorporation of non-canonical hydrogen bonding in a RNA coarse-grained model in order to improve 3D structure prediction. This new model is applied to study the sequence-dependent stability of several RNAs including RNA G-quadruplexes. The final project involves the development of a new single-molecule assay to measure local transitions in nucleic acid structures using ultrashort DNA tethers. This project involves collaboration with an experimental biochemistry group to design the DNA tethers and to prepare single-molecule samples. All projects involve the development of new methods to understand the 3D structure and dynamics of biomolecules.

Contents

Acknowledgements	v
Abstract	vii
List of Figures	xii
Chapter One: Structures and Dynamics of RNA	1
RNA is Not Just a Carrier of Information	1
RNA G-Quadruplexes	1
Navigating RNA's Rugged Free Energy Landscape	3
Coarse-Grained Models of RNA Folding	3
Single Molecule RNA Experiments	4
Research Objectives and Dissertation Outline	5
Chapter Two: Improving the Accuracy and Efficiency of Free Energy Calculations	6
Introduction	6
Polarizable Force Field and Enhanced Sampling	6
Importance of Calculating Hydration Free Energy	7
Materials and Methods	8
Orthogonal Space Random Walk (OSRW)	8
Bennett Acceptance Ratio (BAR)	10
Alchemical Free Energy Cycle	10
Computational Procedure	13
Results	14
Comparison of OSRW and BAR	14
Structural Analysis of Fluorine Containing Ligand	17

Convergence and Efficiency OSRW and BAR	21
Conclusion	24
Chapter Three: Detecting Point-Mutations using a DNA Melting Probe	25
Introduction	25
Experiments and Simulations	25
Detection of Point Mutations using a DNA Nansensor	26
Materials and Methods	27
Parameterization of Duplexes	27
Incorporation of LNA	27
MD Simulation Procedure	28
Analysis Procedure	28
Results	30
Molecular Dynamics Simulations	30
Conclusion	35
Chapter Four: Development and Improvement of the RACER RNA Coarse-Grained Model	36
Introduction	36
Predicting RNA Structure and Energetics with a Coarse-Grained Model	36
Development of RACER 1.0	37
Development of RACER 2.0	37
Materials and Methods	38
5-Bead RACER Model	38
RACER 1.0 and RACER 2.0 Potentials	39
Bonded Potential Energy Functions	40
Non-Bonded Potential Energy Functions	41

Improvements of RACER 2.0	44
Software Changes the TINKER Source Code	46
Calculating Binding Energies using QChem and RACER	49
Optimizing Non-Canonical Hydrogen Bonds	51
Simulated Annealing Procedure	53
Equilibrium Pulling Simulations	53
Results	54
Validating and Testing the RACER 1.0 Model	54
3D Structure Prediction of Small RNAs	54
Folding Free Energy Prediction	56
Validating and Testing the RACER 2.0 Model	57
Adding Non-Canonical Hydrogen Bonds	57
Optimization with Non-Canonical Hydrogen Bonds	60
Testing Parameters on Small and Large RNAs	60
Energy Change Between Folded and Unfolded State	63
Testing RACER 2.0 with Pulling Simulations	64
Conclusion	67
Future Directions	67
Chapter Five: Probing Local Nucleic Acid Transitions with Ultrashort DNA Tethers	68
Introduction	68
Motivation for Developing Complementary Single Molecule Experiments	68
Free Energy Landscapes of Nucleic Acids	68
Notation used in the Chapter	69
Modifications to Experiments for RNA studies	69

State of the Art RNA Experiments	70
Ultrashort Tether Single Molecule Assay	70
Motivation from Biochemical Assay	72
Materials and Methods	72
Biochemical Assay	72
Single Molecule Force Experiments	73
Single Molecule Pulling Procedure	73
Geometrical Amplification Effect	75
Surface Chemistry of the Ultrashort Tether Assay	76
Design of Ultra-Short DNA Tethers	77
Results	79
Testing for the Formation of Ultra-Short DNA Tethers	79
Detecting Specific vs Non-specific Binding	84
Pulling Experiments with Ultrashort DNA Tethers	85
Conclusion	87
Future Directions	88
Appendix A: Chapter 2: Chemical Structures	89
Chemical Structures and Absolute Hydration Free Energies	89
Appendix B: Chapter 5: Secondary Structures of Ultra-short DNA tethers	91
MFold Predictions	91
Bibliography	94

List of Figures

1.1	RNA G-Quadruplex Structure	2
2.1	Hydration Free Energy Alchemical Free Energy Cycle	9
2.2	BAR Free Energy Alchemical Free Energy Cycle	11
2.3	Comparison of Hydration Free Energies from BAR and OSRW Methods	15
2.4	Hydration Free Energy for Selected Compounds Calculated using OSRW	16
2.5	Hydration Free Energy for Selected Compounds Calculated using BAR .	17
2.6	Radial Distribution Functions from BAR simulations	18
2.7	Radial Distribution Functions from OSRW simulations	19
2.8	Sampling Efficiency of λ and F_λ Using the OSRW Method	21
2.9	Free Energy Landscapes Generated Using the OSRW Method	22
3.1	Single-nucleotide polymorphism (SNP) detection using a melting temperature probe containing a single LNA thymidine monomer (t_L)	27
3.2	Molecular dynamics (MD) simulations of 12 25-bp long hybrids, each consisting of a BRAF allele and a melting probe.	31
3.3	Hydrogen-bond (H-bond) probability maps	33
4.1	RACER Model Development Flow Chart	38
4.2	RACER 5-Bead Coarse Grained Model	39
4.3	RACER Potential Energy Function: Torsion and Angle	40
4.4	Comparison of van der Waals Potential Energy Functions	41
4.5	RACER Potential Energy Functions with Distance Dependence	42
4.6	Geometry of a RACER Hydrogen Bond	43
4.7	Example Base Pairs Used for Optimizing Hydrogen Bonds in RACER 2.0	45

4.8	Example RACER 2.0 Key File	47
4.9	Numbering of RACER 2.0 Nucleotides	48
4.10	Calculating Binding Energy using QChem and RACER	50
4.11	Correlation Plot of RACER and QChem Binding Energies for Different Optimization Procedures	51
4.12	Example Umbrella Sampling Distribution for a RNA Hairpin	54
4.13	3D Structure Prediction from Simulated Annealing	55
4.14	Equilibrium Pulling Profiles for small RNAs	56
4.15	Equilibrium Pulling Results with RACER Model	57
4.16	Adding Non-Canonical Hydrogen Bonds to RACER 2.0	58
4.17	RACER Coarse-Grained Representation of Seven Large RNAs	59
4.18	Difference Plot of Binding Energies for small RNAs	61
4.19	Difference Plot of Binding Energies for Large RNAs	62
4.20	$\Delta\Delta E$ with Optimized Non-Canonical Hydrogen Bond Parameters	64
4.21	RACER Model for RNA-G4s	65
4.22	RACER pulling simulations of RNA-G4	66
5.1	Surface Chemistry of Single Molecule Assay	71
5.2	Measurement of DNA G4 (hTelo) and RNA G4 (Terra) lifetimes	74
5.3	Geometrical Amplification Effect for a Bead Tethered Near the Surface	76
5.4	Design for 26nt Two State (2s-2s) DNA Hairpin with Two 13bp DNA Tethers	77
5.5	Design for 74nt Two State (2s-50) DNA Hairpin with Two 25bp DNA Tethers	78
5.6	Design for 94nt Two State DNA (2s-100) Hairpin with Two 50bp DNA Tethers	79
5.7	Native PAGE Showing Formation of 26nt Two State (2s-26) DNA Hairpin Tether	81

5.8	Native PAGE Showing Formation of 74nt Two State (2s-50) DNA Hairpin Tether	82
5.9	Native PAGE Showing Formation of 94nt Two State (2s-100) DNA Hairpin Tether:	83
5.10	2D Position Histogram of a Loosely Bound Bead	84
5.11	Optical Trapping Geometry for Ultrashort Tether Assay	86
5.12	Position histogram for unfolding event of 26nt Two State DNA Hairpin with Extensions.	87
A.1	Calculated Hydration Free Energies of 20 compounds	90
B.1	Mfold Prediction for 26nt Two State DNA Hairpin extended sequences for DNA tethers	91
B.2	Mfold Prediction for 74nt Two State DNA Hairpin extended sequences for DNA tethers	92
B.3	Mfold Prediction for 94 nt Two State DNA Hairpin extended sequences for DNA tethers	93

Chapter One: Structures and Dynamics of RNA

1.0.1 RNA is Not Just a Carrier of Information

RNA is a highly charged biopolymer that has been traditionally viewed as an information carrier in the cell. RNA monomers, or nucleotides, are composed of a negatively charged phosphate, a ribose sugar, and a nitrogenous base. RNA bases come in four types, adenine, guanine, cytosine, and uracil. According to the central dogma of molecular biology, RNA is an intermediate molecule that is transcribed from DNA and translated into protein. Evidence of RNA's role beyond a carrier of information can be found in the diversity of complex three-dimensional RNA structures found in nature. These RNA structures range from a simple RNA hairpin, which is a single strand of RNA that folds onto itself, to a protein-like RNA riboswitch, a RNA sensor that re-folds in the presence of small molecules.

A common feature in these RNA structures is a network of hydrogen-bond patterns that allow the RNA bases to form base pairs with other RNA bases. These hydrogen bonds can form across the Watson-Crick, Hoogsteen, or Sugar edges of an RNA base pair. These base pairs can be categorized into canonical and non-canonical base pairs. Canonical base pairs are formed across the Watson-Crick faces of complementary nitrogenous bases. Three hydrogen bonds are formed in a canonical guanine-cytosine base pair, and two hydrogen bonds are formed in a canonical adenine-uracil base pair. Non-canonical base pairs are all other combinations of base pairs with any other combination of edges. An example of a 3D RNA structure that contains non-canonical hydrogen bonds is an RNA G-quadruplex.

1.0.2 RNA G-Quadruplexes

RNA G-quadruplexes can form in a RNA sequence containing a series of guanines separated by a sequence containing any other nucleotides except guanine, $(GG)_x(N)_y$. The series of guanines, also known as g-runs, assemble into planes called g-quartets. These g-quartets are composed of two pairs of guanines which form a total of eight hydrogen bonds across their Watson-Crick and Hoogsteen edges. Due to the high

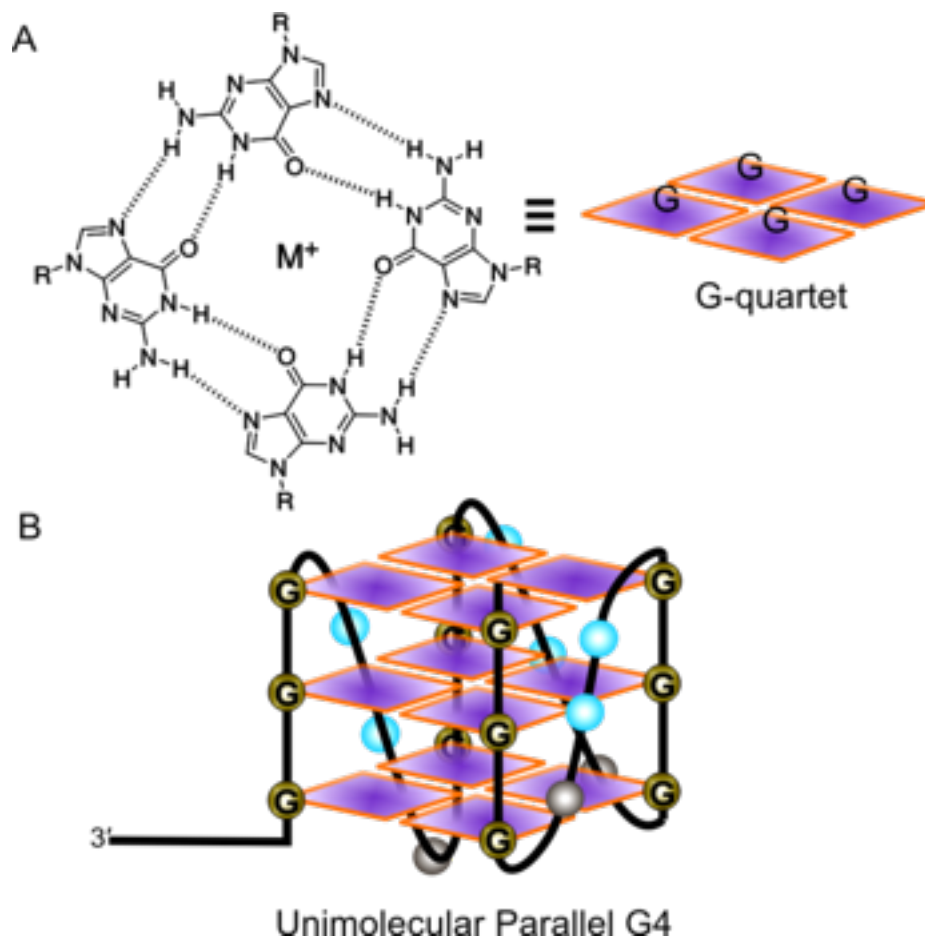


Figure 1.1: **RNA G-Quadruplex Structure:**A. Guanine bases can form Hoogsteen base pairs in a G-quartet, which is stabilized by a monovalent cation (M^+). B. G4 structures (G4s) form by stacking of multiple G-quartets, with the sequences between the runs of G forming loops. RNA favors parallel structures, with all of the G runs sharing a 5' to 3' orientation.

concentration of negatively charged oxygens at the center of the g-quartet, these g-quartets often need a positively charged ion to reside near the center of the quartets in order to neutralize the charge. These g-quartets can stack on top of each other, and the non-guanine nucleotides in the RNA sequence form loops which connect the g-quartets to each other (see Fig. 1.1).

Approximately 375,000 RNA sequences in the human genome have the potential to form RNA G-quadruplexes, which is 10-fold more than expected by chance.[1] Misregulation of structure and dynamics in RNA G-quadruplexes (RNA-G4s) has

been linked to human diseases including Xeroderma pigmentosum, Fragile X syndrome, and several cancers.[1–3] Additionally, these sequences are overrepresented at important regions of the genome including at gene promoters, near exon-intron junctions, and at the ends of chromosomes.[4, 5] In RNA, these G4s are thought to form and be resolved dynamically. Therefore, the formation and disruption of G4s promote switch-like behavior in key steps of gene expression including messenger RNA (mRNA) splicing, polyadenylation, and translation.[6–8] At present, the mechanisms underlying the folding of RNA-G4s from any given sequence to 3D structure are not well understood.

1.0.3 Navigating RNA's Rugged Free Energy Landscape

Predicting the native fold of a RNA sequence is challenging. RNAs can fold locally to form stable secondary structures, allowing the RNA to form stable intermediate folds. As a result, RNA has a rugged free energy landscape with high transition barriers separating intermediate structures. The folding process occurs on a wide range of dynamics occur on a wide range of time scales from picoseconds to seconds.[9] In the picosecond to nanosecond timescales, fine structural rearrangement can occur including sugar puckering and bond rotations and vibrations. At the microsecond timescales basepairs can open up in a RNA secondary structure. At the millisecond to seconds timescales, ions can bind/unbind from RNA structure and RNAs can undergo large scale conformation changes.[10] Coarse-grained modeling is a promising direction to achieve microsecond to millisecond timescales in MD simulations, allowing for studies of RNA folding approaching experimentally relevant timescales.[11]

1.0.4 Coarse-Grained Models of RNA Folding

Coarse-graining is the process of selecting groups of atoms from a molecule of interest, where each group is called a bead or pseudoatom of a certain type. For these beads, a set of potential energy functions is devised, which govern the interactions between those groups.[11] Development of coarse-grained RNA models fall into two categories: theory based potentials a “bottom up approach” and knowledge based potentials a “top down approach”.[12] Theory based potentials use physics-based potential energy functions to approximate the behavior of pseudoatoms with each

other. As a result of reducing the degrees of freedom in the RNA, the dynamics of the RNA are accelerated and the free energy landscape less rugged as compared with an all-atom of RNA. Several theory based potentials have been developed and used to predict 3D structures of RNAs and reproduce thermodynamic results.[13–22] Knowledge based potentials derive parameters from the potential of mean force of distributions derived from experimental structures.[23, 24] These models have been used to study ion effects in RNA and folding processes in RNAs.[14, 15, 25–35] The advantage of knowledge-based potentials is that they include local thermodynamic information about experimental RNA structures. However, these structure represent the RNA sequences than can be crystallized or can be measured at high concentrations in NMR experiments, which can bias the predictions of the any RNA model. Evaluating the accuracy of RNA models can be done by predicting a 3D structure, with little or no structural information, from a sequence and comparing these predictions with experimentally solved structure.[36–38]

Development of RNA coarse-grained models requires thorough testing to evaluate the strengths and weaknesses of the model. Small RNA structures such as RNA hairpins are widely used to benchmark the stability of RNA computational models.[39] RNA hairpins and duplexes are good test structures, because there are libraries of thermodynamic melting data for a wide range of sequences available.[40, 41] Recently, G-quadruplexes have been used to benchmark the performance of nucleic acid models.[42–44] While validation of experimentally solved structures is important, the ultimate goal of developing a computational model is to make predictions about the structure and dynamics of biomolecules that can complement experiments and possibly inspire and lead to future experiments.

1.0.5 Single Molecule RNA Experiments

Single molecule force spectroscopy on RNAs can reveal transition pathways, free energy landscapes, and kinetics of folding.[13, 45–53] These experiments can be performed under equilibrium and non-equilibrium conditions of the molecules, and they can be used to reconstruct free energy.[54] Recent work on RNA G4s indicates that single molecule fluorescence and mechanical unfolding experiments can uncover folded, partially folded, and unfolded states and the effects of salt on stabilizing

G4s.[55–63] Studies using single molecule Förster resonance energy transfer (smFRET) have shown that near-zero force is a reasonable way to probe spontaneous and protein-mediated structural transitions of G4s.[59, 62, 64, 65] A recent work using a combination of smFRET and magnetic tweezers has shown that in very low force regimes, subtle transitions could be measured in DNA-G4s.[57] In single-molecule force experiments, compliant kilobase long DNA tethers are used to to apply force to the single molecule being studied.[54] While these long DNA tethers provide accurate measurements of rates, these long DNA tethers are not ideal for free energy landscape reconstruction because they can miss short-lived intermediate states.[54, 66–68] Recent theoretical work suggests that shorter DNA tethers, that are stiffer than long DNA tethers, provide improved free energy landscape reconstruction.[66, 69–72]

1.0.6 Research Objectives and Dissertation Outline

The goal of this work is to address the three problems of biomolecular modeling: sampling, force field, and experiments.[12] Overcoming the rugged free energy landscape of biomolecules requires developing methods that accurately and efficiently search the free energy landscape of biomolecules. Accurate description of the interactions of the biomolecule with itself and its environment require force fields with parameters that are transferrable i.e. describe the behavior of a biomolecule in a variety of environments. Finally, biomolecular modeling should be complementary to experiments providing insight into the molecular mechanisms underlying experimental observations. Chapter 2 introduces the integration of an enhanced sampling method with the Atomic Multipole Optimized Energetics for Biomolecular Applications (AMOEBA) polarizable force field to calculate the free energy of small organic molecules. Chapter 3 demonstrates the ability of MD simulations to provide a mechanism explain ability of a locked-nucleic acid to enhance detection point mutations in DNA sequences. Chapter 4 describes the development and improvement of the RACER RNA coarse-grained model which can be used to predict 3D structures and folding free energy. Chapter 5 describes the development of a new single-molecule ultra-short tether assay that can be used to probe local interactions in nucleic acids.

Chapter Two: Improving the Accuracy and Efficiency of Free Energy Calculations

2.1 Introduction

2.1.1 Polarizable Force Field and Enhanced Sampling

Improving the accuracy of force fields and the sampling efficiency allows for comparison of computational predictions with experimental results. The AMOEBA polarizable force field treats electrostatics with atom centered multipoles up to quadrupole, allowing for explicit treatment of polarization by point dipole induced dipole scheme. However, the improved accuracy of the AMOEBA force field comes at a higher computational cost. In order to improve the efficiency of free energy calculations performed with the AMOEBA force field, an enhanced sampling method Orthogonal Space Random Walk (OSRW) was implemented. The advantage of this sampling method is that sampling is accelerated continuously along two orthogonal dimensions, smoothing the free energy landscape and driving the system to sample more states. In order to test the accuracy and efficiency of the OSRW method, hydration free energies for 20 small organic molecules are calculated using an alchemical free energy approach.

Alchemical free energy approaches take advantage of the fact that free energy is a state function, thus the path is not important. As a result, alchemical or “unphysical” paths can be used to calculate free energies, as long as the sum of these unphysical paths result in physical initial and final states. These physical states are the gas-phase ligand and the hydrated ligand. In order to test for consistency, the hydration free energies for the same molecules are also calculated using the Bennett Acceptance Ratio (BAR), a perturbative free energy method that samples discrete states along an alchemical free energy path. This work shows that combining enhanced sampling techniques with polarizable force fields allows for accurate and efficient free energy calculations.¹

¹This work was previously published.[73] J.A. and I contributed equally to this work. I parameterized all of the 20 small molecules, performed and analyzed MD simulations, analyzed the structures and energies, and co-wrote the paper.

2.1.2 Importance of Calculating Hydration Free Energy

Water, a substantial component of living organisms, provides an environment where biological processes such as the transportation of ions, the folding of proteins, and the activation/deactivation of signaling pathways, can take place. The interactions between water and physiologically relevant molecules, such as monoatomic ions, small molecules, and macromolecules, are crucial to our efforts of understanding a multitude of biological processes and applications such as protein engineering and drug discovery. Therefore, accurately modeling the hydration process is arguably the first step in modeling these biological processes and developing accurate physical models and robust computational approaches. For instance, the hydration free energy (HFE) is not only a key property in predicting the solubility of organic molecules and their binding to proteins,[74–77] it is also an important measure in the development and evaluation of the accuracy of force fields[78–83] and sampling methods.[84–89] The of a molecule HFE can be calculated by using explicit solvent models, e.g. TIP3P water[90] and AMOEBA water,[91] in combination with alchemical approaches, such as thermodynamic integration (TI) (see review by Kollman[74]), BAR,[92] or OSRW.[77, 93–96] Once the force field is well defined, the accuracy and precision of the alchemical results will become predictable.

Although the importance of including explicit polarization in molecular modeling have been demonstrated in previous studies,[97–99] the routine application of polarizable force fields, such as AMOEBA,[91, 100–103] to obtain accurate thermodynamic properties is still hindered by the high computational cost of traditional alchemical approaches. Thus, enhanced sampling methods such as the OSRW method are more appealing in such simulations. Unlike BAR or TI, which requires a number of arbitrary, fixed order parameter λ to connect the two end states, the OSRW method[77, 93–96] described in later sections utilizes dynamic order parameters, λ , and $\frac{dU}{d\lambda}$, coupled with the metadynamics approach[104] to sample the two dimensions. Where U is the potential. In this way, the alchemical perturbation between the two end states can be performed in a single molecular dynamics simulation and improved efficiency. Previously we have demonstrated that OSRW allows efficient sampling of configurational spaces of molecular crystals.[77] In this paper, the OSRW method is implemented with the polarizable multipole based AMOEBA force field in TINKER, and applied to compute the hydration free energy of several small organic molecules.

The hydration free energy results from OSRW are compared with those computed from the conventional BAR method, which has been utilized to compute free energy of hydration and binding in combination with AMOEBA in previous studies.[78, 103, 105–111]. The results from the two approaches are in excellent agreement (RMSD = 0.49 kcal·mol⁻¹), with OSRW method showing significant advantage in computational efficiency.

2.2 Materials and Methods

2.2.1 Orthogonal Space Random Walk (OSRW)

Because free energy is a path-independent property, a common approach to calculate the free energy difference is to define a mixed potential, so that the potential functions of the two end states of interest can be connected analytically. Such a mixed potential is defined in Equation 2.1. Here U is the mixed potential, r is the coordinate, and the scaling parameter $\lambda = 0$ and 1 corresponds to the two end states, U_0 and U_1 , respectively. The free energy change from one state to the other can thus be given as Figure 2.2, where G is the free energy of each state, ΔG is the change in free energy, and $\langle \rangle_\lambda$ is the ensemble average of each λ state.

$$U(\lambda, r) = (1 - \lambda)U_0 + (\lambda)U_1 \quad (2.1)$$

$$\Delta G = \int_0^1 \left\langle \frac{\partial U}{\partial \lambda} \right\rangle_\lambda d\lambda \quad (2.2)$$

Such construction of the approach, however, relies on the assumption that sufficient conformational sampling can be done as the system adjusts to the new intermediate states. For complex systems, such transition usually requires significant amount of simulation time, especially when there are larger changes in structure. This is often known as the ‘‘Hamiltonian lagging’’ problem, which exists for methods where λ is a continuous and dynamic variable.[112] For methods that perform simulation at discrete λ ‘‘windows’’, a large number of intermediate steps are required and long simulations at each step are needed to ensure sufficient equilibration and sampling at each step. Alternatively, by combining the ideas of the dynamic λ method[112, 113]

and the metadynamics method,[104] Yang and co-workers proposed an efficient free energy sampling approach, which they referred to as OSRW.[77, 93–96]

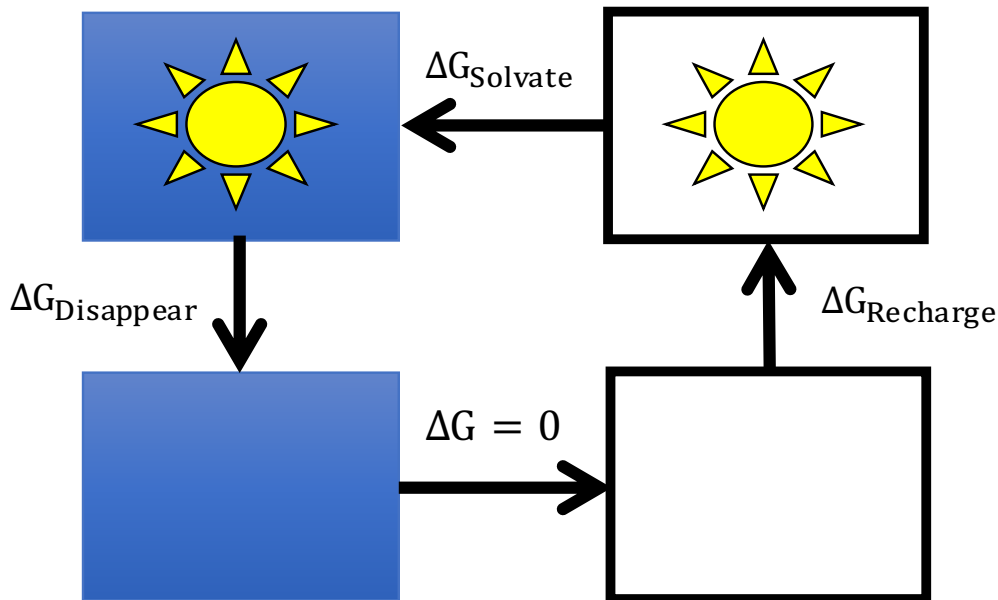


Figure 2.1: **Alchemical Free Energy Cycle:** The initial state A is the ligand (yellow) in gas phase (top right). The final state B is the ligand in solution (top left). Instead of calculating $\Delta G_{\text{Solvate}}$ directly, $\Delta G_{\text{Disappear}}$ and $\Delta G_{\text{Recharge}}$ are calculated. Using $\Delta G_{\text{Solvate}} + \Delta G_{\text{Disappear}} + \Delta G_{\text{Recharge}} = 0$, $\Delta G_{\text{Solvate}}$ can be calculated indirectly.

In this approach, a random walk is performed in two dimensions, λ and its orthogonal generalized force $F_\lambda = \frac{\partial U}{\partial \lambda}$. The use of dynamic λ itself unnecessarily improves the computational efficiency; however, directly basing along the $\frac{\partial U}{\partial \lambda}$ dimension can potentially accelerate the free energy calculation since the integral of $\frac{\partial U}{\partial \lambda}$ is exactly the free energy. By repetitively adding a Gaussian-like repulsive potential to λ and F_λ spaces, the low energy wells can be “flooded” to overcome the energy barriers. The potential of the system in the OSRW can be written (see Eq. 2.3). Here g is the biasing potential that can be defined recursively (see Eq. 2.4). Here h and w are the height and width of the Gaussian, respectively, which can be adjusted to balance the accuracy and efficiency of the method, and t_i is the index of states. The free energy along the reaction coordinates can thus be estimated as $g(\lambda, F_\lambda)$. To move from an initial state to a target state, λ , the free energy change can thus be estimated (see

Eq. 2.5). By adaptively adding a negative $G(\lambda)$ to the system potential as shown in (2.3), the “flooding” of the free energy surface can be accelerated along the λ space.

$$U_{OSRW} = U(\lambda, r) + g(\lambda, F_{lambda}) - G(\lambda) \quad (2.3)$$

$$g(\lambda, F_\lambda) = \sum_{t_i} h \exp\left(\frac{|\lambda - \lambda(t_i)|^2}{2w_1^2}\right) \exp\left(\frac{|F_\lambda - F_\lambda(t_i)|^2}{2w_2^2}\right) \quad (2.4)$$

$$\Delta G = \int_{\lambda_o}^{\lambda} \frac{\partial G}{\partial \lambda} d\lambda = \int_{\lambda_o}^{\lambda} \frac{\sum_{F_\lambda} F_\lambda \exp[\beta g(\lambda, F_\lambda)] \delta(\lambda - \lambda')}{\sum_{F_\lambda} \exp[\beta g(\lambda, F_\lambda)] \delta(\lambda - \lambda')} d\lambda, \quad (2.5)$$

2.2.2 Bennett Acceptance Ratio (BAR)

In BAR calculations, a three-step perturbation approach was applied with the AMOEBA force field.[78, 103, 105] To disappear the solute in the solvent, the electrostatic and polarization interaction were perturbed in 11 windows, scaled by λ equals (1.0, 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2, 0.1, 0.0) respectively. After the electrostatic interactions were scaled to zero, then the vdW's contribution was perturbed using 14 windows with λ equal to (1.0, 0.9, 0.8, 0.75, 0.7, 0.65, 0.6, 0.55, 0.5, 0.4, 0.3, 0.2, 0.1, 0.0) (see Fig. 2.2). 1000 ps NVT simulations at 298 K were performed at each window. The recharging of each solute in the gas phase was modeled using 11 windows (with an interval of 0.1 for λ) of 1000 ps molecular dynamics simulation at 298K, stochastic integrator, and a 0.1 fs time step. Data collected from 50 to 1000 ps range was then processed using the BAR equations. The errors for BAR results were computed as a sum of errors from the individual alchemical perturbation steps.

2.2.3 Alchemical Free Energy Cycle

A hybrid potential based on Eq. 2.1 is implemented to calculate the alchemical free energy using the AMOEBA force field in TINKER. A dual topology approach is used to keep the intramolecular energies of the mutating systems (e.g. solute molecules in solution) throughout the simulation, and the mutating systems A (initial) and B (final) never interact with each other as shown in Fig. 2.1. In case of absolute hydration free energy calculations performed in this study, the two topologies are solute-in-water and water without solute, respectively. In this approach, the mixed potential can be written, with λ as a scaling factor on the non-bonded potentials (see

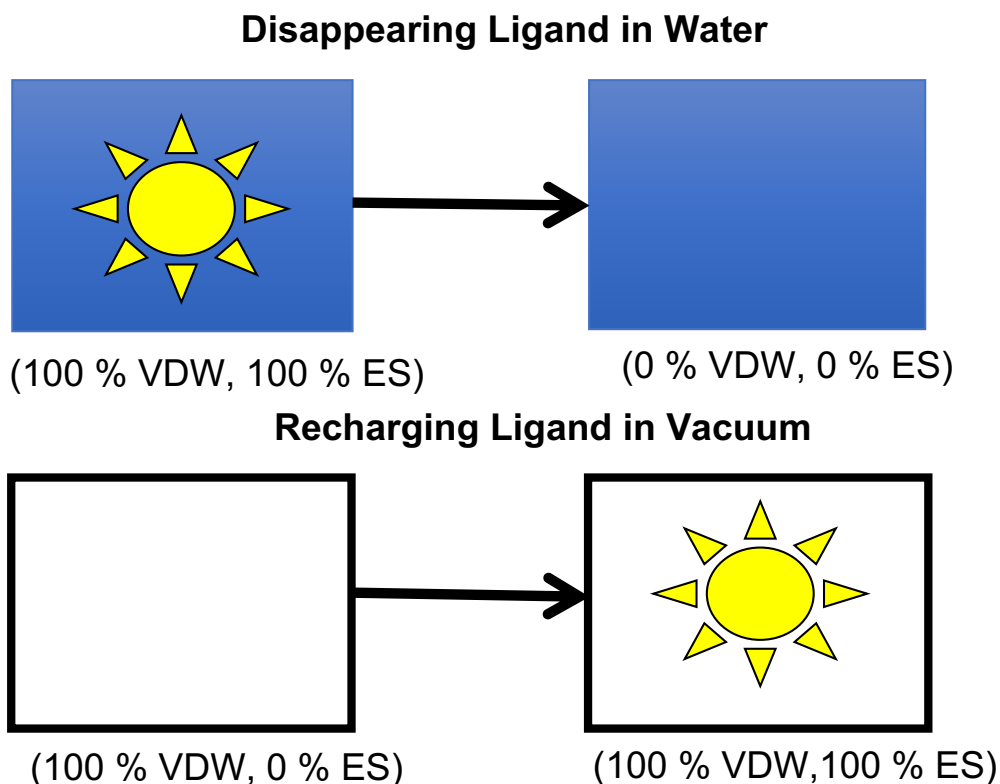


Figure 2.2: **BAR Free Energy Alchemical Free Energy Cycle:** The BAR Free Energy Cycle is divided into two parts: Disappearing Ligand in Water and Recharging the Ligand in Vacuum. First the electrostatic and polarization interactions are scaled down in from 100% strength to 0% then the vdW's interactions are scaled down from 100% strength to 0%.

Eq. 2.6). The subscript dt indicates that the potential is for dual topology, and the bonded term is independent to λ . A softcore van der Waals (vdW) potential[77, 103] has been adapted to replace the original buffered 14-7 potential[114] in AMOEBA in this implementation (see Eq. 2.7). Here i and j are the indices of the atoms, ϵ is the well-depth of the potential, and α is an adjustable constant, $\rho = \frac{r}{r^*}$, and r^* is the equilibrium distance between two atoms. This equation prevents the numerical instability of the system when λ is small, and it reduces to the original buffered 14-7 potential when $\lambda = 1$. The real space electrostatic potential[91] is written in Eq. 2.8. G is the permanent multipole moment, l is the order of the multipole moment, B is the screening function, and f is the modified distance defined in Eq. 2.9. This definition, similar to the softcore potential for vdW term, can prevent the numerical instability

of the system when λ is small and atoms are very close to each other. The final mixed potential for the real space electrostatic potential is defined in Eq. 2.10. Here the superscript *tot* and *mut* respectively indicates the energy of the whole system and energy of the part undergoing alchemical transformation. The reciprocal space part of the permanent multipole interaction energy is mixed linearly (see Eq. 2.11). The polarization energy has a slightly different combination. To eliminate the potential problem[77] of the self-consistent field calculation at unphysically small atomic distances, the polarization is switched off until λ is equal or greater than 0.75, which is referred as $\lambda_{\text{pol}}^{\text{start}}$. An artificial scaling factor, given as $\lambda_{\text{pol},A} = \frac{(1-\lambda_{\text{pol}}^{\text{start}}-\lambda)}{(1-\lambda_{\text{pol}}^{\text{start}})}$ and $\lambda_{\text{pol},B} = \frac{(\lambda-\lambda_{\text{pol}}^{\text{start}})}{(1-\lambda_{\text{pol}}^{\text{start}})}$, is used to smoothly switch the potential across $1 - \lambda_{\text{pol}}^{\text{start}}$ and $\lambda_{\text{pol}}^{\text{start}}$ respectively. The final mixed polarization potential can be written (see Eq. 2.12). Various derivatives, $\frac{\partial U}{\partial \lambda}$, $\frac{\partial^2 U}{\partial^2 \lambda}$, and $\frac{\partial^2 U}{\partial r \partial \lambda}$ have been derived in the same way as reported previously.[77]

$$U_{dt}(\lambda) = U_{\text{bonded}} + U_{vdW}(\lambda) + U_{\text{els}}^{\text{real}}(\lambda) + U_{\text{els}}^{\text{recip}}(\lambda) + U_{\text{pol}}(\lambda) \quad (2.6)$$

$$U_{vdW,ij}(\lambda) = \lambda \epsilon_{ij} \frac{1.07^7}{\alpha(1-\lambda^2) + (\rho + 0.07)^2} \left(\frac{1.12}{\alpha(1-\lambda^2) + (\rho^7 + 0.12)} - 2 \right) \quad (2.7)$$

$$U_{\text{els},ij}^{\text{real}} = \lambda \sum_{l=0}^4 G_{ij}^l(r) B_l(f) \quad (2.8)$$

$$f = (r_{ij}^2 + \alpha(1-\lambda)^2)^{1/2} \quad (2.9)$$

$$U_{\text{els}}^{\text{real}}(\lambda) = U_{\text{real,tot}}^{\text{els},A} (1-\lambda) + \lambda U_{\text{els},B}^{\text{real,tot}}(\lambda) + \lambda U_{\text{els},A}^{\text{gas,mut}} + (1-\lambda) U_{\text{els},B}^{\text{gas,mut}} \quad (2.10)$$

$$U_{\text{els}}^{\text{recip}}(\lambda) = (1-\lambda) U_{\text{els},A}^{\text{recip,tot}} + \lambda U_{\text{els},B}^{\text{recip,tot}} \quad (2.11)$$

$$U_{pol}(\lambda) = \begin{cases} (\lambda_{pol,A}^3 (U_{pol,A}^{tot} - U_{pol}^{env} - U_{pol,A}^{gas,mut}) \\ + U_{pol}^{env} + U_{pol,A}^{gas,mut} + U_{pol,B}^{gas,mut}) & \text{if } 0 \leq \lambda \leq (1 - \lambda_{pol}^{start}) \\ U_{pol}^{env} + U_{pol,A}^{gas,mut} + U_{pol,B}^{gas,mut} & \text{if } (1 - \lambda_{pol}^{start}) < \lambda < \lambda_{pol}^{start} \\ (\lambda_{pol,B}^3 (U_{pol,B}^{tot} - U_{pol}^{env} - U_{pol,B}^{gas,mut}) \\ + U_{pol}^{env} + U_{pol,A}^{gas,mut} + U_{pol,B}^{gas,mut}) & \text{if } \lambda_{pol}^{start} \leq \lambda \leq 1 \end{cases} \quad (2.12)$$

2.2.4 Computational Procedure

In this study, the hydration free energy of 20 small molecule solutes,[115] among which 15 are positively charged compounds and 5 are neutral compounds, were calculated in an explicit solvent model using the polarizable AMOEBA force field[91, 100].Parameters of the compounds were obtained using the program POLTYPE (2014).[110]

Both orthogonal space random walk and Bennett acceptance ratio methods used the same simulation conditions including. These conditions include box size, simulation ensemble, boundary conditions, and non-bonded cutoffs. Thus the comparison between OSRW and BAR methods would not be affected by potential artifacts in the calculated hydration free energy. These artifacts could include different boundary conditions or treatment of electrostatic interactions suggested in previous studies.[115, 116]

All molecular dynamics simulations were conducted with a RESPA integrator,Bussi thermostat, and 2 fs time step using the TINKER software package if not otherwise stated.[117, 118] A cutoff of 12Å was applied to vdW's interaction with $\alpha = 0.07$, while a cutoff of 7.0Å was applied in the real space softcore electrostatic calculations with $\alpha = 2.0$ (see Eqs. 2.7, 2.8,and 2.9). Self-consistent induced dipole moments were converged to below 0.00001 D per atom.

Before the alchemical simulations, all the solutes were first soaked in a 30Å cubic water box followed by a 600ps relaxation using NPT molecular dynamics simulation at 298 K with 2 fs time step. The resulting boxes were used in the subsequent NVT simulations with the density fixed at the average from the NPT simulations.In OSRW,

a 2D grid along the λ and F_λ axes was constructed to store the history of the (λ, F_λ) states visited. Each point on the grid represents a bin with finite dimensions. Our implementation has the width of 0.005 and the F_λ -width of 2.0.

For the λ -axis, λ ranges from 0 to 1, and for mathematical convenience the first and last λ bins are half size (the total number of the λ -bins is 201) and centered at 0 and 1, respectively. For the F_λ -axis, F_λ has no clear range so the range must be dynamically updated if the calculated F_λ falls outside the initial specified range. Also for mathematical convenience, there is always an F_λ bin centered at zero. Throughout the MD simulation, each bin centered at (λ, F_λ) in the grid represents the number of times a particular state with $\lambda - \frac{\Delta\lambda}{2} < \lambda < \lambda + \frac{\Delta\lambda}{2}$ and $F_\lambda - \frac{(\Delta F_\lambda)}{2} < F_\lambda < F_\lambda + \frac{(\Delta F_\lambda)}{2}$ was visited, where $\Delta\lambda$ and ΔF_λ represent the λ -width and F_λ -width, respectively. For each count, a 2D biasing Gaussian centered at (λ, F_λ) is added to the potential. Our implementation uses a Gaussian height of 0.005 with variances of $w_1^2 = (2\Delta\lambda)^2$ and $w_2^2 = (2\Delta F_\lambda)^2$. The Gaussians are cut off after five bins from the central bin. However, simply with the current implementation, the random walk may end up being stuck at the λ end points since λ is between 0 and 1. Thus, mirror conditions are enforced where if a Gaussian has a contributing value outside the λ range, the contribution is mirrored onto a bin with a valid range. For the first and last λ bins, the contributions are automatically doubled since they are centered at 0 and 1, respectively. The statistical error was estimated from 3 repeat simulations of 4 ns for each compound.

2.3 Results

2.3.1 Comparison of OSRW and BAR

The hydration free energy was calculated for the 20 compounds using both OSRW and BAR methods. In general, a good agreement between the two methods was obtained in Figure 2.3. The energy values from the two methods were plotted against each other Figure 2.3, and the R^2 correlation coefficient was calculated to be 0.98 for the charged set and 0.99 for the neutral set. The unsigned average difference in the calculated hydration free energy between the two methods is $0.39 \text{ kcal}\cdot\text{mol}^{-1}$, and the root-mean-square difference is $0.49 \text{ kcal}\cdot\text{mol}^{-1}$ (Figure 2.3).

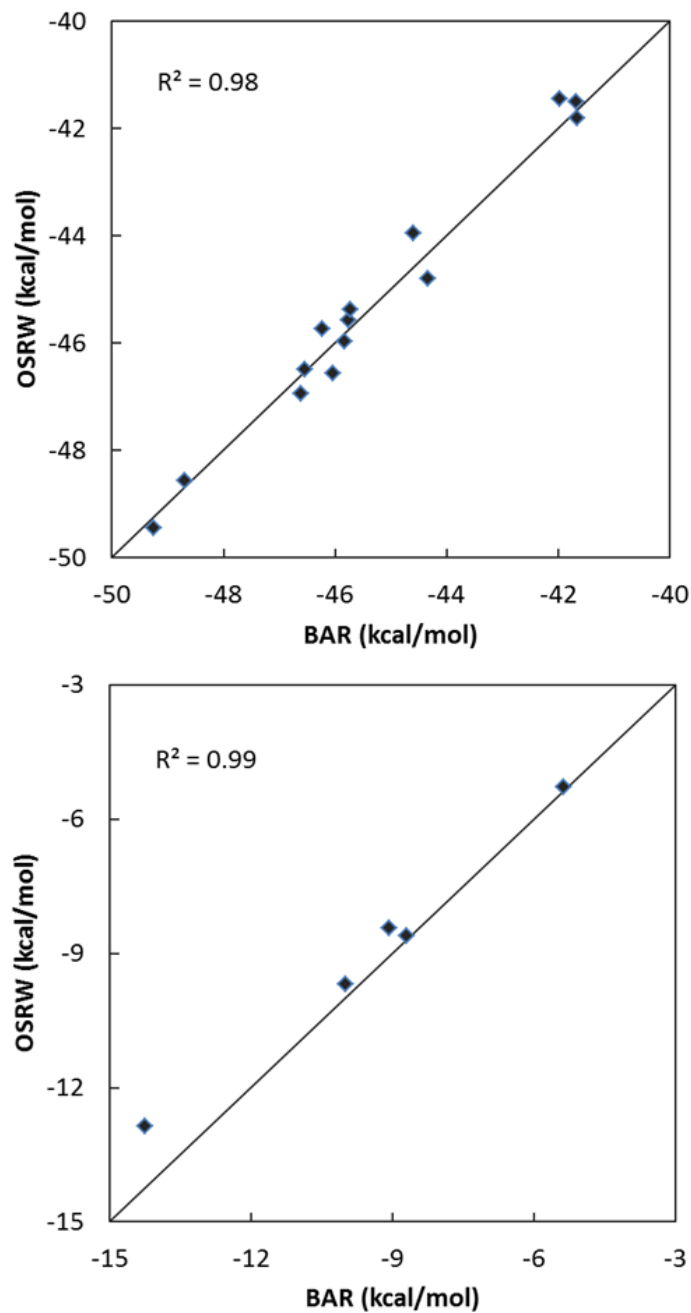


Figure 2.3: **Comparison of Hydration Free Energies from BAR and OSRW Methods:** An excellent agreement between the values from the two methods is obtained with R^2 of 0.98 ± 0.02 and R^2 of 0.99 ± 0.01 for the 15 charged (upper panel) and the 5 neutral (lower panel) respectively

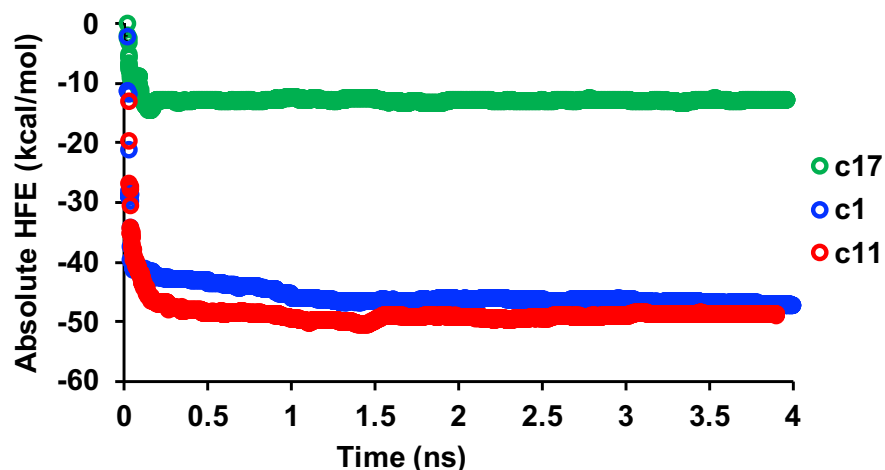


Figure 2.4: **Plots of Hydration Free Energy for Selected Compounds using OSRW:** Compound 1 (blue) Compound 11 (red) Compound 17 (green) For the neutral compound 17, the HFE has reasonably converged in about 300 ps in a single OSRW simulation. For the two charged compounds 1 and 11, it takes about 1.5 ns to reach a similar level of convergence while the absolute value of the HFE is much larger for charged compounds.

For illustration, the hydration free energy over time is plotted for the OSRW Figure 2.4 and BAR Figure 2.5 methods for compounds 1, 11, and 17. Compound 17 shows the largest difference of $1.38 \text{ kcal}\cdot\text{mol}^{-1}$ between the two methods, which reduced to $1.09 \text{ kcal}\cdot\text{mol}^{-1}$ after we significantly extended simulations for both methods (see Fig. A.1). Compound 17 has a “complex” structure with two hydroxyl groups and one fluorine atom all connecting to adjacent carbons in the benzene ring. The interactions among these groups and with water can be rather complex. For example, from both BAR and OSRW simulations, the two hydroxyl groups were seen very flexible with the hydrogen atoms either facing away or forming hydrogen bond with each other albeit with different frequencies. In addition, compound 17 is the only solute with fluorine atom in this set. To verify the fluorine parameters, hydration free energy of fluorobenzene and 2-fluorophenol have also been calculated using OSRW and compared with experiment values. For each of the two compounds, three independent OSRW calculations were performed. The average hydration free energy from the simulations is -0.76 ± 0.33 and $-5.71 \pm 0.16 \text{ kcal}\cdot\text{mol}^{-1}$ for fluorobenzene and 2-fluorophenol respectively. This is in a good agreement with the experimental

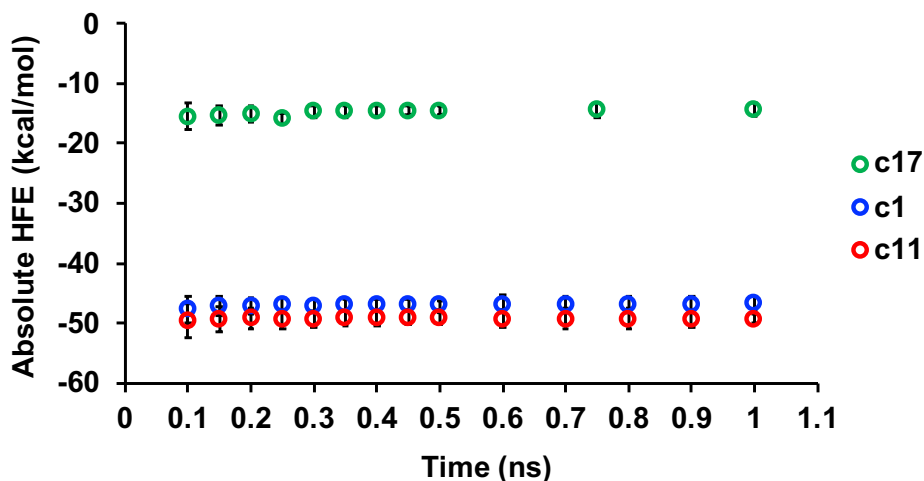


Figure 2.5: **Plots of Hydration Free Energy for Selected Compounds using BAR:** Compound 1 (blue) Compound 11 (red) Compound 17 (green). For all three compounds the HFE has reasonably converged in about 0.5ns. Because 25 independent simulations, were needed for each data point, the combined simulation time for compounds 1 and 11 is 50 ns, and 39 ns for compounds 17 (14 additional windows were added). The total simulation time for all 20 compounds is 106ns.

hydration free energy of -0.80 and -5.29 kcal \cdot mol $^{-1}$ for the two compounds.[82]

2.3.2 Structural Analysis of Fluorine Containing Ligand

We first examined the effect of van der Waals perturbation steps on the hydration free energy calculated from BAR for compound 17. In our implementation of the BAR approach, the electrostatic interaction between solute and environment was first turned off, and then the vdW's interactions are scaled. During this latter stage, water and solute molecules can have significant overlap. As a result there is a large uncertainty in the free energy, evident in the difference between the forward and backward free energy perturbation results. We added additional 6 windows in the middle of the vdW perturbation ($\lambda = 0.775, 0.725, 0.675, 0.625, 0.575, 0.525$). With these additional steps and longer simulations, the hydration free energy of compound 17 is merely increased by 0.2 kcal \cdot mol $^{-1}$.

We have further investigated water structure near the solute sampled during the

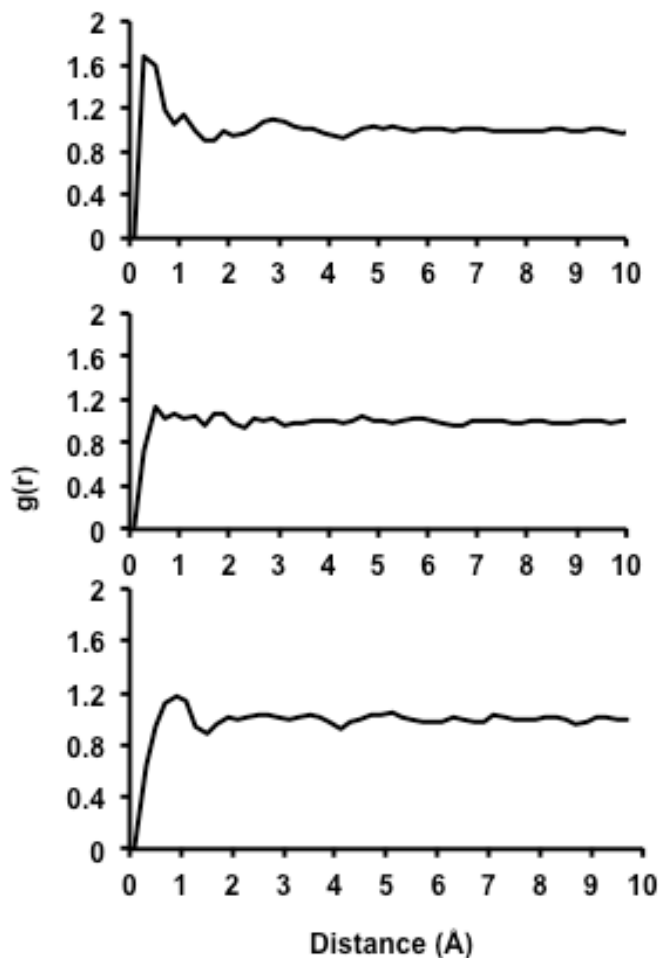


Figure 2.6: **Radial Distribution Functions from BAR simulations:** (top) the oxygen of water and the fluorine of compound 17; (middle) oxygen of water and oxygen of compound 17 in hydroxide group closest to the fluorine; and (bottom) oxygen of water and oxygen of compound 17 in hydroxide group furthest from fluorine. The RDF was evaluated using the simulated structures from all perturbation windows. All three BAR RDFs show water peaks close to the solute at $\approx 1\text{\AA}$. Two possible reasons could explain the differences in the RDFs generated in the BAR simulations and the OSRW simulations. In the BAR simulations, the vdW and electrostatic solute-water interactions are scaled sequentially, whereas in OSRW simulations these interactions are scaled simultaneously. Additionally, in BAR simulations equal numbers of coordinate structures were saved for each lambda window, while in OSRW simulations the number of structures saved at each lambda window were uneven due to the nature of importance sampling.

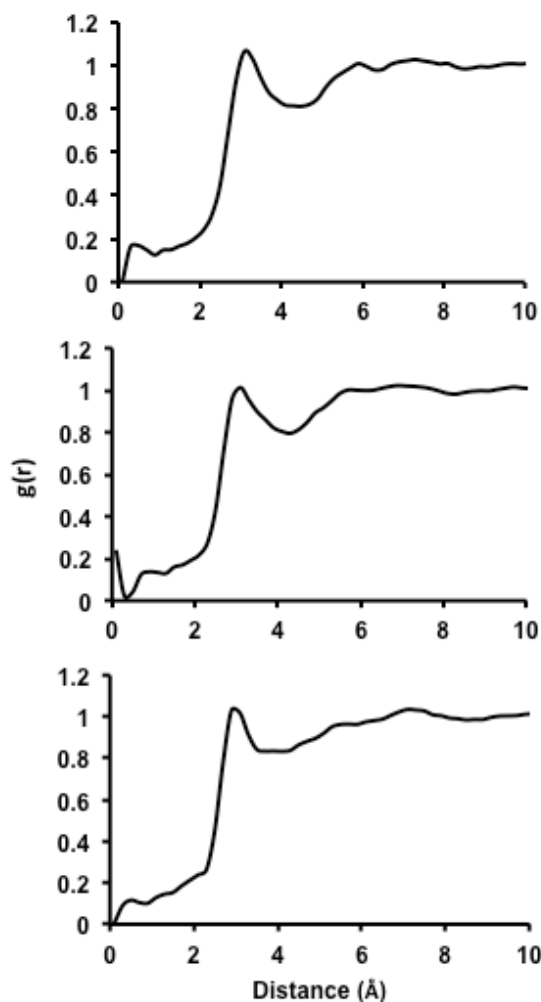


Figure 2.7: **Radial Distribution Functions from OSRW simulations** (top) oxygen of water and fluorine of compound 17; (middle) oxygen of water and oxygen of compound 17 in hydroxide group closest to fluorine in compound 17; (bottom) oxygen of water and oxygen of compound 17 in hydroxide group furthest away from fluorine in compound 17. Based on these RDFs, there is much less structure in the water around the polar groups of the solute during OSRW simulations, which can be attributed to the “flattened” energy surface introduced by the by the biasing potential. An interesting feature of the RDFs from OSRW simulations, is that there is a consistent but faint “peak” at around 3\AA for the OSRW RDFs. This could indicate some ordered water structure around the compounds.

BAR simulations by plotting radial distribution functions (RDFs) between the compound 17 (O atoms in the two hydroxyl groups and F atom) and water (O). It would be the best to compare the RDF at the same λ values. However, this is impossible in this study due to the limitation of our implementation. As explained in the Method section, the BAR method decouples the vdW and electrostatic interactions separately while in the OSRW approach the two are scaled simultaneously. As a result, the same λ in the two methods actually represents different states. Therefore, we plotted the RDF for the three atom pairs using the trajectories of all the lambda windows Figure 2.6. Similar plots for OSRW simulations were shown in Figure 2.7, which naturally included all lambda states visited during the simulations. The first apparent difference between the two methods is that the BAR RDFs have water peaks closer to the solute at around 1\AA . Note that, in addition to the difference in how the vdW and electrostatic solute-water interactions are scaled (sequential in BAR and simultaneous in OSRW), equal number of coordinate structures were saved for each lambda window in BAR simulations while the OSRW had an uneven distribution of lambda values given its nature of importance sampling. Both factors would affect the overall RDF. According to these RDFs, there is much less structure in the water around the polar groups of the solute during OSRW simulations, which we attribute to the “flattened” energy surface by the biasing potential introduced in OSRW. Another interesting feature is that there is a consistent but faint “peak” at around 3\AA for the OSRW RDFs.

As seen from the time evolution of hydration free energy for compound 17 Figure 2.4, the OSRW simulations actually first approached to $-15\text{ kcal}\cdot\text{mol}^{-1}$ in the beginning of the simulation but quickly increased to around $-13\text{ kcal}\cdot\text{mol}^{-1}$. This behavior was also observed in some other repeated OSRW simulations. The results suggest that BAR and OSRW may have sampled different phase space within the limited simulation time in this study. Thus we have further extended the simulations for both methods. For BAR, we increased the simulation length at each window from 1 ns to 5 ns while for OSRW we extended the simulation from 4ns and 8ns and added additional 4 more independent simulation runs for compound 17. The difference between the hydration free energy from the two methods dropped to $1.09\text{ kcal}\cdot\text{mol}^{-1}$ but still greater than the statistical uncertainty. This indicates that, for molecules such as compound 17, substantially long simulations may be necessary to

fully converge the answer.

2.3.3 Convergence and Efficiency OSRW and BAR

To examine the computational efficiency and convergence of OSRW and BAR methods, we analyze the simulations data for compounds 1, 11 and 17 as examples. The cumulative hydration free energies of these compounds calculated by OSRW as a function of time Fig 2.4 clearly demonstrate its efficiency and convergence. For the neutral compound 17, the HFE has reasonably converged in about 300 ps in a single OSRW simulation. For the two charged compounds 1 and 11, it takes about 1.5 ns to reach a similar level of convergence while the absolute value of the HFE is much large for charged compounds.

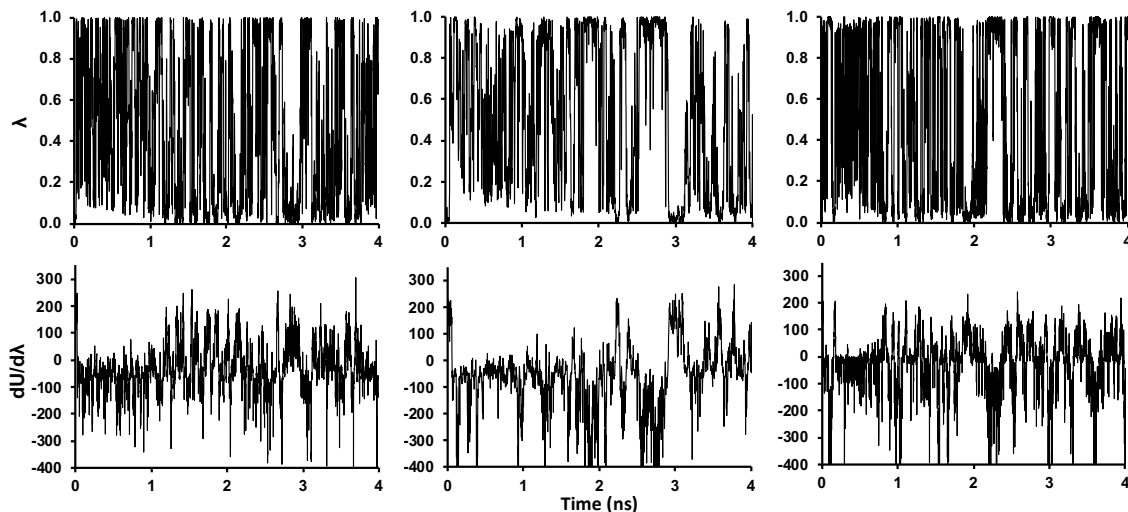


Figure 2.8: **Sampling Efficiency of λ and F_λ Using the OSRW Method:** λ and F_λ values over simulation time from the OSRW simulations of compounds 1, 11 and 17 (data collected at every 5 ps). The spikes in F_λ correspond to barriers crossed by the system. The OSRW method samples the order parameter space multiple times during the simulation. There are frequent peaks and valleys in the $\frac{\partial U}{\partial \lambda}$ plot.

Since the free energy is dependent on the λ and $\frac{\partial U}{\partial \lambda}$, a sufficient sampling of the system in both spaces is crucial to the accuracy of the estimated free energy. Distributions of λ and its generalized force $F_\lambda = \frac{\partial U}{\partial \lambda}$ against time are shown in Figure 2.8. It can be seen that the OSRW method samples the order parameter space multiple times during the simulation. There are frequent peaks and valleys in the $\frac{\partial U}{\partial \lambda}$ plot.

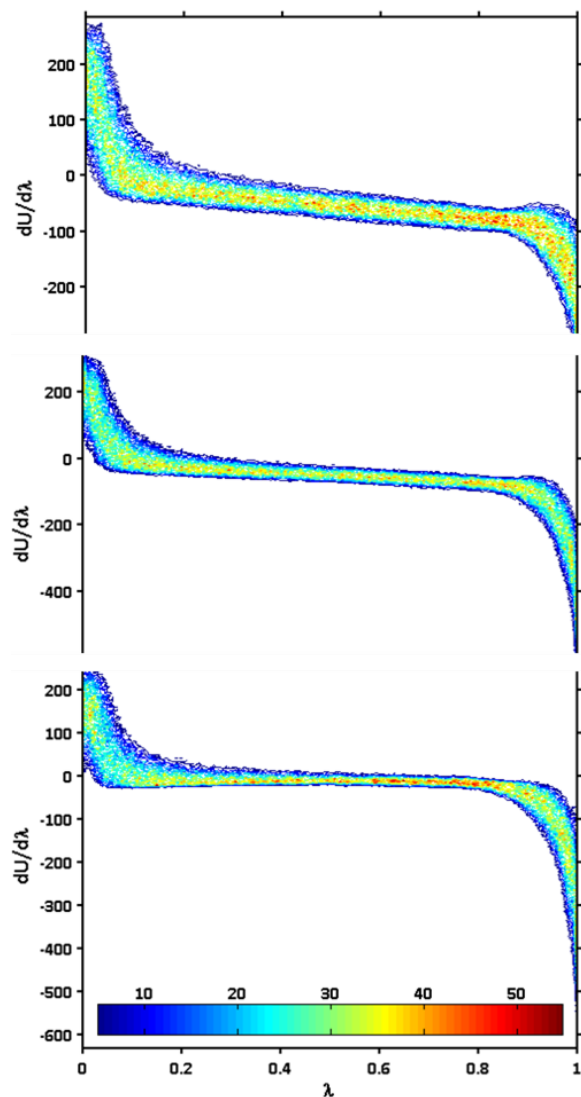


Figure 2.9: **Free Energy Landscapes Generated Using the OSRW Method:** Plots of free energy surfaces for compounds 1 (top), 11 (middle), and 17 (bottom). The color represents the number of visits to a particular state. By examining the free energy surface with respect to $\frac{\partial U}{\partial \lambda}$ and λ in, it is shown that large positive values of $\frac{\partial U}{\partial \lambda}$ correspond to the initial state when the solute is mostly decoupled from the environment while the negative attractive $\frac{\partial U}{\partial \lambda}$ occurs toward the end state when $\lambda = 1$. There is a broad and flat region in between, which corresponds to $\frac{\partial U}{\partial \lambda}$ around zero. A significant feature of the OSRW is that it flattens the energy surface in both λ and $\frac{\partial U}{\partial \lambda}$ dimensions while the latter is the exact quantity needed to compute free energy ($G = \int \langle \frac{\partial U}{\partial \lambda} \rangle d\lambda$).

By examining the free energy surface with respect to $\frac{\partial U}{\partial \lambda}$ and λ in Figure 2.9, it can be seen that the large positive values of $\frac{\partial U}{\partial \lambda}$ correspond to the initial state when the solute is mostly decoupled from the environment while the negative attractive $\frac{\partial U}{\partial \lambda}$ occurs toward the end state when $\lambda = 1$. There is a broad and flat region in between, which corresponds to $\frac{\partial U}{\partial \lambda}$ around zero. A significant feature of the OSRW is that it flattens the energy surface in both λ and $\frac{\partial U}{\partial \lambda}$ dimensions while the latter is the exact quantity needed to compute free energy ($G = \int \langle \frac{\partial U}{\partial \lambda} \rangle d\lambda$). Most other methods that sample only the λ space do not actively handle the energy barriers along $\frac{\partial U}{\partial \lambda}$ at a given λ . Figure 2.9 shows the surface along $\frac{\partial U}{\partial \lambda}$ at a given λ can be complicated. In a typical OSRW simulation, the system quickly reaches the minimum (negative) $\frac{\partial U}{\partial \lambda}$ region and adding biasing Gaussian potentials in both λ and $\frac{\partial U}{\partial \lambda}$ dimensions until it reaches the middle flat region ($\frac{\partial U}{\partial \lambda} \approx 0$) and then glide back and forth along the λ dimension many times. Sufficient sampling near both initial and end states is important as both contribute significantly to the free energy.

It should be noted that our implementation involves the use of metadynamics for the calculation of the biasing potential. The choice of biasing Gaussian heights and λ particle movement has not been fully optimized in this study. In addition, other implementations alternative to metadynamics can be used and have been explored.[94] The convergence of the free energy calculated using BAR, as seen for compounds 1, 11 and 17 in Figure 2.5, generally takes 300 ps simulations for each window to converge to a reasonable level. However, if we take the number of windows needed in BAR simulation into account, it is roughly equivalent to 8 ns of a single molecular dynamics simulation. Thus the superior efficiency of OSRW is apparent. The OSRW simulation is in general easier to set up and process as compared with the BAR method. In the TINKER implementation, only a single molecular dynamics simulation is required while the BAR approach requires the maintenance of ≈ 25 simulations (in this study). A way to improve the sampling ability of the BAR method is to allow configuration exchanges among different simulation windows as in the multistate BAR (mBAR) method.[119]

2.4 Conclusion

We have implemented the OSRW method for the polarizable multipole based AMOEBA force field in TINKER. OSRW method has been shown to exhibit superior efficiency for hydration free energy calculations over BAR, a traditional free energy calculation method. In BAR method, 1ns at each of 25 steps or more have been performed. For OSRW, one simulation of 4ns led to similar results within the statistical uncertainty. We found excellent agreement between the two methods, with an RMSD of $0.49 \text{ kcal}\cdot\text{mol}^{-1}$. A noticeable difference between the OSRW and BAR results was observed for compound 17, which has two hydroxyl groups and a fluorine atom attached to a benzene ring, even after the simulations were extended significantly. This discrepancy shows that even for a relative small solute, the sampling of solute-solvent configurations can be challenging. While further studies are required to thoroughly evaluate the different free energy methods, our results suggest that OSRW is an efficient alternative sampling method that polarizable force fields can benefit from.

Chapter Three: Detecting Point-Mutations using a DNA Melting Probe

3.1 Introduction

3.1.1 Experiments and Simulations

High-resolution melting (HRM) analysis is a closed-tube detection method that can identify point-mutations in DNA sequences by their melting temperature. These melting probes contain a recognition DNA strand that binds to a target DNA strand. When the DNA strands in the melting probe bind, the melting probe fluoresces via an intercalating fluorescent dye. When the temperature is increased, the melting temperature can be interpreted as temperature when the fluorescence intensity is fifty percent of its maximum. Recently developed melting probes can discriminate between a wild-type DNA sequence and a DNA sequence containing a point mutation, or single nucleotide polymorphism (SNP). However, these probes cannot reliably identify among different point mutations, or alleles.

By incorporating a locked nucleic acid (LNA) thymidine, into the melting probe, the melting probe can reliably differentiate the four SNP alleles by four distinct melting temperatures. Molecular dynamics (MD) simulations of a truncated melting probe consisting of control DNA duplexes (no LNA) and hybrid DNA duplexes (containing either LNA thymidine or LNA cytosine) were performed at the experimental melting temperatures, and these simulations correlated with the observed experimental trends. Additionally, hydrogen-bonding profile analysis of the MD simulations revealed that the LNA thymidine enhanced nearest neighbor hydrogen bonding when the target DNA strand contained a matched allele, but the LNA thymidine disrupted nearest neighbor hydrogen bonding when the target DNA strand contained a mismatched allele.¹

¹This work was previously published. I collaborated with J. Obliosca, designed, performed, and analyzed MD simulations. For experimental details see [120]

3.1.2 Detection of Point Mutations using a DNA Nansensor

Genotyping of single-nucleotide polymorphisms (SNPs) is becoming a routine test in clinical laboratory as personalized medicine continues to develop.[121] Whereas many innovative SNP detection techniques have been developed in the past,[122–125] high-resolution melting (HRM) measurement is probably the only method that is becoming a standard procedure in both research and clinical laboratories due to its homogeneous and closed-tube detection format.[126–129] Since the first demonstration of melting curve analysis in conjunction with real-time PCR,[130–132] HRM has been used for scanning of mutations in cancer-related genes[133, 134] and determining HIV diversity[135] in clinical samples.

While HRM (based on FRET dyes[130] or saturating DNA dyes[136]) is a simple, rapid and inexpensive method for in-house SNP testing,[137] a typical DNA melting probe (whether it is a binary probe,[138, 139] singly-labeled probe,[140, 141] unlabeled probe[142] and snapback primer[142]) can only differentiate one fully matched allele from the other three single-mismatch containing alleles (hereafter denoted as mismatched alleles)[139, 143, 144]. Since the melting temperatures (ΔT_m s) associated with the three mismatched probe-allele hybrids are often indistinguishable, additional melting probes are required to differentiate these mismatched alleles. There has not been any report that employs only one melting probe to distinguish the fully matched allele from the three mismatched alleles,[139, 143, 144] and at the same time this same probe can also differentiate the three mismatched alleles among themselves. Although not all SNP homozygous variants are of clinical interest, we believe a melting probe with such an “ultimate discrimination power” is greatly beneficial as it facilitates the discovery of rare genetic mutations at lower cost.

Our probe design consideration is different from that of other HRM researchers. Other researchers only aim to increase the temperature difference (ΔT_m) between the fully matched probe-allele hybrid and the three mismatched hybrids.[143] We, on the other hand, focus on increasing the T_m differences among the three mismatched hybrids themselves (while maintaining the (ΔT_m between matched and mismatched alleles). Our goal is to reliably identify each of the four SNP alleles with a specific T_m . We emphasize that we want to achieve such complete SNP differentiation in a

single test tube using only one unlabeled melting temperature probe and a common DNA binding dye (see Fig. 3.1).

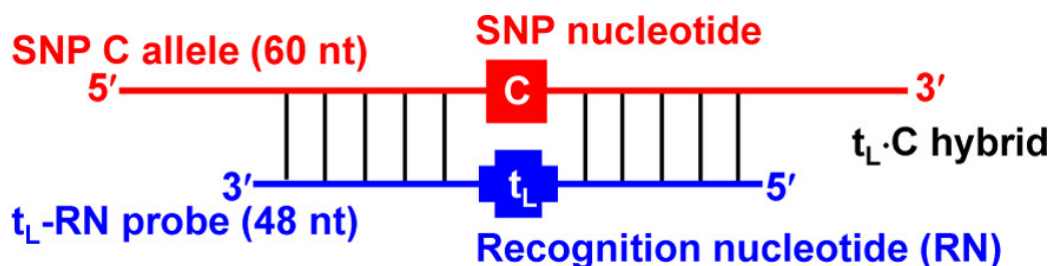


Figure 3.1: **Single-nucleotide polymorphism (SNP) detection using a melting temperature probe containing a single LNA thymidine monomer (t_L)** The probe/allele hybrid contains a single centrally positioned t_L ·C mismatched pair, thus denoted as the “ t_L ·C mismatched hybrid” or simply the “ t_L ·C hybrid”. In our design, the t_L is termed the “recognition nucleotide” (RN) while the C is termed the “SNP nucleotide”. When t_L is employed as the recognition nucleotide, the 48-nt long melting probe is denoted as the “ t_L -RN probe”. Similarly, the 60-nt long SNP allele is denoted as the “C allele”. In this report, the lower case letters represent the RNs while the upper case letters represent the SNP nucleotides. The subscript L represents an LNA monomer.

3.2 Materials and Methods

3.2.1 Parameterization of Duplexes

The force field used for the locked nucleic acid is the AMBER99X force field provided by Yildirim et. al.[145, 146] This force field contains modified chi torsion angle between the sugar and the base for the modified sugar of the locked nucleic acid. These parameters are modifications of the AMBER RNAff99 force field. Based on these parameters, a locked t_L NA was generated from u_L NA coordinate by adding a methyl group.

3.2.2 Incorporation of LNA

First, 25-nt long B-form DNA duplexes were generated using the nucleic acid builder (NAB) program distributed with the AMBER14 software package.[147, 148] The NAB program, allows for specifying sequences of DNA as well as the form of

the DNA duplex. For the LNA-DNA hybrid, the LNA monomers were inserted into the DNA probe. This was done by first generating a B-form DNA duplex, deleting a thymine nucleotide, and using PYMOL to orient the locked nucleic acid into the DNA duplex.[149] Using tLEAP, parameter files were generated for the LNA containing DNA duplex using the AMBER99 force field. Using this procedure, the t_LA , t_LG , t_LT , t_LC , $t \cdot A$, $t \cdot G$, $t \cdot T$ and $t \cdot C$ matched and mismatched hybrids were generated.

3.2.3 MD Simulation Procedure

After incorporating the locked nucleic acid into the center of the DNA duplex, the hybrid DNA duplex is energy minimized over 2,500 cycles using the steepest descent method. SHAKE bond length constraints are used on the hydrogen bonds. Implicit solvent using the Hawkins, Cramer, and Truhlar pairwise generalized Born Model is used with a 0.02 M implicit NaCl concentration.[150] The maximum radius used for calculating effective Born radii is 15Å. Non-periodic boundary conditions were applied, the maximum distance used for calculation of van der Waals, electrostatic, and “off-diagonal” generalized Born interactions is 9,999Å. The minimization was performed at 298K. Langevin thermostat was used for temperature scaling with a collision frequency of 1ps^{-1} . After each LNA-DNA duplex is energy minimized, independent simulations at 30, 40, 50, 60, 65, 70, 75, 80, 85 and 90 C were run. Each simulation is run for a minimum of 500ns up to 3,000 ns at a particular temperature. The same SHAKE bond length constraints, implicit water model, salt concentration, non-periodic boundaries, and Langevin thermostat were employed from the energy minimization procedure. Energies, restart files, coordinates of the MD trajectory are written every 2 ps. The timestep for each MD step is 2 fs.

3.2.4 Analysis Procedure

A duplex was considered melted when 50% of the duplex was double stranded. In other words, at least half of the base pairs in the duplex had in inter-base N1-N3 distance of more than 5Å away from their complementary base. This criterion was applied to all 25 basepairs in the duplex. From MD trajectories, the distance between neighboring nitrogens were calculated for each base pair every 50ps. The index i is defined as the residue number of the nucleotide along the target strand,

where i is indexed starting from the 5' to the 3' end. The complementary base of the i th nucleotide is indexed starting at $50-i+1$ starting from the 3' to the 5' end of the recognition strand.

The last 500 ns were averaged to determine the average dissociation fraction (AF) as a function of time. The association fraction is 1 minus the dissociation fraction. The association fraction (AF) and dissociation fraction (DF) are defined as follows, where $|\vec{d}_i - d_{50-i+1}|$ is the distance between N1 and N3 atoms of residue i and residue $(50i + 1)$, DF_{bp} is the dissociation fraction per base pair.

$$D_{bp} = \begin{cases} 0 & \text{if } n \leq 5\text{\AA} \\ 1 & \text{if } n > 5\text{\AA} \end{cases} \quad (3.1)$$

$$\frac{\sum_{i=1}^{25} DF_{bp}(|\vec{d}_i - d_{50-i+1}|)}{25} \quad (3.2)$$

$$AF = 1 - DF \quad (3.3)$$

After calculation of the average AF at a particular temperature, the melting data is normalized. This is done by taking the ratio of each AF by the maximum AF at a particular melting temperature for each SNP curve. Each SNP melting curve as normalized separately using this procedure. Each melting data set is fit using a logistic curve using MATLAB generalize linear model regression. After fitting all melted duplex data, the melting temperature of each duplex was determined by finding the temperature where the AF was 0.5. Similar approaches for finding the melting temperature were used in previous work [151, 152].

The H-bond probability maps were generated using the VMD hydrogen bond analysis plugin software and MATLAB. The AMBER trajectories for t_L t_LA , tA , t_LC , and tC were loaded into VMD, and a cutoff of distance of 3.0\AA was used for donor-acceptor distances and an angle-cutoff of 20 degrees was used for calculating hydrogen bonds. The hydrogen bonds were calculated for all nucleotides for the last 500 ns of each trajectory. The hydrogen bonds were calculated every 50 ps. From the output of the VMD hydrogen bonds plug-in, the nucleotide number for the donor and acceptor was used to generate a 50 by 50 matrix with values given by the hydrogen

bond probability of that donor-acceptor pair. This 50 by 50 matrix was condensed to a 25 by 25 matrix by considering hydrogen bonds formed between target and probe strands (which are each 25 nt long).

A filled contour map of hydrogen bond probability map was generated using MATLAB, and the nucleotide numbering convention used is shown in Fig 3.2. The nucleotides for the probe strand are on the x-axis and are numbered starting from the 5' to the 3' end. The nucleotides for the target strand are on the y-axis and are numbered starting from the 3' to the 5' end. As a result, the center of the target strand has a nucleotide number of 13, and the center of the probe strand has a nucleotide number of 13. The nearest neighbors of the center are numbers 12 and 14.

3.3 Results

3.3.1 Molecular Dynamics Simulations

We employed molecular dynamics (MD) simulations to investigate the detailed melting processes given by the t-RN and t_L/c_L -RN probes (Fig. 3.2 and Fig. 3.3). In our simulations, the previously published AMBER99X force field[153, 154] was applied to the LNA residue (see Materials and Methods). We calculated the association fraction for the duplex to characterize its melting during the MD simulations. The association fraction (AF) can be considered a measure of duplex formation, where an AF of one means 100% duplex whereas an AF of zero means the duplex is fully melted. The association fraction is calculated as the ratio between the number of melted stable base pairs and the total number of possible base pairs possible (number of residues divided by 2). A base pair is considered melted if the nitrogen-nitrogen distance between a complementary base pair, N1 for purines and N3 for pyrimidines, was greater than 5Å. An example of a base pair in a $t_L \cdot C$ duplex (only nearest neighbor bases are shown for clarity) and melted base pair are shown in Fig. 3.2A. 25- nt BRAF duplexes containing t and t_L with A, G, T, C SNPs were simulated at melting temperature 70 – 75°C, while 25-nt BRAF duplexes containing c_L with A, G, T, C SNPs were simulated at 70 – 72°C. These temperature ranges were chosen based on the melting temperature range measured by the experiments. Each duplex was

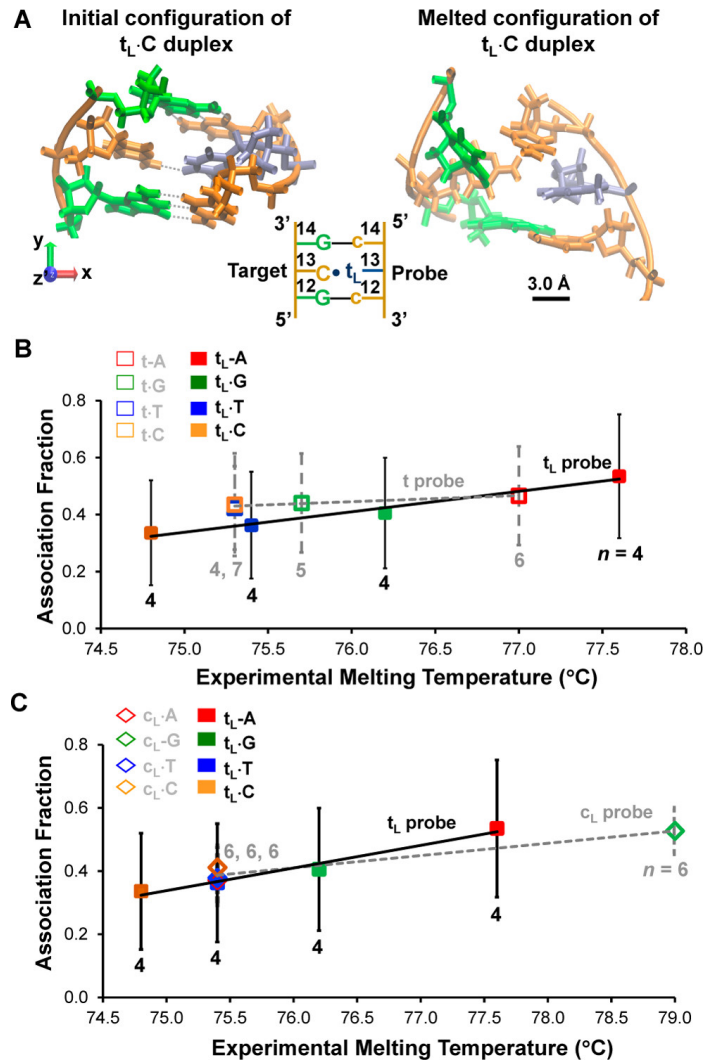


Figure 3.2: Molecular dynamics (MD) simulations of 12 25-bp long hybrids, each consisting of a BRAF allele and a melting probe. (A) 3D views of the mismatch pair and the nearest neighbor bases in the t_L C hybrid at initial and melted states. Inset shows the bases and their identification numbers in the hybrid. (B) Correlations between the simulation-derived association fractions (AF, at 70 and 75°C) and the experiment-obtained melting temperatures (T_m) for the t-RN ($R^2 = 0.93$) and the t_L -RN ($R^2 = 0.97$) probes. (C) Correlations between AF (at 70 and 75°C) and T_m for the c_L -RN ($R^2 = 0.94$) and t_L -RN ($R^2 = 0.97$) probes. n represents the number of repeated 11.5 μ s MD simulations (replicates). Error bars are standard deviations from replicates.

simulated multiple times (4-7 replicates), with each between 1 and 1.5 μ s, to achieve better statistics.

Overall the simulated association fractions for all systems investigated display excellent correlation with the experimental melting temperatures (Fig. 3.2B and Fig. 3.2C). In Fig. 3.2B, both experiments (melting temperature) and simulations (association fraction) show that the t_L -probe has the effect of expanding the range, in contrast to unlocked t , so that each SNP can be better distinguished. The association fraction for locked thymidine displayed the following trend $A > G > T > C$ (Fig. 3.2B), in agreement with the HRM experiments ((see Fig 2B of [120])). In addition, the lower simulated association fraction for $t_L \cdot C$ vs. $t \cdot C$, agrees with the observation of a lower melting temperature for $t_L \cdot C$ vs. $t \cdot C$ from the HRM experiments (see Fig 2 of [120]). Furthermore, association fraction for $t \cdot C$ and $t \cdot T$ overlaps within simulation errors while the HRM experiments indeed suggest that the C-SNP and T-SNP have the same melting temperature and cannot be differentiated by unlocked thymidine. It should be noted that for G and T SNPs, the simulations did not show improved discrimination by t_L vs. t , which is also reflected as slightly different “slopes” of correlation plot in Fig. 3.2B.

The HRM experiments predicted that the locked cytidine could not differentiate between adenine, thymine, or cytosine SNPs (see F(see Fig 2A of [120])). The MD simulations also predicted that the locked cytidine could not differentiate between these SNPs, as seen in the association fraction analysis in Fig. 3.2C. Comparing the locked cytidine (dashed line) and the locked thymidine (solid line), the locked cytidine is only able to distinguish guanine, according to both experiments and computed AF, while the melting profiles of A, T, C SNPs are about the same. Beyond the agreement between the simulated associated fractions with HRM melting temperatures, we have examined MD trajectories to provide atomic insight and explanation for the two extreme cases: $t_L \cdot A$ vs. $t \cdot A$ and $t_L \cdot C$ vs. $t \cdot C$. Both simulations and experiments suggest that t_L increases the duplex stability when paired with adenine while t_L destabilized the duplex when paired with cytosine.

Here, we computed the hydrogen bond probability from the MD trajectories for all nucleotides using distance and an angle cutoff of 3Å and 20° (meaning any pair of nucleotides that meet the criteria is considered hydrogen bonded). The hydrogen

bonds are calculated for all possible base pairs, and a hydrogen bond probability map is generated for each duplex. The hydrogen bond probability is depicted using a color bar where a hydrogen bond probability of 1 means the base pair is hydrogen bonded throughout the analyzed trajectory. A hydrogen bond probability of 0 means the base pair does not form any hydrogen bonds throughout the analyzed trajectory.

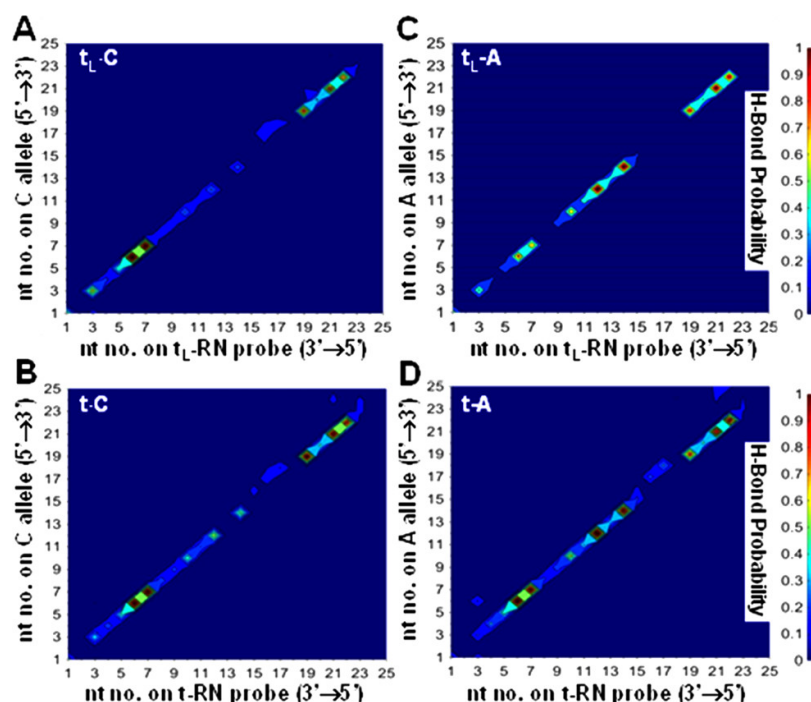


Figure 3.3: **Hydrogen-bond (H-bond) probability maps** (A) $t_L \cdot C$, (B) $t \cdot C$, (C) $t_L \cdot A$, and (D) $t \cdot A$ hybrids. Pair 13 is where the SNP site locates (Figure 3.2A). These maps clearly show how the t_L modification promotes local H-bond probability around the SNP nucleotide A but demotes H-bond probability around C, thus enhancing the SNP discrimination power.

The hydrogen bond probability maps are shown in Fig. 3.3: nucleotide number for the t_L -probe is 13 and two nearest neighbors surrounding the t_L are numbers 12 and 14. For $t_L \cdot C$ the nearest neighbor hydrogen bond probability is low (< 0.3), while the $t \cdot C$ the nearest neighbor hydrogen bond probability is noticeably higher (≈ 0.4). For $t_L \cdot C$ and $t \cdot C$, the hydrogen bond probability between the probe and SNP is near zero (< 0.1) at the center (nucleotide number 13) indicating that the lack of hydrogen bonding at the is disrupted at the center of the duplex due to mismatch (Figs. 3.3C and 3.3D). In contrast, the nearest neighbor hydrogen bonding

probability is enhanced by the presence of the t_L , with a nearest neighbor hydrogen bond probability that is higher for $t_L \cdot A$ (≈ 0.5), than for $t \cdot A$ (≈ 0.35) (Fig. 3.3A and 3.3B). For $t_L \cdot A$ and $t \cdot A$, the hydrogen bond probability at the center is (≈ 0.45) for $t_L \cdot A$ and (≈ 0.4) for $t \cdot A$ indicating hydrogen bonding is also enhanced at the center of the duplex by the LNA probe (Fig. 3.3A and 3.3B).

From these hydrogen bond probabilities, a possible mechanism for the discrimination by the locked-thymidine can be identified. It is known that $t \cdot A$ forms a complementary stronger base pair that is stronger than $t \cdot C$ (and the rest of the SNPs). The chemical linkage in LNA limits the sugar pucker flexibility which in turn restrains the base mobility around χ torsion angle.

As a result, $t_L \cdot A$ is “locked” into the favorable (low energy) configurations, as shown by the enhanced H-bond probability at the center pair and nearest neighbors. Thus, there is a significant gain in the “binding” enthalpy. Correspondingly, the concept of pre-organizing ligand through covalent bonds to increase the protein-ligand binding free energy is a common practice in drug design, for entropic and/or enthalpic reasons.[106] Conversely, $t \cdot C$ is a weaker, non-complementary pair mismatch with very weak base pairing, and the LNA further restricts the ability of the bases to form ensembles of energetically favorable interactions. This in turn allows for higher mobility of nucleotides around the t_L and leads to disruption of nearest neighbor hydrogen bonds. While the weaker interactions could lead to higher entropy in the duplex, here the lost in enthalpy is likely more dominant.

Overall, t_L is able to amplify the difference in stability between A and C SNPs. For c_L , the LNA also clearly increased the stability of the complementary pair; $c_L \cdot G$ has a higher melting temperature than $c \cdot G$. Note that, nonetheless, “t” is able to form wobble base pairs with T and G, while c_L essentially only forms energetically favorable pairs with G (see Fig 4 of [120])). All other pairs formed with c are mismatches and the interactions are so weak that potential for discrimination among A, T and C SNPs is completely lost. Therefore, the SNP discrimination power is encoded in the detailed chemistry and the restriction of sugar pucker by LNA, which subtly “shifts” the pairing strength and thus recognition thresholds.

3.4 Conclusion

Overall, this work describes an integrated experimental and computational project to develop the “4T_m probe”, a hybrid melting probe that has four distinct melting temperatures when hybridizing with four single nucleotide polymorphisms. The molecular dynamics simulations agree with the experimental trends of association fraction versus melting temperature. Further, by analyzing the hydrogen bonding probability along the melting probe, a molecular mechanism was proposed to explain the melting behavior of the “4T_m probe”. This work shows that computation modeling of the structure and dynamics of nucleic acids can complement experimental work.

Chapter Four: Development and Improvement of the RACER RNA Coarse-Grained Model

4.1 Introduction

4.1.1 Predicting RNA Structure and Energetics with a Coarse-Grained Model

Many computational tools are available to predict RNA secondary structure. Two of these methods, Mfold and Vienna, use the RNA sequence and nearest neighbor energies to predict RNA secondary structures.[155–158] The unfolding free energies predicted by these methods agree well with unfolding free energies obtained from mechanical unfolding of RNA hairpins.[45, 47] However, the structure predictions from these programs are not always unique, and these methods are not accurate for large RNA sequences. Tertiary structure prediction methods are available for RNA, and generally fall into two categories, template-based, which use small RNA motifs to build a large RNA structure, and graph-based, which use graph theory to predict RNA structures.[159, 160] While structure prediction methods are valuable for understanding RNA structure, they miss the dynamics of RNA folding.

RNA is a dynamic molecule, undergoing structural changes on timescales from nanoseconds to minutes.[9] Over this wide range of timescales, RNAs can undergo significant structural changes in its secondary and tertiary structure. Therefore, efficient sampling is essential to capture the dynamics of RNA folding. In addition to efficient sampling, accurate physical models of RNA are important. Coarse-grained RNA modeling is possible way to overcome the rugged folding free energy landscape of RNA folding. A newly developed **RNA**Coars**E-GR**ained (RACER) RNA coarse-grained model can be used with the molecular dynamics engine TINKER to predict 3D structure, energetics, and dynamics of RNA folding.¹

¹The RACER 1.0 model was previously published.[161] I contributed to testing of the RACER 1.0 model, manuscript writing, and interpretation of results

4.1.2 Development of RACER 1.0

RACER 1.0 is a recently published coarse-grained RNA physics-based model that can be used in molecular dynamics simulations to predict structure and folding free energies.[161] RACER 1.0 was developed using a “top-down approach”, incorporating experimental structures and thermodynamic melting data to derive parameters for the potential energy functions (see Fig. 4.1). The structure prediction capacities of the RACER 1.0 model were validated using a simulated annealing protocol. The 3D structure predicted from RACER 1.0 was compared to Protein Data Bank (PDB) experimental structure, determined from x-ray crystallography or NMR. Excellent agreement between the RACER 1.0 model and experimental structures was found (average RMSD of $2.93 \pm 0.54 \text{ \AA}$). Finally, the ability of the RACER 1.0 model to predict folding free energies was tested by performing equilibrium pulling simulations (total of 0.86 ms). These folding free energies were compared with thermodynamic melting data of the same sequences and excellent agreement was found ($R^2 = 0.93$). Electrostatics, base-stacking, and hydrogen bonding were found to be the major driving forces of RNA folding. However, RACER 1.0 model only captured three canonical Watson-Crick hydrogen bonds: two for a canonical G-C base pair, and one for a canonical A-U base pair. As a result, the 3D structure and folding free energies of RNAs containing non-canonical hydrogen bonds were not predicted well.

4.1.3 Development of RACER 2.0

In order to apply the RACER model to predict 3D structure and folding pathways of complex RNAs, improvements to the RACER 1.0 model are needed. This work introduces the RACER 2.0 model, which improves the hydrogen bond potential of the RACER 1.0 model. Specifically, definitions and force field parameters for non-canonical hydrogen bonds are now included. The optimized RACER 2.0 model includes definitions for all twelve geometric families of hydrogen bonds.[162] Force field parameters are optimized against quantum mechanical binding energies. Several changes to the TINKER MD source code were made to accommodate the non-canonical hydrogen bond parameters. Major software changes include implementation of neighbor lists up to seven nearest-neighbors in order to prevent hydrogen bonding between stacked nucleotides and sorting user input to determine unique hydrogen bond definitions.

The accuracy and transferrability of the new parameters are tested by performing stability analyses with the RACER 1.0 and RACER 2.0 models with small RNAs (hairpins and duplexes) and larger more complex RNAs (g-quadruplex, tetraplex, hammerhead ribozyme, two domains of a group II self-splicing intron, triplex, twister ribozyme, and adenine-riboswitch). Pulling simulations on a RNA G-quadruplex are performed with the non-canonical hydrogen bond parameters, and the unfolding free energy agrees with free energy differences from single-molecule experiments.

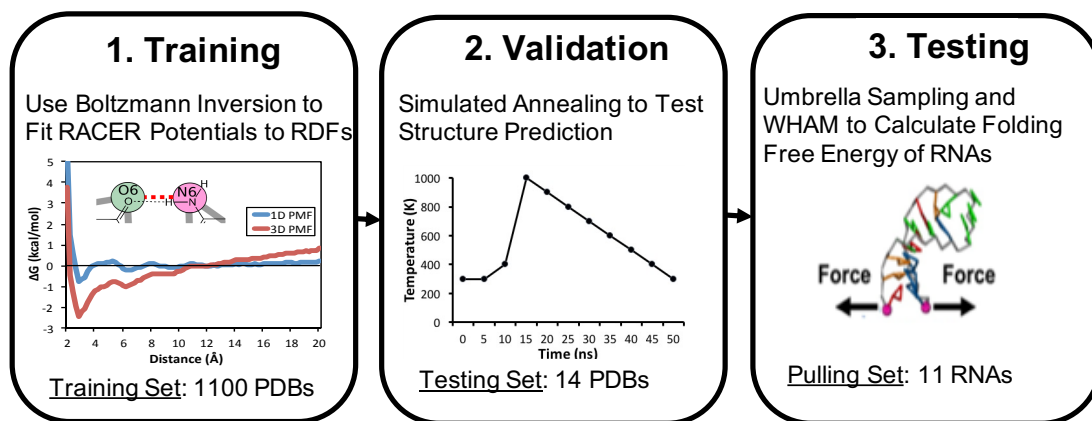


Figure 4.1: **RACER Model Development Flow Chart:** The RACER model was developed in three steps: 1) Training of RACER Potential Energy fitting RACER potentials to 1D RDFs 2) Validating structure prediction using simulated annealing of 14 selected RNA structures 3) Testing the free energy prediction of the RACER model by performing equilibrium pulling simulations on 11 selected RNA sequences.

4.2 Materials and Methods

4.2.1 5-Bead RACER Model

Forward mapping scripts composed of in-house python scripts and shell scripts map an all-atom PDB structure of an RNA and to a coarse-grained RACER structure. The RACER model is a five-bead model, where two pseudoatoms are used for the sugar and phosphate and three pseudoatoms are used to define a planar base (Fig. 4.2). There are a total of sixteen pseudoatom types that groups of atoms from the PDB can be mapped into: sugar pseudoatoms (S1, S2, G1, G2), phosphate pseudoatoms (P1, P2), sugar-connect pseudoatoms (CU, CG), base pseudoatoms (N6,

N2, CA, O2, O6), and ion pseudoatoms (MG, NA, K). The integer after a sugar or phosphate pseudoatom is used to define the direction of an RNA strand. The first sugar pseudoatom at the 5'end is G1 and the last sugar pseudoatom on the 3'end is G2. From the 5'to the 3'end the order of sugar and phosphate pseudoatoms is: G1, P2, S2, P1, S1, P2, S2,..., G2.

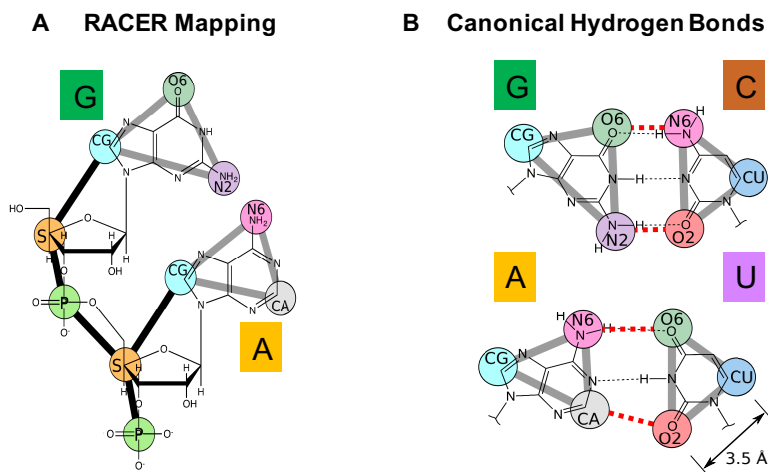


Figure 4.2: **RACER 5-Bead Coarse Grained Model:** A) For all nucleotides, the C4'atom on the ribose sugar is mapped to the center of the sugar pseudoatoms. The phosphorus atom is mapped to the phosphate pseudoatom. For purines, the C8 atom is mapped to a sugar-connect pseudoatom CG. For the pyrimidines, the C6 atom is mapped to a sugar connect pseudoatom CU. Bases are defined by the three pseudoatoms bonded together in a plane: Adenine (N6, CA, CG), Guanine (N2, O6, CG), Cytosine (O2, N6, CU), and Uracil (O2, O6, CU). B) Canonical Hydrogen Bonds between G-C and A-U RACER Nucleotides

4.2.2 RACER 1.0 and RACER 2.0 Potentials

The potentials used in the RACER 1.0 and RACER 2.0 models share the same functional forms. The main difference is that the hydrogen bond potential for RACER 2.0 has different values for U_o and $\sigma_{hb,eq}$ for different hydrogen-bond pairs, whereas RACER 1.0 uses the same U_o and $\sigma_{hb,eq}$ for all hydrogen bond pairs.

4.2.2.1 Bonded Potential Energy Functions

The RACER models include three bonded potentials: bond potential Eq. 4.1, angle potential Eq. 4.2, and torsion potential Eq. 4.3. The first two potentials are harmonic potentials which restrain the bonds between pseudoatoms. The torsion potential takes the form of a truncated Fourier series and governs the twist between pseudoatoms. For visualization,

$$U_{bond}(b) = k_{bond}(b - b_o)^2 \quad (4.1)$$

$$U_{angle}(\theta) = k_{angle}(\theta - \theta_o)^2 \quad (4.2)$$

$$U_{torsion}(\phi) = \sum_{n=1}^3 k_n(1 + \cos(n\phi - \delta_n)) \quad (4.3)$$

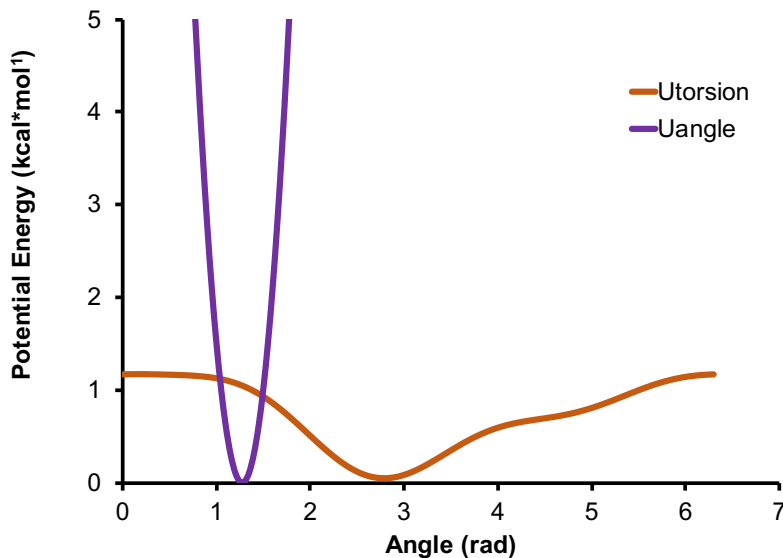


Figure 4.3: **RACER Bonded Potential Energy Function: Torsion and Angle:** Exemplar potential energy curves for Torsion and Angle RACER Potentials. ($k_1 = 0.25 \text{ kcal}\cdot\text{mol}^{-1}$, $k_2 = 0.08 \text{ kcal}\cdot\text{mol}^{-1}$, $k_3 = 0.04 \text{ kcal}\cdot\text{mol}^{-1}$, $\delta_1 = 6.3\text{rad}$, $\delta_2 = 2.1\text{rad}$, $\delta_3 = 4.9 \text{ rad}$, $k_{angle} = 20 \text{ kcal}\cdot\text{mol}^{-1}$, $\theta_{angle} = 1.3 \text{ rad}$)

4.2.2.2 Non-Bonded Potential Energy Functions

The RACER models include three non-bonded potentials: effective van der Waals potential,[163] Debye-Huckel potential,[163] and a hydrogen-bond potential. An earlier version of the RACER 1.0 model used the Buckingham potential, instead of the effective van der Waals potential.[21] During the development of the RACER 1.0 model, the Buckingham potential was found to be too repulsive at short distances.

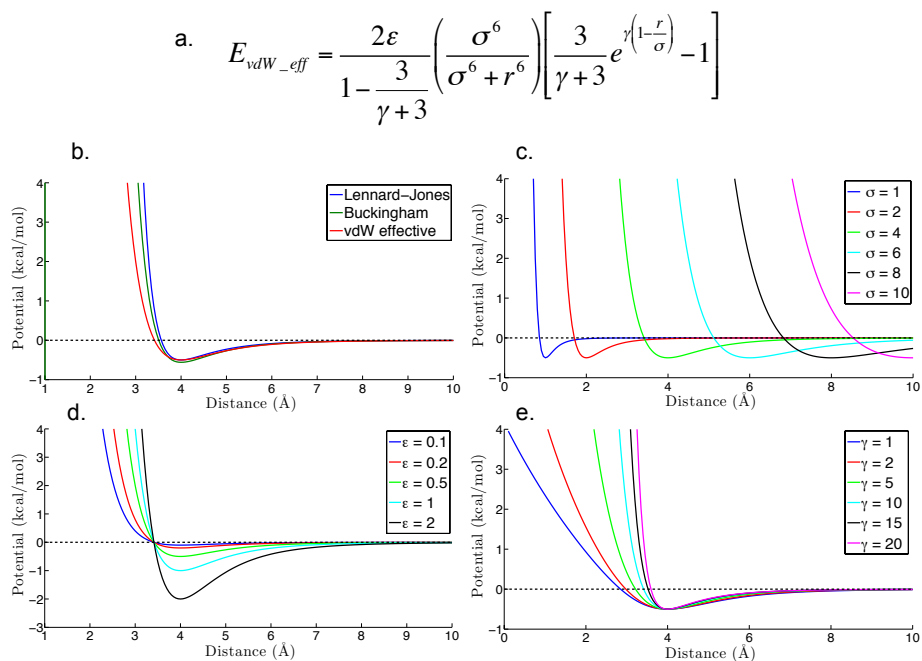


Figure 4.4: **Comparison of van der Waals Potential Energy Functions:** vdW-eff potential. (a.) Functional form of effective van der Waals Potential (b.)Effective potential compared to standard Lennard Jones and Buckingham potentials with minimum energy potential $\epsilon = 0.5 \text{ kcal}\cdot\text{mol}^{-1}$, minimum energy distance, $\sigma = 4\text{Å}$, and gamma of effective potential $\gamma = 10$. (c.) Effect of changing value of minimum energy distance, σ (d.) Effect of changing minimum energy potential, ϵ (e.) Effect of changing the short range behavior with parameter γ . For (c-e), unless stated, $\epsilon = 0.5 \text{ kcal}\cdot\text{mol}^{-1}$, $\sigma = 4\text{Å}$, and $\gamma = 10$.

$$U_{effvdW}(r_{ij}, \sigma_{ij}, \gamma_{ij}) = \frac{2\epsilon}{1 - \frac{3}{\gamma_{ij}+3}} \left(\frac{\sigma_{ij}^6}{\sigma_{ij}^6 + r_{ij}^6} \right) \left[\frac{3}{3 + \gamma_{ij}} e^{\gamma_{ij} \left(1 - \frac{r_{ij}}{\sigma_{ij}}\right)} - 1 \right] \quad (4.4)$$

$$\sigma_{ij} = \frac{\sigma_i + \sigma_j}{2}, \epsilon_{ij} = \epsilon_i \epsilon_j, \gamma_{ij} = \frac{\gamma_i + \gamma_j}{2} \quad (4.5)$$

The effective van der Waals potential uses a damped exponential function to model short-range repulsion Eq. 4.4, where ϵ is the potential well depth and σ is the equilibrium distance, and γ is a scaling parameter that allows for tuning of the slope of the short-range interactions. For comparison, the effective vdW potential, Lennard Jones, and Buckingham potentials are plotted (Fig. 4.4). The effects of varying the adjustable parameters σ , ϵ , and γ on the effective vdW potential for the RACER 1.0 model is shown (Fig. 4.4).

In order to capture electrostatic interactions as well as screening of the charged phosphate groups by the solvent, a Debye-Hückel potential energy form was used. Here, q_i is the charge of phosphate pseudoatom i , r_{ij} is the distance between atom i and atom j , ϵ is the permittivity, and ξ is the Debye length (Fig. 4.5). A permittivity of $25\epsilon_o$ was determined to be optimal for RNA folding under the new model potential, compared to $78\epsilon_o$ from a previous version of the RACER 1.0 model.[21]

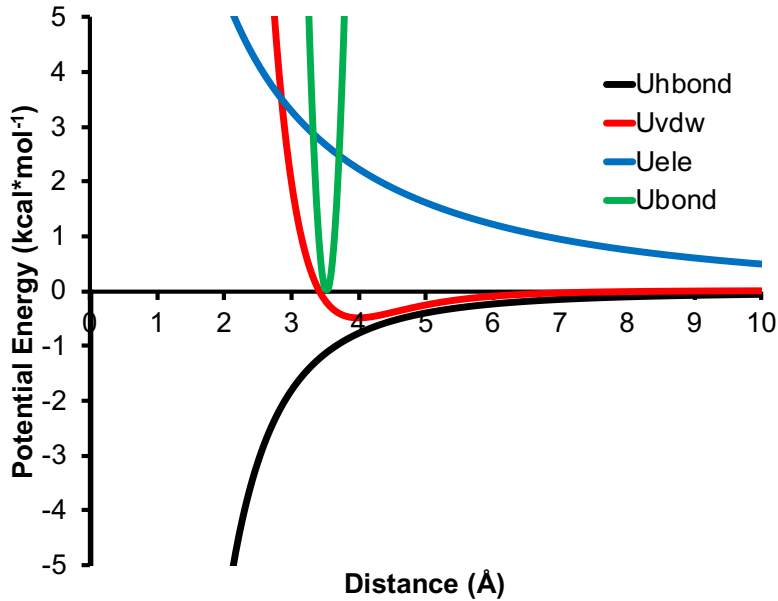


Figure 4.5: **RACER Potential Energy Functions with Distance Dependence:** Bond, Debye-Hückel, effective van der Waals, and Hydrogen Bond. ($\chi_{ele} = 10\text{\AA}$, $\epsilon_{ele} = 25\epsilon_o$, $U_o = 2.0 \text{ kcal}\cdot\text{mol}^{-1}$, $\sigma_{Hbond} = 2.9\text{\AA}$, $\epsilon_{vdW} = 10$, $\sigma_{vdW} = 4.0\text{\AA}$)

$$U_{ele}(r_{ij}) = \frac{q_i q_j e^{-\frac{r_{ij}}{\xi}}}{4\pi\epsilon_{ele}r_{ij}} \quad (4.6)$$

The form of the hydrogen bond potential of RACER 1.0 and 2.0 is the same. However, in the RACER 1.0 hydrogen-bond potential, all pseudoatoms have the same the potential well depth $\epsilon_{hb,max}$, is $2.0 \text{ kcal}\cdot\text{mol}^{-1}$ and the equilibrium distance is $\sigma_{hb,eq}$ of 2.9\AA are varied to allow for variation in hydrogen-bond energies for non-canonical hydrogen bonds. A cutoff of 6\AA is used in RACER 1.0 and RACER 2.0 models. The geometry of the hydrogen bonding pseudoatoms is accounted for by calculating the displacement between two pseudoatoms i and j , $r_{ij}^{\vec{}}$, and by calculating the angle formed by the norm of each base pair with the displacement vector between each base (Fig. 4.6).

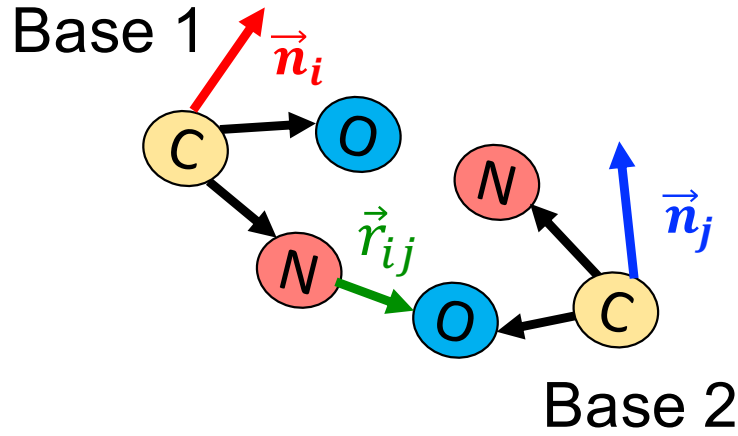


Figure 4.6: **Geometry of a RACER Hydrogen Bond:** First the displacement between two pseudoatoms i and j is calculated $r_{ij}^{\vec{}}$. Then the angle between the norm of to the plane formed by the atoms in Base 1 \vec{n}_i is dotted into the displacement vector $r_{ij}^{\vec{}}$ to obtain θ_i . Similarly the norm of to the plane formed by the atoms in Base 2 \vec{n}_j is dotted into the displacement vector $r_{ij}^{\vec{}}$ to obtain θ_j .

$$U_{\text{hbond}}(r_{ij}, \theta_i, \theta_j) = \frac{-U_o}{2} (1 + \cos 2(\theta_i + \theta_j)) \left(\frac{\sigma}{r_{ij}} \right)^6 \quad (4.7)$$

4.2.3 Improvements of RACER 2.0

In order to generate a new set of hydrogen bond parameters, experimental structures that include non-canonical base-pair geometries were needed. Selection of these non-canonical base pairs was based on definitions for twelve geometric families of hydrogen-bonded RNA structures by Leontis and Westhoff.[162] These families are distinguished by the type of bases forming hydrogen bonds (adenine, guanine, cytosine, or uracil), the edges of the base pairs forming hydrogen bonds (Watson-Crick, Hoogsteen, or Sugar), and the orientation of the glycosidic bonds (cis or trans). For all base pairs from these twelve geometric families, Leontis et. al. extracted exemplar experimental structures from the Protein Data Bank (PDB), available at the time of the publication (pre 2003), or modeled base pairs if an experimental structure was not available.

As a starting point for reparametrizing the RACER model, all base pairs provided in Leontis. et. al. were extracted from the PDB.[162] If there was no experimental structure available for any member of the twelve geometry families, a PDB structure with the same classification was extracted from the RNA Base Pair Catalog available on the Nucleic Acid Database (NDB) website. Exemplar base pairs deposited after the publication of Leontis. et. al. on the Base Pair Catalog were selected (post 2003). Additionally, twenty-four previously published base pair structures were used for optimization.[164] In total 187 base pairs structures were selected as a training set for new RACER hydrogen bond parameters. Using PyMol, these structures were processed by removing the phosphate groups and all but the C1'carbon of the sugar groups. After removing these atoms, hydrogens were added to the base structures.[149] Examples of base structure extracted from Zhang, et. al., Leontis et. al., and the RNA Base Pair Catalog are shown in Figure. 4.7.

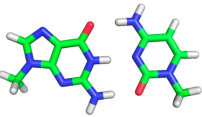
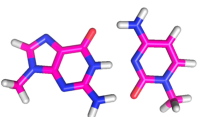

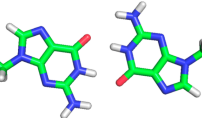
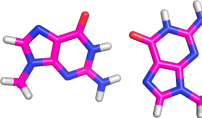
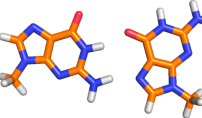
	Zhang JCTC 2017 (optimized)	PDB (pre 2003) (no optimization)	PDB (post 2003) (no optimization)
G-C (1 cis WC-WC)		 PDB 280d	 PDB 3r1e
G-G (3 cis WC-H)		 PDB 1jj2	 PDB 2grb

Figure 4.7: **Example Base Pairs Used for Optimizing Hydrogen Bonds in RACER 2.0:** The green base pair structures were obtained from a previously published work.[164] The base pair structures were optimized in the gas-phase. The pink base pair structures were obtained from a previously published work,[162] where structures were deposited in the PDB before 2003 (pre 2003). These structures were extracted from the Protein Data Bank (PDB) and the PDB ID is given below the structure. These base-structures were not optimized. The orange base pair structures were obtained from the RNA Base Pair Catalog on the Nucleic Acid Data Base (NDB), where structures were deposited in the PDB after 2003 (post 2003). These structures were extracted from the NDB and the PDB ID is given below the structure. These base-structures were not optimized. The first row shows base-base structures for G-C base pairs from the 1 cis Watson-Crick/Watson Crick hydrogen bond family. The first row shows base-base structures for G-G base pairs from the 3 cis Watson-Crick/Hoogsteen hydrogen bond family. Note, that one of the guanines in the G-G base-pair from the 3 cis Watson-Crick/Hoogsteen from Zhang et. al. rotated after optimization where the Watson-Crick Watson-Crick edges are facing each other.

4.2.3.1 Software Changes the TINKER Source Code

In order to accommodate calculation of hydrogen bonds spanning all twelve geometric families of hydrogen bonds, several changes were made to the implementation of the RACER model in the TINKER source code. First, the parameters for U_o , sigma, and cutoff distances were previously hard-coded in the implementation of the hydrogen bond potential (ecghbond.f, ecghbond1.f, ecghbond2.f, and ecghbond3.f). Additionally, U_o , $\sigma_{Hbond,eq}$, and the cutoff distances were the same regardless of the atomtype or geometry of the base pair. The hard-coding was removed from the ecghbond files. In place of the hard-coding, a new program called kcghbond.f reads in hydrogen bond definitions from a parameter file, aka key file.

Hydrogen bond definitions are read in from a user defined key file and sorted to avoid conflicting parameter definitions. In the key file, the keyword CGHBOND at the beginning of the line indicates a hydrogen bond definition is written on that line (see Fig. 4.8). After the CGHBOND, there are six integers followed by three floating point values. The first three integers are the pseudoatom types in the first base participating in the hydrogen bonding (either adenine, guanine, cytosine or uracil). The last three integers are the pseudoatoms types in the second base participating in the hydrogen bond. The first floating point integer is the value of U_o for that base pair. The second floating point is the value of $\sigma_{Hbond,eq}$ for that base pair. The last floating point is the cutoff distance in angstroms, where the hydrogen bond is automatically zero if the distance between a pseudoatom with the second integer type and the pseudoatom with the fifth integer type are more than 6 angstroms away. Per CGHBOND line, there is only one hydrogen bond defined between a pseudoatoms with a type two and six. In the RACER 2.0 model, hydrogen bond definitions read in from a user provided key file, sorted, parameters are stored in an array.

The CGHBOND entries are sorted into a unique definition before they are stored in the array in the kcghbond.f function. As an example, an unsorted CGHBOND definition is "CGHBOND 9 8 7 4 3 2 2.0 2.9 6.0". The sorting algorithm will sort the first three integers 9 8 7 to 7 8 9, where the integer in the first position is less than the integer in the third position (Fig. 4.9). The center integer is left alone, since this pseudotaom type is participating in hydrogen bonding. The sorting algorithm performs the same sorting on the next three integers 4 3 2 to 2 3 4. Finally, to avoid

```

# RACER 2.0: Force Field Parameters for all Potentials
# (Except for Hydrogen Bond Potential)
Parameters RNA_CG_05_10_16_b_1p4eps_2p0tors.prm
# Keywords to use Debye Hückel Electrostatic Model
DEBYE-HUCKEL
# Use dielectric constant of 25
dielectric 25
# Use debye length of 10Å
debye-length 10
# Use the Effective van der Waals Potential
vdwtype LJ-LP

# Example Hydrogen Bond Definition for RACER 2.0
# Pseudoatoms in Base A: 2 3 4
# Pseudoatoms in Base B: 7 8 9
# Defines a hydrogen bond between pseudoatom types 3 and 8
#  $U_o = 2.0 \text{ kcal mol}^{-1}$ 
#  $\sigma_{hb,eq} = 2.9 \text{ \AA}$ 
# cutoff = 6.0Å
CGHBOND 2 3 4 7 8 9 2.0 2.9 6.0

```

Figure 4.8: **Example RACER 2.0 Key File:** Blue text indicates a comment, Bolded black text indicates key words in the RACER 2.0 key file. The first line of the key file defines the force field parameters for all but the hydrogen bond potential. Lines 2-5, specify the kind of non-bonded potential energy functions that will be used for the simulations. In the key file, the keyword CGHBOND at the beginning of the line indicates a hydrogen bond definition is written on that line (see Fig. 4.8). After the CGHBOND, there are six integers followed by three floating point values. The first three integers are the pseudoatom types in the first base participating in the hydrogen bonding (either adenine, guanine, cytosine or uracil). The last three integers are the pseudoatoms types in the second base participating in the hydrogen bond. The first floating point integer is the value of U_o for that base pair. The second floating point is the value of $\sigma_{Hbond,eq}$ for that base pair. The last floating point is the cutoff distance in angstroms, where the hydrogen bond is automatically zero if the distance between a pseudoatom with the second integer type and the pseudoatom with the fifth integer type are more than 6 angstroms away. Per CGHBOND line, there is only one hydrogen bond defined between a pseudoatoms with a type two and six. In the RACER 2.0 model, hydrogen bond definitions read in from a user provided key file, sorted, pareamters are stored in an array.

done when parameterizing non-canonical base pairs. In addition, hydrogen bonds between terminal base pairs were excluded, which would underestimate the stability of the terminal ends of a RNA structure.

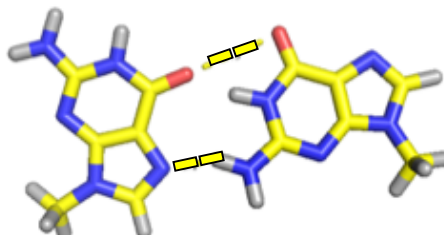
The improved RACER 2.0 model allows intramolecular hydrogen bonds to occur between any pseudoatom type in the base with any other pseudoatom type in another base, while avoiding intramolecular hydrogen bonds between stacked bases. Preventing hydrogen bonds from forming between stacked bases is achieved using neighbor lists (Fig. 4.9). Previously, the RACER model implemented up to 1-5 neighbor lists to use in the effective van der Waals potential. Additional 1-6 and 1-7 neighbor lists were implemented. The tinker source code includes 1-2 through 1-5 neighbor lists in attach.f code. The 1-6 neighbor lists were implemented by searching for 1-2 neighbors among atoms in the 1-6 neighbor lists, and the 1-7 neighbor lists were implemented by searching for the 1-2 neighbors among pseudoatoms in the 1-6 neighbor lists.

4.2.3.2 Calculating Binding Energies using QChem and RACER

In order to determine the strength of non-canonical hydrogen bonds, quantum mechanical interaction energies were calculated for all of the extracted base-pairs. To calculate the interaction energy of each base pair using QChem5 Equation 4.8 was used. First, the energy of the dimer was calculated E_{A+B} . Then, the energy of each monomer was calculated in the presence of the basis set of the other base (E_A and E_B). The energy of the monomers was subtracted from the energy of the dimer to obtain the interaction energy 4.8. All energies were single-point calculations with the smd implicit solvent model. The single-point calculations were performed with the RIMP2 method and basis cc-pvtz and auxillary basis rimp2-cc-pvtz. The same base pair structures were forward mapped to a RACER representation using in-house python scripts. The RACER binding energy is calculated by calling the analyze function and summing the value of the effective vdW potential energy and the hydrogen-bond energy for that base pair.

$$E_{int} = E_{A+B} - E_A - E_B \quad (4.8)$$

G-G Base Dimer Extracted from PDB



Forward Mapping PDB to RACER Model

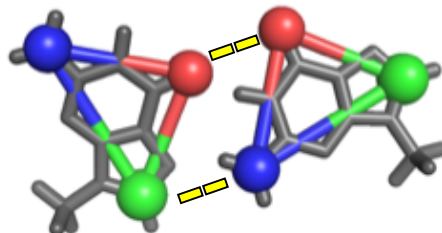


Figure 4.10: **Calculating Binding Energy using QChem and RACER:** Exemplar G-G cis Watson-Crick Hoogsteen base pair extracted from PDB ID: 1JJ2. The hydrogen bonds predicted are shown in yellow (top). The base structure is forward mapped to the RACER representation and the RACER 2.0 hydrogen bonds are shown (bottom). To calculate the interaction energy of each base pair using QChem5 Equation 4.8 was used. First, the energy of the dimer was calculated E_{A+B} . Then, the energy of each monomer was calculated in the presence of the basis set of the other base (E_A and E_B). The energy of the monomers was subtracted from the energy of the dimer to obtain the interaction energy 4.8. All energies were single-point calculations with the smd implicit solvent model. The single-point calculations were performed with the RIMP2 method and basis cc-pvtz and auxiliary basis rimp2-cc-pvtz. The same base pair structures were forward mapped to a RACER representation using in-house python scripts. The RACER binding energy is calculated by calling the analyze function and summing the value of the effective vdW potential energy and the hydrogen-bond energy for that base pair.

4.2.3.3 Optimizing Non-Canonical Hydrogen Bonds

After performing initial binding energy calculations using QChem and the RACER 1.0 model, correlation plots were generated. However, the RACER 1.0 model was found to underestimate the hydrogen bond energy of base structures containing non-canonical hydrogen bonds and overestimate the effective van der Waals potential energy, leading to unfavorable binding energies ($R^2 < 0.01$)(Fig. 4.11A).

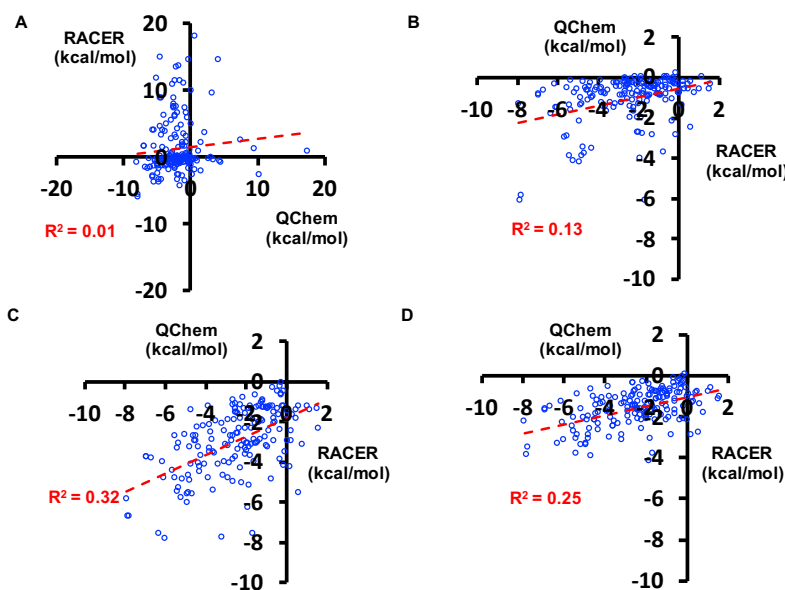


Figure 4.11: **Correlation Plot of RACER and QChem Binding Energies for Different Optimization Procedures:** A) Binding Energy calculated for 187 base-base structures using RACER 1.0 B) Binding Energy calculated for minimized 175 base-base structures using RACER 1.0 (12 outliers removed) C) Binding Energy calculated for minimized 175 base-base structures using 56 hydrogen definitions, no optimization (12 outliers removed) D) Binding energy calculated for minimized 175 base-base structures using 56 hydrogen definitions, with optimization method 2 (12 outliers removed)(see Fig. 4.20)

After performing initial binding energy calculations using QChem and the RACER 1.0 model, correlation plots were generated. However, the RACER 1.0 model was found to underestimate the hydrogen bond energy of base structures containing non-canonical hydrogen bonds and overestimate the effective van der Waals potential energy, leading to unfavorable binding energies ($R^2 < 0.01$)(Fig. 4.11A).

To improve the agreement between QChem binding energies and RACER 2.0 binding energies, an optimization procedure was performed. First, the coarse-grained base-pair structures were coarsely minimized with an RMS gradient per atom criterion of 1. The minimized structures were superposed on the initial structures and an average RMSD of 0.18Å, with the highest RMSD of 0.96Å was found. Twelve base-base structures with QChem binding energies exceeding 2 kcal·mol⁻¹ and negative RACER binding energies (out of 187 structures) were designated as outliers and removed from the structure set (see Figure 4.11A and Figure 4.11B). The remaining 175 minimized structures were used for further optimization.

The optimization procedure was performed using in-house python scripts which minimized the RMSE between the QChem binding energies and the RACER binding energies. The inputs to this optimization program were the QChem binding energies, CGHBOND definitions for canonical and non-canonical hydrogen bonds with initial values for Uo and sigma, and bounds for Uo and sigma values. The minimize function in scipy was used with the Sequential Least Squares Programming (SLSQP) method,[165] ftol of 1e-3, eps of 1e-4.

A total of 83 unique hydrogen bond definitions were found to occur between any pair of pseudoatoms in the bases of a RACER nucleotide. For each of these definitions, Uo and sigma could be defined independently. As a result, a total of 166 parameters could be used to optimize the non-canonical hydrogen bonds. To avoid overfitting, hydrogen bond definitions were grouped together if the second and fifth pseudoatom types were the same. This resulted in 28 groups of hydrogen bond definitions. During the optimization process, the Uo and sigma for these groups were optimized, resulting in total of 56 parameters. Additionally, Uo and sigma were bounded. Sigma was allowed to vary within ±20% of its original value of 2.9Å and Uo was allowed to vary within ±20% of its original values of 2.0 kcal·mol⁻¹. The cutoff distance of 6Å was kept the same for all hydrogen bond groups. The tight bounds for Uo and sigma were enforced in order to maintain the strength of the canonical hydrogen bonds while allowing for hydrogen bonds between different faces the base to be captured.

4.2.4 Simulated Annealing Procedure

Annealing simulations are used to generate a large number of unfolded structures for each RNA. The simulated annealing protocol is a series of 5ns MD simulations with a 1fs time step at temperatures in order of 298(K), 400, 1000, 900, 800, 700, 600, 500, 400, 298 K. Each of these structures is then energy minimized to 0 K. The energy and RMSD (with respect to the native structure) of each structure are used to characterize the energy landscape (see Eq. 4.9).

$$\text{RMSD}(\vec{v}, \vec{w}) = \sqrt{\frac{1}{N} \sum_{i=1}^N \|\vec{v}_i - \vec{w}_i\|^2} \quad (4.9)$$

4.2.5 Equilibrium Pulling Simulations

Single-molecule pulling experiments are modeled by performing equilibrium pulling simulations using the umbrella sampling method.[166] The RNA molecule was pulled apart from its terminal ends. A harmonic potential with a $1 \text{ kcal}\cdot\text{mol}^{-1}\text{\AA}^{-2}$ spring constant restrains the sugar pseudoatoms (C4'sugar atomic site) at the ends of the RNA. Simulations were run with end-to-end extensions from 5.5\AA up to fully extended lengths ($59.5, 76.5, 86.5, 106.5,$ and 307.5\AA for 10, 12, 14, 18, and 52 nt hairpins assuming 5.9\AA per nt contour length) with a spacing of 1\AA between windows.

Duplexes are similarly pulled apart from the sugar pseudoatoms at one terminal end with a $1 \text{ kcal}\cdot\text{mol}^{-1}\text{\AA}^{-2}$ spring constant; the other terminal end is restrained between two terminal sugar pseudoatoms with a $1 \text{ kcal}\cdot\text{mol}^{-1}\text{\AA}^{-2}$ spring constant. Duplex extensions range from 5.5\AA up to fully extended lengths ($80.5, 100.5,$ and 124.5\AA for 6, 8, and 10 base pair duplexes respectively) with umbrella window spacing of 1\AA . For the duplexes and shorter hairpins of size 10 and 18 nt, $1 \mu\text{s}$ of Molecular Dynamics was run for each window. For the TAR hairpin, 100 ns was found to be sufficient given the longer end-to-end extension (more windows) needed. We used a 4 fs time step for pulling simulations. From the umbrella simulations, the free energy landscapes were computed by the Weighted Histogram Analysis Method (WHAM).[167] The energy landscape as computed by a Monte Carlo bootstrap error analysis in the WHAM program by Alan Grossfield.[168] Although exact energy landscapes at equilibrium for both TAR and melting free energy helices are unknown, folding free energies can

be computed according to Equation 4.10. The folded free energy, ΔG , is found by integrating over all folded conformations at end-to-end extension r with free energy $\Delta\omega$. Folded free energy is then normalized to volumetric entropy, with standard state volume V_{ref} of 1660 \AA^3 . The Boltzmann constant multiplied by temperature 298K is $k_B T$.

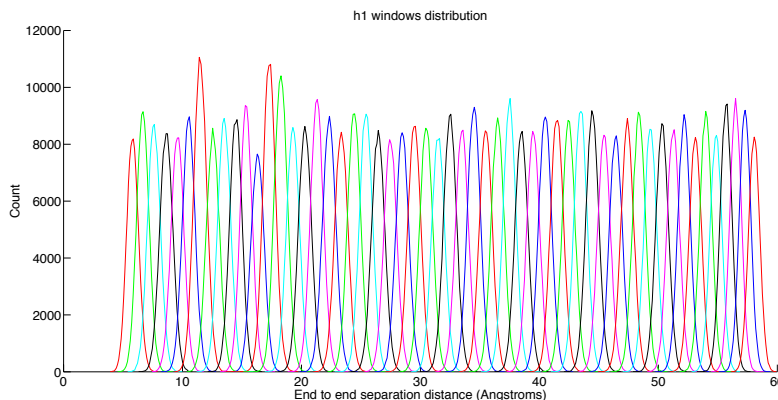


Figure 4.12: **Example Umbrella Sampling Distribution for a RNA Hairpin:** Sampling distribution of each umbrella sampling window for h1 hairpin (sequence 5′ GGCGUAAGCC3′, with bolded sequence forming the loop). The separation distance between windows was 1 \AA for all RNAs.

$$\Delta G = k_B T \ln(8\pi^2 V_{ref}) - k_B T \ln\left(\int (4\pi r^2 dr e^{\frac{-\Delta\omega(r)}{k_B T}})\right) \quad (4.10)$$

4.3 Results

4.3.1 Validating and Testing the RACER 1.0 Model

4.3.1.1 3D Structure Prediction of Small RNAs

In order to test the ability of the RACER 1.0 model to predict 3D structure, a simulated annealing procedure was performed on 14 RNAs with known experimental structures (description in materials and methods). Plotting the potential energy versus RMSD for these fourteen RNAs, revealed that the potential energy landscapes of all fourteen RNAs formed a funnel towards low potential energy and low RMSD. From each potential energy landscape the lowest potential energy structure was extracted,

superimposed on the experimental PDB structure, and the RMSD was calculated. Thirteen out of the fourteen RNAs tested, had an RMSD within 5\AA of the experimental structure, showing that the previous RACER model was capable of 3D structure prediction for small RNAs. However, for structures containing non-canonical hydrogen bonds such as PDB ID: 1F5G, the predicted 3D structure was too compact with RMSD of 8.59\AA . In order to improve the 3D structure prediction capabilities of the RACER model, parameterized non-canonical hydrogen bonds are needed.

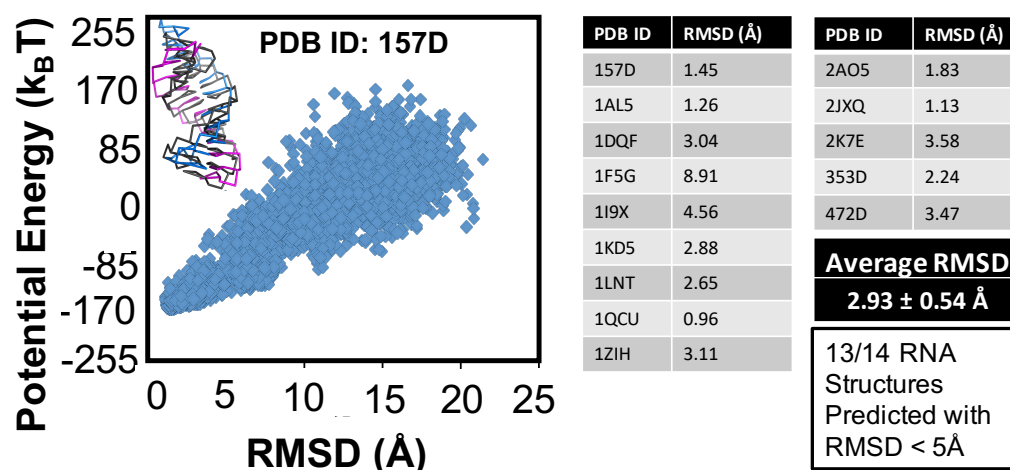


Figure 4.13: **3D Structure Prediction from Simulated Annealing:** Representative energy landscape of an accurately predicted RNA structure 157D. Plotting the potential energy versus RMSD for these fourteen RNAs, revealed that the potential energy landscapes of all fourteen RNAs formed a funnel towards low potential energy and low RMSD. From each potential energy landscape the lowest potential energy structure was extracted, superimposed on the experimental PDB structure, and the RMSD was calculated. The RACER minimum free energy structure is shown in blue and magenta aligned to the PDB structure shown in black. Five thousand structures over 50 ns are shown, with each structure is energy minimized before plotting. Note the funnel toward low energy and low RMSD structures. The RMSD of lowest energy structure for 157D is 1.45\AA . Thirteen out of the fourteen RNAs tested, had an RMSD within 5\AA of the experimental structure, showing that the previous RACER model was capable of 3D structure prediction for small RNAs. However, for structures containing non-canonical hydrogen bonds such as PDB ID: 1F5G, the predicted 3D structure was too compact with RMSD of 8.59\AA . In order to improve the 3D structure prediction capabilities of the RACER model, parameterized non-canonical hydrogen bonds are needed.

4.3.1.2 Folding Free Energy Prediction

The ability of RACER 1.0 to predict folding free energies was tested with equilibrium pulling MD simulations. Using umbrella sampling and weighted histogram analysis (description in materials and methods), eleven RNAs with known experimental melting free energy were unfolded. Plotting the free energy versus end-to-end distance reaction coordinate, the unfolding free energy was determined by taking the difference between the free energy of the folded and unfolded state. For comparison, the folding free energy for each of these eleven RNAs was also calculated using Mfold.[157] Although, the correlation between the folding free energy of the previous RACER model is lower than Mfold ($R^2 = 0.93$ for the RACER model and $R^2 = 0.96$). Mfold overpredicts the stability of these RNAs. Overall, the goal of the RACER 1.0 model is not 2D structure prediction, rather to predict 3D structure and energetics of small to medium RNAs.

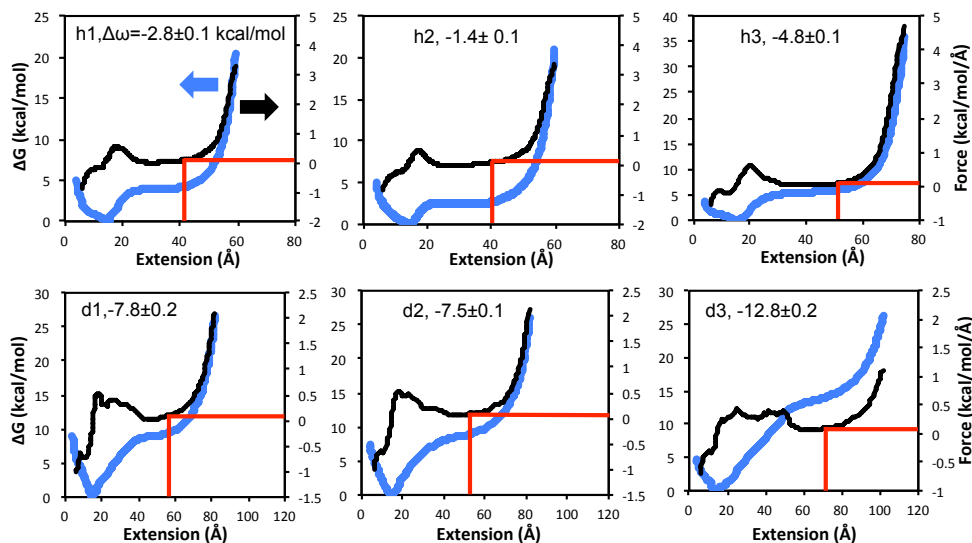


Figure 4.14: **Equilibrium Pulling Profiles for small RNAs:** The equilibrium pulling free energy profile (blue) of hairpins h1-h3 (top) and duplexes d1-d3 (bottom) computed with WHAM using the RACER model. Equilibrium pulling simulations were run for $1\mu\text{s}$ for each window, with a 1\AA window separation. The unfolded state is determined as the state right before the force (derivative of the free energy, curves shown in black) sharply increases from low ($< 0.1 \text{ kcal}\cdot\text{mol}^{-1}\text{\AA}^{-1}$) to high due to overstretching. $0.1 \text{ kcal}\cdot\text{mol}^{-1}\text{\AA}^{-1}$ and the location of the unfolded state are denoted by the red lines. The folding free energy (ΔG_f with units $\text{kcal}\cdot\text{mol}^{-1}$) is included for each RNA. A 4\AA running average of force (black curves) is shown to eliminate noise.

Hairpin	Length (nt)	Expt. ΔG (kcal/mol)	Mfold ΔG (kcal/mol)	RACER ΔG (kcal/mol)	Length per window	Computing Time (hr)
h1	10	-3.5 ± 0.3	-5.3	-2.8 ± 0.12	1 μ s	22
h2	10	-0.3 ± 0.1	+0.9	-1.4 ± 0.11	1 μ s	
h3	12	-4.4 ± 0.2	-3.4	-4.8 ± 0.14	1 μ s	60
h4	14	-2.2 ± 0.08	-2.2	-5.7 ± 0.15	1 μ s	
h5	18	-8.2 ± 0.2	-8.4	-7.9 ± 0.22	1 μ s	
TAR	52	$\approx -21.5 \pm 4.3$	-31.3	-29.7 ± 0.36	0.1 μ s	48
Duplex	Length (bp)					
d1	6	-7.56 ± 0.3	-11.4	-7.8 ± 0.15	1 μ s	30
d2	6	-4.95 ± 0.2	-9.8	-7.5 ± 0.14	1 μ s	
d3	8	-12.32 ± 1.2	-17.0	-12.8 ± 0.17	1 μ s	
d4	8	-10.11 ± 0.2	-14.7	-11.1 ± 0.18	1 μ s	
d5	10	-12.69 ± 0.5	-18.1	-14.0 ± 0.19	1 μ s	74
					Total: 860 μ s	

Figure 4.15: **Equilibrium Pulling Results with RACER Model:** Using umbrella sampling and weighted histogram analysis (description in materials and methods), eleven RNAs with known experimental melting free energy were unfolded. Plotting the free energy versus end-to-end distance reaction coordinate, the unfolding free energy could be determined by taking the difference between the free energy of the folded and unfolded state. For comparison, the folding free energy for each of these eleven RNAs was also calculated using Mfold, a secondary structure prediction program that relies on nearest-neighbor energy parameters.[157] Although, the correlation between the folding free energy of the previous RACER model is lower than Mfold (R^2 of 0.93 for the RACER model and R^2 of 0.96). Mfold overpredicts the stability of these RNAs. Error is taken from a Monte Carlo bootstrap error analysis as implemented in the WHAM program by Grossfield.[168]

4.3.2 Validating and Testing the RACER 2.0 Model

4.3.2.1 Adding Non-Canonical Hydrogen Bonds

Capturing the complexity of 3D RNA structure using computational methods requires accurate physical models of interactions stabilizing the RNA. A first attempt at incorporating non-canonical hydrogen bonds in the RACER model was to simply implement more hydrogen bonds definitions (Fig. 4.16). These 80 new hydrogen bond definitions included hydrogen bonds formed at hoogsteen and sugar edges to be captured. Adding additional hydrogen bonds, without adjusting the U_0 and sigma values, improved the correlation between RACER and QChem as shown in Fig. 4.11C).

To test the effect of adding non-canonical hydrogen bonds, 10 small RNAs consisting of duplexes and hairpins used for equilibrium pulling simulations in the previous RACER model [161] and seven medium RNAs (PDB IDs: 1j6s, 1kxk, 1y26, 2kbp, 2m8k, 3zp8, and 4oji) were selected (Fig. 4.17). All of these structures were mapped into the RACER coarse-grained representation and their binding energy for all twelve structures were calculated with 3 hydrogen bond definitions or 56 hydrogen bond definitions. The difference in binding energies is defined as the binding energy with 56 hydrogen bond definitions minus the binding energy with 3 hydrogen bond definitions.

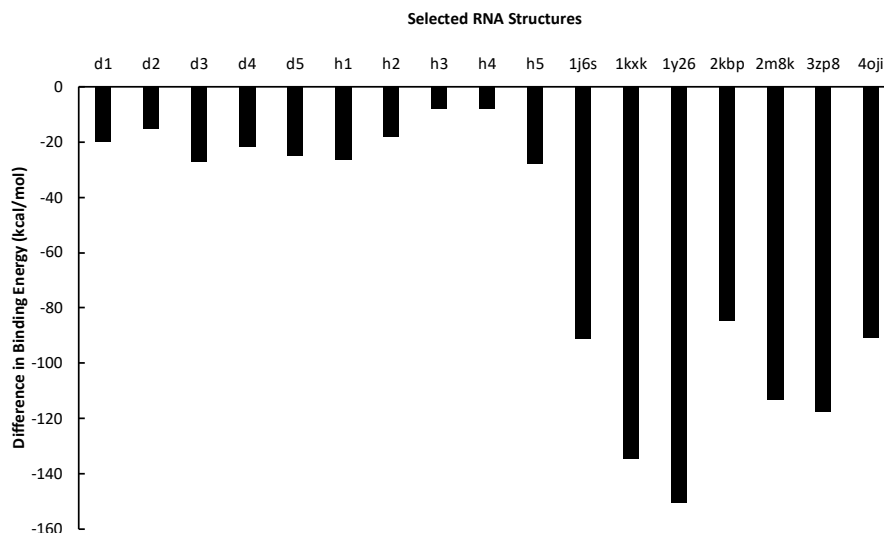


Figure 4.16: **Adding Non-Canonical Hydrogen Bonds to RACER 2.0:** Binding Energies were calculated for 10 small RNAs (d1-d5, h1-h5) consisting of duplexes and hairpins used for equilibrium pulling simulations in the previous RACER 1.0 model [161] and seven large RNAs (PDB IDs: 1j6s, 1kxk, 1y26, 2kbp, 2m8k, 3zp8, and 4oji) were selected (Fig. 4.17). All of these structures were mapped into the RACER coarse-grained representation and their binding energy for all twelve structures were calculated with 3 hydrogen bond definitions or 56 hydrogen bond definitions. The difference in binding energies is defined as the binding energies with 56 hydrogen bond definitions minus the binding energies with 3 hydrogen bond definitions

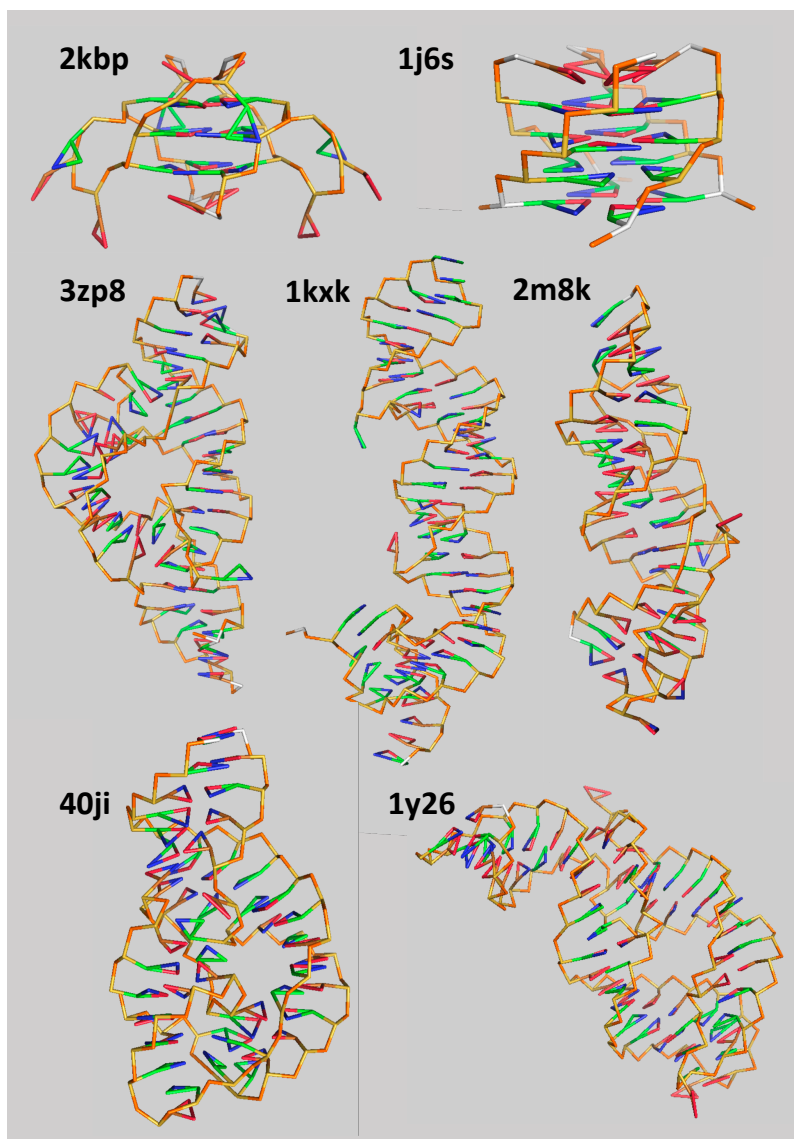


Figure 4.17: **RACER Coarse-Grained Representation of Seven Large RNAs:** PDB IDs for each of the seven large RNAs is given in bold at the top left of each figure. PDB ID: 2kbp is a solution NMR structure of a TERRA RNA G-Quadruplex (Terra RNA-G4),[169] PDBID: 1j6s is a crystal structure of a RNA Tetraplex,[170] PDB ID: 3zp8 is a crystal structure of a hammerhead ribozyme,[171] PDB ID: 1kxk is Domain 5 and Domain 6 of the Yeast ai5gh Group II self-splicing intron,[172] PDB ID: 2m8k solution NMR structure of a RNA triplex,[173] PDB ID: 4oji Crystal structure of a twister ribozyme,[174] and PDB ID: 1y26 crystal structure of a A-riboswitch adenine complex.[175]

4.3.2.2 Optimization with Non-Canonical Hydrogen Bonds

In order to have a baseline to compare the optimized hydrogen bond parameters, binding energies using the RACER 1.0 model, where only canonical hydrogen bonds were defined, were calculated. Plotting using the previous RACER 1.0 model the binding energy of these base-base structures were calculated and plotted against the QChem binding energy of the same base-base structures (see Fig. 4.11A). The RACER model underestimated the strength of these base-base structures, and the previous RACER model predicted that many of these base-base structures would not form (positive RACER binding energies) while QChem predicted that the formation of these base-structures was favorable) as seen in the cluster of data points in the second quadrant of part A. After performing a coarse energy minimization on the RACER base-base structures, the RACER 1.0 model predicted that the majority of these 175 base-base structures would form, but the calculated binding energy of these structures was about half of the binding energy predicted by QChem.

As seen in figure 4.18, including non-canonical hydrogen bonds decreases the stability of the duplexes (d1-d5) by as much as $4 \text{ kcal}\cdot\text{mol}^{-1}$ ($7.2 \text{ k}_B\text{T}$). The trend can be explained by the non-canonical hydrogen bonds being slightly weaker than the canonical hydrogen bonds in the helix. The addition of non-canonical hydrogen bonds in the hairpins increases their stability (h1-h5). This trend can be explained by non-canonical hydrogen bonding in the loop sequence of the hairpins. These loops are non-canonical hydrogen bonded structures which would be missed if these non-canonical hydrogen bonds are not explicitly defined. Overall, adding additional hydrogen bond definitions helps to stabilize non-canonical structures, however due to the additional allowed geometries of hydrogen bonds, there were ≈ 6 times more hydrogen bonds recognized for the small RNAs and ≈ 7 times more hydrogen bonds recognized for the large RNAs. As a result, when adding additional hydrogen bonds, care must be taken to avoid making these hydrogen bonds too strong causing overstabilization of non-canonical structures.

4.3.2.3 Testing Parameters on Small and Large RNAs

Using the optimization procedure described in the materials and methods, multiple optimization schemes were attempted. Initially, the canonical and non-canonical

hydrogen bonds were varied with the same bounds max hydrogen bond strength (U_o) from $0.01 \text{ kcal}\cdot\text{mol}^{-1}$ to $4.0 \text{ kcal}\cdot\text{mol}^{-1}$, with the sigma distance was varied from 2.3 to 3.5\AA and a cutoff distance of 6\AA . However, this wide range of hydrogen bond values led to underestimation of canonical hydrogen bonds ($> 0.1 \text{ kcal}\cdot\text{mol}^{-1}$). In order to preserve the strength of canonical hydrogen bonds, these hydrogen bond definitions were bounded from 1.5 to $2.0 \text{ kcal}\cdot\text{mol}^{-1}$, while non-canonical hydrogen bonds were bounded from 0.1 to $0.5 \text{ kcal}\cdot\text{mol}^{-1}$. As compared with using the same U_o and sigma parameters for canonical and noncanonical hydrogen bonds (Figure 4.11C), the correlation between the binding energies from the optimized RACER model and QChem calculations is slightly worse ($R^2 = 0.32$ vs. $R^2 = 0.25$). However, a better test of these parameters is their ability to predict energetics in 3D RNA structures.

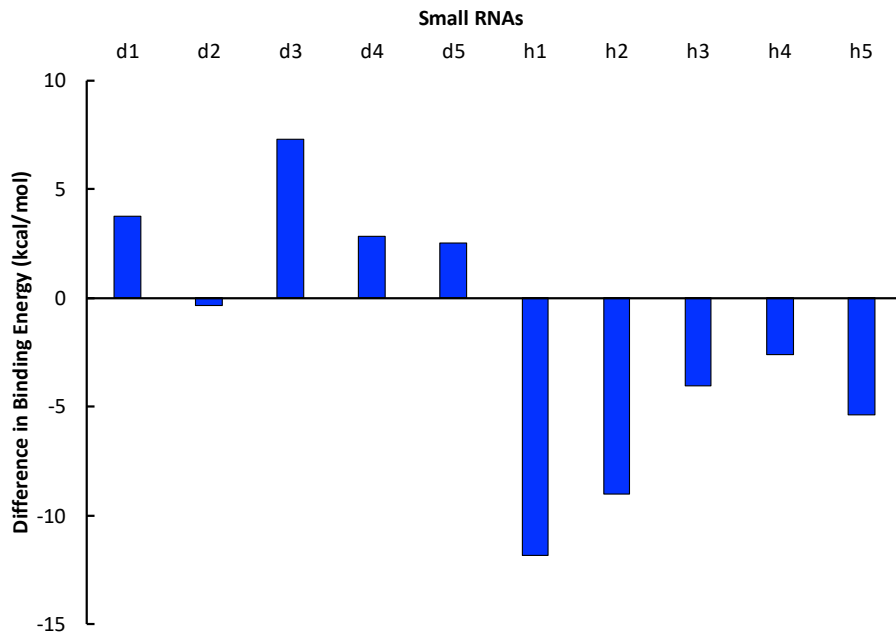


Figure 4.18: **Difference Plot of Binding Energies for small RNAs:** Difference in Binding Energies with Optimized 56 Hydrogen Bond Parameters and Binding Energies with previous RACER Model. The addition of non-canonical hydrogen bonds destabilizes the duplexes (d1-d5) and further stabilizes the hairpins (h1-h5). Figure generated using optimization 2 (see Fig. 4.20)

Calculating the difference in binding energy between the optimized RACER hydrogen bond parameters and the original Scientific Reports hydrogen bond param-

eters is shown for small RNAs d1-d5, h1-h5 (Fig. 4.18, for sequences of d1-d5 and h1-h5 and RACER structures see SI of [161]). Using the optimized hydrogen bond parameters, higher binding energies are predicted with the improved model for duplex sequences (d1-d5) with an average difference in binding energy difference of $9.4 \text{ kcal}\cdot\text{mol}^{-1}$. Since duplex structures are dominated by canonical hydrogen bonding and the optimized parameters are bounded above by $2.0 \text{ kcal}\cdot\text{mol}^{-1}$, it makes sense that the duplexes would be underpredicted with the improved model. However, $9.4 \text{ kcal}\cdot\text{mol}^{-1}$ too high, and this understabilization of duplex structures can be improved by further bounding the canonical hydrogen bonds to below $2 \text{ kcal}\cdot\text{mol}^{-1}$. For the hairpin sequences (h1-h5) the improved model predicts a lower binding energy with an average binding energy difference of $-2.0 \text{ kcal}\cdot\text{mol}^{-1}$. Since the RACER model tends to under predict the stability of the tested RNAs, this trend is in the right direction for improving prediction of energetics from the RACER model.

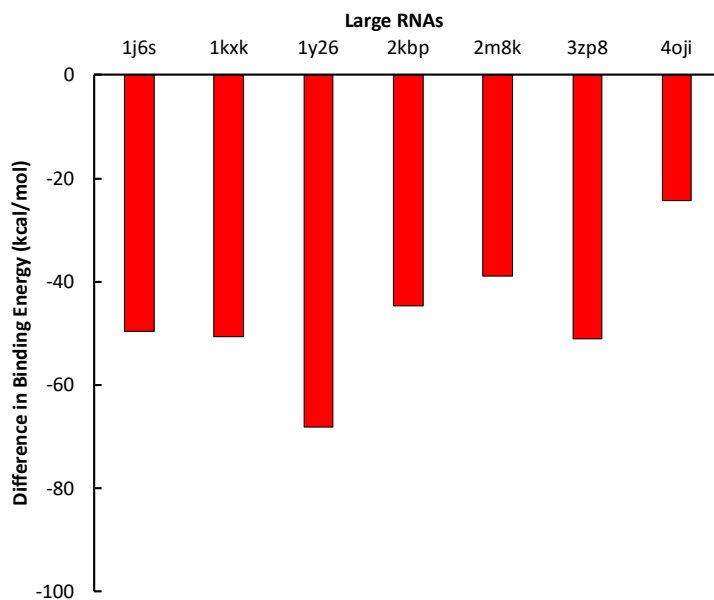


Figure 4.19: **Difference Plot of Binding Energies for Large RNAs:** Difference in Binding Energies with Optimized 56 Hydrogen Bond Parameters and Binding Energies with previous RACER Model. Figure generated using optimization 2 (see Fig. 4.20)

For large RNAs the effect of adding non-canonical hydrogen bonds is more pronounced (Fig. 4.19). Adding non-canonical hydrogen bonds increases the stability of by as much as ≈ 50 kcal·mol⁻¹ for seven large RNAs. For PDB ID: 2kbp, there are three g-quartets (Fig. 4.21C). In each of these quartets there are four pairs of G-G cis Watson-Crick Hoogsteen hydrogen bonds, contributing ≈ 16 kcal·mol⁻¹ the coarse-grained energy per quartet, leading to ≈ 48 kcal·mol⁻¹ increase in stability when adding non-canonical hydrogen bonds. The same arguments can be made for PDB ID: 1j66 which is a RNA tetraplex containing G-tetrad, A-tetrad, U-tetrad, and a G-U octad. The other large RNAs are mostly helical structures. However, the accumulation of non-canonical hydrogen bonds increases their stability as well.

4.3.2.4 Energy Change Between Folded and Unfolded State

The ultimate test of which optimized parameters are better for capturing non-canonical RNA structure and energetics is their performance during with molecular dynamics simulations. As an initial test, the energy (sum of the vdW and hydrogen bond energies) of the folded RNA (coarse grained RNA structure) and the binding energy of the unfolded RNA were calculated for two selected small RNAs. These calculations were performed using seven different parameter sets (Fig. 4.20). Since the structure of the unfolded state is dynamic, the binding energy over the last 100 frames of a 1us trajectory were used. The difference in the folded energy and the unfolded energy was calculated for all seven optimized key files. The energy of the small RNAs calculated using RACER 1.0 parameters was subtracted from all of these energies to get difference of differences in energies. The second and fourth optimization procedures show moderate increase in stability for the hairpin. However, the fourth optimization procedure understabilizes the duplex by 6 kcal·mol⁻¹. The right optimization range for non-canonical and canonical hydrogen bonds likely lies between these two optimization procedures.

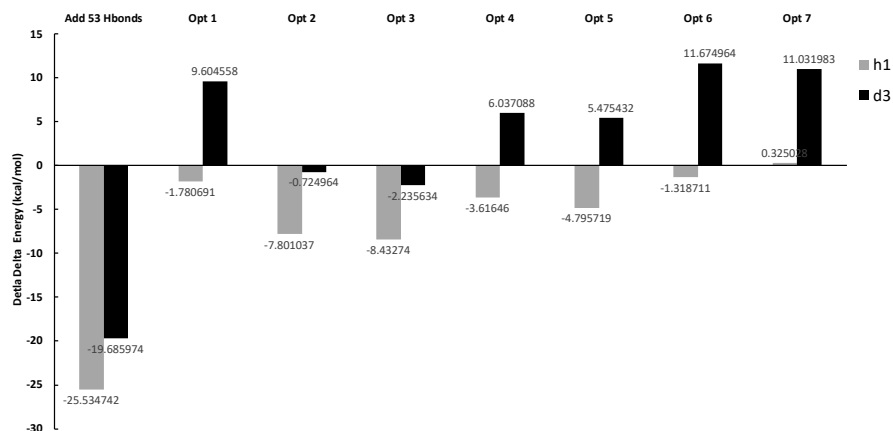


Figure 4.20: $\Delta\Delta E$ with Optimized Non-Canonical Hydrogen Bond Parameters: Difference in Energy of duplex d3 and hairpin h1 for different optimization schemes. Optimization 1 (removing twelve outliers) Optimization 2 (restricting canonical hydrogen bonds between 1.8-2.0 kcal·mol⁻¹), Optimization 3 (restricting all hydrogen bonds between 1.8-2.0 kcal·mol⁻¹), Optimization 4 (restricting canonical hydrogen bonds between 1.5-2.0 kcal·mol⁻¹ and non-canonical hydrogen bonds between 0.1-0.5 kcal·mol⁻¹), Optimization 5 (restricting canonical hydrogen bonds between 1.5-2.0 kcal·mol⁻¹ and non-canonical hydrogen bonds between 0.4-0.6 kcal·mol⁻¹), Optimization 6 (restricting all hydrogen bonds between 0.1-1.0 kcal·mol⁻¹), Optimization 7 (restricting all hydrogen bonds between 0.1-0.5 kcal·mol⁻¹,

4.3.2.5 Testing RACER 2.0 with Pulling Simulations

We performed preliminary computational studies of a RNA G4 using the RACER CG model (Fig 4.21). Our simulations are able to predict an unfolding free energy of -21.95 kcal·mol⁻¹, as compared with an experimentally determined folding free-energies from mechanical unfolding of Terra RNA-G4s of -11.31 kcal·mol⁻¹[58]. Differences between the folding free energies could be explained by different temperatures and buffer conditions used in the experiments and different buffer conditions. Our simulations reproduce the stabilizing effect of potassium ions on RNA-G4s, showing a difference in folding free energy of +7.8 kcal·mol⁻¹ (Fig. 4.22).

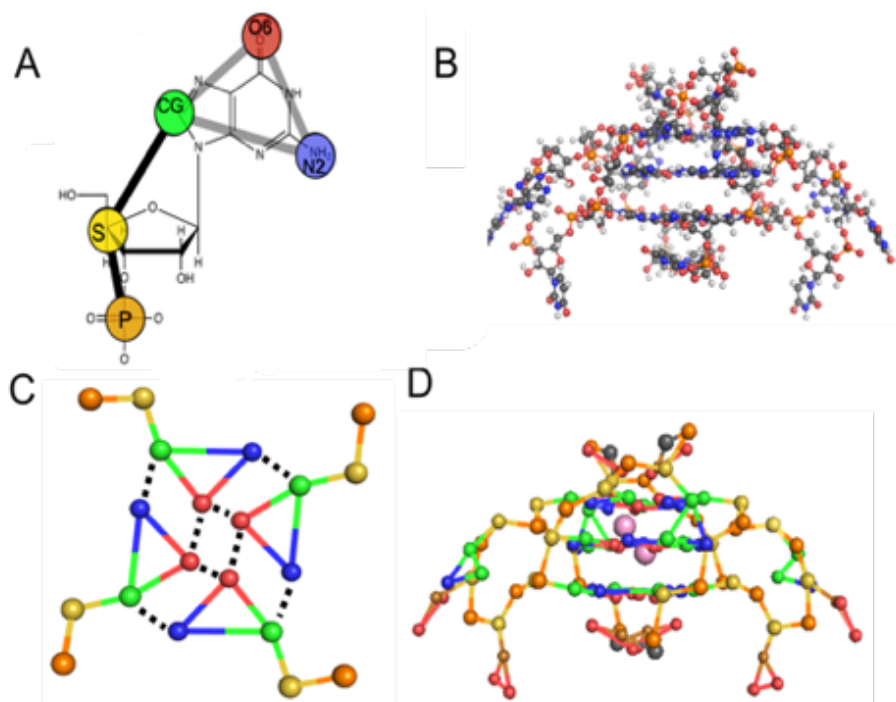


Figure 4.21: **RACER Model for RNA-G4s** A. Mapping from atomic nucleotide to coarse-grained RACER representation. Five pseudoatom types are shown: phosphate (orange), sugar (yellow), sugar-connect (green), oxygen (O6), and nitrogen (blue). B. Atomic structure of a human telomeric RNA (Terra RNA-G4) (PDB ID: 2kbp). C. Top-down image of a coarse-grained g-quartet in the RACER representation with hydrogen bonds shown among sugar connects and nitrogens and between oxygens. D. RACER Coarse-grained representation of Terra RNA-G4 with additional K^+ ions (pink) between quartets.

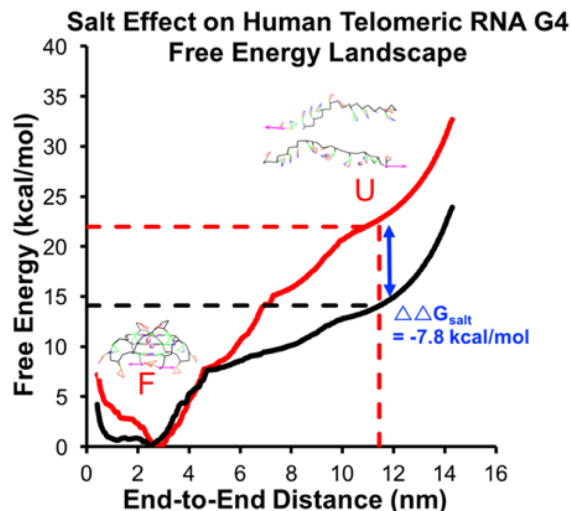


Figure 4.22: **RACER pulling simulations of RNA-G4** The free energy landscape of a RACER coarse-grained human telomeric RNA G4 (PDB ID: 2kbp) with (red) and without (black) potassium ions. The free energy difference between the unfolded states with and without sodium is denoted in blue, $7.8 \text{ kcal}\cdot\text{mol}^{-1}$. Structures extracted from the end-to-end distance corresponding to the folded (top) and unfolded state (bottom), where the dark pink spheres with arrows correspond to the pseudo atoms where force is applied to unfold the RNA G4 ($k = 1 \text{ kcal}\cdot\text{mol}^{-1}$). Each end-to-end distance was sampled for $0.5 \mu\text{s}$ (total simulation time of 0.1 ms). Experimentally determined folding free-energies from mechanical unfolding of Terra RNA-G4s of $-11.31 \text{ kcal}\cdot\text{mol}^{-1}$ [58]. Differences between the folding free energies could be explained by different temperatures and buffer conditions used in the experiments and different buffer conditions. Our simulations reproduce the stabilizing effect of potassium ions on RNA-G4s, showing a difference in folding free energy of $+7.8 \text{ kcal}\cdot\text{mol}^{-1}$.

4.4 Conclusion

Overall, the development and improvement of the RACER model has been discussed. Coarse-grained RNA modeling is a promising approach to addressing the sampling problem facing biomolecular simulations. In addition, more advanced potential energy functions which capture behavior such as hydrogen bonding can improve the accuracy of coarse-grained RNA models. The RACER 1.0 model was able to predict both structure and energetics. The improved RACER 2.0 model has been shown to improve correlation of binding energies with quantum mechanical calculations, and the model has been tested with energetics analysis and pulling simulations of a RNA G-Quadruplex.

4.5 Future Directions

Further testing of the improved RACER model is needed. Equilibrium pulling simulations with structures with non-canonical hydrogen bonding such as tetraloops and pseudoknots could reveal better optimization strategies to gauge the strength of canonical versus non-canonical hydrogen bonds. Additionally, long-time stability tests with large RNAs containing many non-canonical hydrogen bonds would reveal if the additional hydrogen bonds stabilize or destabilize larger RNAs. Finally, mutational analysis of the RNA G-Quadruplex loop sequences could be integrated with single-molecule or biochemical experiments which can measure relative stability of these marginally stable structures.

Chapter Five: Probing Local Nucleic Acid Transitions with Ultrashort DNA Tethers

5.1 Introduction

5.1.1 Motivation for Developing Complementary Single Molecule Experiments

Single molecule experiments are the best way to test the predictions of biomolecular models. These experiments allow for individual biomolecules to be studied, avoiding issues of ensemble averaging. Free energy landscapes can be generated using pulling simulations of biomolecules, and these free energy landscapes can be compared to free energy landscapes generated using single molecule force experiments. This combination of simulations and experiments provides an excellent test of the force fields used in biomolecular modeling. If the force fields are inaccurate, then populations of unphysical structures would be predicted that are not observed in the experiments. This disagreement is invaluable for improving the parameters in the force fields leading to improvement of the predictive power of biomolecular models. Beyond validating the free energy landscapes of the experiments, improved force fields for biomolecules could provide guidance to future experiments, by predicting behavior of biomolecules would could then be tested experimentally. Specifically, predicting which regions of an RNA sequence to mutate in order to influence its folding can be done much more efficiently with biomolecular modeling than with experiments.

5.1.2 Free Energy Landscapes of Nucleic Acids

Single molecule force experiments are capable of measuring transition pathways, free energy landscapes, and kinetics of folding of nucleic acids.[45, 51, 52, 54, 176] In these experiments, pico-Newton size forces are applied to the functionalized beads allowing the nucleic acid to be stretched and subsequently unfolded. By measuring populations of end-to-end extensions, the free energy landscape of the folding/unfolding of the nucleic acid can be generated. In order to measure the free energy landscape

of a nucleic acid along the end-to-end extension reaction coordinate, double stranded DNA tethers are used to attach the nucleic acid to functionalized surfaces. In optical trapping experiments on nucleic acids, these functionalized surfaces can be two functionalized bead surfaces “dumbbell assay”, or a single functionalized bead surface and a functionalized glass surface “surface tether assay”. The advantage of the dumbbell assay is that the molecule of interest is away from any glass surfaces, avoiding surface artifacts. A disadvantage of optical trapping geometry is that kilo-base (kb) long DNA tethers are used. These long DNA tethers can broaden the reconstructed free energy landscape, making it difficult to deconvolute the instrument effects from the signal from the biomolecule.[54]The dumbbell assay is excellent at recovering biomolecular rates, but the soft handles make it difficult to detect transiently formed intermediate states, leading to reduced resolution in the reconstructed free energy landscape.

5.1.3 Notation used in the Chapter

In this chapter nucleotide (nt) is used when discussing single-stranded nucleic acids, such as DNA hairpins, and base pair (bp) is used when discussing double-stranded nucleic acids such as DNA duplexes.

5.1.4 Modifications to Experiments for RNA studies

The pilot work has been performed using DNA almost exclusively, except for one biochemical experiment involving RNA. The reason for using DNA was that single-stranded DNA is more stable than single-stranded RNA. Additionally, new protocols had to be developed to form and test the ultrashort DNA tethers, and purchasing equivalent length RNA sequences would be at least double the cost of purchasing DNA sequences. The design of the DNA sequences used in the ultra-short tether assay can be directly used for studying RNAs, in future experiments. The 2s-26, 2s-50, 2s-100 DNA sequences can be interchanged with a RNA sequence of choice as long as that RNA sequence contains extensions that are complementary to the DNA tethers.

5.1.5 State of the Art RNA Experiments

Recent theoretical work has indicated that shorter linkers can provide improved free energy landscape reconstruction.[66, 70–72, 177, 178] Experimental work using DNA hairpins has shown that shorter double stranded DNA tethers can improve free energy landscape reconstruction. A recent study demonstrated that using 29bp double stranded DNA tethers improved signal-to-noise and slowed the kinetics of the DNA hairpin.[69] This reduction in the kinetics of the hairpin allowed for improved free energy landscape reconstruction. Two studies showed that bundled DNA duplexes, which form a DNA origami, could be used to reconstruct free energy landscapes. While these tethers were not short (1 kilobase long, approximately 500nm long), the bundling of the DNA duplexes resulted in stiffer tethers as compared with a single double stranded DNA tether.[179, 180] These studies demonstrated the resulting free energy landscape was sharper when using the stiff DNA origami tethers as compared with double stranded DNA tethers.

5.1.6 Ultrashort Tether Single Molecule Assay

A novel single molecule assay has been developed to study local nucleic acid transitions in the near zero force regime. This assay features ultrashort double stranded DNA tethers (13bp each) that are functionalized on the 3' and 5' ends. These functionalized ends which allow the tethers to bind to a functionalized bead surface and a functionalized glass surface. A single nucleic acid hybridizes to the tethers forming a three-piece tether to the surface (see Fig. 5.1). As a result, conformational changes in the nucleic acid are directly transmitted to the bead. By tracking the position of the bead, changes in end-to-end distance of the single molecule can be monitored.

Fine features of the energy landscape as small as 1/10 of $k_B T$ (0.6 kcal·mol⁻¹ or 2.5 kJ·mol⁻¹) can be resolved.[182, 183] The depth of the energy landscape that can be probed depends on the bead diameter, length of the tether, viscous drag on the bead, the length of data collection.[184] Additionally, only statistically independent position measurements count for the quality of the energy landscape. Position measurements can be considered statistically independent if the bead had sufficient time to explore its confining potential. For a harmonic potential with spring constant k , this resident time is set by the position autocorrelation time, $\tau \approx \frac{\gamma}{k}$, where $\gamma = 6\pi\eta r$ is the viscous

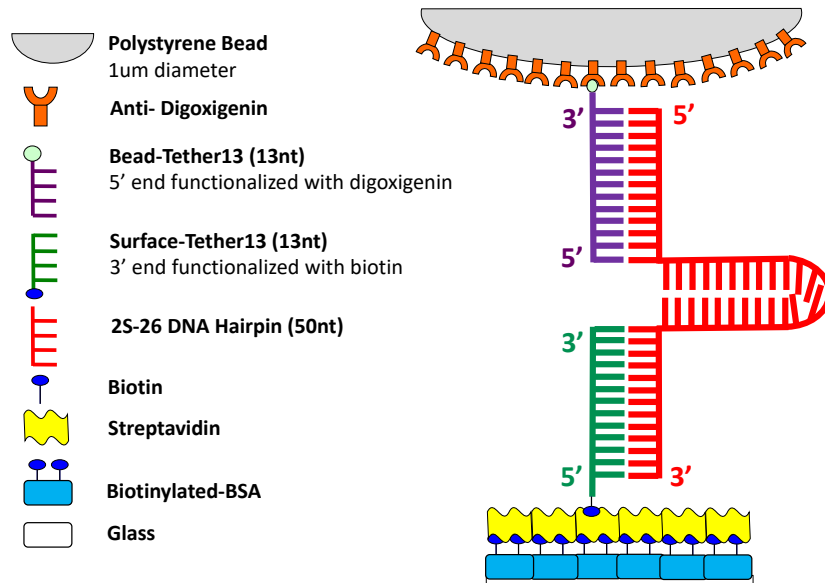


Figure 5.1: **Surface Chemistry of Single Molecule Assay:** Ultra-short tether single molecule assay The assay contains three nucleic acid sequences (Bead-Tether13, Surface-Tether13, and 2s-26nt DNA Hairpin) (see Figure 5.4 for sequences). The DNA tethers (Bead-Tether13, Surface-Tether13) are functionalized with digoxigenin and biotin, respectively. These different functional groups allow the tethers to specifically attach to the anti-digoxigenin coated polystyrene bead or the streptavidin coated glass surface. The 2s-26nt DNA Hairpin has two 13 nt extensions on either side of the hairpin sequence. These 13nt extensions are complementary to the Bead-Tether13 and Surface-Tether13. The advantage of this design is the 2s-26nt DNA Hairpin can be replaced with another sequence without changing any other part of the assay. The sequences for the DNA tethers (Bead-Tether13, Surface-Tether13) were extracted from an RNA sequence from a 3' extension of *Tetrahymena thermophila* ribozyme used in previous FRET studies.[181] The extracted RNA sequence was converted to a DNA sequence by replacing the Uracils with Thymines. The 2s-26 DNA Hairpin was extracted from a two state DNA hairpin used in previous single molecule force experiments.[69]

drag on a bead with radius r . The viscosity η is the viscosity of water. In simple terms, k the width of the potential and sets the speed with which the particle can explore the energy landscape. In the planned experiments, the width of the potential is set by the stiffness of the DNA tether. The shorter the DNA tether the narrower

the potential of the bead, allowing the bead to explore the potential faster and energy landscapes as deep as $10 k_B T$ ($k_B T = 6.0 \text{ kcal}\cdot\text{mol}^{-1}$ or $25 \text{ kJ}\cdot\text{mol}^{-1}$) can be readily measured.

5.1.7 Motivation from Biochemical Assay

Near zero force experiments require short, stiff linker molecules which couple changes in end-to-end distance of the nucleic acid directly to the fluctuations of the reporter bead. These short linker molecules improve the spatial resolution and are appropriate for measuring energy landscapes with the highest spatial resolution. However, the direct coupling via short tethers has the disadvantage that the internal dynamics of the nucleic acid are slowed down. If, for instance, a conformational change leads to a shortening of the nucleic acid end-to-end distance, the molecule will pull the bead with it. Therefore, it is important to have an independent measurement of the conformational dynamics. Kinetic data from biochemical experiments can be compared with the unfolding dynamics measured in single molecule force experiments. They would provide insight into how the attachment of a bead changes the observed unfolding and re-folding dynamics.

5.2 Materials and Methods

5.2.1 Biochemical Assay

A gel-shift mobility assay was performed on two nucleic acid sequences, containing G-quadruplexes (G4). The two sequences tested are a 21nt DNA-G4 hTelo and a 21nt RNA-G4 Terra are shown in Fig. 5.2. Here, hTelo is a human telomeric quadruplex forming DNA sequence composed of contains three G-quartets and three TTA loops. The second sequence is a telomeric repeat-containing RNA (Terra) a long non-coding RNA that is transcribed from telomeric DNA. The Terra sequence contains three G-quartets and three UUA loops. Intramolecular RNA and DNA G4s were radiolabeled using a T4 polynucleotide kinase to add a radiolabeled phosphate to the 5'end. The GQ-Fast thermocycling method was used to form DNA-G4 hTelo and RNA-G4. The GQ-Fast protocol calls for heating at 95°C for 5 minutes to unfold the G4 and cooling at $0.1^\circ\text{C}/\text{s}$ till the temperature reaches 4°C . After thermocycling, 100 mM K^+ and

1 μ M of a 21nt chase DNA was added to the formed G4s. The G4 solutions were then placed in a water bath at 37°C. As the G4 unfolded, the chase DNA traps the unfolded quadruplex, blocking refolding (Figure 5.2A).

The reaction was quenched at six time points for the hTelo DNA-G4 and the Terra RNA-G4. At each of the time points, the G4s and duplex structures were separated by native polyacrylamide gel electrophoresis, where the G4 migrates faster than the duplex, resulting in a shift in the populations of G4 and duplex over time. The time dependence reflects the loss of the G4 and was fit by an exponential curve to give a rate constant for G4 unfolding. The tested G4s with loops of three nucleotides unfold with rate constants of $0.35 \pm 0.02 \text{min}^{-1}$ for DNA G4 (black) and $0.08 \pm 0.002 \text{min}^{-1}$ for Terra RNA-G4 (red). Our controls (results not shown) established that: 1) the rate constant did not depend on the concentration of oligonucleotide chase across a 10-fold range, indicating that the chase oligonucleotide efficiently captured the unfolded G4 and that it did not participate in unfolding of the G4; and 2) the reaction was greatly accelerated by decreasing the K⁺ concentration below 50 mM or by replacing K⁺ with Li⁺, verifying that the G4 was formed at the start of the reaction and its loss was measured.

5.2.2 Single Molecule Force Experiments

5.2.2.1 Single Molecule Pulling Procedure

Tracking the position of the tethered beads was done using a differential interference contrast microscope equipped with an optical trap. An infrared laser with a wavelength of 1064nm with an estimated spring constant of 1 pN/nm was used to trap the beads. Scanning mirrors were used to steer the optical trap, where the minimum step size for the optical trap was determined to be 18 nm by tracking the position of an un-tethered bead in solution as the trap was stepped by the smallest possible displacement. The center position of the beads was measured by DIC video based tracking developed using in-house software.[185, 186] The camera used for these experiments is the UNIQ UP-600-12B Digital CCD Camera. In these experiments, a frame rate of 60 frames/s was used to collect data. At each step of the optical trap, 1000 frames of data were taken. To perform pulling experiments, a bead was first trapped in solution and brought into contact with the surface. The bead was

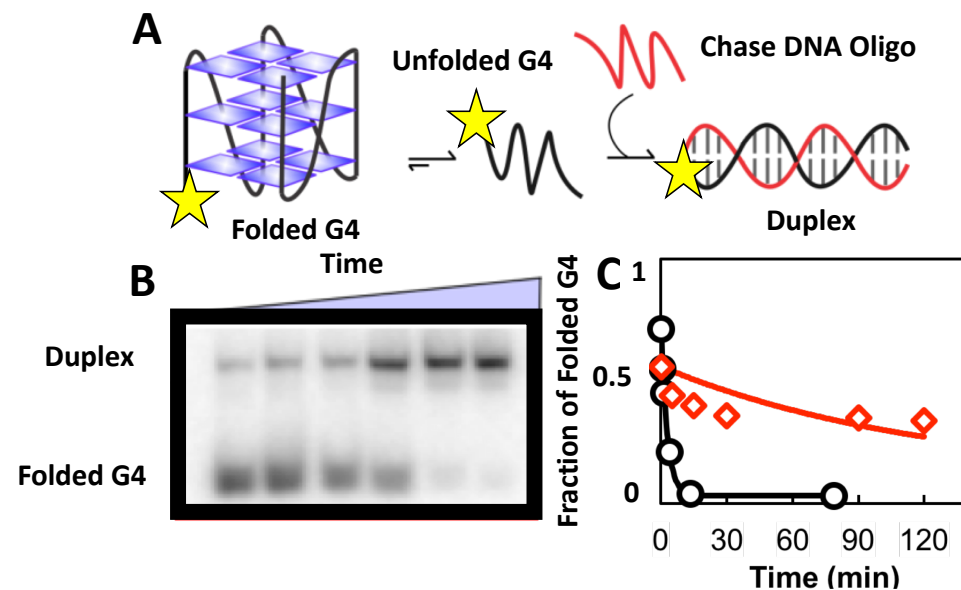


Figure 5.2: **Measurement of DNA G4 (hTelo) and RNA G4 (Terra) lifetimes:** This experiment shows unfolding of the sequences hTelo G-Quadruplex (DNA G4) and Terra G-Quadruplex (RNA G4). hTelo is a human telomeric quadruplex forming DNA sequence. The hTelo sequence contains three G-quartets and three TTA loops. The full sequence of hTelo is $(GGGTTA)_3GGG$. Terra is a long non-coding RNA that is transcribed from telomeric DNA. Terra stands for telomeric repeat-containing RNA. The Terra sequence contains three G-quartets and three UUA loops. The full sequence of Terra is $(GGGUUA)_3GGG$. A. Reaction scheme for measuring G4 Unfolding. The G4 sequence is phosphorylated with a radioactive P32 at its 5' end (yellow star). The G4 is folded using the GQ-Fast protocol using a thermocycler. The GQ-Fast protocol is heating at 95°C for 5 minutes to unfold the G4 and cooling at $-0.1^\circ\text{C}/\text{s}$ till the temperature reaches 4°C. Excess chase DNA oligos which are complementary to the G4 sequence are mixed with the folded G4. The mixed species are placed in a water bath at 37°C. As the G4 unfolds it is trapped by excess chase oligonucleotide (21 mer) (red), which forms a duplex. B. Time dependent unfolding of the hTelo G4 (lower band), with trapping by the duplex (upper band), monitored by native PAGE. C. Unfolding rate is $0.35 \pm 0.02\text{min}^{-1}$ for DNA G4 (black) and $0.08 \pm 0.002\text{min}^{-1}$ for RNA G4.

held at the surface until a tether was formed, as indicated by turning off the trap and observing that the bead did not diffuse away from the surface. After the bead became tethered to the surface, the optical trap was switched off and a video of the fluctuating bead was observed for 1000 frames with a rate of 60 fps. The anchor point on the surface was estimated to be the center of the position distribution, and trap was re-centered to align accordingly.

5.2.2.2 Geometrical Amplification Effect

Because of the special geometry, i.e. a large sphere in contact with a flat surface at a single point, the bead with radius (R) can still move large distance along the surface even for a linker as short as 1 nm (see Fig. 5.3). Thus, the lateral bead fluctuations (Δx_{max}) can be used to calculate the linker length (L) and its changes (see Eq. 5.1). As shown in previous studies changing the length of the spacer molecules between BSA and biotin from 2 nm to 5.6 nm caused a clear shift in the magnitude of lateral fluctuations from below 15 nm to maximally 30 nm, corresponding to an amplification factor of 4.2.[187] The described experiment does not require application of any external force to the molecules. The optical trap was only used to initiate binding and then to measure the thermal position fluctuations of the tethered bead in three-dimensions. Because no external force is used, we call them near zero force experiments. Near zero force instead of zero force because the binding of a bead, or any larger object, to an immobilized molecular complex leads to a rectification of thermal forces away from the surface and thus to a small average pulling force.

$$\Delta x_{max} = \sqrt{(R + L)^2 - R^2} \quad (5.1)$$

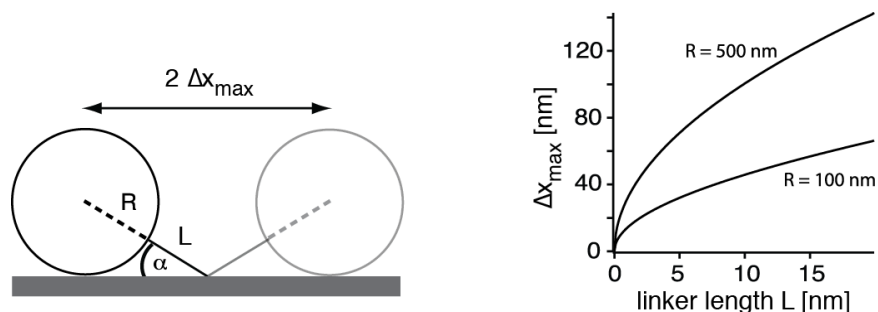


Figure 5.3: **Geometrical Amplification Effect for a Bead Tethered Near the Surface:** a) A simple geometrical model for a bead with radius R (500 nm) tethered by a single molecule with length L is expected to have a maximal lateral fluctuation of Δx_{\max} . The length of the unfolded 2s hairpin can be estimated by multiplying the total number of nucleotides (24) by the length per base pair in single stranded DNA (0.68 nm/nt). From this calculation the unfolded hairpin has a length of 16.32 nm. The linker is the sum of the length of the hairpin and the two 13 bp DNA tethers, which have a length of 8.84 nm. Therefore, the linker length (L) for the folded hairpin is 8.84 nm and the linker length for the unfolded hairpin is 25.16 nm. b) From simulated data, the maximal lateral displacement of the particle as a function of linker length for particles of 500 nm and 100 nm in radius.[183]

5.2.2.3 Surface Chemistry of the Ultrashort Tether Assay

In our experiment, a three-piece tether is spanned between an anti-digoxigenin coated bead and a streptavidin-coated glass coverslip (see Fig. 5.1). The glass surface is first coated with biotinylated bovine-serum albumin (biotinylated-BSA). Streptavidin is added on top of the biotinylated-BSA layer, the streptavidin contains four binding sites for biotin, allowing it to form a layer on top of the biotinylated-BSA. The beads are $1\mu\text{m}$ diameter protein-G beads that have been covalently coupled to anti-digoxigenin fab fragments. Because the protein coatings on the bead surface and the glass coverslip are different, this helps to minimize non-specific interactions between the two surfaces the three-piece tether is spanned between. The surface chemistry of the beads and surfaces has been used in previous single molecule FRET and single-molecule pulling studies.[45, 47, 181] Because BSA can unfold on the glass coverslip, this can cause inhomogeneous protein-coated glass coverslips. Biotinylated-Polyethylene glycol (PEG) can be used instead of biotinylated BSA.[183] The three-piece tether is composed of a single-stranded DNA with a digoxigenin at its 5'end, a single-stranded DNA with a biotin at its 3'end, and a single stranded DNA sequence

with extended 5' and 3' ends that are complementary to the functionalized single-stranded DNA sequences. This single-stranded DNA sequence can be replaced with an RNA sequence as long as the extensions of the RNA sequence are complementary to the functionalized single-stranded DNA tethers.

5.2.2.4 Design of Ultra-Short DNA Tethers

The shortest nucleic acid construct tested was a 50 nt DNA hairpin, composed of a 4 nt loop and a 20 nt stem, and 26 nt extensions, referred to as the 2s-26 DNA Hairpin. The 2s-26nt DNA Hairpin has two 13nt extensions on either side of the hairpin sequence. These 13nt extensions are complementary to the Bead-Tether13 and Surface-Tether13. The hairpin sequence is a truncated sequence from the 2s hairpin.[69] The assay contains three nucleic acid sequences (Bead-Tether13, Surface-Tether13, and 2s-26 DNA Hairpin) (see figure 5.4 for sequences). The advantage of this design is the 2s-26 DNA Hairpin can be replaced with another sequence without changing any other part of the assay.

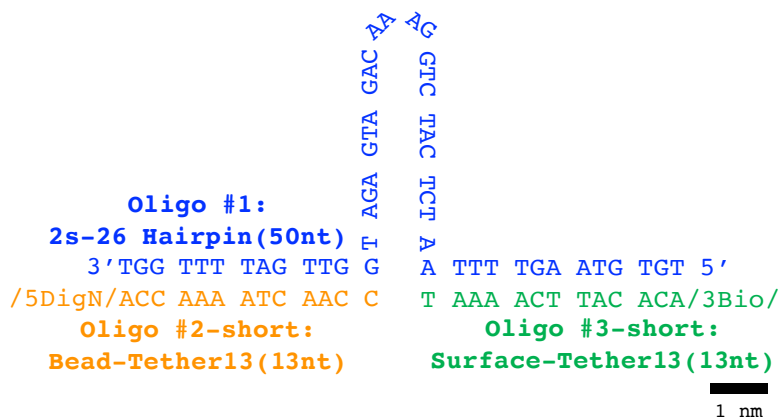


Figure 5.4: **Design for 26nt Two State (26-2s) DNA Hairpin with Two 13bp DNA Tethers** Oligo1: 50nt DNA Hairpin (blue), Oligo2-short: 13nt DNA strand with digoxigenin functionalized to the 5' end (yellow), and Oligo 3-short: 13nt DNA strand with biotin functionalized to the 3' end. The estimated total length of the two 13bp DNA tethers is 8.84 nm

To characterize the effect of surface effects on our assay, two additional DNA tethers were designed that contained the same 2s DNA Hairpin sequence, but had

extended sequences allowing the tether between the glass surface and the bead to be longer. Secondary structure analysis was performed on the extended hairpin sequences to ensure that the extended sequences did not form alternative structures (see most favorable secondary structure predictions in Mfold figures Fig. B.1, B.2, B.3).

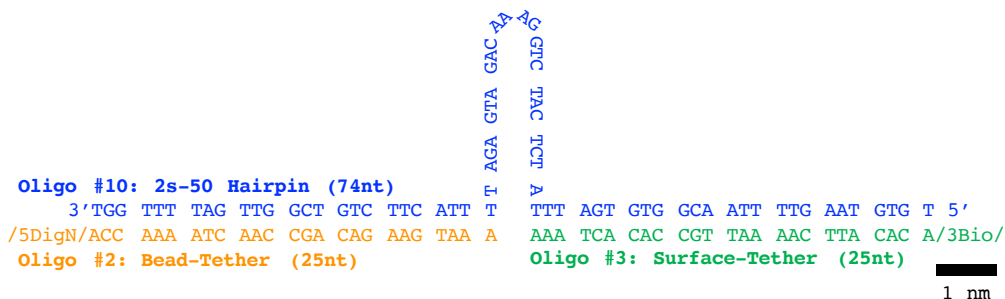


Figure 5.5: **Design for 74nt Two State (2s-50) DNA Hairpin with Two 25bp DNA Tethers:** Oligo10: 74nt DNA Hairpin (blue), Oligo2: 25nt DNA strand with digoxigenin functionalized to the 5' end (yellow), and Oligo3: 25nt DNA strand with biotin functionalized to the 3' end. The estimated total length of the two 25bp DNA tethers is 17 nm

One of these longer tethers is a 74nt DNA hairpin, a 24nt hairpin with two 25 nt DNA extensions on either end (2s-50) (see Figure 5.5 for sequences). The longest tether a 24 nt DNA Hairpin sequence with 33 nt extensions on each end (see figure 5.6 for sequences). Unlike the 2s-26 and 2s-50 tethers, the longest tether was designed to be made with seven sequences rather than three sequences. The two additional sequences were used to extend the 74 nt DNA Hairpin so that they could be attached to Tether-50 to Bead and Tether-50 to surface. The rationale for this was to be able to re-use the functionalized tethers from the 2s-50 designs (Oligo2 and Oligo 3). As a result of the increased number of DNA sequences used to in the longest tether, additional steps were taken to form the longest tether.

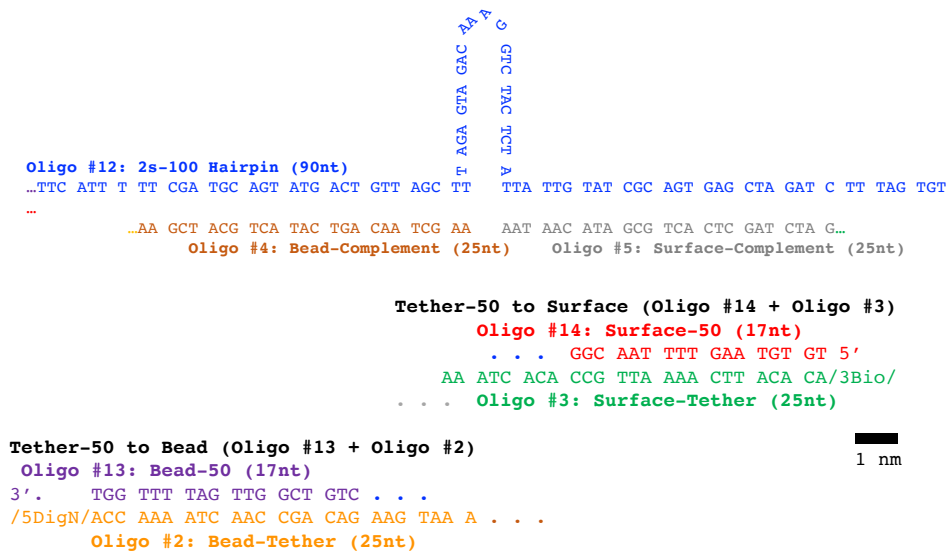


Figure 5.6: **Design for 94nt Two State DNA (2s-100) Hairpin with Two 50bp DNA Tethers:** Oligo12: 94nt DNA Hairpin (blue), Oligo2: 25nt DNA strand with digoxigenin functionalized to the 5' end (yellow), Oligo3: 25nt DNA strand with biotin functionalized to the 3' end, Oligo4: complementary DNA sequence for extended hairpin (brown), and Oligo5: complementary DNA sequence for extended hairpin (grey), and Oligo13: extended hairpin sequence (purple) Oligo14: extended hairpin sequence (red). The estimated total length of the two 50nt tethers is 34 nm

5.3 Results

5.3.1 Testing for the Formation of Ultra-Short DNA Tethers

The three-piece tethers 2s-26 and 2s-50 were formed using the GQ-Fast protocol. The GQ-Fast protocol is performed using a thermocycler. The sequences are heated at 95°C for 5 minutes to and cooled at $-0.1^{\circ}\text{C}/\text{s}$ till the temperature reaches 4°C. All three DNA sequences (2s-26 hairpin, Bead-Tether13, and Surface-Tether13) were added in equal concentration to form the full 2s-26 hairpin Fig. 5.4). Separately, all three DNA sequences (2s-50 hairpin, Bead-Tether, and Surface-Tether) were added

in equal concentration to form the 2s-50 hairpin (Fig. 5.5). The 2s-100 hairpin was formed in three steps. First, the 2s-100 hairpin, Surface-50, Surface-Tether, and Bead-Complement were thermocycled using the GQ-Fast method. Second, the Bead-50 and Bead-Tether were thermocycled. The formed structures were combined and a 40-fold excess of surface complement was added to crowd the formed structures into attaching together. The combined structures were kept at 40°C for 30 minutes.

In order to visualize the formation of the constructs, all unfunctionalized 5'ends of the DNAs were phosphorylated with a radioactive isotope P^{32} . The decay product of this isotope is beta particles. To test for formation of the full tethers, a native polyacrylamide gel electrophoresis PAGE experiment was run. The negative charged DNA is driven towards the positive anode and the location of a band on the gel indicates the mobility of the structure, which is inversely proportional to the logarithm of its molar mass. The gel provides size exclusion where the more compact structures (like a folded hairpin) will migrate slower than a single-stranded DNA (like an unfolded hairpin).

All three tethers were able to be formed. The 2s-26 hairpin is shown in Figure 5.7. The fully formed construct is shifted up from the control DNA ladder and DNA Duplex indicating that the 2s-26 hairpin hybridized with the 13nt DNA tether sequences. The 2s-50 hairpin was fully formed as seen in Figure 5.8. The 2s-50 hairpin hybridized with the 25nt DNA tether sequences and is shifted up from the 2s-50 hairpin within the tethers. The 2s-100 hairpin also fully formed, but the fraction of fully formed 2s-100 is low as seen in Figure 5.9. In Band A, there is no band indicating the formation of an alternative structure, where the 2s-100 hairpin sequence formed a hairpin at its 5p end (see predicted secondary structure Fig. B.3).

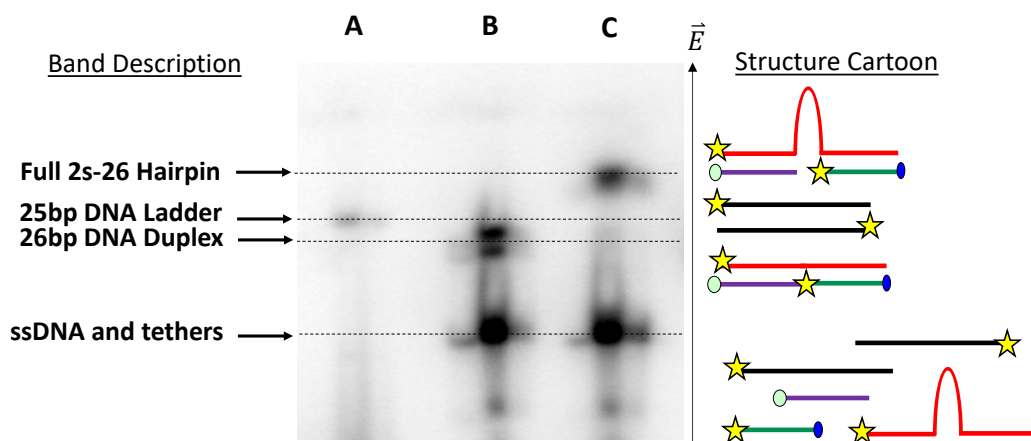


Figure 5.7: **Native PAGE Showing Formation of 26nt Two State (2s-26) DNA Hairpin Tether:** Experiment is performed in a 12% Polyacrylamide gel. All constructs were formed using the GQ-Fast method, and unfuctionalized 5'ends of the DNAs are phosphorylated with P^{32} (yellow star), a radioactive phosphate group that emits beta particles. A) 25 bp DNA duplex (DNA Ladder) used as reference for other bands B) 26bp DNA duplex (DNA control) with complementary sequence to the two 13 nt DNA tethers (Fig. 5.1) C) Top band indicates formation of the full 2s-26 hairpin (hairpin sequence with two 13bp DNA tethers (see Fig. 5.4). The lowest bands in columns B and C are the single-stranded (ss) DNA strands and 2s-26 hairpin that did not associate with other DNA strands.

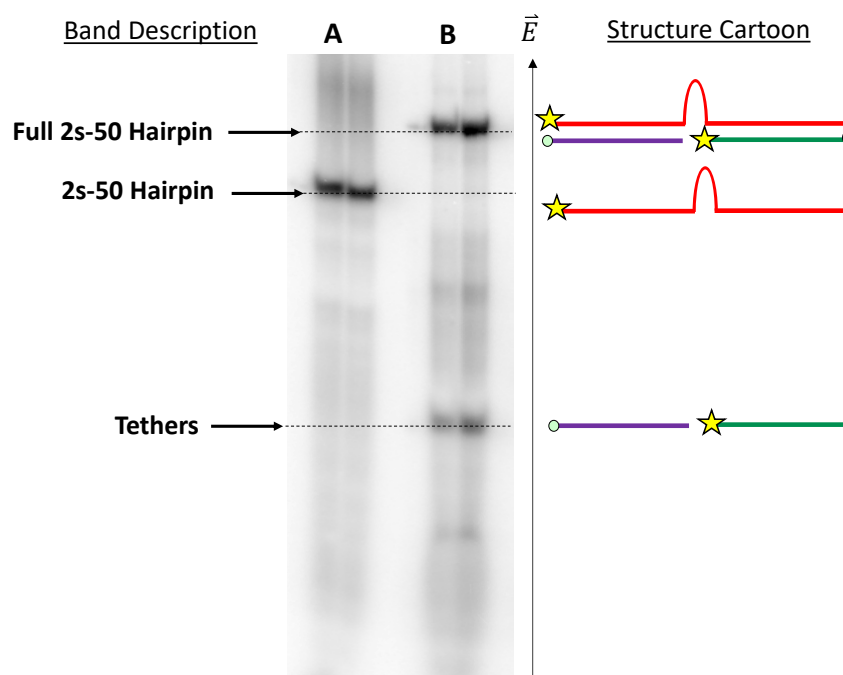


Figure 5.8: **Native PAGE Showing Formation of 74nt Two State (2s-50) DNA Hairpin Tether:** Experiment is performed in a 12% Polyacrylamide gel. All constructs were formed using the GQ-Fast method, and unfuctionalized 5'ends of the DNAs are phosphorylated with P^{32} (yellow star), a radioactive phosphate group that emits beta particles. A) 2s-50 Hairpin (Oligo 15) B) 2s-50 Hairpin with 50bp DNA tethers (see Fig. 5.5).The lowest band in column B is composed of the single stranded DNA sequences that did not associated with the 2s-50 hairpin.

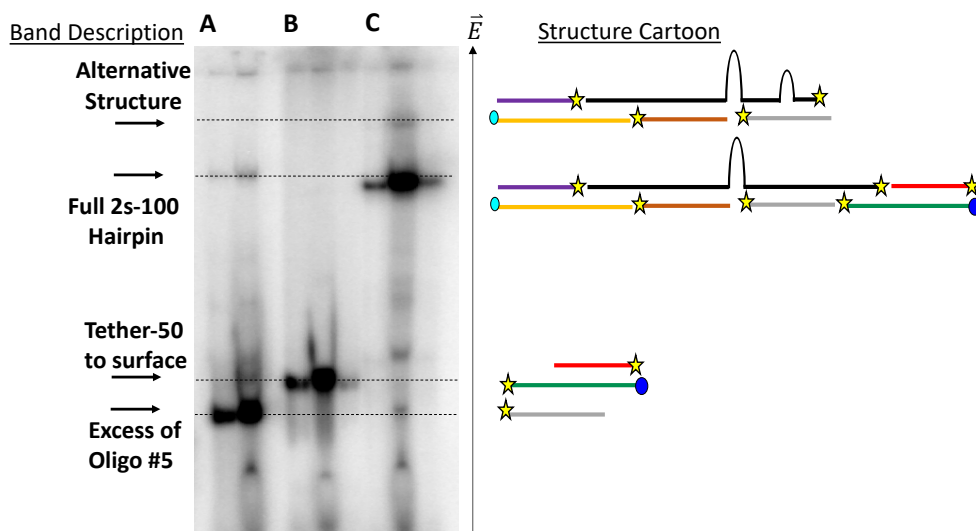


Figure 5.9: Native PAGE Showing Formation of 94nt Two State (2s-100) DNA Hairpin Tether: Experiment is performed in a 12% Polyacrylamide gel. Constructs were formed using the GQ-Fast method, and unfuctionalized 5'ends of the DNAs are phosphorylated with P^{32} (yellow star), a radioactive phosphate group that emits beta particles. Instead of forming the full 2s-100 Hairpin with all 7 pieces, the full 2s-100 Hairpin was assembled in three steps (refer to Fig. 5.6 for terminology). In tube 1, Oligo13, Oligo2, Oligo4, and Oligo12 were mixed an GQ-Fast was run to anneal them. Tube 1 represents all of the full 2s-100 Hairpin except for the tether to surface and Oligo5. Tube 1 contains a 2s-100 hairpin with the Tether-50 to Surface on its 3' end, and a “sticky end” on its 5' end, or a single-stranded extension. In tube 2, Oligo14 and Oligo 3 were mixed and GQ-Fast was run to anneal them. The annealed Tube 2, formed Tether-50 to surface. Tube 3, is a mix of Tube 1 and Tube 2 with Oligo5 added in 40-fold excess to drive association of Tether-50 to the partially formed 2s-100 Hairpin and annealed at 40°C for 30min. A) full 2s-100 Hairpin formed by separately forming 2s-hairpin with stick ends and Tether50 to surface and adding excess of Oligo5 (see Fig. 5.6) (notice lack of an alternative structure band as compared with C) B) Tether-50 to Surface C) Mixed Tube 1 and Tube 2, the main feature is an alternative structure band formed from the at sticky end of hairpin near the surface

5.3.2 Detecting Specific vs Non-specific Binding

To characterize the binding of specifically functionalized beads to specifically functionalized glass surfaces, we measured the lateral fluctuations of surface bound beads using high-resolution DIC video particle tracking. The ultrashort tether assay was assembled with the 2s-26 DNA hairpin sequence, with an estimated length of 8.84nm (see Fig 5.4). Beads were observed for 1 min at a frame rate of 60 fps. The thermally driven lateral position fluctuations were tracked, their average distance calculated and plotted in a histogram. Instead of a uniform distribution, we observed three populations. Tightly bound beads with a standard deviation (sdev) of 5 nm, less strongly bound beads with a sdev of 50 nm and loosely bound beads with sdev of 100 nm. The sdev of 100 nm lateral fluctuation agrees with the tether length of 100 nm, when taking the geometrical amplification effect into account (Fig 5.10). We concluded population with small sdev is nonspecifically bound via short linkers, the population with slightly larger sdev might be bound by multiple specific linker and non-specific interactions. To verify that single-molecule conditions were achieved, pulling experiments were performed.

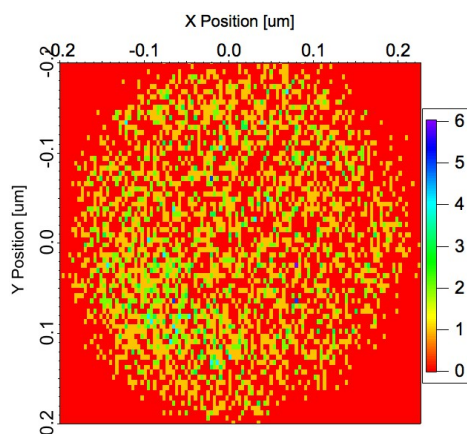


Figure 5.10: **2D Position Histogram of a Loosely Bound Bead:** A symmetric position distribution for an anti-digoxigenin coated bead moving around its center position without preferentially binding to a particular position. The standard deviation of this bead is 100 nm, which is the same standard deviation observed for loosely bound beads. A standard deviation of 100 nm corresponds to a bead that is specifically bound to the surface without formation non-specific attachments due to surface impurities or multiple tethers.

5.3.3 Pulling Experiments with Ultrashort DNA Tethers

To test for the formation of a single-tether, lateral pulling experiments were performed on the tethered bead. If a single tether was formed between the bead and the surface, pulling the bead in the lateral direction would unfold the 2s-26 DNA hairpin. This would lead to the linker length changing from 8.84 nm (length of the two 13bp DNA tethers) to 25.16 nm (sum of the single-stranded 2s-26 DNA hairpin and the length of the two 13bp DNA tethers). With a bead radius of 500 nm and taking into account the geometrical amplification effect, the maximum lateral fluctuation for the 2s-26 DNA hairpin when it is folded is 160 nm and 100 nm when it is unfolded. The difference between the maximum lateral fluctuation for the unfolded 2s-26 DNA hairpin and the maximum lateral fluctuation for the folded 2s-26 DNA hairpin is 60 nm. A lateral pulling experiment with the 2s-26 DNA tether is shown in Fig. 5.11. The 2s-26 DNA hairpin is tethered to a functionalized 500 nm radius bead and a functionalized glass surface. The optical trap is displaced in 18 nm steps in the x-direction, while the functionalized glass surface is immobilized. The graph shows a position histogram of bead positions over four 18 nm steps in the x-direction of the optical trap (see Fig. 5.12). Two dominant populations are found centered around 10 nm and 70 nm. The peak-to-peak distance between these populations is 60 nm. This agrees with the prediction from the geometrical amplification effect that the folded and unfolded state should be separated by 60 nm. This result supports the conclusion that single-molecule conditions were achieved. Specifically that a single 2s-26 DNA hairpin was unfolded, and that the surface chemistry of the ultra-short tether assay works.

Interestingly, this graph shows that the bead did not always follow the position of the optical trap, as seen by only two populations of bead positions. If the bead had been freely diffusing, there should be equally populated distributions at multiples of 18 nm. Instead the bead spent the majority of its time between 0-20 nm and 60-70 nm. Interestingly, there are transient positions found at 25 nm and 40 nm. These small position counts could be the result of transient intermediate unfolded states. The 0-20nm population can be interpreted as the bead position corresponding to a folded 2s-26 hairpin. The 60-70 nm population can be interpreted as the position where the 2s-26 hairpin has become unfolded.

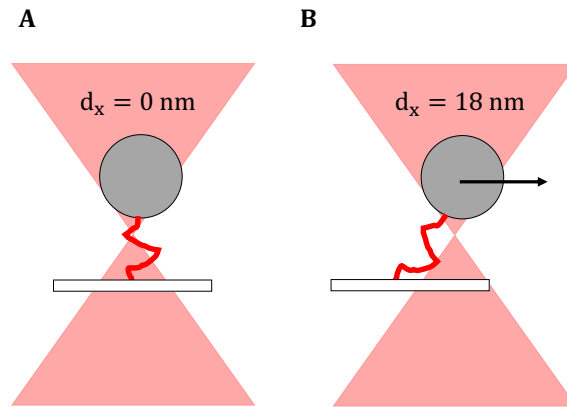


Figure 5.11: **Optical Trapping Geometry for Ultrashort Tether Assay:**The 2s-26 DNA hairpin is tethered to a functionalized 500 nm radius bead and a functionalized glass surface. The optical trap is displaced in 18 nm steps in the x-direction, while the functionalized glass surface is immobilized.

An alternative interpretation is that the 2s-26 hairpin unfolded within one 18nm step. This explanation is supported by the first population of bead positions being centered to the right of 0nm. This shift in the initial population of the bead could mean that the 2s-26 hairpin was already under tension before the optical trap was stepped laterally. This would indicate that the population at 60-70 nm represents a population of bead positions where the bead is no longer attached to the surface. Additionally, because of the slow integration time of the CCD camera, the 40 nm intermediate states could be the average of a folding/unfolding event that cannot be temporally resolved. Further studies are needed to investigate if the alternative interpretations of the position histogram are valid.

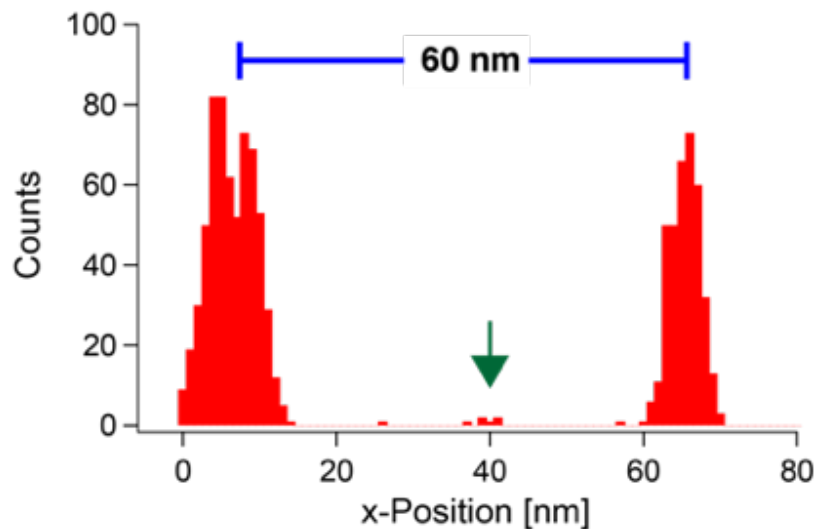


Figure 5.12: **Position histogram for unfolding event of 2s-26 DNA hairpin**
 The x-position refers to the position of the center of the bead as the optical trap is moved in 18 nm steps along the x-direction. The optical trap applies a lateral force on the bead. The optical trap consists of an infrared laser with a wavelength of 1064nm and has an estimated stiffness of 1pN/nm. At each 18nm step, data was collected for 1000 frames (≈ 16 s). Data was collected using a differential interference contrast (DIC) microscope with a camera with a frame rate of 60 frames/second. This position trace contains the position data for a single bead. Two distinct populations are observed centered at ≈ 10 nm and ≈ 70 nm. The peak-to-peak separation is ≈ 60 nm. Transient intermediate states are observed at ≈ 25 nm and ≈ 40 nm. Interestingly, this graph shows that the bead did not always follow the position of the optical trap, as seen by only two populations of bead positions. If the bead had been freely diffusing, there should be equally populated distributions at multiples of 18nm. Instead the bead spent the majority of its time between 0-20 nm and 60-70 nm. Interestingly, there are transient positions found at 25nm and 40 nm. These small position counts could be the result of transient intermediate unfolded states. The 0-20 nm population can be interpreted as the bead position corresponding to a folded 2s-26 hairpin. The 60-70 nm population can be interpreted as the position where the 2s-hairpin has become unfolded.

5.4 Conclusion

In this chapter, the development and testing of ultrashort DNA tethers for single-molecule force experiments were presented. Three DNA tether sequences were designed, and they were demonstrated to form using native PAGE. The shortest DNA tether 2s-26 DNA Hairpin, was successfully integrated into a single molecule assay.


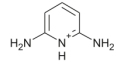
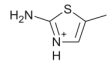
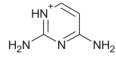
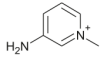

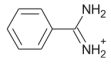

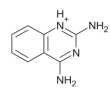
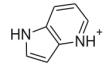
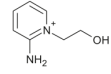
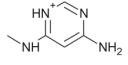
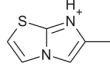
The geometrical amplification effect was used to estimate the lateral fluctuations of a bead near a surface with a known linker length and bead radius. By monitoring thermal fluctuations of the beads using a DIC video particle tracking, non-specific and specific binding interactions were characterized. The standard deviations of the beads measured from these experiments agree with the expected linker length of a folded 2s-26 DNA Hairpin with ultrashort DNA tethers. Pilot single-molecule pulling experiments were performed on the 2s-26 DNA Hairpin, and the observation of two populations in a histogram of bead populations supports the claim that single-molecule conditions were achieved. Future studies are needed to improve the efficiency of the ultra-short assay, such as testing the medium and long DNA tethers which could reduce non-specific interactions due to moving the bead further away from the glass surface.

5.5 Future Directions

The following enhancements can be made to the single molecule setup to improve free energy reconstructions. First, including high bandwidth position detection (≈ 1 MHz) to track the trapped bead, which will prevent blurring of the free energy landscape from low bandwidth.[183, 188, 189] The quality of a free energy landscape calculated from position distributions depends on the number of statistically independent points, which is in practice limited by instrument drift. Instrument drift can be corrected for by tracking markers on the glass surface in three-dimensions with sub-nanometer resolution.[190–192] Ultra-short tethers may lead to stronger interaction of the bead with the surface and possibly to a tilt in the energy landscape. However, we can correct for this external interaction potential. In contrast to folding energy landscapes, the interaction potential between bead and functionalized glass surface is expected to be smooth because of the large contact area between bead and surface. It can be measured independently by analyzing the three-dimensional position distributions of a free bead close to the surface or by analyzing the measured energy landscape in three-dimensions.[184] This is required in our assay for the correct interpretation of thermal fluctuations of tethered beads.[182, 193, 194]

Appendix A: Chapter 2: Chemical Structures

A.1 Chemical Structures and Absolute Hydration Free Energies

Compounds	BAR	OSRW	Absolute deviation
1 	-46.61 ± 0.82	-46.93 ± 0.42	0.32
2 	-44.59 ± 0.52	-43.95 ± 0.28	0.64
3 	-45.76 ± 0.77	-45.57 ± 0.09	0.19
4 	-49.70 ± 0.83	-50.18 ± 0.40	0.48
5 	-41.68 ± 0.79	-41.49 ± 0.31	0.19
6 	-46.04 ± 0.84	-46.55 ± 0.05	0.51
7 	-48.70 ± 0.57	-48.55 ± 0.10	0.15
8 	-41.66 ± 0.56	-41.80 ± 0.20	0.14
9 	-46.22 ± 0.62	-45.73 ± 0.40	0.50
10 	-44.33 ± 0.51	-44.80 ± 0.60	0.47
11 	-49.25 ± 0.15	-49.45 ± 0.79	0.20
12 	-45.73 ± 0.56	-45.37 ± 0.50	0.36
13 	-41.98 ± 0.85	-41.45 ± 0.06	0.53

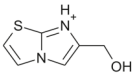
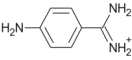


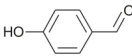
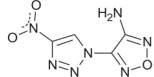
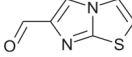
14		-46.55 ± 0.96	-46.49 ± 0.33	0.06
15		-45.83 ± 0.58	-45.98 ± 0.25	0.15
16		-5.38 ± 0.39	-5.28 ± 0.04	0.10
17		-14.25 ± 0.32^a	-12.87 ± 0.33^b	1.38
18		-9.08 ± 0.47	-8.43 ± 0.59	0.65
19		-9.99 ± 0.27	-9.70 ± 0.25	0.30
20		-8.71 ± 0.48	-8.61 ± 0.18	0.10
Mean Unsigned Deviation (without 17)				0.34
Mean Unsigned Deviation				0.40
RMSD (without 17)				0.39
RMSD				0.49

Figure A.1: **Calculated Hydration Free Energies of 20 compounds:**

^a $-14.06 \pm 0.22 \text{ kcal}\cdot\text{mol}^{-1}$ after simulations extended from 1ns to 5ns at each of the 25 steps.

^b $-12.97 \pm 0.39 \text{ kcal}\cdot\text{mol}^{-1}$ after simulations extended from 4ns to 8ns, averaged over 7 independent simulations.

Appendix B: Chapter 5: Secondary Structures of Ultra-short DNA tethers

B.1 MFold Predictions

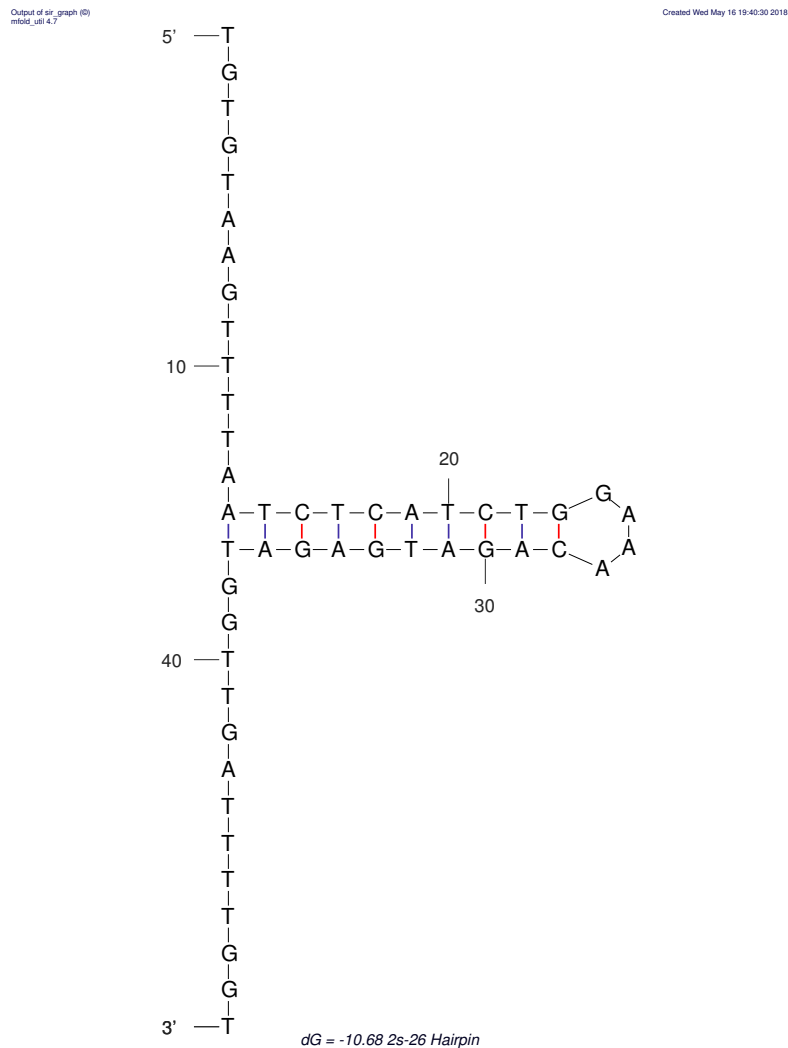
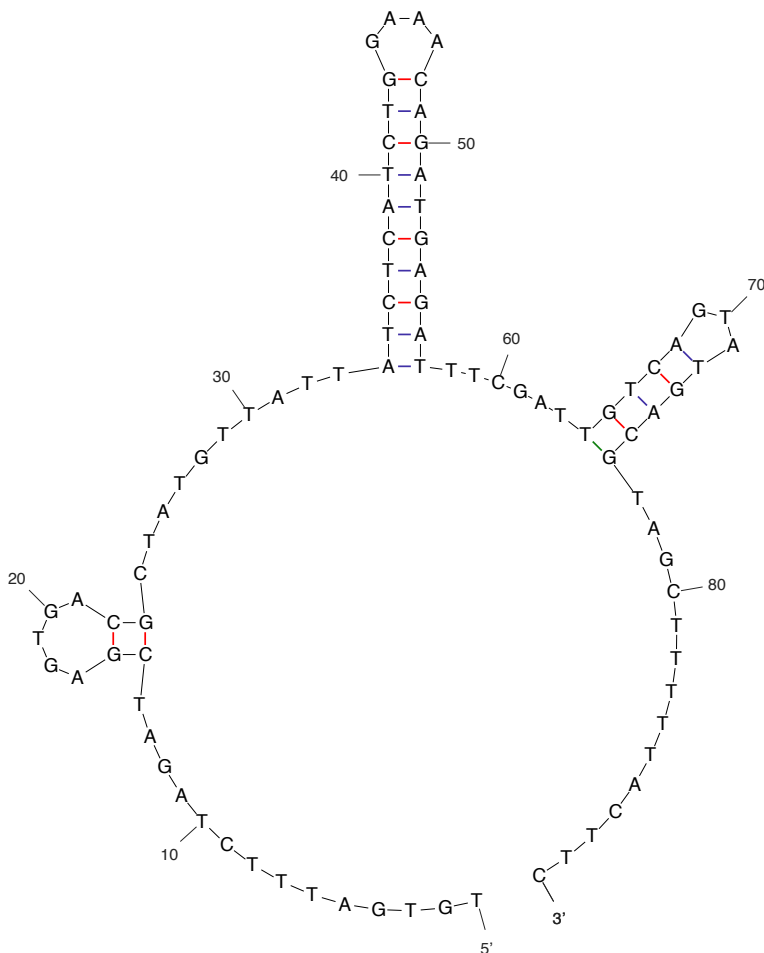


Figure B.1: Mfold Prediction for 26nt Two State DNA Hairpin extended sequences for DNA tethers: Secondary structure predicted for the 26nt Two State DNA Hairpin Tether (Fig. 5.4). Note there is no predicted secondary structure for the extended sequences.



dG = -13.71 2s hairpin 100nt ext [flipped fixed]

Figure B.3: Mfold Prediction for 94 nt Two State DNA Hairpin extended sequences for DNA tethers: Secondary structure predicted for the 94 nt Two State DNA Hairpin Tether (Fig. 5.6). Note there are two hairpins predicted to form in the extended sequences. The 5' hairpin is small and likely unstable due to the short 2 bp stem. The 3' hairpin is slightly larger with a 5 bp stem, but it is less stable than the 10 bp stem in the 2s hairpin sequence

Bibliography

- [1] Huppert J. L. and Balasubramanian S. “Prevalence of quadruplexes in the human genome”. In: *Nucleic Acids Res* 33.9 (2005), pp. 2908–16. ISSN: 1362-4962 (Electronic) 0305-1048 (Linking). DOI: 10.1093/nar/gki609. URL: <https://www.ncbi.nlm.nih.gov/pubmed/15914667>.
- [2] N. Maizels. “G4-associated human diseases”. In: *EMBO Rep* 16.8 (2015), pp. 910–22. ISSN: 1469-3178 (Electronic) 1469-221X (Linking). DOI: 10.15252/embr.201540607. URL: <http://www.ncbi.nlm.nih.gov/pubmed/26150098>.
- [3] A. R. Carter et al. “Stabilization of an optical microscope to 0.1 nm in three dimensions”. In: *Applied Optics* 46.3 (2007).
- [4] J. Eddy and N. Maizels. “Gene function correlates with potential for G4 DNA formation in the human genome”. In: *Nucleic Acids Res* 34.14 (2006), pp. 3887–96. ISSN: 1362-4962 (Electronic) 0305-1048 (Linking). DOI: 10.1093/nar/gkl529. URL: <http://www.ncbi.nlm.nih.gov/pubmed/16914419>.
- [5] J. Eddy and N. Maizels. “Conserved elements with potential to form polymorphic G-quadruplex structures in the first intron of human genes”. In: *Nucleic Acids Res* 36.4 (2008), pp. 1321–33. ISSN: 1362-4962 (Electronic) 0305-1048 (Linking). DOI: 10.1093/nar/gkm1138. URL: <http://www.ncbi.nlm.nih.gov/pubmed/18187510>.
- [6] S. Millevoi, H. Moine, and S. Vagner. “G-quadruplexes in RNA biology”. In: *Wiley Interdiscip Rev RNA* 3.4 (2012), pp. 495–507. ISSN: 1757-7012 (Electronic) 1757-7004 (Linking). DOI: 10.1002/wrna.1113. URL: <http://www.ncbi.nlm.nih.gov/pubmed/22488917>.
- [7] P. Murat and S. Balasubramanian. “Existence and consequences of G-quadruplex structures in DNA”. In: *Curr Opin Genet Dev* 25 (2014), pp. 22–9. ISSN: 1879-0380 (Electronic) 0959-437X (Linking). DOI: 10.1016/j.gde.2013.10.012. URL: <http://www.ncbi.nlm.nih.gov/pubmed/24584093>.

- [8] D. Rhodes and H. J. Lipps. “G-quadruplexes and their regulatory roles in biology”. In: *Nucleic Acids Res* 43.18 (2015), pp. 8627–37. ISSN: 1362-4962 (Electronic) 0305-1048 (Linking). DOI: 10.1093/nar/gkv862. URL: <http://www.ncbi.nlm.nih.gov/pubmed/26350216>.
- [9] A. M. Mustoe, C. L. Brooks, and H. M. Al-Hashimi. “Hierarchy of RNA functional dynamics”. In: *Annu Rev Biochem* 83 (2014), pp. 441–66. ISSN: 1545-4509 (Electronic) 0066-4154 (Linking). DOI: 10.1146/annurev-biochem-060713-035524. URL: <http://www.ncbi.nlm.nih.gov/pubmed/24606137>.
- [10] J. Spöner et al. “RNA Structural Dynamics As Captured by Molecular Simulations: A Comprehensive Overview”. In: *Chem Rev* 118.8 (2018), pp. 4177–4338. ISSN: 1520-6890 (Electronic) 0009-2665 (Linking). DOI: 10.1021/acs.chemrev.7b00427. URL: <https://www.ncbi.nlm.nih.gov/pubmed/29297679>.
- [11] W. K. Dawson et al. “Coarse-grained modeling of RNA 3D structure”. In: *Methods* 103 (2016), pp. 138–156. ISSN: 1046-2023. DOI: <http://dx.doi.org/10.1016/j.ymeth.2016.04.026>. URL: <http://www.sciencedirect.com/science/article/pii/S1046202316301050>.
- [12] H. I. Ingólfsson et al. “The power of coarse graining in biomolecular simulations”. In: *Wiley Interdiscip Rev Comput Mol Sci* 4.3 (2014), pp. 225–248. ISSN: 1759-0876 (Print) 1759-0884 (Linking). DOI: 10.1002/wcms.1169. URL: <http://www.ncbi.nlm.nih.gov/pubmed/25309628>.
- [13] A. Malhotra, R. K. Tan, and S. C. Harvey. “Modeling large RNAs and ribonucleoprotein particles using molecular mechanics techniques”. In: *Biophys J* 66.6 (1994), pp. 1777–95. ISSN: 0006-3495 (Print) 0006-3495 (Linking). DOI: 10.1016/S0006-3495(94)80972-5. URL: <https://www.ncbi.nlm.nih.gov/pubmed/7521223>.
- [14] C. Hyeon and D. Thirumalai. “Mechanical unfolding of RNA hairpins”. In: *Proc Natl Acad Sci U S A* 102.19 (2005), pp. 6789–94. ISSN: 0027-8424 (Print) 0027-8424 (Linking). DOI: 10.1073/pnas.0408314102. URL: <https://www.ncbi.nlm.nih.gov/pubmed/15749822>.

- [15] F. Ding et al. “Ab initio folding of proteins with all-atom discrete molecular dynamics”. In: *Structure* 16.7 (2008), pp. 1010–8. ISSN: 0969-2126 (Print) 0969-2126 (Linking). DOI: 10.1016/j.str.2008.03.013. URL: <https://www.ncbi.nlm.nih.gov/pubmed/18611374>.
- [16] C. M. Gherghe et al. “Native-like RNA tertiary structures using a sequence-encoded cleavage agent and refinement by discrete molecular dynamics”. In: *J Am Chem Soc* 131.7 (2009), pp. 2541–6. ISSN: 1520-5126 (Electronic) 0002-7863 (Linking). DOI: 10.1021/ja805460e. URL: <https://www.ncbi.nlm.nih.gov/pubmed/19193004>.
- [17] Z. Xia et al. “Coarse-Grained Model for Simulation of RNA Three-Dimensional Structures”. In: *Journal of Physical Chemistry B* 114.42 (2010), pp. 13497–13506. ISSN: 1520-6106. DOI: 10.1021/jp104926t. URL: <https://www ISI%3E://WOS:000283110500030>.
- [18] T. Cragolini, P. Derreumaux, and S. Pasquali. “Coarse-Grained Simulations of RNA and DNA Duplexes”. In: *Journal of Physical Chemistry B* 117.27 (2013), pp. 8047–8060. ISSN: 1520-6106. DOI: 10.1021/jp400786b. URL: <https://www ISI%3E://WOS:000321884100001>.
- [19] T. Cragolini et al. “Coarse-Grained HiRE-RNA Model for ab Initio RNA Folding beyond Simple Molecules, Including Noncanonical and Multiple Base Pairings”. In: *Journal of Chemical Theory and Computation* 11.7 (2015), pp. 3510–3522. ISSN: 1549-9618. DOI: 10.1021/acs.jctc.5b00200. URL: <http://dx.doi.org/10.1021/acs.jctc.5b00200>.
- [20] J. J. Uusitalo et al. “Martini Coarse-Grained Force Field: Extension to RNA”. In: *Biophys J* 113.2 (2017), pp. 246–256. ISSN: 1542-0086 (Electronic) 0006-3495 (Linking). DOI: 10.1016/j.bpj.2017.05.043. URL: <https://www.ncbi.nlm.nih.gov/pubmed/28633759>.
- [21] Z. Xia et al. “RNA 3D Structure Prediction by Using a Coarse-Grained Model and Experimental Data”. In: *The Journal of Physical Chemistry B* 117.11 (2013), pp. 3135–3144. ISSN: 1520-6106. DOI: 10.1021/jp400751w. URL: <http://dx.doi.org/10.1021/jp400751w>.

- [22] R. K. Z. Tan, A. S. Petrov, and S. C. Harvey. “YUP: A molecular simulation program for coarse-grained and multiscaled models”. In: *Journal of Chemical Theory and Computation* 2.3 (2006), pp. 529–540. ISSN: 1549-9618. DOI: 10.1021/ct050323r. URL: <https://doi.org/10.1021/ct050323r>. URL: <https://pubs.acs.org/doi/abs/10.1021/ct050323r>.
- [23] M. J. Sippl. “Calculation of conformational ensembles from potentials of mean force. An approach to the knowledge-based prediction of local structures in globular proteins”. In: *J Mol Biol* 213.4 (1990), pp. 859–83. ISSN: 0022-2836 (Print) 0022-2836 (Linking). DOI: 10.1016/0022-2836(90)90125-5. URL: <https://pubmed.ncbi.nlm.nih.gov/2359125/>.
- [24] T. Hamelryck et al. “Potentials of mean force for protein structure prediction vindicated, formalized and generalized”. In: *PLoS One* 5.11 (2010), e13714. ISSN: 1932-6203 (Electronic) 1932-6203 (Linking). DOI: 10.1371/journal.pone.0013714. URL: <https://pubmed.ncbi.nlm.nih.gov/21103041/>.
- [25] M. T. Sykes and M. Levitt. “Describing RNA structure by libraries of clustered nucleotide doublets”. In: *J Mol Biol* 351.1 (2005), pp. 26–38. ISSN: 0022-2836 (Print) 0022-2836 (Linking). DOI: 10.1016/j.jmb.2005.06.024. URL: <https://pubmed.ncbi.nlm.nih.gov/15993894/>.
- [26] Z. J. Tan and S. J. Chen. “RNA helix stability in mixed Na⁺/Mg²⁺ solution”. In: *Biophys J* 92.10 (2007), pp. 3615–32. ISSN: 0006-3495 (Print) 0006-3495 (Linking). DOI: 10.1529/biophysj.106.100388. URL: <https://pubmed.ncbi.nlm.nih.gov/17325014/>.
- [27] R. Das et al. “Structural inference of native and partially folded RNA by high-throughput contact mapping”. In: *Proc Natl Acad Sci U S A* 105.11 (2008), pp. 4144–9. ISSN: 1091-6490 (Electronic) 0027-8424 (Linking). DOI: 10.1073/pnas.0709032105. URL: <https://pubmed.ncbi.nlm.nih.gov/18322008/>.
- [28] M. A. Jonikas, R. J. Radmer, and R. B. Altman. “Knowledge-based instantiation of full atomic detail into coarse-grain RNA 3D structural models”. In: *Bioinformatics* 25.24 (2009), pp. 3259–3266. DOI: 10.1093/bioinformatics/btp576. URL: <http://bioinformatics.oxfordjournals.org/content/25/24/3259.abstract>.

- [29] M. A. Jonikas et al. “Coarse-grained modeling of large RNA molecules with knowledge-based potentials and structural filters”. In: *Rna* 15.2 (2009), pp. 189–199. DOI: 10.1261/rna.1270809. URL: <http://rnajournal.cshlp.org/content/15/2/189.abstract>.
- [30] J. Bernauer et al. “Fully differentiable coarse-grained and all-atom knowledge-based potentials for RNA structure evaluation”. In: *Rna* 17.6 (2011), pp. 1066–1075. DOI: 10.1261/rna.2543711. URL: <http://rnajournal.cshlp.org/content/17/6/1066.abstract>.
- [31] S. Cao and S. J. Chen. “Physics-Based De Novo Prediction of RNA 3D Structures”. In: *Journal of Physical Chemistry B* 115.14 (2011), pp. 4216–4226. ISSN: 1520-6106. DOI: 10.1021/jp112059y. URL: [%3CGo%20to%20ISI%3E://WOS:000289215600050](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3060050/).
- [32] Z. He, Y. Zhu, and S. J. Chen. “Exploring the electrostatic energy landscape for tetraloop-receptor docking”. In: *Phys Chem Chem Phys* 16.14 (2014), pp. 6367–75. ISSN: 1463-9084 (Electronic) 1463-9076 (Linking). DOI: 10.1039/c3cp53655f. URL: <https://www.ncbi.nlm.nih.gov/pubmed/24322001>.
- [33] Z. He and S. J. Chen. “Predicting ion-nucleic acid interactions by energy landscape-guided sampling”. In: *J Chem Theory Comput* 8.6 (2012), pp. 2095–2101. ISSN: 1549-9626 (Electronic) 1549-9618 (Linking). DOI: 10.1021/ct300227a. URL: <https://www.ncbi.nlm.nih.gov/pubmed/23002389>.
- [34] X. Xu and S. J. Chen. “Physics-based RNA structure prediction”. In: *Biophys Rep* 1 (2015), pp. 2–13. ISSN: 2364-3439 (Print) 2364-3439 (Linking). DOI: 10.1007/s41048-015-0001-4. URL: <https://www.ncbi.nlm.nih.gov/pubmed/26942214>.
- [35] Y. Zhu, Z. He, and S. J. Chen. “TBI server: a web server for predicting ion effects in RNA folding”. In: *PLoS One* 10.3 (2015), e0119705. ISSN: 1932-6203 (Electronic) 1932-6203 (Linking). DOI: 10.1371/journal.pone.0119705. URL: <https://www.ncbi.nlm.nih.gov/pubmed/25798933>.
- [36] J. A. Cruz et al. “RNA-Puzzles: a CASP-like evaluation of RNA three-dimensional structure prediction”. In: *RNA* 18.4 (2012), pp. 610–25. ISSN: 1469-9001 (Electronic) 1355-8382 (Linking). DOI: 10.1261/rna.031054.111. URL: <https://www.ncbi.nlm.nih.gov/pubmed/22361291>.

- [37] Z. Miao et al. “RNA-Puzzles Round III: 3D RNA structure prediction of five riboswitches and one ribozyme”. In: *RNA* 23.5 (2017), pp. 655–672. ISSN: 1469-9001 (Electronic) 1355-8382 (Linking). DOI: 10.1261/rna.060368.116. URL: <https://www.ncbi.nlm.nih.gov/pubmed/28138060>.
- [38] Z. Miao et al. “RNA-Puzzles Round II: assessment of RNA structure prediction programs applied to three large RNA structures”. In: *RNA* 21.6 (2015), pp. 1066–84. ISSN: 1469-9001 (Electronic) 1355-8382 (Linking). DOI: 10.1261/rna.049502.114. URL: <https://www.ncbi.nlm.nih.gov/pubmed/25883046>.
- [39] S. J. Chen. “RNA folding: conformational statistics, folding kinetics, and ion electrostatics”. In: *Annu Rev Biophys* 37 (2008), pp. 197–214. ISSN: 1936-122X (Print) 1936-122X (Linking). DOI: 10.1146/annurev.biophys.37.032807.125957. URL: <https://www.ncbi.nlm.nih.gov/pubmed/18573079>.
- [40] T. B. Xia et al. “Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs”. In: *Biochemistry* 37.42 (1998), pp. 14719–14735. ISSN: 0006-2960. DOI: 10.1021/bi9809425. URL: <https://www.ncbi.nlm.nih.gov/pubmed/9809425>.
- [41] D. H. Mathews et al. “Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure”. In: *Proceedings of the National Academy of Sciences* 101.19 (2004), pp. 7287–7292. ISSN: 0027-8424. DOI: 10.1073/pnas.0401799101. URL: <http://www.pnas.org/content/101/19/7287>.
- [42] E. Fadrna et al. “Molecular dynamics simulations of guanine quadruplex loops: Advances and force field limitations”. In: *Biophysical Journal* 87.1 (2004), pp. 227–242. ISSN: 0006-3495. DOI: 10.1529/biophysj.103.034751.
- [43] M. Krepl et al. “Reference simulations of noncanonical nucleic acids with different chi variants of the AMBER force field: quadruplex DNA, quadruplex RNA and Z-DNA”. In: *J Chem Theory Comput* 8.7 (2012), pp. 2506–2520. ISSN: 1549-9626 (Electronic) 1549-9618 (Linking). DOI: 10.1021/ct300275s. URL: <https://www.ncbi.nlm.nih.gov/pubmed/23197943>.
- [44] J. Sponer, X. Cang, and T. E. Cheatham 3rd. “Molecular dynamics simulations of G-DNA and perspectives on the simulation of nucleic acid structures”. In: *Methods* 57.1 (2012), pp. 25–39. ISSN: 1095-9130 (Electronic) 1046-2023

- (Linking). DOI: 10.1016/j.ymeth.2012.04.005. URL: <https://www.ncbi.nlm.nih.gov/pubmed/22525788>.
- [45] J. Liphardt et al. “Reversible Unfolding of Single RNA Molecules by Mechanical Force”. In: *Science* 292.5517 (2001), pp. 733–737. DOI: 10.1126/science.1058498. URL: <http://science.sciencemag.org/content/sci/292/5517/733.full.pdf>.
- [46] W. J. Greenleaf et al. “Passive all-optical force clamp for high-resolution laser trapping”. In: *Phys Rev Lett* 95.20 (2005), p. 208102. ISSN: 0031-9007 (Print) 0031-9007 (Linking). DOI: 10.1103/PhysRevLett.95.208102. URL: <https://www.ncbi.nlm.nih.gov/pubmed/16384102>.
- [47] M. T. Woodside et al. “Direct Measurement of the Full, Sequence-Dependent Folding Landscape of a Nucleic Acid”. In: *Science* 314.5801 (2006), pp. 1001–1004. DOI: 10.1126/science.1133601. URL: <http://www.sciencemag.org/content/314/5801/1001.abstract>.
- [48] M. T. Woodside et al. “Nanomechanical measurements of the sequence-dependent folding landscapes of single nucleic acid hairpins”. In: *Proceedings of the National Academy of Sciences* 103.16 (2006), pp. 6190–6195. DOI: 10.1073/pnas.0511048103. URL: <http://www.pnas.org/content/103/16/6190.abstract>.
- [49] W. J. Greenleaf et al. “Direct observation of hierarchical folding in single riboswitch aptamers”. In: *Science* 319.5863 (2008), pp. 630–3. ISSN: 1095-9203 (Electronic) 0036-8075 (Linking). DOI: 10.1126/science.1151298. URL: <https://www.ncbi.nlm.nih.gov/pubmed/18174398>.
- [50] M. T. Woodside, C. Garcia-Garcia, and S. M. Block. “Folding and unfolding single RNA molecules under tension”. In: *Curr Opin Chem Biol* 12.6 (2008), pp. 640–6. ISSN: 1879-0402 (Electronic) 1367-5931 (Linking). DOI: 10.1016/j.cbpa.2008.08.011. URL: <https://www.ncbi.nlm.nih.gov/pubmed/18786653>.
- [51] A.N. Gupta et al. “Experimental validation of free-energy-landscape reconstruction from non-equilibrium single-molecule force spectroscopy measurements”. In: *Nature Physics* 7.8 (2011), pp. 631–634. ISSN: 1745-2473 1745-2481. DOI: 10.1038/nphys2022.

- [52] A. P. Manuel, J. Lambert, and M. T. Woodside. “Reconstructing folding energy landscapes from splitting probability analysis of single-molecule trajectories”. In: *Proc Natl Acad Sci U S A* 112.23 (2015), pp. 7183–8. ISSN: 1091-6490 (Electronic) 0027-8424 (Linking). DOI: 10.1073/pnas.1419490112. URL: <https://www.ncbi.nlm.nih.gov/pubmed/26039984>.
- [53] K. Neupane and M. T. Woodside. “Quantifying Instrumental Artifacts in Folding Kinetics Measured by Single-Molecule Force Spectroscopy”. In: *Biophys J* 111.2 (2016), pp. 283–6. ISSN: 1542-0086 (Electronic) 0006-3495 (Linking). DOI: 10.1016/j.bpj.2016.06.011. URL: <https://www.ncbi.nlm.nih.gov/pubmed/27369870>.
- [54] M. T. Woodside and S. M. Block. “Reconstructing folding energy landscapes by single-molecule force spectroscopy”. In: *Annu Rev Biophys* 43 (2014), pp. 19–39. ISSN: 1936-1238 (Electronic) 1936-122X (Linking). DOI: 10.1146/annurev-biophys-051013-022754. URL: <https://www.ncbi.nlm.nih.gov/pubmed/24895850>.
- [55] M. Hengesbach et al. “Single-molecule FRET reveals the folding dynamics of the human telomerase RNA pseudoknot domain”. In: *Angew Chem Int Ed Engl* 51.24 (2012), pp. 5876–5879. ISSN: 1362-4962 (Electronic) 0305-1048 (Linking). DOI: 10.1002/anie.201200526. URL: <https://www.ncbi.nlm.nih.gov/pubmed/22544760>.
- [56] S. Dhakal et al. “Structural and mechanical properties of individual human telomeric G-quadruplexes in molecularly crowded solutions”. In: *Nucleic Acids Res* 41.6 (2013), pp. 3915–23. ISSN: 1362-4962 (Electronic) 0305-1048 (Linking). DOI: 10.1093/nar/gkt038. URL: <https://www.ncbi.nlm.nih.gov/pubmed/23396442>.
- [57] X. Long et al. “Mechanical unfolding of human telomere G-quadruplex DNA probed by integrated fluorescence and magnetic tweezers spectroscopy”. In: *Nucleic Acids Res* 41.4 (2013), pp. 2746–2755. ISSN: 1362-4962 (Electronic) 0305-1048 (Linking). DOI: 10.1093/nar/gks1341. URL: <https://www.ncbi.nlm.nih.gov/pubmed/23303789>.
- [58] P. M. Yangyuru et al. “Mechanochemical properties of individual human telomeric RNA (TERRA) G-quadruplexes”. In: *Chembiochem* 14.15 (2013),

- pp. 1931–5. ISSN: 1439-7633 (Electronic) 1439-4227 (Linking). DOI: 10.1002/cbic.201300350. URL: <https://www.ncbi.nlm.nih.gov/pubmed/24038851>.
- [59] W. Xiao Tippana R. and S. Myong. “G-quadruplex conformation and dynamics are determined by loop length and sequence”. In: *Nucleic Acids Res* 42.12 (2014), pp. 8106–8114. DOI: 10.1093/nar/gku464. URL: <https://www.ncbi.nlm.nih.gov/pubmed/24920827>.
- [60] P. M. Yangyuru et al. “Dual binding of an antibody and a small molecule increases the stability of TERRA G-quadruplex”. In: *Angew Chem Int Ed Engl* 54.3 (2015), pp. 910–913. ISSN: 1362-4962 (Electronic) 0305-1048 (Linking). DOI: 10.1002/anie.201408113. URL: <https://www.ncbi.nlm.nih.gov/pubmed/25421962>.
- [61] E. Largy, J. L. Mergny, and V. Gabelica. “Role of Alkali Metal Ions in G-quadruplex Nucleic Acid Structure and Stability”. In: *Met Ions Life Sci* 16 (2016), pp. 203–58. ISSN: 1559-0836 (Print) 1559-0836 (Linking). DOI: 10.1007/978-3-319-21756-7_7. URL: <https://www.ncbi.nlm.nih.gov/pubmed/26860303>.
- [62] R. Tippana et al. “Single-molecule imaging reveals a common mechanism shared by G-quadruplex-resolving helicases”. In: *Proc Natl Acad Sci U S A* 113.30 (2016), pp. 8448–53. ISSN: 1091-6490 (Electronic) 0027-8424 (Linking). DOI: 10.1073/pnas.1603724113. URL: <https://www.ncbi.nlm.nih.gov/pubmed/27407146>.
- [63] P. M. Yangyuru et al. “The G-quadruplex (G4) resolvase DHX36 efficiently and specifically disrupts DNA G4s via a translocation-based helicase mechanism”. In: *J Biol Chem* 293.6 (2018), pp. 1924–1932. ISSN: 1083-351X (Electronic) 0021-9258 (Linking). DOI: 10.1074/jbc.M117.815076. URL: <https://www.ncbi.nlm.nih.gov/pubmed/29269411>.
- [64] M. Aznauryan et al. “A direct view of the complex multi-pathway folding of telomeric G-quadruplexes”. In: *Nucleic Acids Res* 44.22 (2016), pp. 11024–11032. ISSN: 1362-4962 (Electronic) 0305-1048 (Linking). DOI: 10.1093/nar/gkw1010. URL: <https://www.ncbi.nlm.nih.gov/pubmed/27799468>.

- [65] J.W. Parks and Stone MD. “Single-Molecule Studies of Telomeres and Telomerase”. In: *Annu Rev Biophys* 46 (2017), pp. 357–377. ISSN: 1362-4962 (Electronic) 0305-1048 (Linking). DOI: 10.1146/annurev-biophys-062215-011256. URL: <https://www.ncbi.nlm.nih.gov/pubmed/28375735>.
- [66] J. C. Chang, M. de Messieres, and A. La Porta. “Effect of handle length and microsphere size on transition kinetics in single-molecule experiments”. In: *Phys Rev E Stat Nonlin Soft Matter Phys* 87.1 (2013), p. 012721. ISSN: 1550-2376 (Electronic) 1539-3755 (Linking). DOI: 10.1103/PhysRevE.87.012721. URL: <https://www.ncbi.nlm.nih.gov/pubmed/23410373>.
- [67] C. Hyeon, G. Morrison, and D. Thirumalai. “Force-dependent hopping rates of RNA hairpins can be estimated from accurate measurement of the folding landscapes”. In: *Proc Natl Acad Sci U S A* 105.28 (2008), pp. 9604–9. ISSN: 1091-6490 (Electronic) 0027-8424 (Linking). DOI: 10.1073/pnas.0802484105. URL: <https://www.ncbi.nlm.nih.gov/pubmed/18621721>.
- [68] J. D. Wen et al. “Force unfolding kinetics of RNA using optical tweezers. I. Effects of experimental variables on measured results”. In: *Biophys J* 92.9 (2007), pp. 2996–3009. ISSN: 0006-3495 (Print) 0006-3495 (Linking). DOI: 10.1529/biophysj.106.094052. URL: <https://www.ncbi.nlm.nih.gov/pubmed/17293410>.
- [69] N. Forns et al. “Improving signal/noise resolution in single-molecule experiments using molecular constructs with short handles”. In: *Biophys J* 100.7 (2011), pp. 1765–74. ISSN: 1542-0086 (Electronic) 0006-3495 (Linking). DOI: 10.1016/j.bpj.2011.01.071. URL: <https://www.ncbi.nlm.nih.gov/pubmed/21463590>.
- [70] M. de Messieres, B. Brawn-Cinani, and A. La Porta. “Measuring the folding landscape of a harmonically constrained biopolymer”. In: *Biophys J* 100.11 (2011), pp. 2736–44. ISSN: 1542-0086 (Electronic) 0006-3495 (Linking). DOI: 10.1016/j.bpj.2011.03.067. URL: <https://www.ncbi.nlm.nih.gov/pubmed/21641319>.
- [71] P. Cossio, G. Hummer, and A. Szabo. “On artifacts in single-molecule force spectroscopy”. In: *Proc Natl Acad Sci U S A* 112.46 (2015), pp. 14248–53. ISSN:

- 1091-6490 (Electronic) 0027-8424 (Linking). DOI: 10.1073/pnas.1519633112. URL: <https://www.ncbi.nlm.nih.gov/pubmed/26540730>.
- [72] J. Camunas-Soler, M. Ribezzi-Crivellari, and F. Ritort. “Elastic Properties of Nucleic Acids by Single-Molecule Force Spectroscopy”. In: *Annu Rev Biophys* 45 (2016), pp. 65–84. ISSN: 1936-1238 (Electronic) 1936-122X (Linking). DOI: 10.1146/annurev-biophys-062215-011158. URL: <https://www.ncbi.nlm.nih.gov/pubmed/27145878>.
- [73] J. R. Abella et al. “Hydration Free Energy from Orthogonal Space Random Walk and Polarizable Force Field”. In: *J Chem Theory Comput* 10.7 (2014), pp. 2792–2801. ISSN: 1549-9618 (Print) 1549-9618 (Linking). DOI: 10.1021/ct500202q. URL: <http://www.ncbi.nlm.nih.gov/pubmed/25018674>.
- [74] P. Kollman. “Free energy calculations: Applications to chemical and biochemical phenomena”. In: *Chemical Reviews* 93.7 (1993), pp. 2395–2417. ISSN: 0009-2665. DOI: 10.1021/cr00023a004. URL: <http://dx.doi.org/10.1021/cr00023a004>.
- [75] W. L. Jorgensen. “The Many Roles of Computation in Drug Discovery”. In: *Science* 303.5665 (2004), pp. 1813–1818. DOI: 10.1126/science.1096361. URL: <http://www.sciencemag.org/content/303/5665/1813.abstract>.
- [76] D. L. Mobley, J. D. Chodera, and K. A. Dill. “On the use of orientational restraints and symmetry corrections in alchemical free energy calculations”. In: *The Journal of Chemical Physics* 125.8 (2006), pp. 084902–084917. DOI: doi:http://dx.doi.org/10.1063/1.2221683. URL: <http://scitation.aip.org/content/aip/journal/jcp/125/8/10.1063/1.2221683>.
- [77] M. J. Schnieders et al. “The Structure, Thermodynamics, and Solubility of Organic Crystals from Simulation with a Polarizable Force Field”. In: *Journal of Chemical Theory and Computation* 8.5 (2012), pp. 1721–1736. ISSN: 1549-9618. DOI: 10.1021/ct300035u. URL: <http://dx.doi.org/10.1021/ct300035u>.
- [78] Y. Shi et al. “Multipole electrostatics in hydration free energy calculations”. In: *Journal of Computational Chemistry* 32.5 (2011), pp. 967–977. ISSN: 1096-987X. DOI: 10.1002/jcc.21681. URL: <http://dx.doi.org/10.1002/jcc.21681>.

- [79] D. L. Mobley et al. “Small Molecule Hydration Free Energies in Explicit Solvent: An Extensive Test of Fixed-Charge Atomistic Simulations”. In: *Journal of Chemical Theory and Computation* 5.2 (2009), pp. 350–358. ISSN: 1549-9618. DOI: 10.1021/ct800409d. URL: <http://dx.doi.org/10.1021/ct800409d>.
- [80] Michael R. Shirts et al. “Extremely precise free energy calculations of amino acid side chain analogs: Comparison of common molecular mechanics force fields for proteins”. In: *The Journal of Chemical Physics* 119.11 (2003), pp. 5740–5761. DOI: doi:http://dx.doi.org/10.1063/1.1587119. URL: <http://scitation.aip.org/content/aip/journal/jcp/119/11/10.1063/1.1587119>.
- [81] M. R. Shirts and V. S. Pande. “Solvation free energies of amino acid side chain analogs for common molecular mechanics water models”. In: *The Journal of Chemical Physics* 122.13 (2005), pp. 134508–134520. DOI: doi:http://dx.doi.org/10.1063/1.1877132. URL: <http://scitation.aip.org/content/aip/journal/jcp/122/13/10.1063/1.1877132>.
- [82] R. C. Rizzo et al. “Estimation of Absolute Free Energies of Hydration Using Continuum Methods: Accuracy of Partial Charge Models and Optimization of Nonpolar Contributions”. In: *Journal of Chemical Theory and Computation* 2.1 (2005), pp. 128–139. ISSN: 1549-9618. DOI: 10.1021/ct0500971. URL: <http://dx.doi.org/10.1021/ct0500971>.
- [83] C. Baker et al. “Accurate Calculation of Hydration Free Energies using Pair-Specific Lennard-Jones Parameters in the CHARMM Drude Polarizable Force Field”. In: *Journal of Chemical Theory and Computation* 6.4 (2010), pp. 1181–1198. ISSN: 1549-9618. DOI: 10.1021/ct9005773. URL: <http://dx.doi.org/10.1021/ct9005773>.
- [84] M. R. Paliwal H. and Shirts. “A Benchmark Test Set for Alchemical Free Energy Transformations and Its Use to Quantify Error in Common Free Energy Methods”. In: *Journal of Chemical Theory and Computation* 7.12 (2011), pp. 4115–4134. ISSN: 1549-9618. DOI: 10.1021/ct2003995. URL: <http://dx.doi.org/10.1021/ct2003995>.
- [85] Tyler Luchko et al. “Three-Dimensional Molecular Theory of Solvation Coupled with Molecular Dynamics in Amber”. In: *Journal of Chemical Theory*

- and Computation* 6.3 (2010), pp. 607–624. ISSN: 1549-9618. DOI: 10.1021/ct900460m. URL: <http://dx.doi.org/10.1021/ct900460m>.
- [86] D. Sitkoff, K. A. Sharp, and B. Honig. “Accurate Calculation of Hydration Free Energies Using Macroscopic Solvent Models”. In: *The Journal of Physical Chemistry* 98.7 (1994), pp. 1978–1988. ISSN: 0022-3654. DOI: 10.1021/j100058a043. URL: <http://dx.doi.org/10.1021/j100058a043>.
- [87] W. Clark Still et al. “Semianalytical treatment of solvation for molecular mechanics and dynamics”. In: *Journal of the American Chemical Society* 112.16 (1990), pp. 6127–6129. ISSN: 0002-7863. DOI: 10.1021/ja00172a038. URL: <http://dx.doi.org/10.1021/ja00172a038>.
- [88] T. Steinbrecher, I. Joung, and D. A. Case. “Soft-core potentials in thermodynamic integration: Comparing one- and two-step transformations”. In: *Journal of Computational Chemistry* 32.15 (2011), pp. 3253–3263. ISSN: 1096-987X. DOI: 10.1002/jcc.21909. URL: <http://dx.doi.org/10.1002/jcc.21909>.
- [89] M. Feig et al. “Performance comparison of generalized born and Poisson methods in the calculation of electrostatic solvation energies for protein structures”. In: *Journal of Computational Chemistry* 25.2 (2004), pp. 265–284. ISSN: 1096-987X. DOI: 10.1002/jcc.10378. URL: <http://dx.doi.org/10.1002/jcc.10378>.
- [90] W. L. Jorgensen et al. “Comparison of simple potential functions for simulating liquid water”. In: *The Journal of Chemical Physics* 79.2 (1983), pp. 926–935. DOI: [doi:http://dx.doi.org/10.1063/1.445869](http://dx.doi.org/10.1063/1.445869). URL: <http://scitation.aip.org/content/aip/journal/jcp/79/2/10.1063/1.445869>.
- [91] P. Ren and J. W. Ponder. “Polarizable Atomic Multipole Water Model for Molecular Mechanics Simulation”. In: *Journal of Physical Chemistry B* 107.24 (2003), pp. 5933–5947. URL: <http://dx.doi.org/10.1021/jp030486a001>.
- [92] Charles H. Bennett. “Efficient estimation of free energy differences from Monte Carlo data”. In: *Journal of computational physics* 22.2 (1976), pp. 245–268. ISSN: 0021-9991. DOI: 10.1016/0021-9991(76)90078-4.

- [93] L. Zheng, M. Chen, and W. Yang. “Random walk in orthogonal space to achieve efficient free-energy simulation of complex systems”. In: *Proceedings of the National Academy of Sciences* 105.51 (2008), pp. 20227–20232. URL: <http://www.pnas.org/content/105/51/20227.abstract>.
- [94] L. Zheng, M. Chen, and W. Yang. “Simultaneous escaping of explicit and hidden free energy barriers: Application of the orthogonal space random walk strategy in generalized ensemble based conformational sampling”. In: *The Journal of Chemical Physics* 130.23 (2009), pp. 234105–234114. DOI: [doi:http://dx.doi.org/10.1063/1.3153841](http://dx.doi.org/10.1063/1.3153841). URL: <http://scitation.aip.org/content/aip/journal/jcp/130/23/10.1063/1.3153841>.
- [95] D. Min et al. “Practically Efficient QM/MM Alchemical Free Energy Simulations: The Orthogonal Space Random Walk Strategy”. In: *Journal of Chemical Theory and Computation* 6.8 (2010), pp. 2253–2266. ISSN: 1549-9618. DOI: [10.1021/ct100033s](http://dx.doi.org/10.1021/ct100033s). URL: <http://dx.doi.org/10.1021/ct100033s>.
- [96] L. Zheng and W. Yang. “Practically Efficient and Robust Free Energy Calculations: Double-Integration Orthogonal Space Tempering”. In: *Journal of Chemical Theory and Computation* 8.3 (2012), pp. 810–823. ISSN: 1549-9618. DOI: [10.1021/ct200726v](http://dx.doi.org/10.1021/ct200726v). URL: <http://dx.doi.org/10.1021/ct200726v>.
- [97] R. A. Wang Q.and Bryce. “Accounting for non-optimal interactions in molecular recognition: a study of ion-[small pi] complexes using a QM/MM model with a dipole-polarisable MM region”. In: *Physical Chemistry Chemical Physics* 13.43 (2011), pp. 19401–19408. ISSN: 1463-9076. DOI: [10.1039/c1cp21944h](http://dx.doi.org/10.1039/c1cp21944h). URL: <http://dx.doi.org/10.1039/c1cp21944h>.
- [98] C. Hensen et al. “A Combined QM/MM Approach to Protein-Ligand Interactions: Polarization Effects of the HIV-1 Protease on Selected High Affinity Inhibitors”. In: *Journal of Medicinal Chemistry* 47.27 (2004). PMID: 15615516, pp. 6673–6680. DOI: [10.1021/jm0497343](http://dx.doi.org/10.1021/jm0497343). eprint: <https://doi.org/10.1021/jm0497343>. URL: <https://doi.org/10.1021/jm0497343>.
- [99] F. J. Orozco M.and Luque, D. Habibollahzadeh, and J. Gao. “The polarization contribution to the free energy of hydration”. In: *The Journal of Chemical Physics* 102.15 (1995), pp. 6145–6152. DOI: [doi:http://dx.doi.org/10.1063/1.471111](http://dx.doi.org/10.1063/1.471111).

- 1063/1.469348. URL: <http://scitation.aip.org/content/aip/journal/jcp/102/15/10.1063/1.469348>.
- [100] J. W. Ponder et al. “Current Status of the AMOEBA Polarizable Force Field”. In: *The Journal of Physical Chemistry B* 114.8 (2010), pp. 2549–2564. ISSN: 1520-6106. DOI: 10.1021/jp910674d. URL: <http://dx.doi.org/10.1021/jp910674d>.
- [101] P. Ren, C. Wu, and J. W. Ponder. “Polarizable Atomic Multipole-based Molecular Mechanics for Organic Molecules”. In: *Journal of Chemical Theory and Computation* 7.10 (2011), 3143–3161 PMID: PMC3196664.
- [102] Y. Shi et al. “The Polarizable Atomic Multipole-based AMOEBA Force Field for Proteins”. In: *J Chem Theory Comput* 9.9 (2013), pp. 4046–4063. ISSN: 1549-9626 (Electronic) 1549-9618 (Linking). DOI: 10.1021/ct4003702. URL: <https://www.ncbi.nlm.nih.gov/pubmed/24163642>.
- [103] D. Jiao et al. “Trypsin Ligand Binding Free Energies from Explicit and Implicit Solvent Simulations with Polarizable Potential”. In: *Journal of Computational Chemistry* 30.11 (2009), 1701–1711 PMID: PMC2752704. ISSN: 0192-8651. URL: [%3CGo%20to%20ISI%3E://000267504400010](https://pubmed.ncbi.nlm.nih.gov/24163642/).
- [104] A. Laio and M. Parrinello. “Escaping free-energy minima”. In: *Proc Natl Acad Sci U S A* 99.20 (2002), pp. 12562–12566. ISSN: 0027-8424 (Print) 0027-8424 (Linking). DOI: 10.1073/pnas.202427399. URL: <http://www.ncbi.nlm.nih.gov/pubmed/12271136>.
- [105] D. Jiao et al. “Calculation of Protein-Ligand Binding Free Energy by Using a Polarizable Potential”. In: *Proceedings of the National Academy of Sciences of the USA* 105.17 (2008), pp. 6290–6295. ISSN: 0027-8424. URL: [%3CGo%20to%20ISI%3E://000255534100016](https://pubmed.ncbi.nlm.nih.gov/12271136/).
- [106] Y. Shi et al. “Probing the Effect of Conformational Constraint on Phosphorylated Ligand Binding to an SH2 Domain Using Polarizable Force Field Simulations”. In: *Journal of Physical Chemistry B* 116.5 (2012), 1716–27 PMID: PMC3277292. ISSN: 1520-5207 (Electronic) 1520-5207 (Linking). DOI: 10.1021/jp210265d. URL: <http://www.ncbi.nlm.nih.gov/pubmed/22214214>.

- [107] J. Zhang et al. “Modeling Structural Coordination and Ligand Binding in Zinc Proteins with a Polarizable Potential”. In: *Journal of Chemical Theory and Computation* 8 (2012), 1314–1324 PMID: PMC3383645.
- [108] A. Grossfield, P. Ren, and J. W. Ponder. “Ion solvation thermodynamics from simulation with a polarizable force field”. In: *Journal of the American Chemical Society* 125.50 (2003), pp. 15671–15682. ISSN: 0002-7863. URL: <https://doi.org/10.1021/ja000187a006>.
- [109] D. Jiao et al. “Simulation of Ca²⁺ and Mg²⁺ solvation using polarizable atomic multipole potential”. In: *Journal of Physical Chemistry B* 110.37 (2006), pp. 18553–18559. ISSN: 1520-6106. URL: <https://doi.org/10.1021/jp061007z>.
- [110] J. C. Wu, G. Chattree, and P. Ren. “Automation of AMOEBA polarizable force field parameterization for small molecules”. In: *Theoretical Chemistry Accounts* 131.3 (2012), pp. 1–11. ISSN: 1432-881X. DOI: 10.1007/s00214-012-1138-6. URL: <http://dx.doi.org/10.1007/s00214-012-1138-6>.
- [111] X. Mu et al. “Modeling Organochlorine Compounds and the sigma-Hole Effect Using a Polarizable Multipole Force Field”. In: *The Journal of Physical Chemistry B* DOI: 10.1021/jp411671a (2014). ISSN: 1520-6106. DOI: 10.1021/jp411671a. URL: <http://dx.doi.org/10.1021/jp411671a>.
- [112] D. A. Pearlman and P. A. Kollman. “The lag between the Hamiltonian and the system configuration in free energy perturbation calculations”. In: *The Journal of Chemical Physics* 91.12 (1989), pp. 7831–7839. DOI: doi:http://dx.doi.org/10.1063/1.457251. URL: <http://scitation.aip.org/content/aip/journal/jcp/91/12/10.1063/1.457251>.
- [113] X. Kong and C.L.Brooks. “Lambda-dynamics: A new approach to free energy calculations”. In: *The Journal of Chemical Physics* 105.6 (1996), pp. 2414–2423. DOI: doi:http://dx.doi.org/10.1063/1.472109. URL: <http://scitation.aip.org/content/aip/journal/jcp/105/6/10.1063/1.472109>.
- [114] T. A. Halgren. “The representation of van der Waals (vdW) interactions in molecular mechanics force fields: potential form, combination rules, and vdW parameters”. In: *Journal of the American Chemical Society* 114.20 (1992),

- pp. 7827–7843. ISSN: 0002-7863. DOI: 10.1021/ja00046a032. URL: <http://dx.doi.org/10.1021/ja00046a032>.
- [115] G. J. Rocklin et al. “Calculating the binding free energies of charged species based on explicit-solvent simulations employing lattice-sum methods: An accurate correction scheme for electrostatic finite-size effects”. In: *Journal of Chemical Physics* 139.18 (2013), pp. 184103–184134. ISSN: 0021-9606. DOI: [10.1063/1.4826261](https://doi.org/10.1063/1.4826261). URL: <http://scitation.aip.org/content/aip/journal/jcp/139/18/10.1063/1.4826261>.
- [116] M. Kastenholz and P.H. Hunenberger. “Computation of methodology independent ionic solvation free energies from molecular simulations. II. The hydration free energy of the sodium cation”. In: *The Journal of Chemical Physics* 124.22 (2006), pp. 224501–224520. DOI: [doi:http://dx.doi.org/10.1063/1.2201698](https://doi.org/10.1063/1.2201698). URL: <http://scitation.aip.org/content/aip/journal/jcp/124/22/10.1063/1.2201698>.
- [117] M. Tuckerman, B. J. Berne, and G. J. Martyna. “Reversible multiple time scale molecular dynamics”. In: *The Journal of Chemical Physics* 97.3 (1992), pp. 1990–2001. DOI: [doi:http://dx.doi.org/10.1063/1.463137](https://doi.org/10.1063/1.463137). URL: <http://scitation.aip.org/content/aip/journal/jcp/97/3/10.1063/1.463137>.
- [118] G. Bussi, D. Donadio, and M. Parrinello. “Canonical sampling through velocity rescaling”. In: *The Journal of Chemical Physics* 126.1 (2007), pp. 014101–014107. DOI: [doi:http://dx.doi.org/10.1063/1.2408420](https://doi.org/10.1063/1.2408420). URL: <http://scitation.aip.org/content/aip/journal/jcp/126/1/10.1063/1.2408420>.
- [119] M. R. Shirts and J. D. Chodera. “Statistically optimal analysis of samples from multiple equilibrium states”. In: *The Journal of Chemical Physics* 129.12 (2008), pp. 124105–124114. ISSN: 1089-7690 (Electronic) 0021-9606 (Linking). DOI: [10.1063/1.2978177](https://doi.org/10.1063/1.2978177). URL: <http://www.ncbi.nlm.nih.gov/pubmed/19045004>.
- [120] J. M. Obliosca et al. “LNA Thymidine Monomer Enables Differentiation of the Four Single-Nucleotide Variants by Melting Temperature”. In: *J Am Chem Soc* 139.20 (2017), pp. 7110–7116. ISSN: 1520-5126 (Electronic) 0002-7863 (Link-

- ing). DOI: 10.1021/jacs.7b03395. URL: <https://www.ncbi.nlm.nih.gov/pubmed/28463488>.
- [121] G. S. Ginsburg and J. J. McCarthy. “Personalized medicine: revolutionizing drug discovery and patient care”. In: *Trends. Biotechnol.* 19.12 (2001), pp. 491–496. ISSN: 0167-7799.
- [122] P. Y. Kwok. “Methods for genotyping single nucleotide polymorphisms”. In: *Annu. Rev. Genomics. Hum. Genet.* 2 (2001), pp. 235–258. ISSN: 1527-8204. URL: <https://www.ncbi.nlm.nih.gov/pubmed/11712042>.
- [123] K. Nakatani. “Chemistry challenges in SNP typing”. In: *ChemBioChem* 5.12 (2004), pp. 1623–1633. ISSN: 1439-4227. DOI: 10.1002/cbic.200400161. URL: <https://onlinelibrary.wiley.com/store/10.1002/cbic.200400161/asset/1623ftp.pdf?v=1&t=hsf3ogka&s=18c17eb276fc89d91b3f05934397de82995f4cba>.
- [124] A. Syvanen. “Accessing genetic variation: genotyping single nucleotide polymorphisms”. In: *Nat. Rev. Genet.* 2.12 (2001), pp. 930–942. ISSN: 1471-0056. URL: <http://www.nature.com/nrg/journal/v2/n12/pdf/nrg1201-930a.pdf>.
- [125] H.-C. Yeh et al. “A fluorescence light-up Ag nanocluster probe that discriminates single-nucleotide variants by emission color”. In: *J. Am. Chem. Soc.* 134.28 (2012), pp. 11550–8. ISSN: 1520-5126 (Electronic) 0002-7863 (Linking). DOI: 10.1021/ja3024737. URL: <http://www.ncbi.nlm.nih.gov/pubmed/22775452><http://pubs.acs.org/doi/pdfplus/10.1021/ja3024737>.
- [126] K. Knez et al. “Emerging technologies for hybridization based single nucleotide polymorphism detection”. In: *Analyst* 139.2 (2014), pp. 353–370. URL: <http://pubs.rsc.org/en/content/articlepdf/2014/an/c3an01436c>.
- [127] M. P. Audrezet et al. “Validation of high-resolution DNA melting analysis for mutation scanning of the cystic fibrosis transmembrane conductance regulator (CFTR) gene”. In: *Journal of Molecular Diagnostics* 10.5 (2008), pp. 424–434. ISSN: 1525-1578. DOI: 10.2353/jmoldx.2008.080056. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2518737/pdf/main.pdf>.

- [128] M. Erali, K. V Voelkerding, and C. T Wittwer. “High-resolution melting applications for clinical laboratory medicine”. In: *Experimental and molecular pathology* 85.1 (2008), pp. 50–58. ISSN: 0014-4800. URL: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2606052/pdf/nihms67161.pdf>.
- [129] G. H Reed, J. O Kent, and C. T. Wittwer. “High-resolution DNA melting analysis for simple and efficient molecular diagnostics”. In: *Pharmacogenomics* 8.6 (2007), pp. 597–608. ISSN: 1462-2416.
- [130] M. J. Lay and C. T. Wittwer. “Real-time fluorescence genotyping of factor V Leiden during rapid-cycle PCR”. In: *Clinical Chemistry* 43.12 (1997), pp. 2262–2267. ISSN: 0009-9147. URL: <http://www.clinchem.org/content/43/12/2262.full.pdf>.
- [131] K. M. Ririe, R. P. Rasmussen, and C. T. Wittwer. “Product differentiation by analysis of DNA melting curves during the polymerase chain reaction”. In: *Analytical Biochemistry* 245.2 (1997), pp. 154–160. ISSN: 0003-2697. DOI: DOI10.1006/abio.1996.9916. URL: <http://www.sciencedirect.com/science/article/pii/S00032697969916>.
- [132] C. T. Wittwer et al. “The LightCycler(TM) a microvolume multisample fluorimeter with rapid temperature control”. In: *Biotechniques* 22.1 (1997), pp. 176–181. ISSN: 0736-6205. URL: <http://www.biotechonix.com/doi/10.1016/j.biotech.1997.03.030>.
- [133] H. Millward et al. “Homogeneous amplification and mutation scanning of the p53 gene using fluorescent melting curves”. In: *Clinical Chemistry* 48.8 (2002), pp. 1321–1328. ISSN: 0009-9147. URL: <http://www.clinchem.org/content/48/8/1321.full.pdf>.
- [134] L. L. Hsieh et al. “Characteristics and prevalence of KRAS, BRAF, and PIK3CA mutations in colorectal cancer by high-resolution melting analysis in Taiwanese population”. In: *Clinica Chimica Acta* 413.19-20 (2012), pp. 1605–1611. ISSN: 0009-8981. DOI: 10.1016/j.cca.2012.04.029. URL: <http://www.sciencedirect.com/science/article/pii/S000989811200032>.
- [135] M. M. Cousins et al. “Comparison of a High-Resolution Melting Assay to Next-Generation Sequencing for Analysis of HIV Diversity”. In: *Journal of Clinical Microbiology* 50.9 (2012), pp. 3054–3059. ISSN: 0095-1137. DOI: 10.1128/Jcm.

- 01460-12. URL: %3CGo%20to%20ISI%3E://WOS:000307941900031%20http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3421787/pdf/zjm3054.pdf.
- [136] C. T. Wittwer. “High-resolution DNA melting analysis advancements and limitations”. In: *Human mutation* 30.6 (2009), pp. 857–859. ISSN: 1098-1004.
- [137] C. N. Gundry et al. “Base-pair neutral homozygotes can be discriminated by calibrated high-resolution melting of small amplications”. In: *Nucleic Acids Research* 36.10 (2008), pp. 3401–3408. ISSN: 0305-1048. DOI: 10.1093/nar/gkn204. URL: %3CGo%20to%20ISI%3E://WOS:000257183200023%20http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2425497/pdf/gkn204.pdf.
- [138] P. S. Bernard, G. H. Pritham, and C. T. Wittwer. “Color multiplexing hybridization probes using the apolipoprotein E locus as a model system for genotyping”. In: *Analytical Biochemistry* 273.2 (1999), pp. 221–228. ISSN: 0003-2697. DOI: DOI10.1006/abio.1999.4217. URL: %3CGo%20to%20ISI%3E://WOS:000082713000011.
- [139] K. S. J. Elenitoba-Johnson et al. “Multiplex PCR by multicolor fluorimetry and fluorescence melting curve analysis”. In: *Nature Medicine* 7.2 (2001), pp. 249–253. ISSN: 1078-8956. DOI: Doi10.1038/84708. URL: %3CGo%20to%20ISI%3E://WOS:000166756300045%20http://www.nature.com/nm/journal/v7/n2/pdf/nm0201_249.pdf.
- [140] A. O. Crockett and C. T. Wittwer. “Fluorescein-labeled oligonucleotides for real-time PCR: Using the inherent quenching of deoxyguanosine nucleotides”. In: *Analytical Biochemistry* 290.1 (2001), pp. 89–97. ISSN: 0003-2697. DOI: DOI10.1006/abio.2000.4957. URL: %3CGo%20to%20ISI%3E://WOS:000167283300011.
- [141] S. Kurata et al. “Fluorescent quenching-based quantitative detection of specific DNA/RNA using a BODIPY (R) FL-labeled probe or primer”. In: *Nucleic Acids Research* 29.6 (2001). ISSN: 0305-1048. DOI: ARTNe3410.1093/nar/29.6.e34. URL: %3CGo%20to%20ISI%3E://WOS:000207967600005%20http://nar.oxfordjournals.org/content/29/6/e34.full.pdf.
- [142] L. M. Zhou et al. “Snapback primer genotyping with saturating DNA dye and melting analysis”. In: *Clinical Chemistry* 54.10 (2008), pp. 1648–1656. ISSN: 0009-9147. DOI: 10.1373/clinchem.2008.107615. URL: %3CGo%20to%20ISI%

- 3E://WOS:000259939900011%20http://www.clinchem.org/content/54/10/1648.full.pdf.
- [143] P. Mouritzen et al. “Single nucleotide polymorphism genotyping using locked nucleic acid (LNA((TM)))”. In: *Expert Review of Molecular Diagnostics* 3.1 (2003), pp. 27–38. ISSN: 1473-7159. DOI: Doi10.1586/14737159.3.1.27. URL: %3CGo%20to%20ISI%3E://WOS:000188437500004%20http://www.tandfonline.com/doi/abs/10.1586/14737159.3.1.27.
- [144] L. S. Chou et al. “Unlabeled oligonucleotide probes modified with locked nucleic acids for improved mismatch discrimination in genotyping by melting analysis”. In: *Biotechniques* 39.5 (2005), pp. 644–650. ISSN: 0736-6205. DOI: 10.2144/000112050. URL: %3CGo%20to%20ISI%3E://WOS:000233297000005.
- [145] I. Yildirim et al. “Interplay of LNA and 2'-O-Methyl RNA in the Structure and Thermodynamics of RNA Hybrid Systems: A Molecular Dynamics Study Using the Revised AMBER Force Field and Comparison with Experimental Results”. In: *Journal of Physical Chemistry B* 118.49 (2014), pp. 14177–14187. ISSN: 1520-6106. DOI: 10.1021/jp506703g. URL: %3CGo%20to%20ISI%3E://WOS:000349059700022%20http://pubs.acs.org/doi/pdfplus/10.1021/jp506703g.
- [146] D. E Condon et al. “Optimization of an AMBER Force Field for the Artificial Nucleic Acid, LNA, and Benchmarking with NMR of L (CAAU)”. In: *The Journal of Physical Chemistry B* 118.5 (2014), pp. 1216–1228. ISSN: 1520-6106. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3917691/pdf/jp408909t.pdf.
- [147] T. J. Macke and D. A. Case. “Modeling Unusual Nucleic Acid Structures”. In: *Molecular Modeling of Nucleic Acids*. Chap. 24, pp. 379–393. DOI: 10.1021/bk-1998-0682.ch024. eprint: https://pubs.acs.org/doi/pdf/10.1021/bk-1998-0682.ch024. URL: https://pubs.acs.org/doi/abs/10.1021/bk-1998-0682.ch024.
- [148] D. A. Case et al. *Amber 14*. University of California, San Francisco. 2014.
- [149] Schrödinger, LLC. “The PyMOL Molecular Graphics System, Version 1.8”. 2015.

- [150] G. D. Hawkins, C.J. Cramer, and D. G. Truhlar. “Parametrized Models of Aqueous Free Energies of Solvation Based on Pairwise Descreening of Solute Atomic Charges from a Dielectric Medium”. In: *J. Phys. Chem* 100 (1996), pp. 19824–19839.
- [151] W. J. Wang et al. “Holliday Junction Thermodynamics and Structure: Coarse-Grained Simulations and Experiments”. In: *Scientific Reports* 6 (2016). ISSN: 2045-2322. DOI: Artn2286310.1038/Srep22863. URL: %3CGo%20to%20ISI%3E://WOS:000371875900001.
- [152] R. Owczarzy. “Melting temperatures of nucleic acids: discrepancies in analysis”. In: *Biophys Chem* 117.3 (2005), pp. 207–15. ISSN: 0301-4622 (Print) 0301-4622 (Linking). DOI: 10.1016/j.bpc.2005.05.006. URL: <https://www.ncbi.nlm.nih.gov/pubmed/15963627>.
- [153] F. Aboulela et al. “Base-Base Mismatches - Thermodynamics of Double Helix Formation for Dca3xa3g+Dct3yt3g (X, Y = a,C,G,T)”. In: *Nucleic Acids Research* 13.13 (1985), pp. 4811–4824. ISSN: 0305-1048. DOI: DOI10.1093/nar/13.13.4811. URL: %3CGo%20to%20ISI%3E://WOS:A1985AMH3100014%20http://nar.oxfordjournals.org/content/13/13/4811.full.pdf.
- [154] M. D. Poulson and C. T. Wittwer. “Closed-tube genotyping of apolipoprotein E by isolated-probe PCR with multiple unlabeled probes and high-resolution DNA melting analysis”. In: *Biotechniques* 43.1 (2007), pp. 87–91. ISSN: 0736-6205. DOI: 10.2144/000112459. URL: %3CGo%20to%20ISI%3E://WOS:000248207100019.
- [155] I. Hofacker. “RNA Consensus Structure Prediction With RNAalifold”. In: *Comparative Genomics*. Ed. by NicholasH Bergman. Vol. 395. Methods in Molecular Biology. Humana Press, 2008. Chap. 33, pp. 527–543. ISBN: 978-1-58829-693-1. DOI: 10.1007/978-1-59745-514-5_33. URL: http://dx.doi.org/10.1007/978-1-59745-514-5_33.
- [156] I. L. Hofacker et al. “Fast Folding and Comparison of RNA Secondary Structures”. In: *Monatshefte Fur Chemie* 125.2 (1994), pp. 167–188. ISSN: 0026-9247. DOI: 10.1007/bf00818163. URL: %3CGo%20to%20ISI%3E://WOS:A1994MY85400005.

- [157] M. Zuker. “Mfold web server for nucleic acid folding and hybridization prediction”. In: *Nucleic Acids Research* 31.13 (2003), pp. 3406–3415. DOI: 10.1093/nar/gkg595. URL: <http://nar.oxfordjournals.org/content/31/13/3406.abstract>.
- [158] N. R. Markham and M. Zuker. “UNAFold: software for nucleic acid folding and hybridization”. In: *Methods Mol Biol* 453 (2008), pp. 3–31. ISSN: 1064-3745 (Print) 1064-3745 (Linking). DOI: 10.1007/978-1-60327-429-6_1. URL: <https://www.ncbi.nlm.nih.gov/pubmed/18712296>.
- [159] Y. Xin et al. “Annotation of tertiary interactions in RNA structures reveals variations and correlations”. In: *RNA* 14.12 (2008), pp. 2465–77. ISSN: 1469-9001 (Electronic) 1355-8382 (Linking). DOI: 10.1261/rna.1249208. URL: <https://www.ncbi.nlm.nih.gov/pubmed/18957492>.
- [160] C. Laing and T. Schlick. “Computational approaches to 3D modeling of RNA”. In: *Journal of Physics-Condensed Matter* 22.28 (2010), p. 18. ISSN: 0953-8984. DOI: 10.1088/0953-8984/22/28/283101. URL: <http://www.isinet.org/WOS:000279257300002>.
- [161] D. R. Bell et al. “Capturing RNA Folding Free Energy with Coarse-Grained Molecular Dynamics Simulations”. In: *Scientific Reports* (2017).
- [162] N. B. Leontis, J. Stombaugh, and E. Westhof. “The non-Watson-Crick base pairs and their associated isostericity matrices”. In: *Nucleic Acids Res* 30.16 (2002), pp. 3497–531. ISSN: 1362-4962 (Electronic) 0305-1048 (Linking). URL: <https://www.ncbi.nlm.nih.gov/pubmed/12177293>.
- [163] L. Wang, J. Chen, and T. Van Voorhis. “Systematic Parametrization of Polarizable Force Fields from Quantum Chemistry Data”. In: *Journal of Chemical Theory and Computation* 9.1 (2013), pp. 452–460. ISSN: 1549-9618. DOI: 10.1021/ct300826t. URL: <http://www.isinet.org/WOS:000313378700048>.
- [164] C. Zhang et al. “Polarizable Multipole-Based Force Field for Aromatic Molecules and Nucleobases”. In: *J Chem Theory Comput* 13.2 (2017), pp. 666–678. ISSN: 1549-9626 (Electronic) 1549-9618 (Linking). DOI: 10.1021/acs.jctc.6b00918. URL: <https://www.ncbi.nlm.nih.gov/pubmed/28030769>.
- [165] D. Kraft. “A software package for sequential quadratic programming”. In: *Tech. Rep.* DFVLR-FB 88-28 (1988).

- [166] J. Kästner. “Umbrella sampling”. In: *Wiley Interdisciplinary Reviews: Computational Molecular Science* 1.6 (2011), pp. 932–942. ISSN: 17590876. DOI: 10.1002/wcms.66.
- [167] S. Kumar et al. “The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules .1. The Method”. In: *Journal of Computational Chemistry* 13.8 (1992), pp. 1011–1021. URL: <https://doi.org/10.1002/jcc.10011>.
- [168] A. Grossfield. *WHAM: the weighted histogram analysis method*. Computer Program. Version XXXX. URL: <http://membrane.urmc.rochester.edu/content/wham>.
- [169] H. Martadinata and A. Phan. “Structure of Propeller-Type Parallel-Stranded RNA G-quadruplexes, Formed by Human Telomeric RNA Sequences in K⁺ Solution”. In: *Journal of the American Chemical Society* 131.7 (2009). PMID: 19183046, pp. 2570–2578. DOI: 10.1021/ja806592z. eprint: <https://doi.org/10.1021/ja806592z>. URL: <https://doi.org/10.1021/ja806592z>.
- [170] B. Pan et al. “Crystal Structure of an RNA Purine-Rich Tetraplex Containing Adenine Tetrads”. In: *Structure* 11.7 (2003), pp. 815–823. ISSN: 09692126. DOI: 10.1016/s0969-2126(03)00107-2.
- [171] M. Anderson et al. “Active-site monovalent cations revealed in a 1.55-Å-resolution hammerhead ribozyme structure”. In: *J Mol Biol* 425.20 (2013), pp. 3790–8. ISSN: 1089-8638 (Electronic) 0022-2836 (Linking). DOI: 10.1016/j.jmb.2013.05.017. URL: <https://www.ncbi.nlm.nih.gov/pubmed/23711504>.
- [172] L. Zhang and J. A. Doudna. “Structural Insights into Group II Intron Catalysis and Branch-Site Selection”. In: *Science* 295.5562 (2002), pp. 2084–2088. ISSN: 0036-8075. DOI: 10.1126/science.1069268. eprint: <http://science.sciencemag.org/content/295/5562/2084.full.pdf>. URL: <http://science.sciencemag.org/content/295/5562/2084>.
- [173] Darian D. Cash et al. “Pyrimidine motif triple helix in the *Kluyveromyces lactis* telomerase RNA pseudoknot is essential for function in vivo”. In: *Proceedings of the National Academy of Sciences* 110.27 (2013), pp. 10970–10975. ISSN: 0027-8424. DOI: 10.1073/pnas.1309590110. eprint: <http://www.pnas.org>.

- org/content/110/27/10970.full.pdf. URL: <http://www.pnas.org/content/110/27/10970>.
- [174] Y. Liu et al. “Crystal structure and mechanistic investigation of the twister ribozyme”. In: *Nat Chem Biol* 10.9 (2014), pp. 739–44. ISSN: 1552-4469 (Electronic) 1552-4450 (Linking). DOI: 10.1038/nchembio.1587. URL: <https://www.ncbi.nlm.nih.gov/pubmed/25038788>.
- [175] A. Serganov et al. “Structural basis for discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs”. In: *Chem Biol* 11.12 (2004), pp. 1729–41. ISSN: 1074-5521 (Print) 1074-5521 (Linking). DOI: 10.1016/j.chembiol.2004.11.018. URL: <https://www.ncbi.nlm.nih.gov/pubmed/15610857>.
- [176] K. Nuepane et al. “Direct observation of transition paths during the folding of proteins and nucleic acids”. In: *Science* 352.6282 (2016), pp. 239–242.
- [177] J. L. Fiore and D. J. Nesbitt. “An RNA folding motif: GNRA tetraloop-receptor interactions”. In: *Q Rev Biophys* 46.3 (2013), pp. 223–264. ISSN: 1362-4962 (Electronic) 0305-1048 (Linking). DOI: 0.1017/S0033583513000048. URL: <https://www.ncbi.nlm.nih.gov/pubmed/23915736>.
- [178] G. M. Nam and D. E. Makarov. “Extracting intrinsic dynamic parameters of biomolecular folding from single-molecule force spectroscopy experiments”. In: *Protein Sci* 25.1 (2016), pp. 123–34. ISSN: 1469-896X (Electronic) 0961-8368 (Linking). DOI: 10.1002/pro.2727. URL: <https://www.ncbi.nlm.nih.gov/pubmed/26088347>.
- [179] E. Pfitzner et al. “Rigid DNA beams for high-resolution single-molecule mechanics”. In: *Angew Chem Int Ed Engl* 52.30 (2013), pp. 7766–71. ISSN: 1521-3773 (Electronic) 1433-7851 (Linking). DOI: 10.1002/anie.201302727. URL: <https://www.ncbi.nlm.nih.gov/pubmed/23794413>.
- [180] P. C. Nickels et al. “Molecular force spectroscopy with a DNA origami-based nanoscopic force clamp”. In: *Science* 354.6310 (2016), pp. 305–307. ISSN: 1095-9203 (Electronic) 0036-8075 (Linking). DOI: 10.1126/science.aah5974. URL: <https://www.ncbi.nlm.nih.gov/pubmed/27846560>.

- [181] X. Zhuang et al. “A Single-Molecule Study of RNA Catalysis and Folding”. In: *Science* 288.5473 (2000), pp. 2048–2051. ISSN: 00368075 10959203. DOI: 10.1126/science.288.5473.2048.
- [182] S. Jeney et al. “Mechanical properties of single motor molecules studied by three-dimensional thermal force probing in optical tweezers”. In: *Chemphyschem* 5.8 (2004), pp. 1150–8. ISSN: 1439-4235 (Print) 1439-4235 (Linking). DOI: 10.1002/cphc.200301027. URL: <https://www.ncbi.nlm.nih.gov/pubmed/15446737>.
- [183] T. F. Bartsch et al. “Detecting sequential bond formation using three-dimensional thermal fluctuation analysis”. In: *Chemphyschem* 10.9-10 (2009), pp. 1541–7. ISSN: 1439-7641 (Electronic) 1439-4235 (Linking). DOI: 10.1002/cphc.200900211. URL: <https://www.ncbi.nlm.nih.gov/pubmed/19466704>.
- [184] A. Keidel, T. F. Bartsch, and E. L. Florin. “Direct observation of intermediate states in model membrane fusion”. In: *Sci Rep* 6 (2016), p. 23691. ISSN: 2045-2322 (Electronic) 2045-2322 (Linking). DOI: 10.1038/srep23691. URL: <https://www.ncbi.nlm.nih.gov/pubmed/27029285>.
- [185] A. Keidel. “Exploiting high precision single particle diffusion measurements to probe cellular functions in vitro and in vivo”. PhD thesis. University of Texas at Austin, 2014.
- [186] C. C. Ibeneche. “Cell Freezing in Response to Advanced Glucose Starvation: A Novel Cytoplasmic State in Fission Yeast”. PhD thesis. University of Texas at Austin, 2012.
- [187] T. F. Bartsch et al. “Detecting sequential bond formation using three-dimensional thermal fluctuation analysis”. In: *Chemphyschem* 10.9-10 (2009), pp. 1541–7. ISSN: 1439-7641 (Electronic) 1439-4235 (Linking). DOI: 10.1002/cphc.200900211. URL: <https://www.ncbi.nlm.nih.gov/pubmed/19466704>.
- [188] T. F. Bartsch et al. “Nanosopic imaging of thick heterogeneous soft-matter structures in aqueous solution”. In: *Nat Commun* 7 (2016), p. 12729. ISSN: 2041-1723 (Electronic) 2041-1723 (Linking). DOI: 10.1038/ncomms12729. URL: <https://www.ncbi.nlm.nih.gov/pubmed/27596919>.

- [189] R. Huang et al. “Direct observation of the full transition from ballistic to diffusive Brownian motion in a liquid”. In: *Nature Physics* 7.7 (2011), pp. 576–580. ISSN: 1745-2473 1745-2481. DOI: 10.1038/nphys1953.
- [190] Speidel M. and Alexandr J. and Florin E.L. “Three-dimensional tracking of fluorescent nanoparticles with subnanometer precision by use of off-focus imaging”. In: *Optics Letters* 28.2 (2003), pp. 69–71.
- [191] A.R. Carter et al. “Stabilization of an optical microscope to 0.1 nm in three dimensions”. In: *Applied Optics* 46.3 (2007).
- [192] D. T. Edwards et al. “Optimizing 1-us-Resolution Single-Molecule Force Spectroscopy on a Commercial Atomic Force Microscope”. In: *Nano Letters* 15.10 (2015). PMID: 26421945, pp. 7091–7098. DOI: 10.1021/acs.nanolett.5b03166. eprint: <https://doi.org/10.1021/acs.nanolett.5b03166>. URL: <https://doi.org/10.1021/acs.nanolett.5b03166>.
- [193] S. Jeney et al. “Three-dimensional mechanical properties of the molecular motor kinesin studied by the photonic force microscope”. In: *Biophys. J.* 76.1 (1999), A16–A16. ISSN: 1362-4962 (Electronic) 0305-1048 (Linking).
- [194] E. L. Florin et al. “The elastic properties of kinesin studied by three-dimensional particle tracking on the single molecule level”. In: *Mol Bio Cell* 9 (1998), 28A–28A. ISSN: 1362-4962 (Electronic) 0305-1048 (Linking).