STRUCTURAL STUDIES OF MODEL RNA SYSTEMS IN SOLUTION USING FÖRSTER RESONANCE ENERGY TRANSFER

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Julie Louise Sutton

May 2016

© 2016 Julie Louise Sutton

ALL RIGHTS RESERVED

STRUCTURAL STUDIES OF MODEL RNA SYSTEMS IN SOLUTION USING FÖRSTER RESONANCE ENERGY TRANSFER

Julie Louise Sutton, Ph.D.

Cornell University 2016

New biological roles for RNA are continually being discovered, reinforcing the importance of understanding the fundamental forces governing RNA structure. Open questions remain about the structure of RNA because of the complex interplay between base pairing, base stacking, and counterion-mediated interactions. Förster resonance energy transfer (FRET) provides a powerful tool for measuring RNA conformational changes in solution. In this work we present three structural studies of model RNA systems using FRET. In the first two studies, we use single-molecule FRET to measure the subtle conformational changes on model single- and double-stranded RNA junctions induced by salt concentration. For single-stranded junctions at physiological salt concentrations both helix-helix repulsion and junction sequence are important factors determining RNA conformation. Furthermore, the sequence dependence of double stranded junctions is particularly important in solutions containing Mg^{2+} ions. In the third study, we measure sub-millisecond conformational changes of a small independently folding RNA induced by a jump in Mg^{2+} concentration in a microfluidic mixer. We show that mutations within a specific ion binding site can alter the folding rates of a small RNA sub-domain. We propose that understanding how subtle changes in ion concentration and RNA sequence can bias overall structure is necessary for a full picture of RNA folding.

BIOGRAPHICAL SKETCH

Julie was born and raised in Montreal, Quebec, Canada. She earned her Bachelors of Applied Science in Engineering Science from the University of Toronto in May 2009. As an undergrad, she completed summer research projects in particle, thin film and cold atoms physics. After joining the Applied Physics Ph.D. program at Cornell University in August 2009, she was persuaded to give biophysics a try. She has been a Graduate Research Assistant in Lois Pollack's lab since the spring of 2010.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor Lois for inviting me to join her group and introducing me to the field of biophysics. I am grateful for her advice and unwavering support. To my committee members Warren Zipfel and Peng Chen, thank you for your feedback and guidance along the way. I would also like to thank my current colleagues in the Pollack Lab: Suzette Pabit, Josh Tokuda, Yujie Chen, Andrea Katz, George Calvey, Alex Plumridge, Yunsheng Wang, Alex Mauney, Abhijit Lavania, Yen-Lin Chen and Josue San Emeterio, as well as former members Steve Meisburger, Huimin Chen, Li Li, Chris Jones, and Josh Blose. I have learned something new from each of you and am lucky to have had the chance to connect with such wonderful people.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Far from being a mere 'messenger', RNA can regulate gene expression [1], catalyze splicing and protein synthesis reactions [2], and may even be responsible for epigenetic inheritance [3]. New biological roles for RNA are continually discovered [4]. As with most biological molecules, understanding the structure of RNA can provide valuable insights into the underlying mechanisms that dictate its function.

RNA structure is typically described by three levels of complexity. At the primary level, RNA is a linear chain with four nucleotide building blocks: C, G, A and U. Watson-Crick pairing between C-G and A-U bases creates a double helix structure. The RNA helix is A-form — shorter and wider than the DNA B-form helix. In comparison to DNA, a single strand of RNA is more likely to base-pair with itself, generating an intricate network of hairpins, loops and bulges. Base pairing determines RNA secondary structure. Finally, the arrangement of the base-paired elements in 3-dimensions is described as its tertiary structure. The tertiary structure can be influenced by several factors, including the sequence and length of non-base-paired junctions between helical regions, stabilizing base-stacking interactions, specific Mg^{2+} ion binding, and base-pairing between distal secondary structure elements [5].

While the above hierarchical description of RNA structure appears complete, it is still not possible to predict the 3-dimensional structure of an RNA molecule given its primary sequence. This so-called 'RNA folding problem' is distinct from its protein counterpart because RNA is a polyelectrolyte — it contains one negative charge per nucleotide along its phosphate backbone. To form

compact structures, RNA must overcome electrostatic repulsion in part by attracting positively charged counterions. These ions can coordinate to specific regions of the RNA backbone, or form a diffuse ion 'cloud' shielding charges and reducing overall repulsion [6]. While progress has been made [7–9], there is still a need for further understanding on the effects of ions on RNA structure.

In this thesis, we tackle the RNA folding problem by breaking down RNA into isolated structural elements. By measuring how different ionic conditions affect each element, we aim to gain insight into the intricate relationship between ions and RNA structure. We use Förster Resonance Energy Transfer (FRET) as a probe for structural changes, which we monitor either for single molecules or in bulk using a microfluidic mixer. The central theme of this thesis is that each structural element contributes uniquely to overall RNA conformation, and can be used to 'tune' RNA interactions for specific biological roles.

An outline of the thesis is as follows. In Chapter 2 we introduce FRET, the principal technique used in the thesis. We expand on its theoretical concepts and the instrumentation needed to monitor individual macromolecules. Chapters 3-5 represent three separate projects that fit into the central theme of the thesis. While they are not represented chronologically, the projects are ordered in increasing RNA structural complexity. The chapters show a progression from simple models of single- and double-stranded RNA junctions to a biologically derived RNA subdomain.

In Chapter 3 we describe the use of a model RNA system consisting of two helices connected by a single strand to examine the contributions of junction sequence and helix length to RNA structural flexibility. Unexpectedly, we find a non-monotonic dependence of FRET with salt concentration. We com-

pare changes observed when the junction is mutated from poly(U) to poly(A) or when helix lengths are changed from 12 to 24 base pairs. Helix repulsion dominates at low salt, whereas junction sequence dominates at high salt. Interestingly, these regimes overlap near physiological monovalent concentrations, suggesting a biologically relevant means for tuning RNA conformation.

In Chapter 4 we build upon the framework developed in Chapter 3 and introduce another degree of sequence and structural complexity. We modify the topology of the RNA constructs to contain double-, rather than single-stranded junctions and compare two alternate junction sequences. We find that the two sequences have a similar bend angle but different twist in the presence of KCl, while the difference between sequences is further exaggerated in the presence of $MgCl_2$. The manner in which sequence and salt concentration can bias RNA junction conformation is important for determining how an RNA molecule can reach its folded state.

In Chapter 5 we examine the contribution of specific Mg^{2+} ion binding to RNA compaction using the tP5abc subdomain of the *Tetrahymena* ribozyme as a model system. After a rapid jump in salt concentration, we observe that the FRET changes in ~ 0.85ms — faster than the formation of tertiary contacts. The timescale for conformational changes in the wild-type is faster than that of a mutant which lacks a specific Mg^{2+} binding site. This result suggests that site-binding may occur quickly, and may aid in tuning the energy landscape towards properly folded states.

Finally, in Chapter 6 we summarize our results and discuss future work to use smFRET to gain a deeper structural understanding of tP5abc and to improve atomic models of flexible RNA.

References

[1] A. Serganov and E. Nudler, "A decade of riboswitches," *Cell*, vol. 152, no. 1, pp. 17–24, 2013.

[2] S. A. Strobel and J. C. Cochrane, "Rna catalysis: ribozymes, ribosomes and riboswitches," *Current opinion in chemical biology*, vol. 11, no. 6, pp. 636–643, 2007.

[3] A. B. Rodgers, C. P. Morgan, N. A. Leu, and T. L. Bale, "Transgenerational epigenetic programming via sperm microrna recapitulates effects of paternal stress," *Proceedings of the National Academy of Sciences*, vol. 112, no. 44, pp. 13 699–13 704, 2015.

[4] T. R. Cech and J. A. Steitz, "The noncoding rna revolution - trashing old rules to forge new ones.," *Cell*, vol. 157, no. 1, pp. 77–94, 2014.

[5] S. E. Butcher and A. M. Pyle, "The molecular interactions that stabilize rna tertiary structure: rna motifs, patterns, and networks," *Accounts of Chemical Research*, vol. 44, no. 12, pp. 1302–1311, 2011.

[6] P. B. Moore, "The rna folding problem," in *The RNA world*, Cold Spring Harbor Laboratory Press, 1999, pp. 381–401.

[7] D. E. Draper, "A guide to ions and rna structure," *RNA*, vol. 10, no. 3, pp. 335–343, 2004.

[8] S. A. Woodson, "Metal ions and rna folding: a highly charged topic with a dynamic future," *Current opinion in chemical biology*, vol. 9, no. 2, pp. 104–109, 2005.

[9] J. C. Bowman, T. K. Lenz, N. V. Hud, and L. D. Williams, "Cations in charge: magnesium ions in rna folding and catalysis," *Current opinion in structural biology*, vol. 22, no. 3, pp. 262–272, 2012.

CHAPTER 2

METHODS

FRET has emerged in the past decade as an invaluable tool for studying biomolecular structure. A non-radiative dipole-dipole coupling between two fluorophores can be used to report on their separation in space. In the context of our work, labeling RNA molecules with fluorophores gives us a structural probe to monitor changes in RNA conformation. While Theodore Förster laid out the theoretical framework 70 years ago [1], several recent innovations to implement single-molecule detection alongside FRET [2–4], have enabled researchers to observe conformational heterogeneity rather than ensemble averages. In conjunction with technical advances in instrumentation, the introduction of new stable fluorophores and the availability of custom oligonucleotide synthesis techniques have made smFRET accessible to a wide community of scientists. Timely to our lab's interests in the physical properties of nucleic acids, we built a low-cost smFRET microscope and combined it with custom RNA constructs in order to study the effects of ions on RNA structure in solution. While Chapters 3-5 in this thesis each contain their own details on the methods used, we provide here a brief overview of the key concepts behind FRET as well as an overview of its instrumentation and other sample-related considerations.

2.1 A brief theoretical overview of FRET

Fluorescence is the emission of a photon after the absorption of a photon of shorter wavelength. FRET occurs when there is a resonance between the emission energies of a donor (D) fluorophore and the absorption energy of an acceptor (A) fluorophore (Figure 2.1(a)).

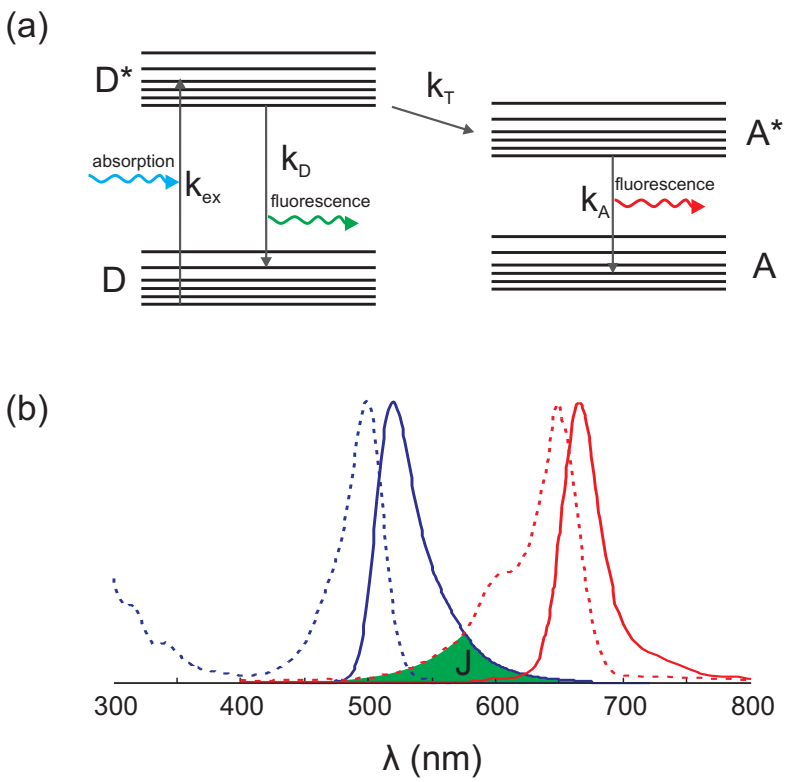

Figure 2.1: (a) Jablonski diagram representing FRET. k_D, k_A and k_T represent the rate constants for donor fluorescence, acceptor fluorescence, and energy transfer, respectively. (b) Absorption (dashed line) and emission (solid line) spectra of Alexa Fluor 488 (blue) and Cy5 (red). The area in green represents the overlap integral, J from equation 2.3

The efficiency of energy transfer, E_T, is defined in equation 2.1, and represents the probability that the donor will relax to its ground state through energy transfer to the acceptor. The rate of transfer, k_T, depends on the distance between the donor and acceptor, R, and the fluorescence lifetime of the donor, $\tau_D = \frac{1}{k_D+k_{nr}}$, as shown in equation 2.2. Here, k_{nr} is the sum of all rate constants for non-radiative processes that compete with k_D and k_T such as internal conversion, intersystem crossing or other types of quenching [5] (not shown in Figure 2.1). While a full derivation will not be given here, the basis for the R^6 dependence comes from the dipole-dipole coupling of the fluorophores. A com-

plete derivation from both classical and quantum-mechanical perspectives can be found in reference [6].

$$E_T = \frac{k_T}{k_T + k_D + k_{nr}} \tag{2.1}$$

$$k_T = \frac{1}{\tau_D}\left(\frac{R_0}{R}\right)^6 \tag{2.2}$$

The Förster radius, R_0, is defined as distance at which $E_T = 1/2$. R_0 depends on the overlap between the donor emission and acceptor absorption spectra (J, Figure 2.1(b)), the donor quantum yield, ϕ_D, and the solution index of refraction, n. R_0 can be calculated using equation 2.3.

$$R_0^6 = \frac{9000(\ln 10)\kappa^2\phi_D J}{128\pi^5 N n^4}, J = \int_0^\infty f_D(\lambda)\varepsilon_A(\lambda)\lambda^4 d\lambda \tag{2.3}$$

Here, N is Avogadro's number, $\varepsilon_A(\lambda)$ is the absorption spectrum of the acceptor and $f_D(\lambda)$ is the normalized emission spectrum of the donor. The factor κ^2 takes into account the relative orientations of the fluorophores. For freely rotating fluorophores, $\kappa^2 = 2/3$, however, as discussed in Chapter 5, interactions between the fluorophore and biomolecule can invalidate this assumption.

Combining equations 2.1 and 2.2 we can relate R and E_T through equation 2.4. An important consequence of equation 2.4 is the non-linear dependence of E_T on R/R_0. The dependence effectively reduces the useable range of FRET to systems where the R is roughly within $0.5-1.5\times R_0$ (Figure 2.2(a)). For commonly used fluorophores, R_0 is typically in the 20-60Å range [7], which is, thankfully, an

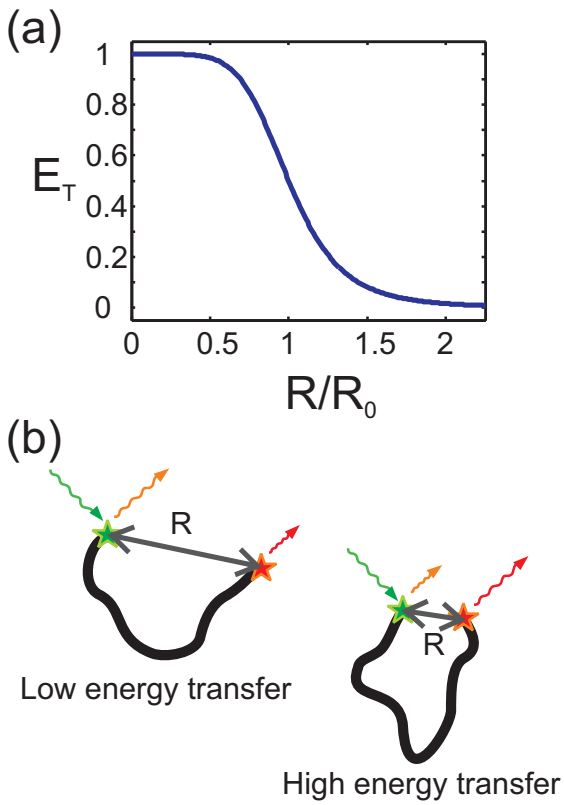

Figure 2.2: (a) E_T is most sensitive when R is between $0.5 - 1.5 \times R_0$. (b) Schematic representing FRET. Fluorescence intensities of donor and acceptor dyes depend on their proximity. The closer their proximity, the more energy is transferred to the acceptor.

ideal length scale for studying biomolecules like RNA. That being said, careful choice of dye positions must be made to capture molecular motions of interest.

$$E_T = \frac{1}{1 + \left(\frac{R}{R_0}\right)^6} \tag{2.4}$$

We can measure FRET in the laboratory by detecting the fluorescence emitted from donor and acceptor fluorophores after the donor is excited. Equations 2.5 and 2.6 describe the dependence of the fluorescence intensities of the donor (F_D) and acceptor (F_A) on E_T. As depicted Figure 2.2(b), F_A will increase and F_D decrease when the fluorophores get closer together.

$$F_D = N_D^{ex} \eta_D \phi_D (1 - E_T) \tag{2.5}$$

$$F_A = N_D^{ex} \eta_A \phi_A E_T \tag{2.6}$$

In equations 2.5 and 2.6 N_D^{ex} is the number of excited donor molecules and $\eta_{D(A)}$ is the detection efficiency of the donor (acceptor). We can then rearrange these equations to get E_T as a function of F_D and F_A:

$$E_T = \frac{F_A}{F_A + \gamma F_D} \tag{2.7}$$

where $\gamma = \frac{\phi_A \eta_A}{\phi_D \eta_D}$. The ratio of detection efficiencies, η_A/η_D, is specific to the microscope used to measure F_A and F_D. The quantum yields, $\phi_{D(A)}$, and η_A/η_D can be measured in the laboratory. In principle, once γ is determined, we can measure F_A and F_D and use equations 2.4 and 2.7 to determine the inter-fluorophore distance R.

While it is simple to measure E_T for bulk samples using a standard fluorimeter or fluorescence microscope, several factors will complicate the seemingly simple conversion from E_T to R. For example, molecules present with incomplete labeling will skew the results to lower E_T. Additionally, any inhomogeneity in the sample conformations will be averaged out. For these reasons, FRET measurements of bulk solutions are better suited to monitoring *changes* in structure, rather than accurate distances. For example, in Chapter 5 of this thesis, we use FRET to measure the timescale of structural changes in RNA after a rapid jump in salt concentration. The issues with bulk FRET can be circumvented by

measuring the E_T of single molecules, a technique we use in Chapters 3 and 4. This type of measurement requires slightly more complex instrumentation than the bulk measurement, which we describe in more detail in the following section.

2.2 Single-molecule FRET instrumentation

There are two commonly used strategies for measuring smFRET. The first, is to tether the FRET-labeled molecules of interest to a glass coverslide, and use total internal reflection microscopy for excitation. Fluorescence is collected and imaged on a CCD, and each molecule can be distinguished spatially. This approach was first used to study RNA conformational changes by Kim et al. in 2002 [4]. This strategy has the benefit of tracking the fluorescence of individual molecule over long time scales (~10s of seconds). A major downside of this approach, however, is that the proximity of the coverglass has the potential to perturb the subtle electrostatic interactions between ions and RNA that are of interest to us. The second strategy, first developed by Deniz et al. [3], is the one we focus on in this thesis. The strategy uses confocal microscopy to measure the fluorescence of dilute solutions of freely-diffusing FRET-labeled molecules. In solutions dilute enough to have fewer than one molecule in the focal volume on average, we can measure bursts of fluorescence from individual molecules, and build up a population histogram. This approach limits the amount of time that a single molecule is observed to its diffusion time (typically miliseconds), but it eliminates any potential interactions between the coverglass and the molecules of interest.

A schematic of the confocal microscope built in our lab is shown in Figure 2.3, with a picture shown in Figure 2.4. A detailed protocol for smFRET measurements can be found in Appendix A. The setup was designed for the Alexa 488 - Cy5 dye pair, however other dyes can be used by changing the laser and filters. Briefly, the sample is epi-illuminated by a 488nm Coherent Sapphire laser through an Olympus 60x 1.2NA objective. The laser beam is expanded to overfill the objective's back-aperture. The fluorescence emission is focused by a tube lens and separated into donor and acceptor components by a dichroic mirror. The fluorescence of each channel is cleaned up using the appropriate emission filters, then coupled into optical fibers. The fibers serve as confocal pinholes to filter out background fluorescence from outside the tight focal volume of the objective. Finally, the photons are collected using single-photon counting avalanche photodiodes (APDs). The signal from the APDs is converted into photon arrival times using a correlator card. The arrival times are then converted to fluorescence intensity vs. time using MATLAB (Figure 2.3(b)). For dilute solutions these traces consist of a series of bursts corresponding to individual molecules diffusing through the focal volume. Bursts above a certain total-intensity threshold are collected into a histogram (Figure 2.3(c)), where populations with distinct E_T values can be distinguished.

2.3 Sample considerations for FRET measurements

A major factor in designing a successful FRET experiment is the appropriate choice of fluorophore identity and labeling location. Fluorophore choice will determine R_0, and set the length scale for E_T sensitivity. While the typical R_0 for frequently used dye pairs can be found online, it may not be accurate because of

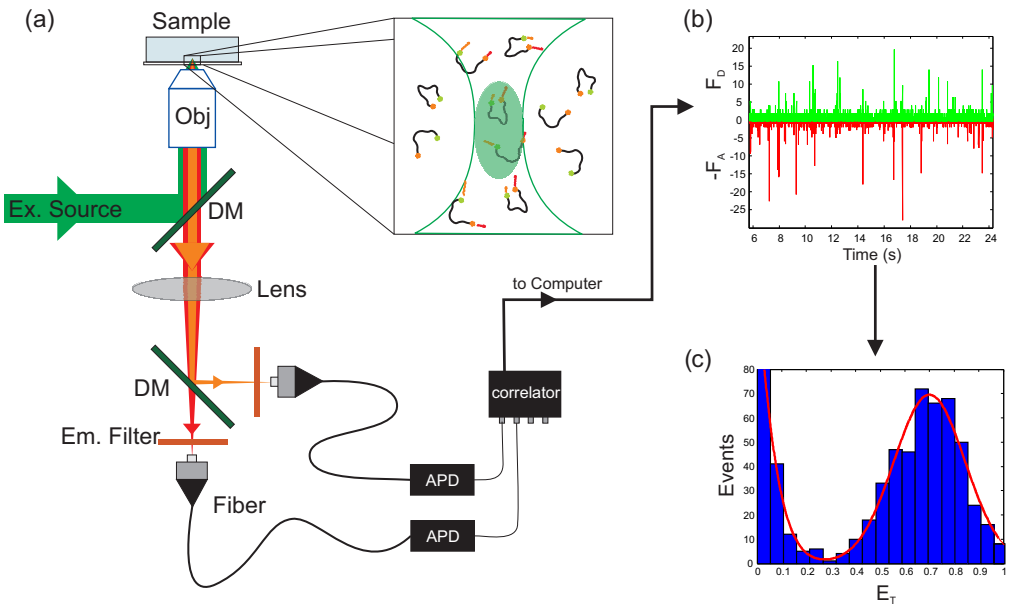

Figure 2.3: (a) Schematic of the confocal microscope used to measure smFRET on freely diffusing molecules. Inset: depiction of molecules diffusing through the focal volume. Fluorescence from molecules diffusing outside the focal volume is filtered out by the confocal pinhole. (b) Example of F_D and F_A vs. time for dilute solution of FRET-labeled molecules. (c) Example E_T histogram of single-molecule fluorescence events.

specific fluorophore-biomolecule interactions. For example, Cy3 and Cy5 tend to stack on the end of RNA and DNA helices, invalidating the $\kappa^2 = 2/3$ assumption [8], whereas attaching Alexa 488 and Cy5 internally can be assumed to be freely rotating [9].

When choosing the locations of the fluorophore attachment points to the RNA of interest, one must consider whether the expected E_T will be within its sensitive range. Occasionally, a crystal structure can be used as a reference. Since fluorophores are typically attached to the molecule via a flexible linker, a good estimate of E_T must take this flexibility into account. Software developed by Muschielok et al. [10] models the sterically accessible volume (AV) of a fluorophore given a pdb structure (Figure 2.5(a)). One can then calculate the ex-

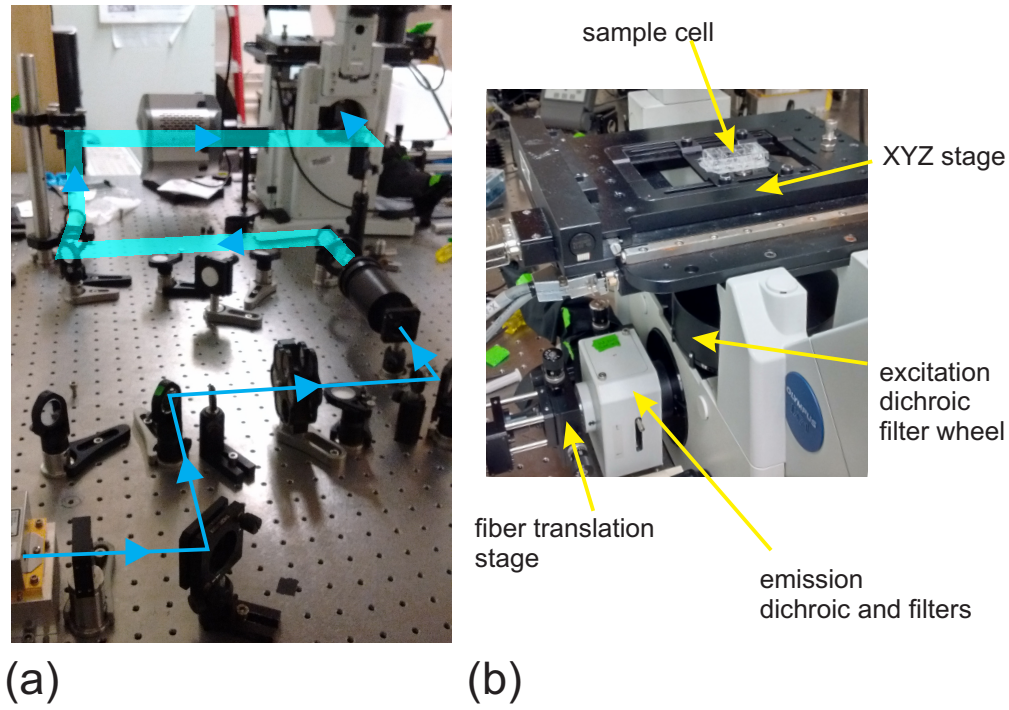

Figure 2.4: (a) Image of optics used to steer laser beam into Olympus microscope body. A 488nm Coherent Sapphire laser is expanded by 10x and directed into the microscope. A filter wheel containing a variety of neutral density filters allows the laser power to be reduced as needed. (b) Image of sample stage and detection path. Samples are loaded into 8-well #1 coverglass chambers. The laser enters from the back of the microscope into the excitation dichroic filter wheel and is reflected up into the microscope objective (not visible). Fluorescence passes through the excitation dichroic and is reflected horizontally into the output port of the microscope. The emission dichroic separates donor and acceptor signals which are collected through their respective 50μm optical fibers, which can be positioned using translation stages.

pected E_T based on a pdb structure of the molecule, and parameters describing the fluorophore geometry. A detailed example of the use AV simulations with helical RNA is discussed in reference [9]. Users must still be cautious, however, since this approach does not address any possibility for interactions between fluorophore and biomolecule, including electrostatics and stacking. Regardless, the AV approach is useful for providing an initial estimate on expected E_T values.

Another important limitation when choosing labeling positions is the availability of chemical modifications to incorporate a fluorophore at the desired site. Fortunately, there are a variety of available options to covalently link fluorophores with nucleic acids. Companies like Integrated DNA Technologies (Coralville, IA) and Dharmacon (Lafayette, CO) offer fluorophore modifications that can be incorporated into their custom DNA or RNA synthesis. Modifications for the 3' and 5' ends are more widely available, but internal modifications can also be found. Sometimes it can be simpler to label two separate strands of RNA and then anneal them together to form a double-labeled complex. A detailed description of RNA labeling strategies can be found in reference [11], and a protocol for labeling amino-modified RNA can be found in Appendix B.

A final important factor to consider when designing RNA sequences for any *in vitro* experiments is the possibility for alternative secondary structures. Several online tools can be used to evaluate the probability of alternative secondary structures. For example, the MC-fold web server [12] can be used to predict the secondary structure of a single strand of RNA, including non-canonical base pairs. Alternatively, the DINAMelt web server [13] provides the thermodynamic parameters and melting temperatures of associated with a given RNA

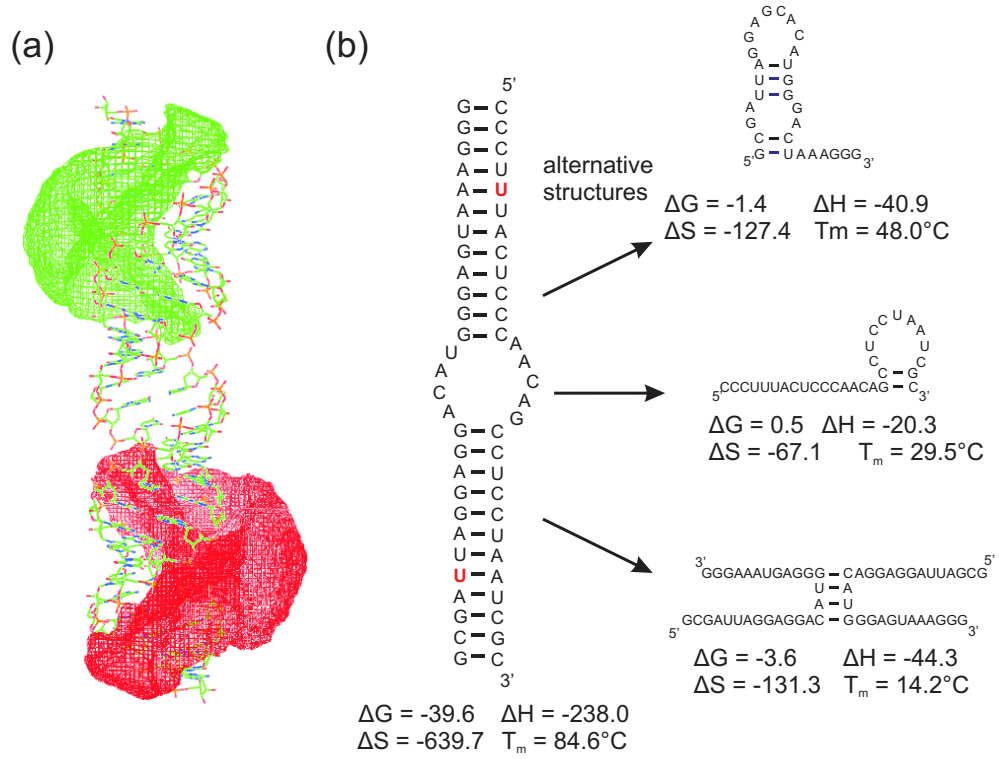

Figure 2.5: (a) Representative AV simulation of fluorophores attached to an RNA duplex. Red and green mesh surfaces denote the space that a fluorophore can explore without steric clashes with the RNA (b) Schematic of RNA sequence design showing the sequence of a double-stranded RNA junction. The desired base-paired structure is shown on the left, with donor and acceptor locations highlighted in red. Alternative structures found using the DINAMelt web server and their associated thermodynamic parameters are shown on the right. Notice that both ΔG and T_m are favorable for the desired structure.

or DNA sequence (see Figure 2.5(b)). The web server can be used to predict the most likely secondary structures formed by a single RNA strand and between two separate strands. A lower ΔG and higher T_m both denote a more favorable structure. The major limitation of this approach is that RNA calculations can only be done under 1M NaCl solution conditions.

References

[1] T. Forster, "Zwischenmolekulare energiewanderung und fluoreszenz," *Annalen der Physik*, vol. 2, no. 1-2, pp. 55–75, 1948.

[2] T. Ha, T. Enderle, D. F. Ogletree, D. S. Chemla, P. R. Selvin, and S. Weiss, "Probing the interaction between two single molecules: fluorescence resonance energy transfer between a single donor and a single acceptor.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 93, no. 13, pp. 6264–6268, 1996.

[3] A. A. Deniz, M. Dahan, J. R. Grunwell, T. Ha, A. E. Faulhaber, D. S. Chemla, S. Weiss, and P. G. Schultz, "Single-pair fluorescence resonance energy transfer on freely diffusing molecules: observation of förster distance dependence and subpopulations.," *Proceedings of the National Academy of Sciences*, vol. 96, no. 7, pp. 3670–3675, 1999.

[4] H. D. Kim, G. U. Nienhaus, T. Ha, J. W. Orr, J. R. Williamson, and S. Chu, "Mg2+-dependent conformational change of rna studied by fluorescence correlation and fret on immobilized single molecules," *Proceedings of the National Academy of Sciences*, vol. 99, no. 7, pp. 4284–4289, 2002.

[5] C. R. Cantor and P. R. Schimmel, *Biophysical Chemistry Part II: Techniques for the study of biological structure and function*. New York: W. H. Freeman and Company, 1980.

[6] B. R. Masters and P. T. C. So, "Forster resonance energy transfer (fret)," in *Biomedical Nonlinear Optical Microscopy*, B. R. Masters and P. T. C. So, Eds., 2008, pp. 557–598.

[7] J. R. Lakowicz, *Principles of fluorescence spectroscopy*, 3rd. New York: Springer, 2006, p. 954.

[8] A. Iqbal, S. Arslan, B. Okumus, T. J. Wilson, G. Giraud, D. G. Norman, T. Ha, and D. M. J. Lilley, "Orientation dependence in fluorescent energy transfer between cy3 and cy5 terminally attached to double-stranded nucleic acids," *Proceedings of the National Academy of Sciences*, vol. 105, no. 32, pp. 11 176–11 181, 2008.

[9] S. Sindbert, S. Kalinin, H. Nguyen, A. Kienzler, L. Clima, W. Bannwarth, B. Appel, S. Muller, and C. A. Seidel, "Accurate distance determination of nucleic acids via forster resonance energy transfer: implications of dye linker length and rigidity," *Journal of the American Chemical Society*, vol. 133, no. 8, pp. 2463–2480, 2011.

[10] A. Muschielok, J. Andrecka, A. Jawhari, F. Brückner, P. Cramer, and J. Michaelis, "A nano-positioning system for macromolecular structural analysis," *Nature methods*, vol. 5, no. 11, pp. 965–971, 2008.

[11] S. Solomatin and D. Herschlag, *Methods of site-specific labeling of RNA with fluorescent dyes*, 1st ed. Elsevier Inc., 2009, vol. 469, pp. 47–68.

[12] M. Parisien and F. Major, "The mc-fold and mc-sym pipeline infers rna structure from sequence data," *Nature*, vol. 452, no. 7183, pp. 51–55, 2008.

[13] N. R. Markham and M. Zuker, "Dinamelt web server for nucleic acid melting prediction," *Nucleic acids research*, vol. 33, no. Web Server issue, W577–581, 2005.

CHAPTER 3

TUNING RNA FLEXIBILITY WITH HELIX LENGTH AND JUNCTION SEQUENCE

Julie L. Sutton[1], and Lois Pollack[1]

1. School of Applied and Engineering Physics, Cornell University, Ithaca, New York

Reprinted from Biophysical Journal, Vol 109, Issue 12, Pages 2644-2653, Copyright (2015), with permission from Biophysical Society.

Abstract

The increasing awareness of RNA's central role in biology calls for a new understanding of how RNAs, like proteins, recognize biological partners. Because RNA is inherently flexible, it assumes a variety of conformations. This conformational flexibility can be a critical aspect of how RNA attracts and binds molecular partners. Structurally, RNA consists of rigid base paired duplexes, separated by flexible non-base-paired regions. Here, using an RNA system consisting of two short helices, connected by a single-stranded (non-base-paired) junction, we explore the role of helix length and junction sequence in determining the range of conformations available to a model RNA. Single-molecule FRET reports on the RNA conformation as a function of either mono- or divalent ion concentration. Electrostatic repulsion between helices dominates at low salt concentration, while junction sequence effects determine the conformations

at high salt concentration. Near physiological salt concentrations, RNA conformation is sensitive to both helix length and junction sequence, suggesting a means for sensitively tuning RNA conformations.

3.1 Introduction

Recognition of RNAs expanding biological roles demands an understanding of its interactions with partner molecules. To date, most knowledge of how macromolecules recognize partners is derived from studies of proteins. Two alternative mechanisms have emerged [1]. In the induced fit model, the ligand alters the conformation of its macromolecular binding partner. At the other extreme, the conformational capture model requires that the unbound macromolecule fluctuate through various conformations. The ligand binds only when the macromolecule fluctuates through the 'correct' structure. For protein systems the conformational capture mechanism effectively describes several important classes of ligand binding [1, 2]. Similar questions are now being asked about mechanisms for RNA-ligand binding. In particular, a number of recent papers suggest that the conformational capture mechanism is exploited in the initial recognition of ligands by riboswitches [3–7]. Although a combination of both induced fit and conformational capture is often required to describe full molecular recognition pathways, understanding the conformational capture mechanism in particular is crucial for gaining insight into the early stages of RNA-ligand binding.

Since double-stranded RNA is relatively rigid, the flexibility required for conformational fluctuations arises from non-base-paired regions in the struc-

ture [8]. For example, unpaired junctions lacking stable structures can serve as flexible hinges that bring together two sides of a tertiary contact [9, 10]. In other cases, the dynamics of single stranded chains can drive conformational switching, such as in RNA splicing or riboswitch gene regulation [11]. To further explore the factors that support conformational fluctuations in RNA, it is critical to understand the link between sequence, structure and the equilibrium conformational fluctuations of RNA. However, few traditional experiments can probe these flexible states, and MD simulations, which can assist in modelling dynamic regions, are limited in accuracy [12].

A major goal of this work is to begin to understand the connection between RNA structural motifs such as duplexes, the nucleotide sequence of non-base-paired regions, and the range of conformations through which an RNA can fluctuate. To explore these effects we employ a model RNA construct consisting of two RNA helices connected by a single-stranded RNA linker that serves as a junction, representing a model helix-junction-helix (HJH) construct [13].

Before discussing the behavior of the HJH constructs, it is first useful to briefly review the response of each independent motif, helix or strand, to changing ionic conditions. The negative charge associated with the sugar-phosphate backbone is an important factor for determining any RNA structure. In salt-containing solutions, positively charged counterions are attracted to the RNA and reduce the overall electrostatic potential. Studies of isolated, short duplexes clearly illustrate ion valence and concentration dependent changes in RNA's ion atmosphere [14–16], which, in turn mediate the repulsion between RNA elements [14, 17]. These latter interactions depend on helix length because end effects alter the potential around duplexes [18, 19], on length scales on or-

der of 5bp [20]. Although most past studies focused on longer duplexes, e.g. 25 bp, end effects may also be important because the duplexes found in functional, biologically relevant RNAs tend to be short (only ~4bp [21–23]). Single stranded RNA (or DNA) is also affected by ions through charge screening. Intra-strand repulsion is reduced as cation concentrations are increased, effectively rendering the molecule more flexible [24–26]. Base stacking also strongly influences the behavior of single stranded regions. For example, stacking interactions present in poly(A), but not poly(U), make it more rigid at a given salt concentration [26–28]. Often, electrostatic effects and stacking cannot be separated: ions can affect single strand flexibility by decreasing electrostatic repulsion or by stabilizing base stacking interactions. Finally the specific association of ions may alter single strand conformation, for example Mg^{2+} ions can bind directly to poly(A) [29, 30], stabilizing its helical conformation. The growing interest in ion dependent conformations of single stranded RNA is underscored by numerous studies measuring the effects of ions on homopolymers [26, 28–30].

Here, we examine the role of both duplex length and junction sequence in determining the overall conformations of the HJH construct. Single-molecule Förster resonance energy transfer (smFRET) studies of double-labeled, freely diffusing RNAs sensitively report ion-dependent conformational changes. By examining the changing conformations of these simple HJH motifs as a function of increasing salt, and comparing these changes with known responses of the elements in isolation, we gain insight into the rich behaviors that arise from their unique combinations. At low salt, we find that RNA conformation is sensitive to neither helix length nor junction sequence, while at high salt it depends only on junction sequence. In solutions containing near-physiological salt con-

centrations, however, we find an unexpected decrease in the efficiency of energy transfer for our RNA constructs, which depends on both junction sequence and helix length. We propose a model consistent with the observed trends. Our results suggest that trade-offs between helix length and junction sequence provide a way to tune the unfolded state conformations of RNA, critical for any process that relies on conformational fluctuations.

3.2 Materials and Methods

3.2.1 Sequence design

RNA constructs consist of three strands annealed together to form two duplexes connected by a 5-nucleotide single-stranded linker (Figure 3.1). The full construct forms a helix-junction-helix type structure (HJH). Varying helix length to be either 12bp or 24bp and junction sequence to be either poly(U)$_5$ or poly(A)$_5$ provides four distinct RNA constructs. The DINAMelt web server [31] was used to ensure that unwanted hairpin and dimer formation was less favorable than the desired duplex assembly. To provide structural information, we measure FRET between two fluorophores incorporated into the RNA. Sites for fluorescent labels were selected to be 8 base pairs away from the junction region to keep E_{FRET} values in the linear regime for the particular dye pair in use. Although slight differences exist between the simulated accessible volume of the dyes for 12 and 24 bp helix lengths, differences in the measurements do not appear to be significant (discussed in more detail in Appendix C Supplementary Information, and depicted in Figure C.1), therefore choosing sites to be a fixed

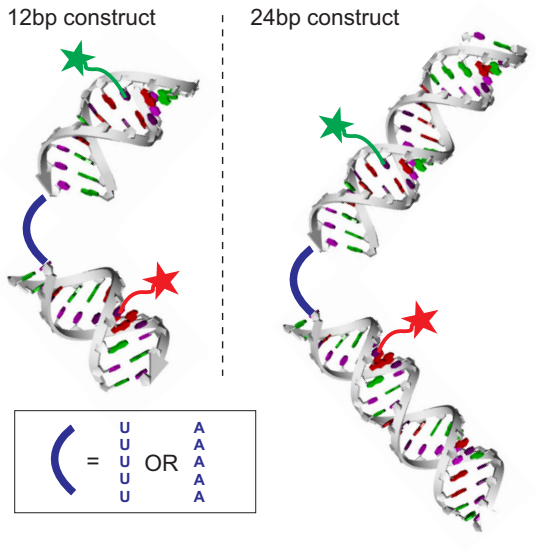

Figure 3.1: Model RNA constructs consist of three strands annealed together to form two helices connected by a single-stranded junction. Individual helices are 12bp or 24bp in length and the junction consists of 5nt of either poly(U) or poly(A). Fluorescent label sites are illustrated by green (donor) and red (acceptor) stars.

distance from the junction region distinguishes effects caused by the label's environment from effects caused by extending the helices. We chose the Alexa Fluor - Cy5 FRET pair since previous studies using this pair attached internally to RNA duplexes validated the assumption that the dyes are freely rotating [32]. Finally, label sites were chosen so that incomplete annealing or the presence of unlabelled strands results in either zero FRET or no observable signal. Refer to the Supplementary Information for details about duplex sequence and label sites.

3.2.2 Sample preparation

RNA molecules were purchased desalted and HPLC purified from IDT (Coralville, IA). An internal amino-C6-dT nucleobase was included in the RNA sequence at the desired labelling site. Modified RNA strands were reconstituted in 100mM phosphate buffer (pH 8.3). Fluorescent dyes were dissolved in DMSO and promptly mixed with the appropriate RNA strand in a 10:1 dye to RNA molar ratio (Alexa Fluor 488 TFP (Life Technologies, Grand Island, NY) was used as donor and Cy5 NHS ester (GE Healthcare) as the acceptor). This mixture was left overnight on a rocker at 4°C. The third unmodified strand was reconstituted in 50mM KMOPS buffer containing 900mM KCl and 0.1mM EDTA. RNA strands were mixed together in a 1:1.5:2 ratio of donor to unlabelled to acceptor strand, annealed at 95°C for 2 minutes then cooled slowly in a water bath for 40 minutes. Samples were run on a Micro Bio-Spin 6 chromatography column (BioRAD, Hercules, CA) to separate RNA from unreacted dye. RNA samples were concentrated to 5-10μM, then divided into aliquots and stored at -20°C.

On the day of each experiment, one aliquot was thawed and diluted by 150x in buffer containing 100mM KCl, 50mM KMOPS, pH 7 and 20μM EDTA. For constructs with a poly(A) junction, the sample was annealed at 90°C for two minutes, and then slowly cooled to room temperature over 50 minutes. For poly(U) junctions, the annealing had no effect on measurement outcomes. This daily stock sample was diluted 1000 fold into the desired buffer for each smFRET measurement. Measurement buffers consisted of 50mM KMOPS, pH 7, 20μM EDTA and the desired amount of added KCl, or $MgCl_2$. Values of ion concentrations quoted in this paper do not include the additional 24mM K^+ ions contributed from the buffer, unless stated otherwise.

3.2.3 Single-molecule FRET

Single-molecule measurements on freely diffusing RNA were performed on a house-built microscope with confocal detection. Fluorescent samples were loaded into a chambered coverglass and illuminated with a 488nm laser through a 60x 1.2 NA objective. The fluorescence emission was collected through the same objective and split into donor and acceptor channels using a 550nm longpass dichroic. The donor channel contained an additional 530/30 bandpass emission filter, while the acceptor channel contained a 630 longpass emission filter. Two 50nm optical fibres provided confocal detection and photons were detected by two avalanche photodiodes (SPCM-AQR-14, Perkin Elmer, Santa Clara, CA). Data were acquired using Flex-2kD correlator card (Correlator.com) in photon counting mode for which photon arrival times were sampled in 25ns intervals, and then processed further using house-written MATLAB (MathWorks, Natick, MA) scripts. For each salt concentration, single-molecule data were collected for 30 minutes. At least two independent measurements were taken to estimate the variance in the measurement. The raw photon counts were converted into the number of photons measured per 1ms in the donor and acceptor channels, and only events above a threshold on total intensity of 20 counts/ms. The efficiency of energy transfer (E_{FRET}) was calculated as the background-subtracted ratio of the acceptor intensity to the total intensity. A histogram of E_{FRET} values was fitted to three gaussians, the middle of which represents the signal of interest (Figures C.2-C.9 in the Supplementary Information show typical histograms collected over 30 minutes for each salt concentration and RNA construct). The low- E_{FRET} peak represents either donor-only constructs or ones with an inactive acceptor. A small population with high- E_{FRET} (0.8-0.9) was also observed, but the salt dependence was minimal. Therefore in

this paper we focus on the mid-FRET peak which has a strong salt dependence.

3.3 Results

3.3.1 Effect of helix length

We first investigated the effects of RNA helix length on the conformation of our helix-junction-helix (HJH) construct. The short double helices we employ can be considered rigid due to the long persistence length of duplex RNA (approximately 250 bp [33, 34]); however, as a result of end effects their electrostatic properties may vary with helix length [18]. To examine the impact of helix length on RNA conformation, we use smFRET to compare RNA constructs in which a poly$(U)_5$ junction is flanked by either 12bp or 24bp helices. This junction sequence represents the simplest case for these studies since base-stacking interactions are absent in poly(U). For both 12 and 24bp helix constructs, the donor and acceptor dyes were attached 8bp away from the junction region, one on each flanking helix (Figure 3.1). Figure 3.2 shows the E_{FRET} of attached dyes in each of these constructs, measured as a function of KCl in solution. Unexpectedly, the E_{FRET} is non-monotonic with [KCl], in contrast to the monotonic changes observed for isolated poly$(U)_{40}$ reported in reference [26]. In addition, the convergence of the two curves at both low and high salt suggests no dependence on helix length in these regimes. The impact of helix length on RNA conformation emerges at intermediate salt concentrations, between 30-100 mM monovalent (KCl) concentrations, as the E_{FRET} trends from increasing to decreasing. The reversal in E_{FRET} occurs at lower salt when the construct contains

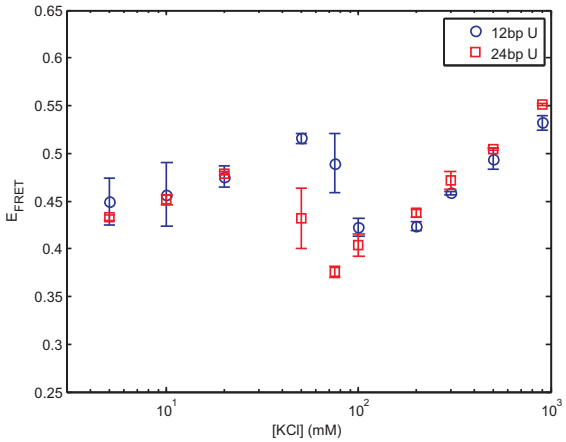

Figure 3.2: Helix length comparison of E_{FRET} for poly(U) junction constructs in KCl.

longer 24bp helices as opposed to 12bp helices. While end-fraying may be a concern for such short duplexes, we expect this effect to be more pronounced for the 12 bp construct, where the dye is much closer to the blunt end (fraying at the end closest to the junction is less likely due to the stabilizing effect of the linker [35]). The agreement between the E_{FRET} values for both constructs, at the extremes of salt, suggests that fraying does not contribute substantially to the measured effects.

3.3.2 Effect of junction sequence

Next we investigate the sequence dependence of RNA flexibility in solutions containing KCl. The change from a poly(U) junction to poly(A) introduces the possibility of base-stacking interactions [27, 36]. To investigate the effects of base-stacking on RNA conformation, we measured the KCl dependence of E_{FRET} on RNA constructs with a poly(A) junction and compared results with data shown in Figure 3.2 for poly(U). Figure 3.3a and b, respectively show re-

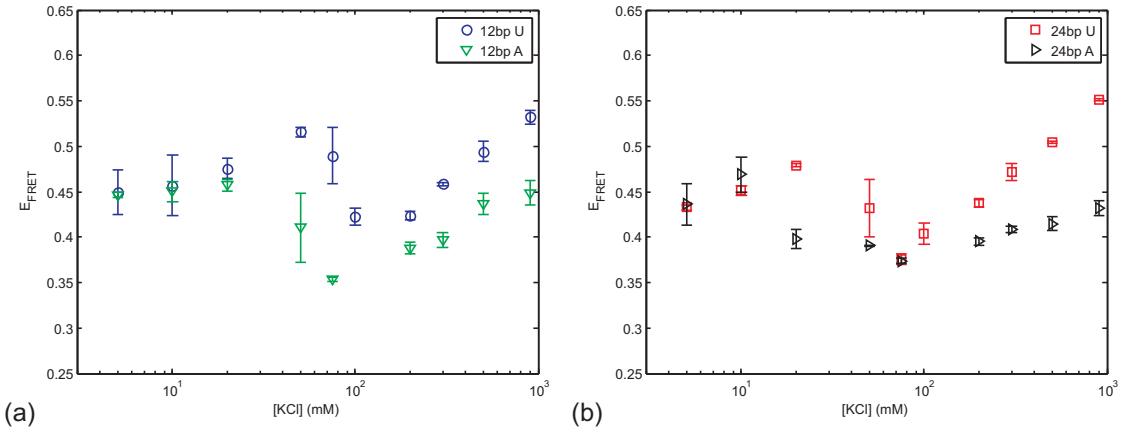

Figure 3.3: Junction sequence comparison in KCl for (a) 12bp and (b) 24bp helix lengths.

sults for the 12 and 24bp helix lengths. Note that the poly(U) data are repeated from Figure 3.2 for ease of comparison. Again, we observe the striking non-monotonic behavior for the poly(A) junctions that is present in the poly(U) data, however the new junction sequence alters the E_{FRET} in several ways. At low salt the E_{FRET} values are equivalent for a given helix length, suggesting that junction sequence has little effect when salt concentrations are low (<10-30mM KCl). At intermediate salt concentrations, the turnover in E_{FRET} of the poly(A) junction construct is shifted to lower salt when compared to the construct containing the poly(U) junction. Finally, at even higher salt (>200mM KCl), the E_{FRET} once again increases, however the E_{FRET} for the poly(A) construct is always lower than for the poly(U) construct. Except for the values at high salt, the poly(A) constructs exhibit similar trends to those seen with the poly(U) linker: helix length effects manifest at intermediate salt concentrations, in this case between 20-100mM KCl (Figure C.10 in the Supplementary Information shows the poly(A) helix length comparison).

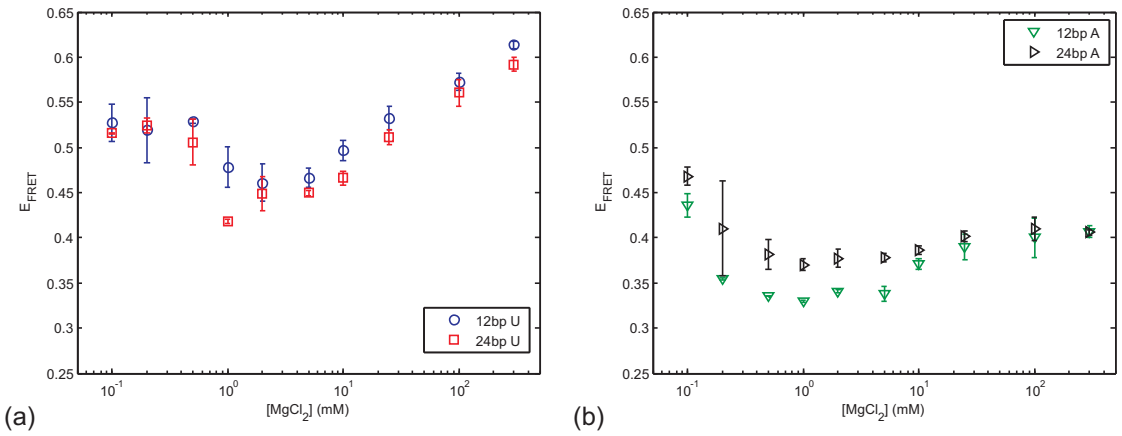

Figure 3.4: Helix length comparison of E_{FRET} in $MgCl_2$ for (a) poly(U) and (b) poly(A) junctions.

3.3.3 Effects of $MgCl_2$

Mg^{2+} ions stabilize RNA tertiary structure and are essential for RNA functions such as ribozyme catalysis [37]. To investigate the effect of Mg^{2+} ions on RNA structure, we measured HJH conformations as a function of increasing [$MgCl_2$] for all four of our RNA constructs: 12bp or 24bp helices and poly(U) or poly(A) junctions. The only monovalent ions present here are the 24mM K^+ present in the buffer. With poly(U) junctions, the effects of $MgCl_2$ echo those in KCl (Figure 3.4a), but are shifted to lower salt concentration: at low salt, the E_{FRET} of both 12 and 24bp constructs start at the same values, undergo a decrease at 0.5-1mM $MgCl_2$, and increase again at higher [$MgCl_2$]. At high salt the E_{FRET} values appear independent of helix length. Furthermore, while the differences are less pronounced in $MgCl_2$ than in KCl, the decrease in E_{FRET} occurs at slightly lower [$MgCl_2$] for the longer helix than for the shorter helix. Helix length effects disappear above 2mM $MgCl_2$.

In the poly(A) constructs, E_{FRET} begins to decrease at the lowest $MgCl_2$ con-

centrations measured (~0.2mM, Fig. 4b), following the trend observed in KCl that the reversal in E_{FRET} occurs at lower salt for the poly(A) linker than for the poly(U). Helix length effects manifest differently for poly(A) in $MgCl_2$ in that they occur over a larger range (0.1-10mM $MgCl_2$). Above 10mM $MgCl_2$, however, these length effects disappear.

3.3.4 Results Summary

Single molecule FRET data were acquired for four different RNA constructs in buffered solutions containing either K^+ or Mg^{2+} counterions, over a wide range of ionic concentrations. General trends are most clearly identified in the monovalent ion series. At low [KCl], the E_{FRET} value is independent of both helix length and junction sequence for all constructs studied. As [KCl] increases, E_{FRET} displays non-monotonic behavior. Finally at higher [KCl], the increasing E_{FRET} is independent of helix length, but does depend on junction sequence. Similar trends are observed as a function of increasing [$MgCl_2$].

To facilitate comparison between different constructs, we introduce C_{cr}, the critical salt concentration where the first reversal in E_{FRET} is observed (or the salt concentration where E_{FRET} reaches its local maximum.) We postulate in the following section that this abrupt reversal in E_{FRET} reflects the initiation of a transition where the molecules reorient to adopt a different conformational state, hence C_{cr} provides a metric for comparison across conditions. Figure 3.5 illustrates the effect of both junction sequence and helix length on this conformational rearrangement and motivates the following key observations. First, C_{cr} can be reduced by either increasing helix length or by introducing base-

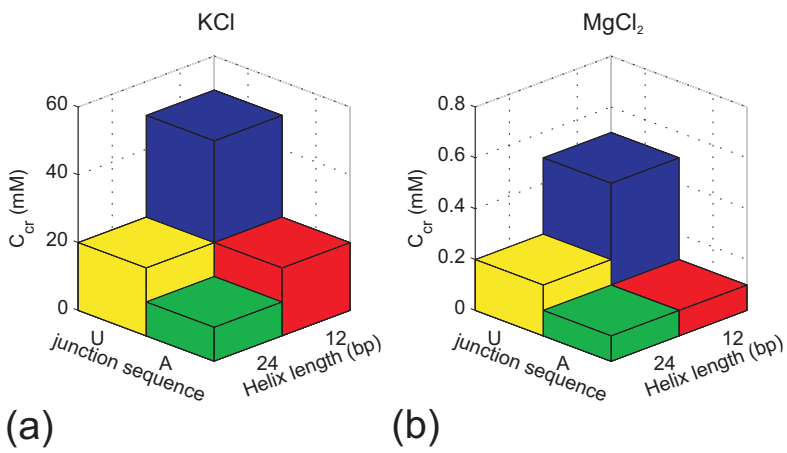

Figure 3.5: Summary of the critical salt concentration (C_{cr}) dependence on helix length and junction sequence in (a) KCl and (b) $MgCl_2$.

stacking to the junction (changing it from poly(U) to poly(A)). In monovalent salt solutions, these effects appear to be of comparable scale. Second, we observe that helix length effects for poly(U) are similar but less pronounced in $MgCl_2$ than in KCl. Finally poly(A) constructs in $MgCl_2$ have indistinguishable C_{cr}, possibly because salt concentrations below 0.1mM $MgCl_2$ were not measured. With the concept of C_{cr}, we organize the RNA constructs in terms of their propensity to initiate this transition. In KCl we have: 24A<12A~24U<12U. The order is similar in $MgCl_2$, except that the sequence effects are more pronounced giving: 24A~12A<24U<12U.

3.4 Discussion

3.4.1 Salt concentration-dependent behavior of isolated RNA elements

All of the RNA constructs used in this study contain two structural elements: helices and single strands. Our hypothesis is that each element contributes in a distinct way to overall RNA conformation, tuning the behavior of the system. Before we discuss the implications of our results from the combined system, it is important to review the effects of salt concentration on each individual component. We therefore start by expanding on the introduction and discussing the salt dependence of isolated helix repulsion and on the sequence dependence of RNA homopolymers.

Interactions between isolated short helices

The interaction potential between particles in solution can be described by the second virial coefficient (A_2). When $A_2 = 0$, the particles are non-interacting, $A_2 > 0$ represents repulsion, and $A_2 < 0$ denotes an attraction. The salt dependence of A_2 was reported previously for isolated RNA helices [14, 17], showing that repulsion between isolated helices in solution decreases with increasing salt concentration. A_2 for 25bp RNA helices drops to zero at ~150mM KCl, after which increasing salt concentration leads to attraction in the form of end-to-end stacking [17]. End-to-end stacking of two separate RNA molecules is not a concern in the dilute solutions used in the present study, and the likelihood of tethered helices stacking within one molecule is low (discussed in the Supple-

mentary Information). Nevertheless, the 150mM threshold represents a useful cutoff describing where repulsion no longer dominates inter-helical interactions for long RNA helices.

Inter-helical interactions are also affected by helix length via end effects [19]. These effects are particularly relevant because nature uses short helices in functional RNAs [21–23]. Modelling DNA as a uniformly charged cylinder, Allison [20] showed that potentials along the cylinder's surface are reduced for short helices but the differences between helix lengths become less significant as the salt concentration is increased. Simulations of ion distributions around DNA helices show that helix length effects emerge at a length < 24bp [38].

To verify whether similar end effects exist for RNA helices we used the Adaptive Poisson-Boltzmann Solver (APBS) [39] to simulate electrostatic potentials around isolated RNA helices of 12 and 24bp in length (see Supplementary Information for details, and results in Figure C.11). While the non-linear Poisson Boltzmann equation has certain caveats when applied to nucleic acids (i.e. it underestimates the number of ions close to the surface [40]), we simply use it to gain insight into the electrostatics of end-effects. Furthermore, we focus on monovalent salt conditions since predictions using the Poisson-Boltzmann equation are believed to be more accurate for monovalent than divalent ions [16]. These simulations show similar end effects in RNA as were observed for DNA: the potential of 12bp RNA helices is smaller in magnitude that that of 24bp helices, and end effects become less pronounced with increasing salt concentration. In addition, the simulations suggest that these differences arise from a reduction in radial rather than axial potential. The potential along the helical axis is essentially unaffected by helix length (Figure C.12 in the Supplementary

Information).

Unpaired single strands

Past work on poly(U) homopolymers suggests that it follows random coil behavior, and molecular flexibility increases with salt [26, 28]. In contrast, references [27, 41, 42] show the propensity of adjacent nucleotides of poly(A) to stack, causing it to form short-range helical structures. The stacking propensity of poly(A) may also be influenced by the salt concentration [43]. The stark differences between conformations of poly(A) and poly(U) homopolymers discussed in the introduction suggest that the junction sequence will profoundly affect the overall conformation of the RNA construct.

3.4.2 Unique behaviour emerges as structural elements are combined in monovalent salt

When we combine RNA structural elements using our HJH construct, the unexpected behavior shown in Figures 3.2 and 3.3 emerges. To interpret, we divide the data into three experimentally distinct regimes: low salt (increasing E_{FRET}), mid-salt (decreasing E_{FRET}) and high salt (increasing E_{FRET}). The boundaries between these regimes depend in a non-trivial way on the construct. Because these regions are most distinctly separated in solutions containing KCl, we will focus the first part of the discussion on the monovalent ion dependence of HJH conformations. We propose a model that is consistent with our observations, and suggests how a unique combination of helix length, junction sequence and

ion identity can affect RNA conformation.

Individual structural elements dominate at limiting [KCl]

We begin by examining the limiting cases of low and high [KCl], respectively. At low salt (<10mM - 20mM KCl) neither junction sequence nor helix length affects the E_{FRET} values (Figures 3.2 and 3.3, and C.2). In this regime, helix repulsion is significant, and the overall electrostatic potential is minimized in a conformation in which the two helices are pushed as far from each other as possible. Despite any potential end effects, the strength of the helix repulsion is so high that the junction is fully stretched, limiting the available helix conformations to the same subset of states independent of helix length or junction sequence.

Next we describe the high salt limiting case. Above 150mM KCl, E_{FRET} is independent of helix length (Figure 3.2, Figure C.10). This observation is not surprising because end effects diminish at high salt ([20] and Figure C.12) thus local electrostatic potentials near the junction region should not depend on helix length. Additionally, in light of past measurements of A_2, which show that inter-helical repulsion is insignificant above 150mM KCl, albeit for longer helices [17], therefore we expect helix repulsion itself to be negligible at high salt concentrations.

In this regime, the data show the dramatic impact of junction sequence on E_{FRET} (Figure 3.3). Most obviously, the poly(A) containing constructs are always more extended (lower E_{FRET}) than those containing poly(U). To compare the salt-dependence of E_{FRET} for the poly(U) junction with that previously measured for (isolated) single-stranded poly$(U)_{40}$ [26], we approximate the data as

linear with the logarithm of salt concentration. The comparison confirms that junction conformation dominates in the high salt regime. These two data sets display the same power law dependence of E_{FRET} on [salt] (Figure C.13 in the Supplementary Information), confirming that the salt dependent E_{FRET} changes at high salt, where helix repulsion is negligible, are caused by the ion-dependent changes in the persistence length of poly(U) [26].

We cannot make the same comparison for poly(A) junctions because we do not have corresponding FRET results as we did in the case for poly(U). However, our data show that the fluorescent dyes are always closer together for the poly(U) constructs than for the poly(A) constructs at high salt, consistent with an increase in poly(A) base-stacking with salt concentration [43], which drives the flanking helices towards coaxial conformations. In contrast the flexible poly(U) linker allows the flanking helices (and associated fluorophores) to approach one another more closely. Thus, our data are consistent with a picture where the junction sequence determines the RNA conformation at high salt.

Decreasing E_{FRET} at physiological monovalent concentrations explained by flip in relative helix orientation

At low salt, we have shown that helix repulsion dominates RNA conformation; while at high salt, junction sequence dominates. Next we address the biologically important, mid salt regime, which includes physiological monovalent concentrations (~100mM [44]). Helix repulsion will be reduced relative to the low salt case, and junction dynamics may become visible, allowing helices to explore more conformational space and making end effects more prevalent. Therefore this regime is characterized by the interplay between all three con-

tributing structural factors: helix repulsion, end effects and junction dynamics.

In this regime we observe a striking decrease in E_{FRET}, which, assuming freely rotating dyes (see Supplementary Information for a discussion of this assumption), indicates that the dyes are moving farther apart. We propose a model in which the decrease in E_{FRET} results from two types of interchanging helix conformations, one with a larger E_{FRET}. The population of conformations with higher E_{FRET} are not favored at low salt, and only become accessible as salt is increased above C_{cr}. Since E_{FRET} histograms show a shifting peak position rather than two separate peaks, these populations must be rapidly interchanging on a timescales faster than the measurement time [45]. One potential scenario is that helices can 'flip' from a subset of skewed to parallel conformations at C_{cr} (Figure 3.6). This 'flip' causes the average distance between dyes to increase. A complete picture of how salt may affect the combined HJH RNA is described in the caption for Figure 3.6.

Our proposed model relies on the assumption that 'skewed' helix orientations are preferred at low salt. Such asymmetric conformations are not unprecedented. In previous work on DNA helices tethered by an uncharged synthetic linker, the off-center attachment points of the linker cause equilibrium distributions of helix orientation to fluctuate around a conformation where the interhelical angle is 150° [46]. We expect our RNA constructs to exhibit a similar angular bias since the 'attachment' of the RNA single stranded junction is also offset from the helical axis. While the DNA helix can be modelled as a right-angle cylinder (its bases are tilted by only -6°), the cross section of the RNA helix is more trapezoidal (its bases are tilted by 16-19° [47], illustrated in Figure 3.6). This tilt of the RNA bases may cause the discrete transition observed

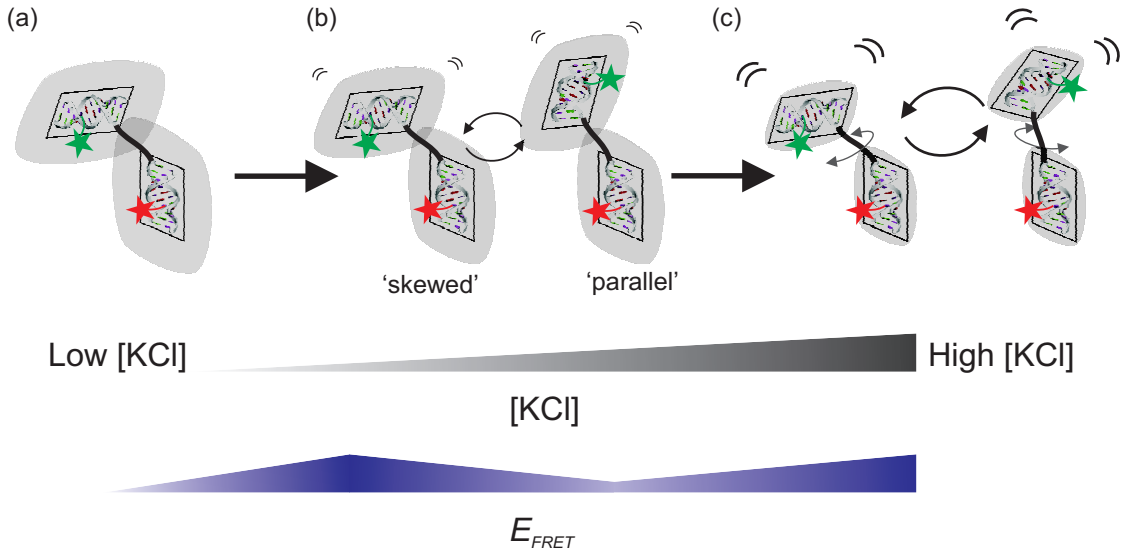

Figure 3.6: Schematic of proposed model describing mid-salt decrease in E_{FRET}. In this picture, colored stars represent dye labels and shaded grey areas represent the boundary where electrostatic potential of individual helices drops to a finite cutoff. (a) At low [KCl], helix repulsion is high (large overlap between shaded areas). Consequently, the junction is extended to its maximum. Helices are fixed in skewed conformations due to the off-center attachment of the single-stranded RNA linker. (b) As salt is increased to C_{cr}, repulsion is reduced enough for junction dynamics to let the helices flip between skewed and parallel conformations. The non-right angled nature of the RNA helix cross-section may cause these two distinct subsets of helix orientations. The 'parallel' conformations have dyes farther apart and thus a lower E_{FRET} than the 'skewed' conformations. (c) Increasing salt even further screens helix repulsion completely and junction dynamics alone determine RNA conformation.

at intermediate salt concentrations (Figure 3.6(b)), rather than the continuous one observed for tethered DNA helices [46]. Verification is needed to confirm whether 'skewed' conformations are favored at low salt, and whether a free energy barrier exists between skewed and parallel states. However, a detailed analysis of the energetics of our RNA constructs is beyond the scope of this work.

3.4.3 Proposed model is consistent with dependence of E_{FRET} on junction sequence, helix length and $MgCl_2$

The proposed model can explain the observed trends in E_{FRET} as both junction sequence and helix length are varied: C_{cr} is reduced for poly(A) relative to poly(U) and for 24 vs 12bp helices. Sequence dependence is straightforward: stacking of adjacent poly(A) nucleotides drives the junction towards a helical conformation. Since stacking favors parallel helix conformations, less salt is required to flip the helices out of the 'skewed' conformation, explaining the reduced C_{cr} for poly(A) constructs relative to poly(U).

APBS simulations show that end effects alter the radial electrostatic potential of a helix, while the axial potentials are relatively unchanged (Figure C.12). If helices are skewed at low salt, they will interact with one another along their radial directions (i.e. refer to Figure 3.6a). Longer helices have a larger radial potential, so the skewed conformation will be slightly disfavored compared to the shorter helices. Thus as salt concentration is increased and overall repulsion is reduced, the longer helices will prefer to 'flip' earlier in order to maximize the distance between the long sides of either helix, causing a reduction in C_{cr}.

Mg^{2+} ions affect RNA flexibility in at least two ways. First, at a given salt concentration, they screen RNA's negatively charged backbone more efficiently than do monovalent ions. Second, they stabilize poly(A) stacking. All the same trends observed for poly(U) in KCl are also detected in $MgCl_2$. Not surprisingly, much less Mg^{2+} is required than K^+ to initiate the transition from skewed to flipped, consistent with many previous observations showing greater effectiveness for charge screening of Mg^{2+}. On the contrary, for poly(A), C_{cr} is indistinguishable for both helix lengths. This result is consistent with the high sensitivity of poly(A) base stacking to Mg^{2+} ions. Since stacking is strongly preferred in the presence of Mg^{2+}, helices are driven to the parallel conformation, regardless of end effects. In fact, the measured E_{FRET} of poly(A) constructs agree remarkably well with the simulated E_{FRET} computed from a model where a fully A-form junction connects two 24bp helices (see Supplementary Information).

3.4.4 Helix length and junction sequence can bias conformations of unfolded RNA

In the conformational capture picture of riboswitch ligand binding, the distribution of conformations in the unfolded state determines the function of the RNA molecule. Environmental factors that shift the distribution towards its ligand-bound state will increase binding rates, while factors that shift the distribution towards an unbound state will reduce the likelihood of binding. For example, recent work has shown that Mg^{2+} stabilizes a 'pseudoknot-like' structure in the SAM-II riboswitch, favoring SAM binding [4]. Furthermore, the A-rich tail of the class I prequeuosine riboswitch exhibits a dynamic helical structure in so-

lution that the authors propose increases the efficiency of ligand binding with respect to a disordered structure [48]. 'Preorganization' of binding partners is not unique to RNA. Evidence from smFRET measurements suggests that NaCl facilitates DNA duplex formation in solution by favouring pre-formed helical structures of single strands, which can then form a duplex through a 'docking' type interaction [49, 50]. In general, this type of preorganization may be critical in tipping the free energy balance between bound and unbound states of RNA. In this work, we report that helix length and junction sequence affect the available conformations of a model unfolded RNA. We can bias the conformations of RNA (decreasing C_{cr}) by either decreasing the helix length or by mutating the junction residues from U to A. We propose that these two factors can be exploited to pre-organize the structures of unfolded RNA and may be used to influence the interactions between RNA and its binding partners. For example, structural elements of riboswitches may contribute to biasing their conformations towards either ligand bound or unbound states.

It is interesting to note that the average duplex length in long RNAs is only 4bp [21–23], highlighting the importance of short duplexes in biology. We see differences arise between the 12 and 24bp duplexes at just below physiological monovalent concentrations (20-100mM KCl). Since decreasing helix length pushes the E_{FRET} transition to higher salt, we might expect that differences between lengths of even shorter helices would be further propelled towards physiologically relevant salt conditions.

While the biological importance of single-stranded RNA linkers is clearly important [51–53], double-stranded junctions are pervasive [54–56]. Changing the topology of the junction can dramatically alter the available conformations

of the RNA [55–58]. We anticipate that our observation that RNA conformation is influenced by the length of its helices and the sequence of its junctions would be generally true for all junctions, single or double stranded. The studies of single-stranded junctions described here provide a first step in understanding the underlying behavior of junctions in RNA. Future experiments will continue to explore the link between topology and RNA junction structure.

3.5 Conclusions

We described here the use of smFRET to compare the salt-dependence of RNA conformations with two different helix lengths and junction sequences. These measurements show that different structural elements determine the conformations at extreme monovalent salt concentrations: helix repulsion dominates at low salt while junction sequence dominates at high salt. At intermediate, near physiological concentrations, both the junction sequence and the length of the helices determine available RNA conformations. In the presence of Mg^{2+} ions, the critical salt concentration dividing low and high salt regions is ~100 times lower than in the presence of K^{+} ions, and the increased sensitivity of base-stacking in poly(A) causes a dramatic sequence dependence. The influence of helix length and junction sequence on RNA conformation at near physiological salt concentrations suggests a biologically relevant mechanism for 'tuning' the available conformations of unfolded RNA. In light of the conformational capture mechanism, helix length and junction sequence may both be important factors in directing interactions between RNA and its binding partners.

Acknowledgement

The authors would like to thank the Pollack lab members George Calvey, Yujie Chen, Andrea Katz, Suzette Pabit, Alex Plumridge and Josh Tokuda for useful discussions. This work was supported by the National Institutes of Health [R01-GM085062].

References

[1] D. D. Boehr, R. Nussinov, and P. E. Wright, "The role of dynamic conformational ensembles in biomolecular recognition," *Nature chemical biology*, vol. 5, no. 11, pp. 789–796, 2009.

[2] P. Csermely, R. Palotai, and R. Nussinov, "Induced fit, conformational selection and independent dynamic segments: an extended view of binding events," *Trends in biochemical sciences*, vol. 35, no. 10, pp. 539–546, 2010.

[3] C. D. Stoddard, R. K. Montange, S. P. Hennelly, R. P. Rambo, K. Y. Sanbonmatsu, and R. T. Batey, "Free state conformational sampling of the sam-i riboswitch aptamer domain," *Structure (London, England : 1993)*, vol. 18, no. 7, pp. 787–797, 2010.

[4] A. Haller, U. Rieder, M. Aigner, S. C. Blanchard, and R. Micura, "Conformational capture of the sam-ii riboswitch," *Nature chemical biology*, vol. 7, no. 6, pp. 393–400, 2011.

[5] B. Heppell, S. Blouin, A. M. Dussault, J. Mulhbacher, E. Ennifar, J. C. Penedo, and D. A. Lafontaine, "Molecular insights into the ligand-controlled organization of the sam-i riboswitch," *Nature chemical biology*, vol. 7, no. 6, pp. 384–392, 2011.

[6] R. C. Wilson, A. M. Smith, R. T. Fuchs, I. R. Kleckner, T. M. Henkin, and M. P. Foster, "Tuning riboswitch regulation through conformational selection," *Journal of Molecular Biology*, vol. 405, no. 4, pp. 926–938, 2011.

[7] A. Serganov and D. J. Patel, "Molecular recognition and function of riboswitches," *Current opinion in structural biology*, vol. 22, no. 3, pp. 279–286, 2012.

[8] J. Nowakowski and I. T. Jr., "Rna structure in solution," in, ser. Oxford handbook of nucleic acid structure, S. Neidle, Ed., Oxford ;New York: Oxford University Press, 1999, p. 567.

[9] A. A. Szewczak and T. R. Cech, "An rna internal loop acts as a hinge to facilitate ribozyme folding and catalysis," *RNA*, vol. 3, no. 8, pp. 838–849, 1997.

[10] J. C. Schlatterer, L. W. Kwok, J. S. Lamb, H. Y. Park, K. Andresen, M. Brenowitz, and L. Pollack, "Hinge stiffness is a barrier to rna folding," *Journal of Molecular Biology*, vol. 379, no. 4, pp. 859–870, 2008.

[11] E. A. Dethoff, J. Chugh, A. M. Mustoe, and H. M. Al-Hashimi, "Functional complexity and regulation through rna dynamics," *Nature*, vol. 482, no. 7385, pp. 322–330, 2012.

[12] J. Šponer, P. Banáš, P. Jurečka, M. Zgarbová, P. Kührová, M. Havrila, M. Krepl, P. Stadlbauer, and M. Otyepka, "Molecular dynamics simulations of nucleic acids. from tetranucleotides to the ribosome," *The Journal of Physical Chemistry Letters*, vol. 5, no. 10, pp. 1771–1782, 2014.

[13] D. Herschlag, B. E. Allred, and S. Gowrishankar, "From static to dynamic: the need for structural ensembles and a predictive model of rna folding and function," *Current opinion in structural biology*, vol. 30, pp. 125–133, 2015.

[14] S. A. Pabit, X. Qiu, J. S. Lamb, L. Li, S. P. Meisburger, and L. Pollack, "Both helix topology and counterion distribution contribute to the more effective charge screening in dsrna compared with dsdna," *Nucleic acids research*, vol. 37, no. 12, pp. 3887–3896, 2009.

[15] S. Kirmizialtin, S. A. Pabit, S. P. Meisburger, L. Pollack, and R. Elber, "Rna and its ionic cloud: solution scattering experiments and atomically detailed simulations," *Biophysical journal*, vol. 102, no. 4, pp. 819–828, 2012.

[16] S. Kirmizialtin, A. R. Silalahi, R. Elber, and M. O. Fenley, "The ionic atmosphere around a-rna: poisson-boltzmann and molecular dynamics simulations," *Biophysical journal*, vol. 102, no. 4, pp. 829–838, 2012.

[17] S. A. Pabit, J. L. Sutton, H. Chen, and L. Pollack, "Role of ion valence in the submillisecond collapse and folding of a small rna domain," *Biochemistry*, vol. 52, no. 9, pp. 1539–1546, 2013.

[18] J. D. Ballin and G. M. Wilson, "Role and applications of electrostatic effects on nucleic acid conformational transitions and binding processes," in, ser. Application of Thermodynamics to biological and materials science, M. Tadashi, Ed., InTech, 2011, pp. 129–174.

[19] T. Nicolai and M. Mandel, "The ionic strength dependence of the second virial coefficient of low molar mass dna fragments in aqueous solutions," *Macromolecules*, vol. 22, no. 1, pp. 438–444, 1989.

[20] S. A. Allison, "End effects in electrostatic potentials of cylinders - models for dna fragments," *Journal of Physical Chemistry*, vol. 98, no. 46, pp. 12 091–12 096, 1994.

[21] W. Fontana, D. A. Konings, P. F. Stadler, and P. Schuster, "Statistics of rna secondary structures," *Biopolymers*, vol. 33, no. 9, pp. 1389–1404, 1993.

[22] A. M. Yoffe, P. Prinsen, A. Gopal, C. M. Knobler, W. M. Gelbart, and A. Ben-Shaul, "Predicting the sizes of large rna molecules," *Proceedings of the National Academy of Sciences*, vol. 105, no. 42, pp. 16 153–16 158, 2008.

[23] L. T. Fang, A. M. Yoffe, W. M. Gelbart, and A. Ben-Shaul, "A sequential folding model predicts length-independent secondary structure properties of long ssrna," *The journal of physical chemistry.B*, vol. 115, no. 12, pp. 3193–3199, 2011.

[24] B.-Y. Ha and D. Thirumalai, "Electrostatic persistence length of a polyelectrolyte chain," *Macromolecules*, vol. 28, no. 2, pp. 577–581, 1995.

[25] M. Fixman, "Electrostatic persistence length.," *The journal of physical chemistry. B*, vol. 114, no. 9, pp. 3185–3196, 2010.

[26] H. Chen, S. P. Meisburger, S. A. Pabit, J. L. Sutton, W. W. Webb, and L. Pollack, "Ionic strength-dependent persistence lengths of single-stranded rna and dna," *Proceedings of the National Academy of Sciences*, vol. 109, no. 3, pp. 799–804, 2012.

[27] Y. Seol, G. Skinner, K. Visscher, A. Buhot, and A. Halperin, "Stretching of homopolymeric rna reveals single-stranded helices and base-stacking," *Physical Review Letters*, vol. 98, no. 15, pp. 1–4, 2007.

[28] Y. Seol, G. M. Skinner, and K. Visscher, "Elastic properties of a single-stranded charged homopolymeric ribonucleotide," *Physical Review Letters*, vol. 93, no. 11, p. 118 102, 2004.

[29] B. I. Kankia, "Binding of mg2+ to single-stranded polynucleotides: hydration and optical studies," *Biophysical chemistry*, vol. 104, no. 3, pp. 643–654, 2003.

[30] B. Kankia, "Inner-sphere complexes of divalent cations with single-stranded poly(ra) and poly(ru)," *Biopolymers*, vol. 74, no. 3, pp. 232–239, 2004.

[31] N. R. Markham and M. Zuker, "Dinamelt web server for nucleic acid melting prediction," *Nucleic acids research*, vol. 33, no. Web Server issue, W577–581, 2005.

[32] S. Sindbert, S. Kalinin, H. Nguyen, A. Kienzler, L. Clima, W. Bannwarth, B. Appel, S. Muller, and C. A. Seidel, "Accurate distance determination of nucleic acids via forster resonance energy transfer: implications of dye linker length and rigidity," *Journal of the American Chemical Society*, vol. 133, no. 8, pp. 2463–2480, 2011.

[33] J. A. Abels, F. Moreno-Herrero, T. van der Heijden, C. Dekker, and N. H. Dekker, "Single-molecule measurements of the persistence length of double-stranded rna," *Biophysical journal*, vol. 88, no. 4, pp. 2737–2744, 2005.

[34] P. Kebbekus, D. E. Draper, and P. Hagerman, "Persistence length of rna," *Biochemistry*, vol. 34, no. 13, pp. 4354–4357, 1995.

[35] P. J. Romaniuk, D. W. Hughes, R. J. Gregoire, T. Neilson, and R. A. Bell, "Stabilizing effect of dangling bases on a short rna double helix as determined by proton nuclear magnetic resonance spectroscopy," *Journal of the American Chemical Society*, vol. 100, no. 12, pp. 3971–3972, 1978.

[36] H. Eisenberg and G. Felsenfeld, "Studies of the temperature-dependent conformation and phase separation of polyriboadenylic acid solutions at neutral ph," *Journal of Molecular Biology*, vol. 30, no. 1, pp. 17–37, 1967.

[37] D. E. Draper, "A guide to ions and rna structure," *RNA*, vol. 10, no. 3, pp. 335–343, 2004.

[38] M. C. Olmsted, C. F. Anderson, and M. T. R. Jr, "Monte carlo description of oligoelectrolyte properties of dna oligomers: range of the end effect and

the approach of molecular and thermodynamic properties to the polyelectrolyte limits," *Proceedings of the National Academy of Sciences*, vol. 86, no. 20, pp. 7766–7770, 1989.

[39] N. A. Baker, D. Sept, S. Joseph, M. J. Holst, and J. A. McCammon, "Electrostatics of nanosystems: application to microtubules and the ribosome," *Proceedings of the National Academy of Sciences*, vol. 98, no. 18, pp. 10 037–10 041, 2001.

[40] S. P. Meisburger, S. A. Pabit, and L. Pollack, "Determining the locations of ions and water around dna from x-ray scattering measurements.," *Biophysical journal*, vol. 108, no. 12, pp. 2886–2895, 2015.

[41] D. N. Holcomb and I. Tinoco, "Conformation of polyriboadenylic acid: ph and temperature dependence," *Biopolymers*, vol. 3, no. 2, pp. 121–133, 1965.

[42] M. Leng and G. Felsenfeld, "A study of polyadenylic acid at neutral ph," *Journal of Molecular Biology*, vol. 15, no. 2, pp. 455–466, 1966.

[43] T. G. Dewey and D. H. Turner, "Laser temperature-jump study of stacking in adenylic acid polymers," *Biochemistry*, vol. 18, no. 26, pp. 5757–5762, 1979.

[44] H. Lodish, A. Berk, S. L. Zipursky, P. Matsudaira, D. Baltimore, and J. Darnell, *Molecular cell biology*, en, 2000.

[45] I. V. Gopich and A. Szabo, "Theory of single-molecule fret efficiency histograms," in *Single-Molecule Biophysics: Experiment and Theory*, ser. Single-Molecule Biophysics, T. Komatsuzaki, M. Kawakami, S. Takahashi, H. Yang, and R. J. Silbey, Eds., vol. 146, John Wiley & Sons, Inc., 2011, pp. 245–297.

[46] Y. Bai, V. B. Chu, J. Lipfert, V. S. Pande, D. Herschlag, and S. Doniach, "Critical assessment of nucleic acid electrostatics via experimental and computational investigation of an unfolded state ensemble," *Journal of the American Chemical Society*, vol. 130, no. 37, pp. 12 334–12 341, 2008.

[47] W. Saenger, *Principles of nucleic acid structure*, C. R. Cantor, Ed. New York: Springer-Verlag, 1984.

[48] C. D. Eichhorn, J. Feng, K. C. Suddala, N. G. Walter, C. L. B. 3rd, and H. M. Al-Hashimi, "Unraveling the structural complexity in a single-stranded rna tail: implications for efficient ligand binding in the prequeuosine riboswitch," *Nucleic acids research*, vol. 40, no. 3, pp. 1345–1355, 2012.

[49] J. A. Holbrook, M. W. Capp, R. M. Saecker, and M. T. R. Jr, "Enthalpy and heat capacity changes for formation of an oligomeric dna duplex: interpretation in terms of coupled processes of formation and association of single-stranded helices," *Biochemistry*, vol. 38, no. 26, pp. 8409–8422, 1999.

[50] N. F. Dupuis, E. D. Holmstrom, and D. J. Nesbitt, "Single-molecule kinetics reveal cation-promoted dna duplex formation through ordering of single-stranded helices," *Biophysical journal*, vol. 105, no. 3, pp. 756–766, 2013.

[51] M. Mandal, M. Lee, J. E. Barrick, Z. Weinberg, G. M. Emilsson, W. L. Ruzzo, and R. R. Breaker, "A glycine-dependent riboswitch that uses cooperative binding to control gene expression," *Science*, vol. 306, no. 5694, pp. 275–279, 2004.

[52] M. F. Lodeiro, C. V. Filomatori, and A. V. Gamarnik, "Structural and functional studies of the promoter element for dengue virus rna replication," *Journal of virology*, vol. 83, no. 2, pp. 993–1008, 2009.

[53] P. Nguyen, X. Shi, S. T. Sigurdsson, D. Herschlag, and P. Z. Qin, "A single-stranded junction modulates nanosecond motional ordering of the substrate recognition duplex of a group i ribozyme," *Chembiochem : A European journal of chemical biology*, vol. 14, no. 14, pp. 1720–1723, 2013.

[54] T. A. Goody, S. E. Melcher, D. G. Norman, and D. M. Lilley, "The kink-turn motif in rna is dimorphic, and metal ion-dependent," *RNA*, vol. 10, no. 2, pp. 254–264, 2004.

[55] A. M. Mustoe, M. H. Bailor, R. M. Teixeira, C. L. B. 3rd, and H. M. Al-Hashimi, "New insights into the fundamental role of topological constraints as a determinant of two-way junction conformation," *Nucleic acids research*, vol. 40, no. 2, pp. 892–904, 2012.

[56] M. H. Bailor, X. Sun, and H. M. Al-Hashimi, "Topology links rna secondary structure with global conformation, dynamics, and adaptation," *Science*, vol. 327, no. 5962, pp. 202–206, 2010.

[57] V. B. Chu, J. Lipfert, Y. Bai, V. S. Pande, S. Doniach, and D. Herschlag, "Do conformational biases of simple helical junctions influence rna folding stability and specificity?" *RNA*, vol. 15, no. 12, pp. 2195–2205, 2009.

[58] R. S. Tang and D. E. Draper, "Bend and helical twist associated with a symmetric internal loop from 5s ribosomal rna," *Biochemistry*, vol. 33, no. 33, pp. 10 089–10 093, 1994.

CHAPTER 4

INTERPLAY BETWEEN SEQUENCE AND SALT ON DOUBLE-STRANDED RNA JUNCTION CONFORMATION

Julie L. Sutton[1], and Lois Pollack[1]

1. School of Applied and Engineering Physics, Cornell University, Ithaca, New York

Abstract

Studying the influence of enviromnental factors on isolated RNA elements can provide useful insights into the fundamental forces that drive RNA conformational changes. We use the j5/5a junction from the P4-P6 domain of the *Tetrahymena* ribozyme as a representative flexible hinge and compare it to a U_4-U_5 mutant. Using single-molecule FRET, we compare the conformations of the isolated RNA junctions in solutions containing varying concentrations of KCl and $MgCl_2$ concentrations. Our results suggest that sequence has a profound effect on junction structure where subtle changes lead directly to different bend and twist angles of two-helix junctions. The interplay between sequence and salt concentration is an important factor in determining the conformation of RNA.

4.1 Introduction

The flexibility of RNA, coupled with its base stacking and base pairing interactions, means that numerous conformational states can have similar energet-

ics. Subtle changes in environment can tip the balance between different states leading to large conformational changes. This rough energy landscape makes predicting how RNA will respond to environmental cues a difficult problem [1–4]. Because it is not yet possible to predict folding pathways from the RNA primary sequence alone, researchers have employed different strategies to try to understand RNA folding. One strategy is to choose a particular RNA of interest and study its folding pathway under a variety of conditions. The *Tetrahymena* ribozyme and its P4-P6 domain in particular have been studied in detail (see [5–9] among others). However, by isolating the individual components that contribute to RNA folding, structural measurements can provide specific insight into the fundamental forces governing RNA folding [10–12].

In vitro RNA folding experiments typically start under conditions where secondary structure is formed, but there are no tertiary contacts. In this context, RNA can be considered a series of rigid helix elements connected by non-base-paired junctions [13]. Folding can be induced by Mg^{2+} ions, for example, which may increase the flexibility of the junctions [9]. With the increased flexibility, these regions can act as hinges, bringing together distal sides of a tertiary contact [14]. Understanding how these hinges respond to changing conditions is one step towards a full understanding of RNA folding.

Recent work from our group has investigated the role of sequence and salt in RNA junctions consisting of a single strand connecting two helix elements [15]. Others have considered the role of junction connectivity or topology [16]. For example Chu et al. [10] demonstrated that connecting DNA helices with two rather than one flexible linker can significantly restrict the available relative helix conformations. This idea has been extended to RNA junctions in references

[12, 17], where topological constraints were able to describe the limited range of helix-helix orientations in the PDB.

The pioneering work described in [17] aims to build a framework for predicting junction structure *ab initio*. In addition to topological constraints, Mustoe et al. [17] demonstrate that junction sequence serves to maximize non-canonical base pairs, however they do not discuss the effects of ions. One drawback to this work is that it was done by surveying junction structures in the PDB. While crystal structures provide unmatched resolution, they may not always be representative of RNA structure in solution or *in vivo*. Conversely, many biologically interesting RNA structures cannot be crystalized because of their inherent dynamics. Additionally, solving crystal structures often requires heavy, multivalent ions such as cobalt(III) hexammine, which are not physiological [18]. Because of RNA's high negative charge density, the influence of counterions is a critical part of understanding RNA structure. Based on our previous work [15], we hypothesize that the interwoven effects of ions and RNA sequence may also play a role in these double-stranded RNA junctions. We aim to expand on this framework to include the sequence and salt-dependence of RNA junction conformation.

Single-molecule Förster resonance energy transfer (smFRET) is a powerful tool to study changes in RNA conformation because it provides a sensitive measure of distance between two RNA-attached dyes. While extracting exact conformational detail is difficult due to the flexibility of the fluorophore linkers, the sensitivity of the technique is best exploited when measuring changes under varying external conditions. Specifically, since smFRET is sensitive to the inter-dye distance, it can report on subtle bends and twists of the RNA. While one

FRET value can arise from multiple RNA conformations, examining molecules with additional labeling positions can help reduce the degeneracy.

In this work, we examine the combined effects of sequence, ion concentration, and ion type between two RNA helices connected by a double-stranded RNA junction using smFRET. We compare the conformations of two different junction sequences: (i) the j5/5a hinge that is an important component of the P4-P6 domain of the *Tetrahymena* ribozyme [19] and (ii) a U_4U_5 mutant containing the same number of nucleotides as j5/5a which we call jU_4U_5. We find that the two sequences have similar bend angles but different twists in solutions containing KCl. The effects of $MgCl_2$ concentrations are larger, such that the structural differences between RNA sequences include both bend and twist angles. Furthermore, structural differences between j5/5a and jU_4U_5 grow with salt concentration, especially in $MgCl_2$. Finally, by comparing our results to a base-paired control, we provide evidence that the jU_4U_5 junction may be fully paired in high $MgCl_2$ concentrations. We propose that sequence-specific ion effects are important in guiding RNA junctions towards properly folded states.

4.2 Materials and Methods

4.2.1 RNA construct design

RNA constructs consist of two strands annealed together, each with their own fluorophore attachment (Figure 4.1(a)). The annealed sample contains the same two RNA helices which are connected by an asymmetric junction of 4 and 5 nt. The junction consists of one of three different sequence variations: j5/5a (based

on the sequence of the j5/5a hinge in the *Tetrahymena* ribozymes P4-P6 domain), the jU_4U_5 mutant (which contains the same number of nucleotides as j5/5a), and a base pair control (where the second strand is the same as the j5/5a, but the first strand is complementary to the second). The dye positions were chosen to ensure that the environment is the same for all constructs, making direct comparison across constructs possible (e.g. even if there is salt dependence). The acceptor label site (A in Figure 4.1(a)) was chosen to be 8nt from the junction region, while the donor label site was either at the 5′ end of the helix or 8bp away on the opposite side of the junction. We label the donor positions D1 and D2 for the end and internal label positions (refer to Figure 4.1(a)). RNA molecules were purchased desalted and HPLC purified from IDT with an amino-modified dT at the desired label site for internal labels or an amine group attached to a C12 linker for the end-labeled site.

4.2.2 Sample preparation

RNA strands were first labeled with the desired fluorophore (Alexa Fluor 488 TFP ester for the donor and Cy5 NHS ester for the acceptor) as in [15]. Excess free dye was removed from the sample by ethanol precipitation, the two strands were annealed together in a 1:2 donor:acceptor strand ratio at 92°C for 2 minutes and then slow cooled. Samples were then buffer exchanged to 100mM KCl in 50mM K-MOPS buffer. This process also helped remove trace amounts of free Alexa 488 which was less efficiently removed by the ethanol precipitation. Samples were aliquoted in small quantities and stored at -20°C until needed.

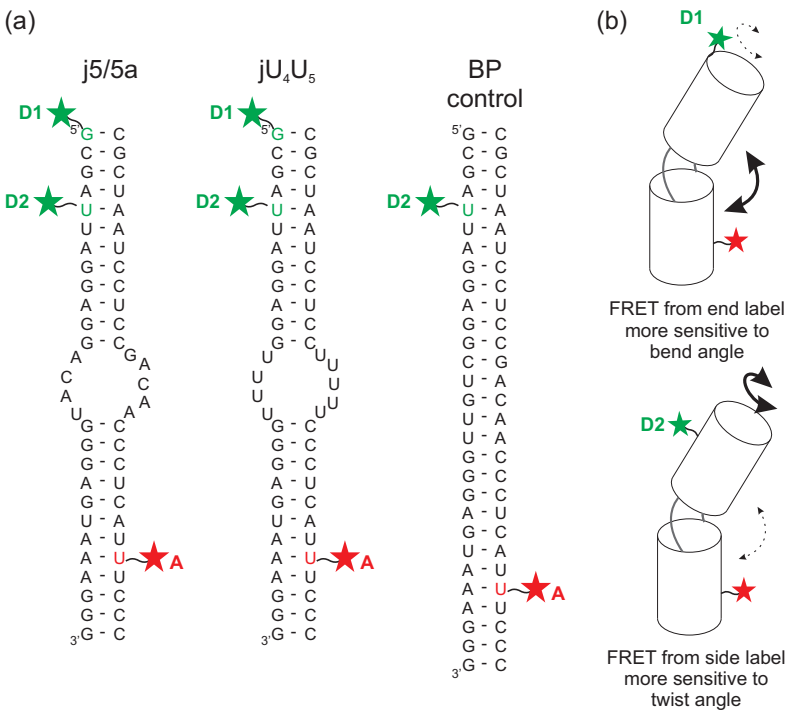

Figure 4.1: (a) Schematic of RNA constructs. Constructs consist of two RNA oligonucleotides which are individually labeled with a fluorophore, and annealed together to form a duplex. Green and red stars represent the dye label positions, which are achieved from either a 5′ C12-amino modifier (D1) or an amino-modified dT replacing the U-nucleotide in the sequence (D2 and A). (b) A cartoon representing the RNA junction constructs and the associated label positions. D1 labels are closer to the helical axis and are thus more sensitive to the overall bend angle, while D2 labels are off center and are more sensitive to relative twist.

4.2.3 Single-molecule FRET

On the day of each experiment, a fresh aliquot of sample was thawed, diluted, then annealed in a water bath for 2 minutes at 90°C after which the heat was turned off and the heat block was cooled slowly over 50 minutes. Samples were then diluted by a factor of at least 1000 into the desired measurement buffer. smFRET measurements were performed as described previously [15]. In this

case we calculated $E_{FRET} = \frac{I_A}{I_A+\gamma I_D}$, where I_A and I_D are the fluorescence measured in the acceptor and donor channels, corrected for bleedthrough (β) and background. The factor γ accounts for the difference in quantum yields and detection efficiencies of donor and acceptor fluorophores. For our microscope, $\beta = 0.027$ and $\gamma = 1.2$. Histograms were fit with three gaussians. The zero-FRET peak represents molecules with an absent or inactive acceptor. The high-FRET peak is relatively insensitive to salt, and is present even in the BP control, thus it is likely caused by a small population of dyes in a confined or restricted environment. The peak of interest is the mid-FRET peak whose position responds to changes in salt concentration. Each measurement was repeated twice and the errorbars represent the combination of the 95% confidence interval for the fit parameters and the standard deviation of the two independent measurements.

4.2.4 Accessible volume simulations

Software developed by Muschielok et al. [20] was used to estimate the accessible volume (AV) of the fluorophores for the BP control molecule. The software inputs are dye and linker parameters, a pdb structure of the molecule being labeled, and the attachment point of the dye linker on the pdb structure. The program computes the space which the dye can explore without steric clashes with the input molecule. We used the Nucleic Acid Builder web server [21] to build an A-form RNA helix with the same sequence as the BP control in Figure 4.1. The dye parameters and the procedure used were the same as those used in references [15, 22]. Using the AV simulations, we calculated the average expected E_{FRET} assuming fast dye reorientation, but slow conformational dynamics and compared it to the measured E_{FRET} using the appropriate cor-

rections (β and γ). We also compare data and simulations using the computed FRET-averaged distance $R_{FRET} = R_0(1/E_{FRET} - 1)^{\frac{1}{6}}$, using $R_0 = 52$Å for this dye pair [22].

4.3 Results

4.3.1 End- and side-labeled junctions

Figure 4.2 compares the E_{FRET} of j5/5a and jU_4U_5 junctions reported by the D1-A dye pair as a function of salt concentration. For this dye pair, the donor fluorophore is attached to the end of one RNA helix. Since the donor is closer to the helical axis, it is less sensitive to the twisting angle between the two helices compared to the D2-A dye pair and thus most sensitive to the bend angle as indicated in Figure 4.1(b). In solutions containing KCl (Figure 4.2(a)), only subtle differences are detected between the E_{FRET} of the two sequences, suggesting that the intra-dye distances are similar for the two molecules. In contrast, the E_{FRET} for j5/5a is consistently larger than that of the jU_4U_5 in solutions containing $MgCl_2$ (Figure 4.2(b)). Furthermore, it is intriguing that the E_{FRET} for the j5/5a increases with salt concentration, while for the jU_4U_5 it appears relatively insensitive. This trend is visible in KCl, but is exaggerated in $MgCl_2$.

To complement the D1 donor position measurements, we also measure the E_{FRET} for each RNA sequence using the D2 donor position (illustrated in Figure 4.1(a)). Whereas the D1 position keeps the donor close to the helical axis, the D2 donor position offsets the dye from the helical axis, increasing the sensitivity of the D2-A dye pair to the relative twist angle between the two helices. The

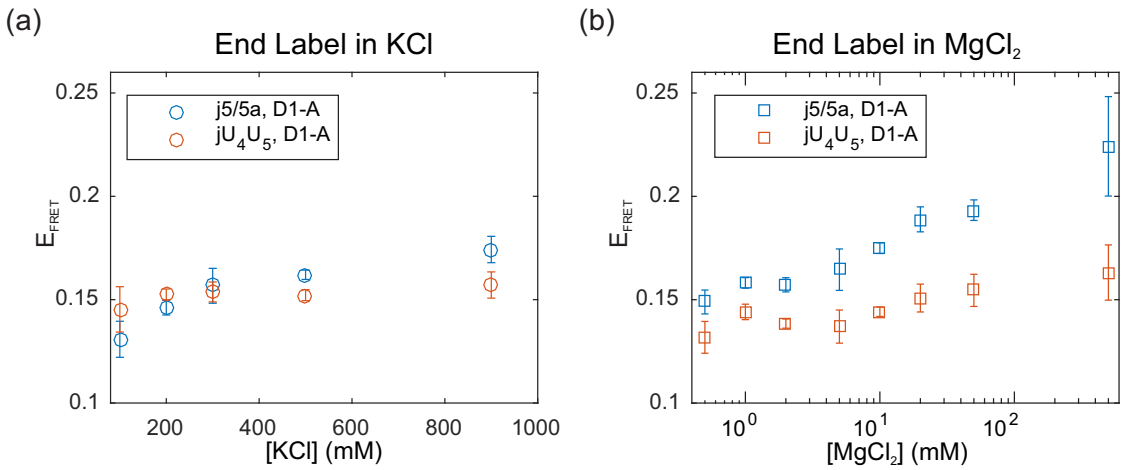

Figure 4.2: Measured effects on E_{FRET} of end-label donor dye position (D1-A dye pair from Figure 4.1) as a function of (a) [KCl] and (b) [$MgCl_2$].

E_{FRET} data for the D2-A labeled j5/5a and jU_4U_5 junctions are shown in Figure 4.3. Where the D1-A dye pair showed similar conformations between the j5/5a and jU_4U_5 junctions in KCl, the D2-A dye pair instead shows that the dyes are consistently closer together for the j5/5a sequence. Additionally, the difference between the two sequences increases with KCl concentration. In $MgCl_2$, the E_{FRET} values also increase with concentration, and the sequence dependence is further exaggerated. The difference between sequences is again larger in $MgCl_2$, consistent with the D1-A label observations. For both the D1-A and D2-A labeled RNA junctions, it is intriguing to note that in KCl, the j5/5a appears more sensitive to salt concentration than the jU_4U_5. In $MgCl_2$, there is a large offset between the D2-A labeled sequences, however the difference in salt dependence is less obvious.

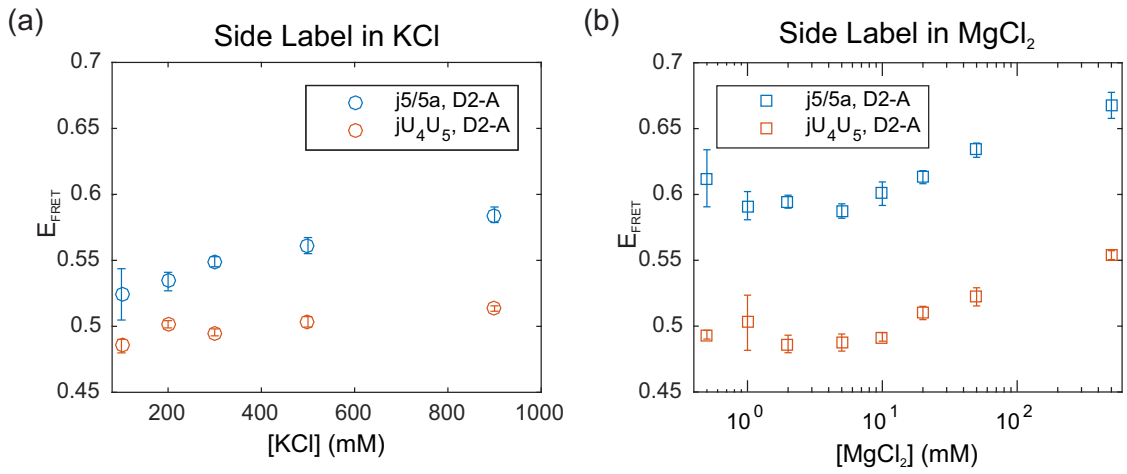

Figure 4.3: Measured effects on E_{FRET} of side-label donor dye position (D2-A dye pair from Figure 4.1) as a function of (a) [KCl] and (b) [$MgCl_2$].

4.3.2 Comparison with BP control molecule

Both Alexa 488 and Cy5 dyes are negatively charged and are connected to the RNA by flexible linkers. One caveat that we have not yet considered is the possibility of a changing dye environment with salt concentration. To address this concern, we measured the E_{FRET} for the D2-A dye pair with a fully base-paired RNA helix, for which we expect no structural changes with salt concentration. We converted the results to the FRET-averaged distance R_{FRET} to gain insight into the lengthscales involved in these changes. Figure 4.4 (a) and (b) show R_{FRET} as a function of KCl and $MgCl_2$ concentration, respectively. Note that the data for the j5/5a and jU_4U_5 are the same as shown in Figure 4.3, but converted to R_{FRET}. (The corresponding conversion for the D1-A label positions can be found in Appendix D Supplementary Information Figure D.1.) Each sequence displays distinct R_{FRET} values with the BP control having the largest distance between fluorophores, followed by the jU_4U_5 and the j5/5a sequences. Further-

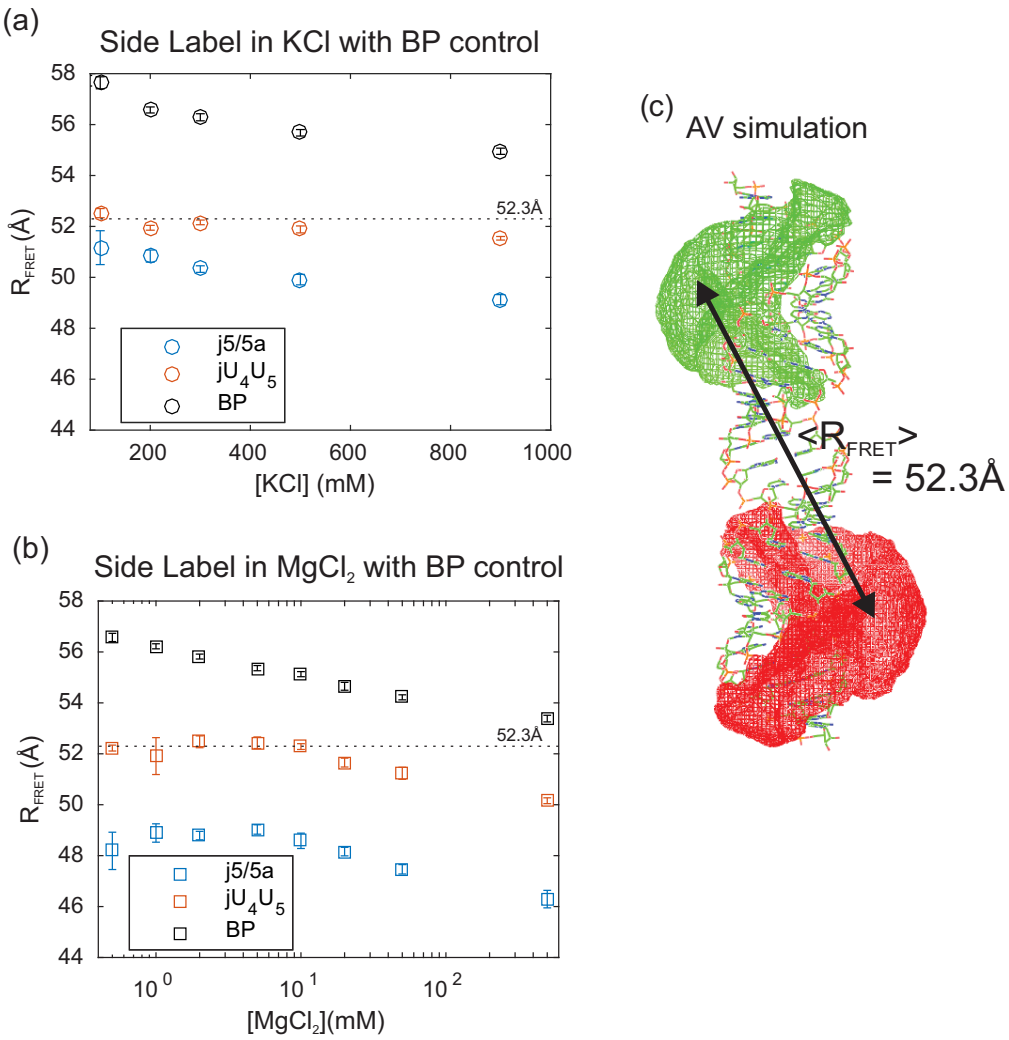

Figure 4.4: Comparison between R_{FRET} for BP control and RNA junctions in (a) KCl and (b) $MgCl_2$. (c) Representative accessible volumes of donor (green) and acceptor (red) dyes around the BP RNA control. Expected R_{FRET} from the simulation is 52.3Å (dotted line).

more, each sequence displays an overall decrease in R_{FRET} with salt, including the BP control. Results from the AV simulations for the BP control predict an R_{FRET} of 52.3 Å (Figure 4.4(c)). It is interesting to note that the this value is only approached in extremely high concentrations of $MgCl_2$ (Figure 4.4 (a) and (b) dotted line).

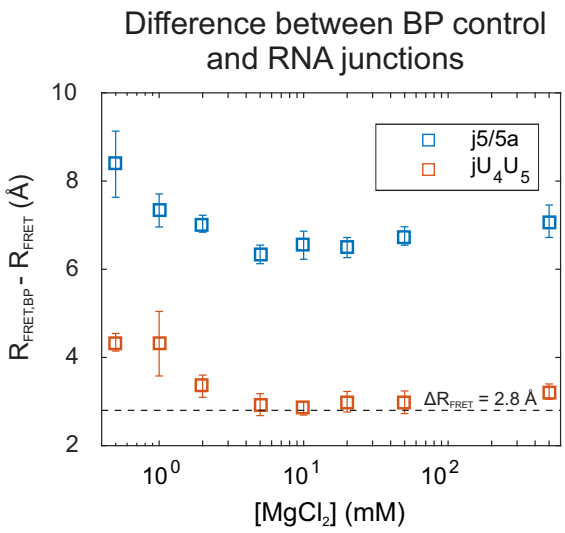

Figure 4.5: Difference between R_{FRET} for the BP control and RNA junctions j5/5a and jU_4U_5 in $MgCl_2$. The salt dependence of the jU_4U_5 matches that of the BP control at and above 5mM $MgCl_2$, and is within error of 2.8 Å or one base pair difference.

Since the change in R_{FRET} in $MgCl_2$ for the BP control is of comparable order to the changes observed for j5/5a and jU_4U_5, we next compare the difference between them in Figure 4.5. (The difference between the junctions and the control in KCl solutions is shown in Supplementary Information Figure D.2.) It is interesting to note that the difference between the R_{FRET} of jU_4U_5 compared to the BP control stabilizes around 2.8Å above 5 mM $MgCl_2$. A difference of 2.8Å corresponds to the change in length of an A-form helix from one extra base pair. The j5/5a, in turn, has a much larger difference between it and the BP control, and also does not stabilize at high salt.

4.4 Discussion

We first consider the unexpected salt-dependence of the BP control that is apparent in Figure 4.4, as this has implications for the rest of our interpretation. Next we will provide a simple model to explain the bending and twisting of j5/5a and jU_4U_5. Finally, we will discuss the applications of our model system to P4-P6 and RNA folding in general.

4.4.1 Salt-dependence of the BP control

Despite our expectation that the BP control would have no conformational changes, we detect a change in R_{FRET} comparable to that of the other RNA junctions (Figure 4.4). There are two potential explanations for the observed decreasing trend in R_{FRET}. One possibility is that the RNA double helix is shortened with salt, rather than behaving as a rigid rod. Recent work showed that the presence of cobalt(III) hexammine may distort and compress the RNA helix [23], although there is no evidence for the same effect caused by Mg^{2+} or K^+ ions. A second likely explanation for the salt-dependence of the BP control is that the fluorophore environment is affected by the salt concentration. Both Alexa Fluor 488 and Cy5 fluorophores are negatively charged. Since R_{FRET} for the BP control approaches the AV-predicted R_{FRET} at high $MgCl_2$ concentrations, it is possible that electrostatic effects restrict the volume accessible to the fluorophores to a subset of states obtained from the simulations. Repulsion between the RNA phosphate backbone and the fluorophores could restrict their environment at low salt. AV simulations would not capture this type of behavior because they neglect any interactions between the fluorophores and the RNA other than steric

clashes. We predict that repulsion would restrict the dyes to the environment around the ends of the helix at low salt. Increasing the salt concentration may allow the dyes to more closely approach the RNA helix, ultimately allowing them to move inwards and decreasing their average separation. While we did not measure the D1-A dye pair for the BP control, we speculate that this type of effect would still be an issue for the D1 donor position, such that the fluorophore is restricted to larger separation at low salt.

The BP control can provide us specific insights into the structure of the jU_4U_5 mutant. Figure 4.5 shows that the difference between jU_4U_5 and the control approaches 2.8 Å above 5 mM $MgCl_2$. This is consistent with a picture where jU_4U_5 is fully paired and is roughly coaxial. Because jU_4U_5 is asymmetric, a fully paired junction would only have four base pairs, which is one fewer than the control. This observation may be evidence that Mg^{2+} stabilizes base stacking interactions within the junction.

Despite the salt-dependence of the BP control, we can still directly compare the E_{FRET} for different RNA sequences. Because the dyes are attached to identical flanking helices, the dye environment is exactly the same for each RNA construct, thus the salt dependence does not depend on the specific junction sequence. Therefore, differences in E_{FRET} between the j5/5a and jU_4U_5 represent sequence-dependent RNA conformations. We must exercise caution in interpreting the trends observed with increasing salt concentration, as these include the dye effects in addition to junction effects.

4.4.2 RNA junction bend and twist angles are determined by sequence and salt concentration

By combining the information from two distinct labeled RNAs, we can start to build a picture of how sequence affects RNA junction conformation. Having two separate donor label positions allows us to roughly distinguish twisting from bending of the junction. In increasing KCl, E_{FRET} from the D1 position reports a comparable conformation between the j5/5a and jU_4U_5, while the D2 positions are different, consistent with a picture where the bend angle is the same, but each sequence has a distinct relative twist. As the KCl concentration is further increased, the difference between sequences grows. In increasing amounts of $MgCl_2$, the bend angle of the j5/5a increases more dramatically than jU_4U_5. Additionally, even in the smallest concentration of $MgCl_2$ measured (0.5 mM) there is a larger difference between twist angles than seen even in 900mM KCl. These results show the pronounced effect of Mg^{2+} ions on the conformation of the j5/5a junction.

In Figure 4.6, we present a model of our results, focusing on solutions containing $MgCl_2$ since the sequence effects are exaggerated relative to KCl. At low salt (0.5 mM $MgCl_2$), j5/5a is more twisted than jU_4U_5, however the bend angle is similar. With increasing $MgCl_2$ concentration, the difference between sequences is amplified. The j5/5a junction is very sensitive to $MgCl_2$, and both bend and twist angles increase. In contrast, Mg^{2+} ions cause jU_4U_5 to adopt a completely stacked structure.

Stacking of jU_4U_5 in high concentrations of $MgCl_2$ is consistent with the work reported in reference [17], suggesting that junction sequence serves to

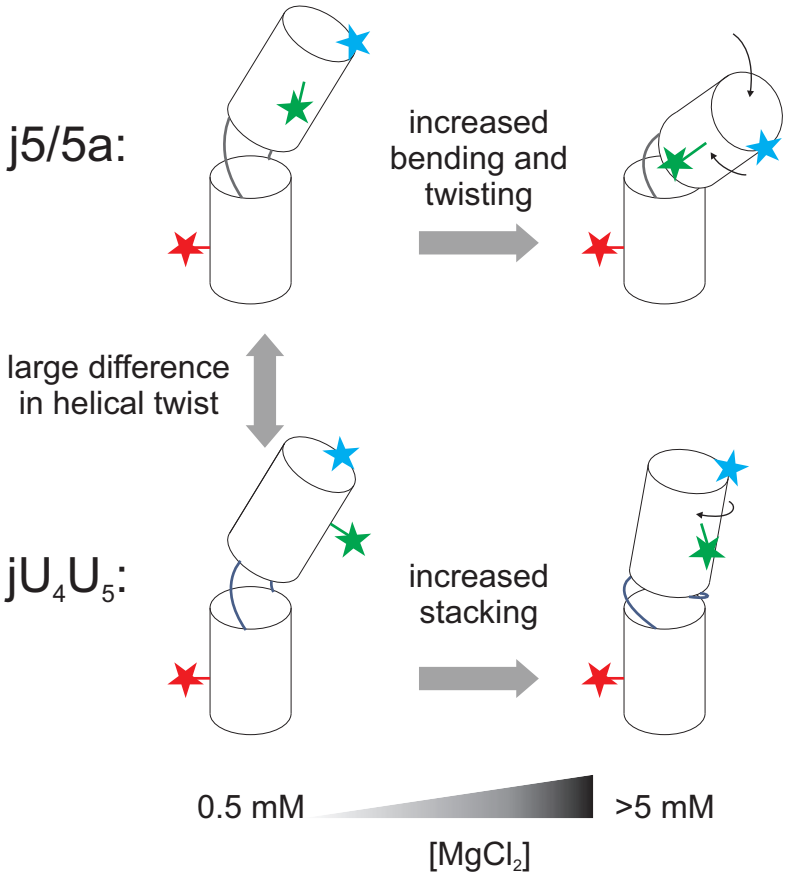

Figure 4.6: Schematic representing salt concentration and sequence effects of RNA junction conformations. At 0.5 mM $MgCl_2$, j5/5a and jU_4U_5 have different helical twists as shown by the side label E_{FRET} data, but similar bend as shown in the end label E_{FRET} data. As [$MgCl_2$] is increased to above 5 mM, the difference between sequences also increases, demonstrating an interplay between sequence and salt dependence.

maximize the number of non-canonical basepairs. The U-U mismatch (along with the G-A mismatch) is more stable than other non-canonical base pairs [24], and may therefore be more likely to continue the A-form RNA helix. The most likely junction structures as predicted by MC-fold [25] suggest that all but one U nucleotide in jU_4U_5 is paired, while three nucleotides are unpaired in j5/5a. The specific base-pairing patterns may describe why there is such a big difference in twist angle between the two sequences, consistent with [17]. However, this pic-

ture does not describe the salt-dependence of the junction conformations.

We can think of two possible explanations for the different salt dependencies of j5/5a and jU_4U_5. Mg^{2+} ions may bind to specific sites in the j5/5a junction stabilizing a larger bend. Juneau et al. [26] suggest that a Mg^{2+} ion binding site adjacent to the j5/5a within P4-P6 may help to stabilize the bend, however, our isolated j5/5a construct does not contain all the nucleotides coordinated to the Mg^{2+} ion in the P4-P6 crystal structure. Another explanation for the salt-dependence can arise from base stacking propensity of single- and double-stranded RNA. Measurements by Zacharias et al. [27] showed that the bend angle within a duplex containing bulged U nucleotides was reduced significantly in 2mM Mg^{2+}. In contrast, the bend angles for A bulges were only marginally affected by Mg^{2+}. They propose that the U bulges can 'loop out' allowing the flanking RNA helices to stack, while the unpaired A nucleotides stack with themselves, preventing the flanking helices from stacking with each other. The combination of our results and those of Zacharias et al. points towards the salt dependence of base stacking as a driver in RNA junction conformation. While single stranded poly(U) RNA has a low base stacking propensity, the U-U mismatch may bias the RNA towards an A-form helix. The base stacking propensity of the unpaired nucleotides may then determine whether they will 'loop out' letting adjacent helices to stack or to 'loop in' causing a bend. Most importantly, our results show that this stacking can be tuned by the salt concentration. Our results also show a KCl-dependent RNA junction conformation, highlighting that the effect occurs in the presence of both mono- and divalent ions.

4.4.3 Insights into P4-P6 folding

The j5/5a junction acts as a hinge allowing two halves of the P4-P6 domain in the *Tetrahymena* ribozyme to come together to form tertiary contacts. Mutating the sequence of the junction can alter the folding properties of P4-P6 [14, 28]. By isolating the junction in our smFRET experiments we were able to separate the effects of junction structure from tertiary contact formation. Szewczak et al. mutated the j5/5a junction to a poly(U) analogue and found that a higher Mg^{2+} concentration was required for folding [14]. The insights from our study on isolated j5/5a and jU_4U_5 in varying salt conditions provide a window into the mechanism for this behavior. Consistent with Szewczak, we found that the jU_4U_5 system is both less compact and more resistant to changes in salt than the j5/5a wild type. Furthermore, a fully base-paired jU_4U_5 junction in the presence of Mg^{2+} can explain why the U-mutant resists folding within P4-P6. The j5/5a junction in turn, is more bent, and may better position the two sides of the tertiary contacts when Mg^{2+} is added [10].

The U mutant is also more likely to fold than a P4-P6 mutant with a junction containing complementary base pairs [14]. Thus the U-U mismatches may be more dynamic and less stable than canonical pairs. Non-canonical base pairs are important for determining RNA structure and for protein recognition [29, 30]. Little is known about the effects of salt on the stability of non-canonical base pairs and further research into these effects would be extremely valuable.

More broadly, our measurements highlight the importance of both salt concentration and sequence in determining the conformation of RNA during folding. Salt and sequence are not independent factors, but there is a coupling between their effects on RNA structure. By studying isolated RNA junctions with

smFRET, our method provides a new window into quantifying this interplay. Moving forward, exploring the salt-dependence of non-canonical base pair stability and base stacking interactions will be essential for understanding and predicting RNA folding pathways.

Acknowledgement

The authors would like to thank the Pollack lab members George Calvey, Yujie Chen, Andrea Katz, Alex Mauney, Suzette Pabit, Alex Plumridge and Josh Tokuda for useful discussions. This work was supported by the National Institutes of Health [R01-GM085062].

References

[1] P. B. Moore, "The rna folding problem," in *The RNA world*, Cold Spring Harbor Laboratory Press, 1999, pp. 381–401.

[2] R. Schroeder, A. Barta, and K. Semrad, "Strategies for rna folding and assembly," *Nature reviews molecular cell biology*, vol. 5, no. 11, pp. 908–919, 2004.

[3] H. M. Al-Hashimi and N. G. Walter, "Rna dynamics: it is about time," *Current opinion in structural biology*, vol. 18, no. 3, pp. 321–329, 2008.

[4] S. V. Solomatin, M. Greenfeld, S. Chu, and D. Herschlag, "Multiple native states reveal persistent ruggedness of an rna folding landscape," *Nature*, vol. 463, no. 7281, pp. 681–684, 2010.

[5] P. P. Zarrinkar and J. R. Williamson, "The kinetic folding pathway of the tetrahymena ribozyme reveals possible similarities between rna and protein folding," *Nature Structural Biology*, vol. 3, pp. 432–438, 1996.

[6] S. K. Silverman, M. L. Deras, S. a. Woodson, S. a. Scaringe, and T. R. Cech, "Multiple folding pathways for the p4-p6 rna domain.," *Biochemistry*, vol. 39, no. 40, pp. 12 465–12 475, 2000.

[7] R. Russell, I. S. Millett, M. W. Tate, L. W. Kwok, B. Nakatani, S. M. Gruner, S. G. J. Mochrie, V. Pande, S. Doniach, D. Herschlag, and L. Pollack, "Rapid compaction during rna folding," *Proceedings of the National Academy of Sciences*, vol. 99, no. 7, pp. 4266–4271, 2002.

[8] R. Das, L. W. Kwok, I. S. Millett, Y. Bai, T. T. Mills, J. Jacob, G. S. Maskel, S. Seifert, S. G. J. Mochrie, P. Thiyagarajan, S. Doniach, L. Pollack, and D. Herschlag, "The fastest global events in rna folding : electrostatic relax-

ation and tertiary collapse of the tetrahymena ribozyme," *Journal of molecular biology*, vol. 332, no. 2, pp. 311–319, 2003.

[9] J. C. Schlatterer, L. W. Kwok, J. S. Lamb, H. Y. Park, K. Andresen, M. Brenowitz, and L. Pollack, "Hinge stiffness is a barrier to rna folding," *Journal of Molecular Biology*, vol. 379, no. 4, pp. 859–870, 2008.

[10] V. B. Chu, J. Lipfert, Y. Bai, V. S. Pande, S. Doniach, and D. Herschlag, "Do conformational biases of simple helical junctions influence rna folding stability and specificity?" *RNA*, vol. 15, no. 12, pp. 2195–2205, 2009.

[11] J. L. Fiore, B. Kraemer, F. Koberling, R. Edmann, and D. J. Nesbitt, "Enthalpy-driven rna folding : single-molecule thermodynamics of tetraloop-receptor tertiary interaction," *Biochemistry*, vol. 48, no. 11, pp. 2550–2558, 2009.

[12] M. H. Bailor, X. Sun, and H. M. Al-Hashimi, "Topology links rna secondary structure with global conformation, dynamics, and adaptation," *Science*, vol. 327, no. 5962, pp. 202–206, 2010.

[13] S. E. Butcher and A. M. Pyle, "The molecular interactions that stabilize rna tertiary structure: rna motifs, patterns, and networks," *Accounts of Chemical Research*, vol. 44, no. 12, pp. 1302–1311, 2011.

[14] A. A. Szewczak and T. R. Cech, "An rna internal loop acts as a hinge to facilitate ribozyme folding and catalysis," *RNA*, vol. 3, no. 8, pp. 838–849, 1997.

[15] J. Sutton and L. Pollack, "Tuning rna flexibility with helix length and junction sequence," *Biophysical Journal*, vol. 109, no. 12, pp. 2644–2653, 2015.

[16] M. H. Bailor, A. M. Mustoe, C. L. Brooks, and H. M. Al-Hashimi, "Topological constraints: using rna secondary structure to model 3d conformation, folding pathways, and dynamic adaptation," *Current Opinion in Structural Biology*, vol. 21, no. 3, pp. 296–305, 2011.

[17] A. M. Mustoe, M. H. Bailor, R. M. Teixeira, C. L. B. 3rd, and H. M. Al-Hashimi, "New insights into the fundamental role of topological constraints as a determinant of two-way junction conformation," *Nucleic acids research*, vol. 40, no. 2, pp. 892–904, 2012.

[18] A. Y. Keel, R. P. Rambo, R. T. Batey, and J. S. Kieft, "A general strategy to solve the phase problem in rna crystallography.," *Structure*, vol. 15, no. 7, pp. 761–772, 2007.

[19] J. H. Cate, R. L. Hanna, and J. A. Doudna, "A magnesium ion core at the heart of a ribozyme domain," *Nature structural biology*, vol. 4, no. 7, pp. 553–558, 1997.

[20] A. Muschielok, J. Andrecka, A. Jawhari, F. Brückner, P. Cramer, and J. Michaelis, "A nano-positioning system for macromolecular structural analysis," *Nature methods*, vol. 5, no. 11, pp. 965–971, 2008.

[21] T. J. Macke and D. A. Case, "Modeling unusual nucleic acid structures," in, vol. 682, American Chemical Society, 1997, pp. 379–393.

[22] S. Sindbert, S. Kalinin, H. Nguyen, A. Kienzler, L. Clima, W. Bannwarth, B. Appel, S. Muller, and C. A. Seidel, "Accurate distance determination of nucleic acids via forster resonance energy transfer: implications of dye linker length and rigidity," *Journal of the American Chemical Society*, vol. 133, no. 8, pp. 2463–2480, 2011.

[23] S. A. Pabit, A. M. Katz, I. S. Tolokh, A. Drozdetski, N. Baker, A. V. Onufriev, and L. Pollack, "Understanding nucleic acid structural changes by comparing wide-angle x-ray scattering (waxs) experiments to molecular dynamics simulations," *Under review*, 2016.

[24] S. Schroeder, J. Kim, and D. H. Turner, "G.a and u.u mismatches can stabilize rna internal loops of three nucleotides," *Biochemistry*, vol. 35, no. 50, pp. 16 105–16 109, 1996.

[25] M. Parisien and F. Major, "The mc-fold and mc-sym pipeline infers rna structure from sequence data," *Nature*, vol. 452, no. 7183, pp. 51–55, 2008.

[26] K. Juneau, E. Podell, D. J. Harrington, and T. R. Cech, "Structural basis of the enhanced stability of a mutant ribozyme domain and a detailed view of rna-solvent interactions," *Structure*, vol. 9, no. 3, pp. 221–231, 2001.

[27] M. Zacharias and P. J. Hagerman, "Bulge-induced bends in rna: quantification by transient electric birefringence.," *Journal of molecular biology*, vol. 247, no. 3, pp. 486–500, 1995.

[28] S. Matsumura, Y. Ikawa, and T. Inoue, "Biochemical characterization of the kink-turn rna motif," *Nucleic acids research*, vol. 31, no. 19, pp. 5544–5551, 2003.

[29] T. Hermann and E. Westhof, "Non-watson-crick base pairs in rna-protein recognition," *Chemistry & Biology*, vol. 6, no. 12, R335–R343, 1999.

[30] N. B. Leontis and E. Westhof, "Analysis of rna motifs," *Current Opinion in Structural Biology*, vol. 13, no. 3, pp. 300–308, 2003.

CHAPTER 5

THE ROLE OF ION VALENCE IN THE SUBMILLISECOND COLLAPSE AND FOLDING OF A SMALL RNA DOMAIN

Suzette A. Pabit[1,*], Julie L. Sutton[1,*], Huimin Chen[1] and Lois Pollack[1]

1. School of Applied and Engineering Physics, Cornell University, Ithaca, NY 14853, USA

*These authors contributed equally.

Reprinted with permission from Biochemistry, Volume 52, Issue 9, 2013, Pages 1539-1546. Copyright 2013 American Chemical Society.

Abstract

Following the addition of ions to trigger folding, RNA molecules transition from rigid, extended states to a compact ensemble. Determining the time scale for this collapse provides important insights into electrostatic contributions to RNA folding; however it can be challenging to isolate the effects of purely non-specific collapse, e.g. relaxation due to backbone charge compensation, from the concurrent formation of some tertiary contacts. To solve this problem, we decoupled non-specific collapse from tertiary folding using a single point mutation to eliminate tertiary contacts in the small RNA subdomain known as tP5abc. Microfluidic mixing with microsecond time resolution and FRET detection provides insight into the ionic strength dependent transition from extended to compact ensembles. Differences in reaction rates are detected when

folding is initiated by monovalent or divalent ions, consistent with equilibrium measurements illustrating the enhanced screening of divalent ions relative to monovalent ions at the same ionic strength. Ion-driven collapse is fast and a comparison of the collapse time of the wild type and mutant tP5abc suggests that site binding of Mg^{2+} occurs on submillisecond time scales.

5.1 Introduction

RNA plays important biological roles in translation, splicing and enzymatic/catalytic reactions [1]. A recent focus on the role of RNA in the control of gene expression indicates that RNA molecules can be exploited for biotechnology applications [2, 3]. Growing interest in the use of RNA aptamers and riboswitches as therapeutic and analytic agents [4] calls for a process of designing molecules based on insights from RNA folding kinetic mechanisms [5].

Structurally, RNA is a collection of short base-paired helices connected by non-base paired regions that include loops, bulges, hinges and junctions [6, 7]. Because the RNA backbone carries a high negative charge, strong repulsive electrostatic forces must be overcome for the molecule to fold. RNA folding is induced in vitro by the addition of ions. Crystal structures reveal a small number of site-bound ions in some RNAs [8]; however, the majority of counterions form a diffuse cloud around the macromolecule [9]. In low salt unfolded states, the helices repel and molecular conformations are extended. Following the addition of counterions to trigger folding, the backbone charge is more locally screened and the molecules relax to compact states. Recent equilibrium studies suggest that this 'electrostatic relaxation' is anisotropic; the junctions direct folding by

entraining helix motions along certain well-defined pathways [10, 11]. Native contacts can then form when the two sides of a tertiary contact come into close proximity. However, studies of short base-paired helices suggest that the negative duplex charge is not fully compensated on intramolecular length scales [12]. An outward electrostatic pressure opposes tight compaction in the absence of tertiary contacts, even at moderate to high ionic strength. Thus RNA folding is a balance between weakened, but non-negligible repulsive electrostatic forces and attractive forces, e.g. hydrogen bonding between the two sides of a tertiary contact. The primary goal of this study is to complement the increasing number of RNA folding and kinetic studies [13, 14] by focusing on the process of collapse upon addition of charge compensating ions. How does the rapid, initial collapse depend on the valence and concentration of counterions used to trigger it?

Previous small-angle X-ray scattering (SAXS) studies of the *Tetrahymena* ribozyme and selected mutants reveal a rapid compaction upon the addition of ions [15]. Concurrent time-resolved hydroxyl radical footprinting experiments show that the majority (but not all) of tertiary contacts in the molecule remain unformed within the time scale of rapid collapse. However, collapse occurred within the mixing dead times of those kinetic measurements so only an upper limit for the collapse time (milliseconds) was obtained. More extensive time-resolved SAXS studies of the collapse and folding of the *Azoarcus* ribozyme [16] were carried out to focus on this initial rapid collapse. This group I intron displayed heterogeneous folding kinetics when folding was initiated by Mg^{2+} [16]. Some subpopulations collapse rapidly with tertiary contacts formed, others undergo non-specific collapse before slower structural rearrangements can occur. Thus, the millisecond time scales reported for this system do not distinguish

pure non-specific collapse due to charge compensation from specific collapse. Other attempts to measure pure electrostatic collapse in simplified systems were obscured by the presence of a stiff hinge joining two helical domains that precluded relaxation to a compact ensemble [17].

An experiment to measure the time scale of ion-mediated electrostatic collapse in RNA requires a clear delineation between non-specific collapse (purely electrostatic driven) and specific collapse (containing native or non-native tertiary contacts). We accomplish this by choosing a molecule that collapses but is incapable of forming tertiary contacts: the A186U mutant of the tP5abc subdomain of the *Tetrahymena* ribozyme. In this well-characterized construct, mutation of a specific contact residue for Mg^{2+} binding [18] prevents secondary structure rearrangement, which is a prerequisite for the final steps of tertiary folding [19]. Nuclear magnetic resonance (NMR) studies have shown that the A186U mutant has an extended structure similar to that of wild type tP5abc, but does not undergo tertiary folding [20], therefore it is used as a control for unfolded tP5abc or the P4-P6 domain [21–23]. A recent publication focusing on the later folding steps of wild type tP5abc speculates about the tight connection between secondary structure rearrangement and tertiary contact formation, suggesting that they occur concomitantly [23]. Use of the A186U mutation ensures that we probe only non-specific collapse without the participation of tertiary contacts, allowing us to elucidate the nature of structures within the rapidly formed, compact ensemble.

Here, we describe the dependence of the earliest events in RNA folding on ion type, valence and concentration, using the A186U mutation in tP5abc RNA to monitor the formation of the collapsed state. Fluorescence correlation spec-

troscopy (FCS) measurements verify that this construct collapses to a compact state in high salt solutions. To establish the electrostatic contributions of different ions in facilitating collapse, we quantify the strength of electrostatic interactions using second virial coefficients (A_2) of short RNA duplexes determined by SAXS. Since theoretical studies suggest that ion-driven non-specific collapse in RNA should occur in submillisecond time scales [24, 25], we use a microfluidic mixer with microsecond time resolution [26] combined with Förster resonance energy transfer (FRET) detection in a confocal microscope to measure the kinetics of RNA collapse due to a rapid jump in salt concentration. Between solutions with the same counterion valence, we measure collapse times that decrease as the ionic strength increases, reflecting the increasing entropy of the collapsed state with ionic strength as suggested by equilibrium measurements [27]. Near physiological monovalent ionic strengths, the initial non-specific collapse is fast and comparable to the dead time of our rapid mixing microfluidic device ($\sim$ 200 μs). When Mg^{2+} is used to initiate folding, collapse times are faster in the wild type than in the mutant, suggesting that site binding of Mg^{2+} occurs within the first millisecond of folding, faster than previously implied using techniques with slower time resolution [28, 29]. Our data suggest that specific contact formation accelerates the earliest steps of folding, and 'directs' folding along specific pathways.

5.2 Materials and Methods

5.2.1 Materials

RNA molecules were synthesized, desalted and purified by Dharmacon RNAi Technologies (Lafayette, CO). The sequence of the 25 base pair (bp) double-stranded RNA used for second virial coefficient measurements is GCA UCU GGG CUA UAA AAG GGC GUCG as in previous studies [30]. We used the truncated P5abc (tP5abc) construct and its A186U mutant described by Wu and Tinoco [19] for fluorescence correlation spectroscopy (FCS) and rapid mixing experiments. We have added a uracil to the 5′ end in order to prevent interactions between the donor and the nearby G-C base pair. The 5′ end was labeled with the donor fluorophore (fluorescein) while the A171 nucleotide in the P5c stem-loop was labeled with the acceptor fluorophore (Dy547, with spectral characteristics similar to Cy3). The positions of the labels were chosen such that a change in Förster resonance energy transfer (FRET) is measurable between the extended and collapsed states of tP5abc. The tP5abc RNA molecules were stored in 50 mM K^+-MOPS buffer, pH 7, with 0.1 mM EDTA. Prior to measurements, they were annealed in 20 mM K^+-MOPS buffer, pH 7, with 8 mM EDTA at 50°C for 5 minutes then slowly cooled to room temperature. The molecules were buffer-exchanged to 20 mM K^+-MOPS buffer, pH 7 for all experiments and different salt solutions were added to change ion conditions. Monovalent and divalent ion solutions were prepared from Chloride salts unless stated otherwise. All chemicals were purchased from Sigma (St. Louis, MO). Rotational motion of the dye labels were investigated using fluorescence anisotropy measurements. Results are described in the Supplementary Information (Appendix

E) and shown in Figure E.1.

5.2.2 Second virial coefficient measurements

The strength of intermolecular interactions between 25 bp RNA duplexes was assessed by extracting the second virial coefficient (A_2) from SAXS profiles measured as a function of RNA concentration at varying cation valence and salt concentrations. A_2 is a measure of the interaction potential between short RNA helices in solution and takes into account the contributions from excluded volume, electrostatic repulsion and interhelical attraction [12, 30–32]. The SAXS experiments to measure A_2 were carried out at the Cornell High Energy Synchrotron Source (CHESS) and are described extensively in previous publications [12, 30].

5.2.3 Equilibrium Fluorescence measurements

FCS was used to measure the changes in diffusion times of tP5abc under different ionic conditions. Molecules in the collapsed state are more compact and diffuse faster than molecules in extended and unfolded configurations. The standard confocal setup described by Chen et al. [33] used a 488 nm laser for fluorescein excitation. Data processing and fitting were performed with Origin software (Microcal, Northampton, MA). Equilibrium FRET was calculated from fluorescence spectra taken with a Cary Eclipse Fluorescence Spectrophotometer (Varian Inc., Australia). The effective efficiency of energy transfer (E_{FRET}) was calculated as $A/(A + D)$, where A and D are the peak intensities of the acceptor

and donor emission, respectively.

5.2.4 Rapid Mixing to measure RNA collapse

We used a microfluidic mixer to rapidly change the ionic conditions of the tP5abc RNA [26, 34]. The mixer uses hydrodynamic focusing [35, 36] to facilitate fast diffusion of ions into the RNA sample. The experimental conditions are described in the Supplementary Information with a schematic of the device shown in Figure E.2. The mixing dead time for K^+ and Rb^+ ions was 140 μs, and that for Mg^{2+} and Sr^{2+} ions was 235 μs. Additional details of the manufacture, characterization and use of the mixing device were published in previous work [26]. Devices were fabricated at the Cornell NanoScale Science and Technology Facility (CNF).

5.3 Results

5.3.1 Dependence of charge screening on ion valence for RNA duplexes

To characterize how ions affect charge screening interactions between RNA helices, we used SAXS to measure A_2 between 25 bp RNA duplexes as a function of ion type and ionic strength. End effect contributions to A_2 are minimal when duplex lengths exceed ~25 bp [32], allowing us to focus on how ions affect screening lengths measured from the cylindrical axis of the double helix.

We find that the charge screening efficiency depends on ion valence but not ion type (Figure 5.1A). Measured values of A_2 in Na^+, K^+ and Rb^+ agree within experimental error while values of A_2 in Mg^{2+} and Sr^{2+} are indistinguishable. Furthermore, these results are consistent with previous experiments showing that divalent ions provide more efficient charge screening than monovalent ions at the same ionic strengths [9, 27, 37–41]. A cartoon illustrating the ionic strength dependence of charge screening efficiency is shown in Figure 5.1B. Since A_2 can be expressed in units of volume [31], we associate an electrostatic excluded volume with each duplex at every ionic strength, denoted by the shaded balloon in the figure. At the lowest ionic strength, where A_2 has its largest value and intermolecular repulsion dominates, this excluded volume is large. As ionic strength increases and the duplex negative charge is more efficiently screened, both A_2 and the excluded volume decrease, allowing the helices to approach each other more closely. At even higher ionic strengths A_2 becomes negative and interhelical attraction consistent with end-to-end stacking is observed [12, 30]. In the cartoon representation (Figure 5.1B) for $A_2 < 0$, the electrostatic excluded volume is small enough to allow other interaction forces to take effect. The A_2 measurements shown in Figure 5.1A enable a quantitative comparison of the screening effectiveness of different ionic solutions. Clearly, ionic strength alone is insufficient to explain the RNA charge screening efficiency of ions with different valences, but comparison between ions of the same valence display similar electrostatic effects.

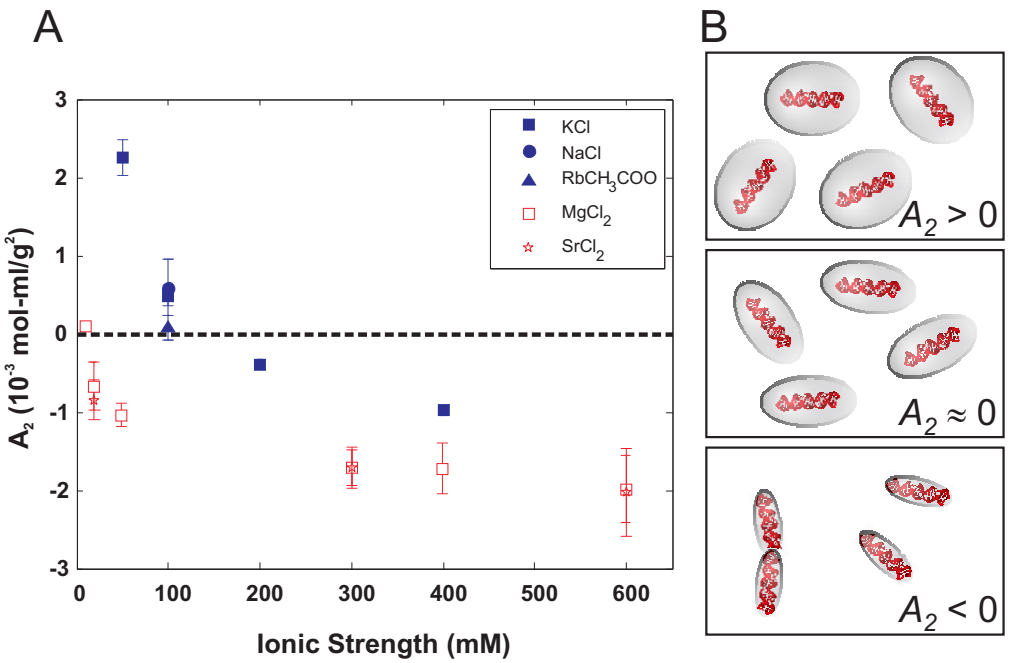

Figure 5.1: We use the second virial coefficient (A_2) to quantify the ionic strength dependent intermolecular interaction potential of 25 bp RNA helices. Positive values of A_2 denote intermolecular repulsion while negative values of A_2 can reflect duplex association via end-to-end stacking [12, 30] (A) A_2 measurements of 25 bp RNA duplexes obtained from SAXS data. As ionic strength increases the interhelix repulsion decreases due to greater screening of RNA charges by counterions. A_2 in solutions with monovalent cations (KCl (■), NaCl (•), and RbCH3COO (▲)) are indistinguishable, as are those with divalent cations (open symbols: $MgCl_2$ (□) and $SrCl_2$ (*)). However, for a given ionic strength, divalent ions are more effective in screening the charges than monovalent ions. (B) Cartoon illustrating the electrostatic excluded volume at increasing ionic strengths. When $A_2 > 0$, large repulsive forces between helices are represented by large electrostatic excluded volumes. When $A_2 \approx 0$, the RNA-RNA interactions are negligible. $A_2 < 0$ represents RNA association via end-to-end stacking [30].

5.3.2 Equilibrium fluorescence measurements of tP5abc

The tP5abc RNA consists of three helices (P5a, P5b and P5c) connected at a junction by single stranded regions as shown in Figure 5.2A [19]. At low salt concentrations the three helices of tP5abc are arranged in an extended state (Figure 5.2B). Increasing the bulk salt concentration reduces the repulsive forces among the three helices, therefore we expect the RNA to collapse to more compact conformations as salt is added. Since a more extended molecule has a larger hydrodynamic radius and diffuses more slowly, compaction is signaled by a reduction in diffusion time of the RNA.

We used FCS to monitor the A186U mutant, which has the same extended-state secondary structure as the wild type, but does not form tertiary contacts and thus does not fold to the native state [18, 19]. A single diffusion constant (τ_D) and two exponential time constants (τ_A, τ_B) were needed to adequately fit the FCS data. The presence of only one diffusion component indicates the absence of conformational states with substantially different diffusion times and thus different hydrodynamic radii. In addition to the characteristic diffusion time, two fast characteristic fluctuation time components are required to accurately fit the FCS data. We tentatively attribute these time scales to a combination of internal molecular motions leading to fluctuations in FRET, and dye photophysics. At increased ionic strengths, FCS data report a dramatic decrease in τ_D (Figure 5.2C and D). Because τ_D is proportional to the hydrodynamic radius of the molecule, the different values measured in the initial (low salt concentration, 9 mM K^+) buffer and the final solution (containing an additional 30 or 160 mM K^+ (Figure 5.2C) or 1, 3, or 25 mM Mg^{2+} (Figure 5.2D)) indicate collapse to a compact state. FCS measurements of wild type tP5abc show the same

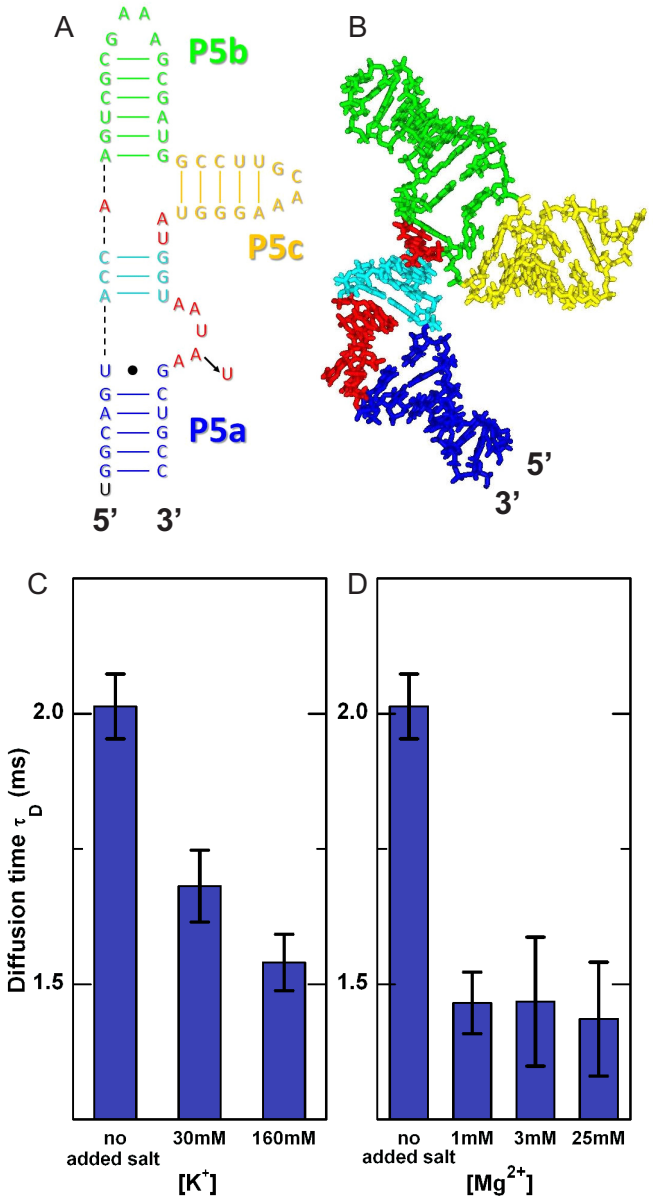

Figure 5.2: tP5abc collapses with increasing ionic strength. (A) The extended state secondary structure of tP5abc [19]. The A186U mutation is indicated with an arrow. (B) tP5abc extended state conformation in solution [20], shown with the same color scheme as in (A). (C, D) Diffusion times of A186U mutant obtained from FCS measurements in K^+ and Mg^{2+}, respectively. A dramatic decrease in τ_D is observed upon the addition of salt, demonstrating tP5abc compaction with the increase in bulk counterion concentration.

trends and comparable diffusion times as the mutant (Figure E.3, Supplementary Information); however, because of its lack of sensitivity FCS cannot further distinguish the folded (wild type) from the collapsed (mutant) states. Figure E.4 (Supplementary Information) shows the diffusion times of the mutant in solutions containing different ionic species. Similar to second virial measurements, we see no significant dependence of diffusion time on ion type.

Equilibrium FRET measurements were also performed at different ionic conditions. In our case, FRET cannot be used as a quantitative probe of the size of the molecule because fluorescence anisotropy measurements show that the internally labeled acceptor dye was not freely rotating in solution (see Supplementary Information for details). Figure E.5 (Supplementary Information) shows the effective equilibrium E_{FRET} of the tP5abc wild type and mutant in different ionic conditions. We see no significant dependence of E_{FRET} on monovalent ion type but there is a slight difference between Mg^{2+} and Sr^{2+} ions. Since FCS and A_2 measurements cannot detect this difference, it is possibly due to a change in the local dye environment during divalent ion aided collapse. Nonetheless, a normalized change in FRET efficiency still signals a conformational change upon addition of ions and can be used to compare time scales of collapse for each ion type. Additional equilibrium fluorescence control experiments are discussed in the Supplementary Information and shown in Figure E.6.

5.3.3 Measurement of collapse times at different ionic conditions

The equilibrium measurements discussed above suggest that tP5abc and its mutant, held in low salt concentration solution, become compact following the addition of ions. To gain insight into the energy landscape of collapse, we measured the collapse time (τ_c) of the A186U non-folding mutant under several different experiment conditions using microfluidic mixing [26]. Mixing of ions is facilitated by diffusion after hydrodynamic focusing [35, 36] as described in the Supplementary Information. Initially, RNA is equilibrated in a low salt buffer (9 mM K^+), where FCS indicates that it populates extended states. To focus on the collapse transition, we measured how the effective E_{FRET} changed in time when the A186U mutant was rapidly mixed with buffers at different ionic conditions. The signal was normalized by the measured E_{FRET} from a control experiment in which the RNA was mixed with the same low salt buffer as in its initial state. The resulting normalized E_{FRET} vs. time was fit to a single exponential decay with a constant offset to extract τ_c. Figure 5.3 shows representative, normalized kinetic traces at different concentrations of added divalent ions.

The time scale for collapse (Figure 5.4) displays a marked decrease with increasing ionic strength. At the highest ionic strengths measured, the collapse time is comparable to the dead times of the mixing instrument (indicated by cross hatched regions in Figure 5.4). To test its importance on ion type, we measured the collapse timescale in solutions containing 30 mM K^+ or 30 mM Rb^+ and 1 mM of Mg^{2+} or 1 mM Sr^{2+}. These lower ionic strengths enhanced our sensitivity to differences by increasing the separation between mixing and collapse time scales.

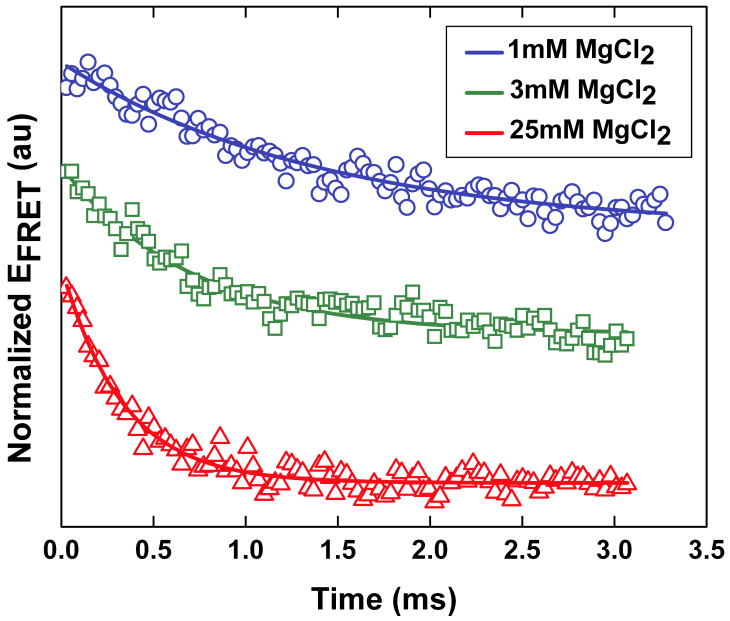

Figure 5.3: Typical normalized kinetic traces of the tP5abc A186U mutant after mixing with different Mg^{2+} concentrations. Solid lines represent single exponential fits to the data. Offsets have been added to the traces to separate them.

Results for the non-folding A186U mutant show no significant dependence of collapse time on ion type. Thus for both monovalent and divalent ions, non-specific collapse depends on the valence and concentration, but not the identity of the ion. Similarly, as we expect, there are no differences between the collapse times of the wild type and the mutant in monovalent salt solutions (Figure 5.4A).

Previous studies using hydroxyl radical footprinting have found folding times for the isolated P5abc domain to be slow, on the time scale of tens to hundreds of milliseconds [17, 42]. Other studies used stopped-flow fluorescence to measure a folding time of 240 ms for wild type tP5abc in 1 mM Mg^{2+} and show that secondary structure rearrangement is rate-limiting for the tertiary folding

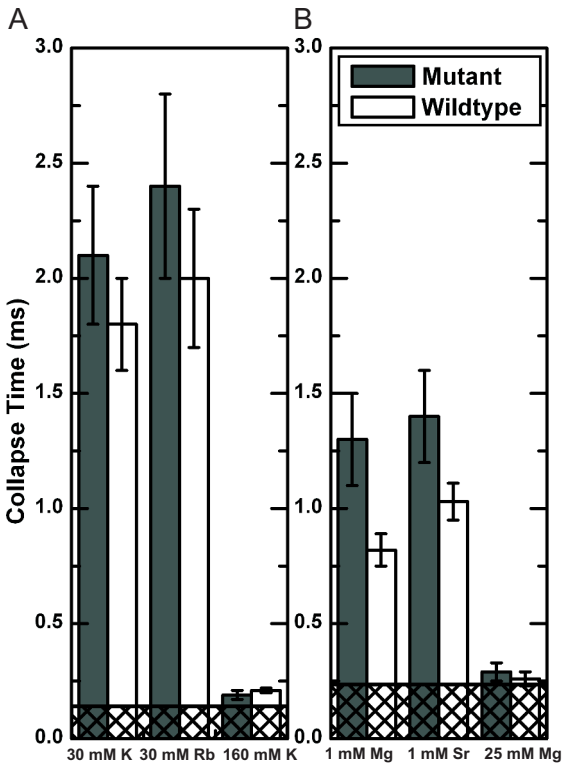

Figure 5.4: Collapse times of tP5abc when rapidly mixed with (A) monovalent and (B) divalent ions. Collapse of the mutant does not depend on ion type (30 mM K^+ and Rb^+ or 1 mM Mg^{2+} and Sr^{2+}). The wild type collapses faster than the mutant in 1 mM Mg^{2+}, suggesting early binding of Mg^{2+} to the specific binding site. Cross-hatched regions represent the mixer dead times (140 μs in (A) and 235 μs in (B)).

of wild type tP5abc [19, 20, 23]. Since our rapid mixing results gives us faster time resolution than the previous measurements cited, we are observing a transition of the extended wild type molecule to a compact set of states in which tertiary contacts and secondary structure rearrangement have not yet occurred.

To discern the importance of specific Mg^{2+} binding to rapid compaction, we compare the collapse time of the mutant with that of the wild type tP5abc. At 1 mM Mg^{2+}, the collapse time of the wild type is 0.82 ± 0.07 ms compared with 1.3 ± 0.2 ms for the mutant (Figure 5.4B). The small but significant difference

between collapse times of the wild type and the non-folding mutant suggests that something other than electrostatic screening stabilizes the collapsed state of the wild type. We also investigated the effect of ion type in collapse of wild type tP5abc and measured 1.03 ± 0.08 ms in 1 mM Sr^{2+}, which is slower than the collapse in 1 mM Mg^{2+} (0.82 ± 0.07 ms). The difference in collapse time of the wild type due to different divalent ions is small and just outside the error bars. While it is only in the wild type where we can confidently distinguish between collapse times due to ions of different types, we note the same general trend for both mutant and wildtype.

5.4 Discussion

5.4.1 Impact on collapse times of ion valence and type

Measurements of second virial coefficients of model systems detect minimal differences in intermolecular interactions based on ion type. At a given ionic strength, K^+, Rb^+ and Na^+ have identical second virial coefficients, as do Mg^{2+} and Sr^{2+} (Figure 5.1A). We also tested the effectiveness of various ions in inducing the collapse transition in the A186U mutant. No ion type effects were observed in non-specific collapse: τ_c does not distinguish 30 mM K^+ from 30 mM Rb^+, or 1 mM Mg^{2+} from 1 mM Sr^{2+}. However, these effects do not scale across solutions of different valence ions. The FCS, A_2 and previous persistence length measurements [38] confirm that charge screening of the RNA phosphate backbone is more efficient in divalent than monovalent ions. This is, in fact, what we observe in the kinetic measurements (see Figure 5.4).

From Debye-Hückel (DH) arguments alone, we expect non-specific electrostatic effects to be determined solely by ionic strength. However, the anomalous behavior of divalent relative to monovalent ions is a frequent theme in studies of nucleic acid electrostatics [22, 27, 37, 38, 43]. Consistent with literature reports, in the kinetic experiments reported here mixing with divalent ions leads to a shorter collapse time than mixing with monovalent ions at a given ionic strength. Although some studies may suggest that spatially correlated counterions can lead to enhanced screening of polyelectrolytes, these effects seem to be more evident when the counterion valence is greater than two [44, 45]. For the case of divalent ions, some insight can be gleaned from recent MD simulations of ions around RNA [46]. Notably, simulation results agree with DH models at large distances from the RNA: screening is a function of ionic strength at distances larger than 16 Å from the central axis of the molecule. However, to achieve this agreement requires adjustment of the surface charge density near the RNA surface. Close to the surface, both nucleic acid topologies as well as the distinctive features of counterions (and co-ions) are important. Mean field models are inapplicable because they cannot account for atomic detail *ab initio*. Our measurements suggest that the surface potential is smaller in the presence of divalent ions, due to their tighter localization and the larger degree of charge compensation. As a result, there is simply less charge to screen at larger distances where the mean field theory can be applied. In terms of the model shown in Figure 5.1B, the electrostatic excluded volume is smaller for divalent than monovalent ions at a given ionic strength, because the 'near charge' on the duplex is more completely neutralized. Thus differences result from the ions that are in closest proximity to the RNA surface.

5.4.2 Ionic strength dependence of collapse reflects increasing conformational entropy in the collapsed ensemble

Our results can be placed into the context of recent reports suggesting that tP5abc folding is a sequential process with at least two barriers [23]. While the previous work focused on understanding the transition from the collapsed intermediate to the folded state [23], our use of a submillisecond mixing device and a non-folding mutant allowed us to focus on the formation of the collapsed state from an extended conformation. For a given valence of ion, we find that collapse times decrease with increasing salt concentration (Figure 5.3 and 5.4). To explain our results, we propose a simple model based on the concept of electrostatic excluded volume, derived from the second virial coefficients (see Results and Figure 5.1B). Within this model, the ionc strength dependence of τ_c is readily explained. With no added salt, the A186U mutant exists in an extended (or unfolded) state (U) due to the large electrostatic excluded volume of its three helices. With the addition of charge compensating ions, the electrostatic excluded volume decreases as illustrated in Figure 5.5A and more conformations are available to the molecule. The more flexible, collapsed collection of states (I) have increased conformational entropy relative to U, as suggested by Bai et al. [27]. In the absence of tertiary contacts, there are minimal changes to the enthalpy, so the free energy decreases with respect to the extended state. Therefore population of the collapsed state(s) is more favorable at increased ionic strength and the kinetic transition time τ_c decreases. It is interesting to note a recent report discussing other favorable entropic effects of tertiary folding as a function of increasing [Mg^{2+}] [47]. Clearly, the role of thermodynamic contributions in ion-dependent RNA folding is still an open question and should be

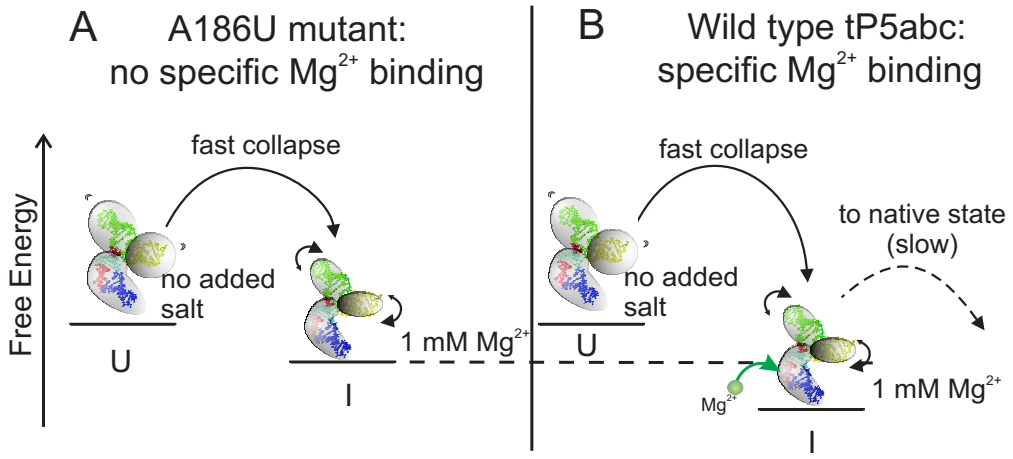

Figure 5.5: Possible schematic of the early stages of tP5abc collapse. (A) With no added salt, the A186U mutant exists in an extended (or unfolded) state (U) due to the large electrostatic excluded volume of its three helices. When rapidly mixed with ions, the molecule collapses to a compact intermediate state (I). The increase in ion concentration reduces the electrostatic excluded volume allowing the molecule to become more flexible. This change increases its conformational entropy and reduces its free energy, which all lead to faster collapse times. (B) The same arguments as in (A) apply to the wild type tP5abc, however, specific binding of Mg^{2+} ions reduces the enthalpy of the collapsed state, which can manifest as a faster collapse time as observed in the mixing experiments. Once specific binding has occurred, the wild type can proceed to fold to its native state, although this step is slow due to the large free energy barrier of secondary structure rearrangement.

subject to further investigations. Future directions should involve temperature dependent kinetic experiments with microsecond time resolution to examine Arrhenius-type dependence of ion concentration.

5.4.3 Site bound Mg^{2+} ions

To determine the effect of specific interaction on the initial stages of RNA collapse, we compared the collapse time of mutant and wild type tP5abc in 1 mM Mg^{2+} and 1 mM Sr^{2+}. The crystal structure of the *Tetrahymena thermophila* ribozyme's full P4-P6 domain shows 5 specific binding sites for Mg^{2+} ions within P5abc [18]. The A186U mutation disrupts one of these sites and prevents secondary structure rearrangement [19]. It is possible that the altered secondary structure is necessary for full coordination of the remaining bound Mg^{2+} ions, thus this single mutation may disrupt all site binding. We speculate that the shorter collapse time in 1 mM Mg^{2+} reflects a more stable compact intermediate state of the wild type relative to the mutant (Figure 5.5B). We attribute the reduced free energy of the collapsed state of the wild type to the lower enthalpy contributed by binding of at least one Mg^{2+} ion. Thus we suggest that Mg^{2+} binding occurs prior to either tertiary contact formation or secondary structure rearrangement, since these are known to occur more slowly [17, 23, 42]. Recent results by Koculi et al. showed that the slow step of wild type tP5abc folding does not require binding of additional Mg^{2+} ions to the RNA [23]. This also suggests that Mg^{2+} binds early in the folding process. The submillisecond time scale measured here is much faster than previously suggested for Mg^{2+} binding [28, 29], however those studies only had millisecond time resolution.

Our measurements indicate that the wild type collapses faster in 1 mM Mg^{2+} than in 1 mM Sr^{2+}. Evidence that Sr^{2+} cannot occupy the Mg^{2+} binding sites of P5abc [48] supports our proposal that the faster collapse of the wild type in Mg^{2+} is due to bound ions. We note that while the difference in the collapse times of the wild type RNA due to 1 mM Mg^{2+} and Sr^{2+} ions is just outside

the error bars, the same trend (albeit statistically insignificant) can also be discerned in the mutant. An alternate explanation for this observation may be found in previous theoretical work which suggests that the presence of smaller ions (like Mg^{2+}) in the ion atmosphere increase the stability of the collapsed ensemble relative to larger ions with the same valence (Sr^{2+}) [49]. However, the size of diffuse ions alone is not sufficient to account for the small difference in collapse timescales observed between the wild type and the mutant in the presence of 1 mM Sr^{2+}, where no specific binding occurs. Experimental techniques that specifically probe binding of Mg^{2+} and Sr^{2+} ions are needed to fully understand the contribution of Sr^{2+} ions in tertiary folding.

5.4.4 A model of tP5abc collapse and folding

Taken together, and in the context of other folding measurements on the same molecule, our data present a unified picture for how this small domain acquires tertiary structure. In the initial, low salt state, the molecule possesses the secondary structure that minimizes its free energy. Following the addition of ions to reduce the electrostatic repulsion between helices, the molecule favors more compact conformations. Mixing experiments indicate that formation of this collapsed ensemble occurs on the submillecond time scales predicted by theoretical studies of non-specific collapse of RNA [24, 25]. In the wild type molecule, our results suggest that Mg^{2+} ion interactions with specific nucleotides in the A-rich bulge are concurrent with the rapid collapse. It is interesting to speculate that the early binding of Mg^{2+} may be the cause of the energetically expensive shift in secondary structure that distinguishes non-specifically collapsed from folded states. The bound Mg^{2+} ion coordinated to A186 also interacts specifically with

nucleotides in the nearby strands of the P5c stem loop (Figure 5.2A). This interaction, coupled with dynamic fluctuations of the non-Watson-Crick base pairs near the three helix junction, may assist with the secondary structure shift measured in fully folded tP5abc. The A186U mutation disrupts the binding of this Mg^{2+} ion, and the barrier for rearrangement can never be overcome. The very large barrier associated with the folding transition may account for the wide separation in time of the collapse and folding transitions, and enables their separate study. Interestingly, the proposed Mg^{2+} binding site in the VS ribozyme is also near the sequence where a secondary structure shift occurs [50], hinting that the binding of Mg^{2+} may facilitate energetically expensive shifts in RNA secondary structure.

5.5 Conclusions

We used a number of different experimental techniques (equilibrium SAXS and FCS, rapid mixing with FRET detection) to relate the electrostatic properties of isolated RNA elements (double-stranded and non-base paired regions) to the kinetic folding behavior of a small, RNA domain. We find that non-specific electrostatic interactions are important in determining the conformational space accessed by a compact molecule that lacks tertiary contacts, which in turn is related to the time scale for collapse. Divalent ions are very effective in electrostatic screening and generate comparable rapid collapse rates to monovalent ions at much higher ionic strength. We propose that divalent ions are more effective than monovalent ions in charge screening due to their tighter localization around the RNA helix. Comparison of the collapse times of the wild type and A186U mutant of tP5abc RNA in divalent ions show that folding RNA se-

quences collapse faster than their non-folding mutants, supporting the concept that specific interactions lead to increased efficiency of RNA collapse. We suggest that specific binding of Mg^{2+} ions can occur even during the initial collapse and bias the reaction. Our results highlight the differing roles of counterions in RNA folding kinetics and may aid the design and use of synthetic RNA for biotechnology applications.

Acknowledgements

We acknowledge useful discussions with Steve Meisburger, Ron Elber and Serdal Kirmizialtin, and experimental assistance from Ken Finkelstein, Arthur Woll, Watt Webb, Warren Zipfel, Avtar Singh, Christopher Jones, Hye Yoon Park, David Rigie and Natalie Paquette. We thank Ignacio Tinoco for sharing extended state structures.

This work was supported by National Institutes of Health grant number [R01-GM085062]. CHESS is supported by the National Science Foundation and the National Institutes of Health/National Institute of General Medical Sciences NSF & NIH/NIGMS via NSF award DMR-0936384. CNF, a member of the National Nanotechnology Infrastructure Network, is supported by the National Science Foundation [ECS-0335765].

References

[1] J. A. Doudna and T. R. Cech, "The chemical repertoire of natural ribozymes," *Nature*, vol. 418, no. 6894, pp. 222–228, 2002.

[2] S. R. Eddy, "Non-coding rna genes and the modern rna world," *Nature Reviews Genetics*, vol. 2, no. 12, pp. 919–929, 2001.

[3] J. S. Mattick and I. V. Makunin, "Non-coding rna," *Human molecular genetics*, vol. 15 Spec No, R17–29, 2006.

[4] K. H. Link and R. R. Breaker, "Engineering ligand-responsive gene-control elements: lessons learned from natural riboswitches," *Gene therapy*, vol. 16, no. 10, pp. 1189–1201, 2009.

[5] E. Westhof, B. Masquida, and L. Jaeger, "Rna tectonics: towards rna design," *Folding & design*, vol. 1, no. 4, R78–88, 1996.

[6] S. E. Butcher and A. M. Pyle, "The molecular interactions that stabilize rna tertiary structure: rna motifs, patterns, and networks," *Accounts of Chemical Research*, vol. 44, no. 12, pp. 1302–1311, 2011.

[7] J. A. Cruz and E. Westhof, "The dynamic landscapes of rna architecture," *Cell*, vol. 136, no. 4, pp. 604–609, 2009.

[8] J. C. Bowman, T. K. Lenz, N. V. Hud, and L. D. Williams, "Cations in charge: magnesium ions in rna folding and catalysis," *Current opinion in structural biology*, vol. 22, no. 3, pp. 262–272, 2012.

[9] D. E. Draper, "A guide to ions and rna structure," *RNA*, vol. 10, no. 3, pp. 335–343, 2004.

[10] V. B. Chu, J. Lipfert, Y. Bai, V. S. Pande, S. Doniach, and D. Herschlag, "Do conformational biases of simple helical junctions influence rna folding stability and specificity?" *RNA*, vol. 15, no. 12, pp. 2195–2205, 2009.

[11] M. H. Bailor, X. Sun, and H. M. Al-Hashimi, "Topology links rna secondary structure with global conformation, dynamics, and adaptation," *Science*, vol. 327, no. 5962, pp. 202–206, 2010.

[12] L. Li, S. A. Pabit, J. S. Lamb, H. Y. Park, and L. Pollack, "Closing the lid on dna end-to-end stacking interactions," *Applied Physics Letters*, vol. 92, no. 22, pp. 223 901–2 239 013, 2008.

[13] H. M. Al-Hashimi and N. G. Walter, "Rna dynamics: it is about time," *Current opinion in structural biology*, vol. 18, no. 3, pp. 321–329, 2008.

[14] S. A. Woodson, "Compact intermediates in rna folding," *Annual review of biophysics*, vol. 39, pp. 61–77, 2010.

[15] L. W. Kwok, I. Shcherbakova, J. S. Lamb, H. Y. Park, K. Andresen, H. Smith, M. Brenowitz, and L. Pollack, "Concordant exploration of the kinetics of rna folding from global and local perspectives," *Journal of Molecular Biology*, vol. 355, no. 2, pp. 282–293, 2006.

[16] J. H. Roh, L. Guo, J. D. Kilburn, R. M. Briber, T. Irving, and S. a. Woodson, "Multistage collapse of a bacterial ribozyme observed by time-resolved small-angle x-ray scattering.," *Journal of the American Chemical Society*, vol. 132, no. 29, pp. 10 148–10 154, 2010.

[17] J. C. Schlatterer, L. W. Kwok, J. S. Lamb, H. Y. Park, K. Andresen, M. Brenowitz, and L. Pollack, "Hinge stiffness is a barrier to rna folding," *Journal of Molecular Biology*, vol. 379, no. 4, pp. 859–870, 2008.

[18] J. H. Cate, R. L. Hanna, and J. A. Doudna, "A magnesium ion core at the heart of a ribozyme domain," *Nature structural biology*, vol. 4, no. 7, pp. 553–558, 1997.

[19] M. Wu, I. Tinoco, and I. T. Jr., "Rna folding causes secondary structure rearrangement.," *Proceedings of the National Academy of Sciences*, vol. 95, no. 20, pp. 11 555–11 560, 1998.

[20] M. Zheng, M. Wu, I. T. Jr, and I. Tinoco, "Formation of a gnra tetraloop in p5abc can disrupt an interdomain interaction in the tetrahymena group i ribozyme," *Proceedings of the National Academy of Sciences*, vol. 98, no. 7, pp. 3695–3700, 2001.

[21] S. K. Silverman, M. Zheng, M. Wu, and I. Tinoco, "Quantifying the energetic interplay of rna tertiary and secondary structure interactions," *RNA*, vol. 5, no. 12, pp. 1665–1674, 1999.

[22] M. Greenfeld, S. V. Solomatin, and D. Herschlag, "Removal of covalent heterogeneity reveals simple folding behavior for p4-p6 rna," *The Journal of biological chemistry*, vol. 286, no. 22, pp. 19 872–19 879, 2011.

[23] E. Koculi, S. S. Cho, R. Desai, D. Thirumalai, and S. A. Woodson, "Folding path of p5abc rna involves direct coupling of secondary and tertiary structures," *Nucleic acids research*, vol. 40, no. 16, pp. 8011–8020, 2012.

[24] N. Lee and D. Thirumalai, "Dynamics of collapse of flexible polyampholytes," *Journal of Chemical Physics*, vol. 113, no. 13, pp. 5126–5129, 2000.

[25] D. Thirumalai, N. Lee, S. A. Woodson, and D. K. Klimov, "Early events in rna folding," *Annual Review of Physical Chemistry*, vol. 52, pp. 751–762, 2001.

[26] H. Y. Park, X. Qiu, E. Rhoades, J. Korlach, L. W. Kwok, W. R. Zipfel, W. W. Webb, and L. Pollack, "Achieving uniform mixing in a microfluidic device: hydrodynamic focusing prior to mixing," *Analytical Chemistry*, vol. 78, no. 13, pp. 4465–4473, 2006.

[27] Y. Bai, V. B. Chu, J. Lipfert, V. S. Pande, D. Herschlag, and S. Doniach, "Critical assessment of nucleic acid electrostatics via experimental and computational investigation of an unfolded state ensemble," *Journal of the American Chemical Society*, vol. 130, no. 37, pp. 12 334–12 341, 2008.

[28] S. K. Silverman and T. R. Cech, "An early transition state for folding of the p4-p6 rna domain," *RNA*, vol. 7, no. 2, pp. 161–166, 2001.

[29] H. D. Kim, G. U. Nienhaus, T. Ha, J. W. Orr, J. R. Williamson, and S. Chu, "Mg2+-dependent conformational change of rna studied by fluorescence correlation and fret on immobilized single molecules," *Proceedings of the National Academy of Sciences*, vol. 99, no. 7, pp. 4284–4289, 2002.

[30] S. A. Pabit, X. Qiu, J. S. Lamb, L. Li, S. P. Meisburger, and L. Pollack, "Both helix topology and counterion distribution contribute to the more effective charge screening in dsrna compared with dsdna," *Nucleic acids research*, vol. 37, no. 12, pp. 3887–3896, 2009.

[31] F. Bonnete and D. Vivares, "Interest of the normalized second virial coefficient and interaction potentials for crystallizing large macromolecules," *Acta crystallographica.Section D, Biological crystallography*, vol. 58, no. Pt 10 Pt 1, pp. 1571–1575, 2002.

[32] T. Nicolai and M. Mandel, "The ionic strength dependence of the second virial coefficient of low molar mass dna fragments in aqueous solutions," *Macromolecules*, vol. 22, no. 1, pp. 438–444, 1989.

[33] H. Chen, E. Rhoades, J. S. Butler, S. N. Loh, and W. W. Webb, "Dynamics of equilibrium structural fluctuations of apomyoglobin measured by fluorescence correlation spectroscopy," *Proceedings of the National Academy of Sciences*, vol. 104, no. 25, pp. 10 459–10 464, 2007.

[34] H. Y. Park, S. A. Kim, J. Korlach, E. Rhoades, L. W. Kwok, W. R. Zipfel, M. N. Waxham, W. W. Webb, and L. Pollack, "Conformational changes of calmodulin upon ca2+ binding studied with a microfluidic mixer.," *Proceedings of the National Academy of Sciences*, vol. 105, no. 2, pp. 542–547, 2008.

[35] J. B. Knight, A. Vishwanath, J. P. Brody, and R. H. Austin, "Hydrodynamic focusing on a silicon chip: mixing nanoliters in microseconds," *Physical Review Letters*, vol. 80, no. 17, pp. 3863–3866, 1998.

[36] L. Pollack, M. W. Tate, N. C. Darnton, J. B. Knight, S. M. Gruner, W. A. Eaton, and R. H. Austin, "Compactness of the denatured state of a fast-folding protein measured by submillisecond small-angle x-ray scattering," *Proceedings of the National Academy of Sciences*, vol. 96, no. 18, pp. 10 115–10 117, 1999.

[37] X. Qiu, K. Andresen, L. W. Kwok, J. S. Lamb, H. Y. Park, and L. Pollack, "Inter-dna attraction mediated by divalent counterions," *Physical Review Letters*, vol. 99, no. 3, p. 38 104, 2007.

[38] H. Chen, S. P. Meisburger, S. A. Pabit, J. L. Sutton, W. W. Webb, and L. Pollack, "Ionic strength-dependent persistence lengths of single-stranded rna and dna," *Proceedings of the National Academy of Sciences*, vol. 109, no. 3, pp. 799–804, 2012.

[39] S. A. Woodson, "Metal ions and rna folding: a highly charged topic with a dynamic future," *Current opinion in chemical biology*, vol. 9, no. 2, pp. 104–109, 2005.

[40] V. B. Chu, Y. Bai, J. Lipfert, D. Herschlag, and S. Doniach, "A repulsive field: advances in the electrostatics of the ion atmosphere," *Current opinion in chemical biology*, vol. 12, no. 6, pp. 619–625, 2008.

[41] J. Lipfert, A. Y. Sim, D. Herschlag, and S. Doniach, "Dissecting electrostatic screening, specific ion binding, and ligand binding in an energetic model for glycine riboswitch folding," *RNA*, vol. 16, no. 4, pp. 708–719, 2010.

[42] M. L. Deras, M. Brenowitz, C. Y. Ralston, M. R. Chance, and S. a. Woodson, "Folding mechanism of the tetrahymena ribozyme p4-p6 domain," *Biochemistry*, vol. 39, no. 36, pp. 10 975–10 985, 2000.

[43] D. B. McIntosh and O. A. Saleh, "Salt species-dependent electrostatic effects on ssdna elasticity," *Macromolecules*, vol. 44, no. 7, pp. 2328–2333, 2011.

[44] B.-Y. Ha and D. Thirumalai, "Bending rigidity of stiff polyelectrolyte chains: a single chain and a bundle of multichains," *Macromolecules*, vol. 36, no. 25, pp. 9658–9666, 2003.

[45] K. Andresen, X. Qiu, S. A. Pabit, J. S. Lamb, H. Y. Park, L. W. Kwok, and L. Pollack, "Mono- and trivalent ions around dna: a small-angle scattering study of competition and interactions," *Biophysical journal*, vol. 95, no. 1, pp. 287–295, 2008.

[46] S. Kirmizialtin and R. Elber, "Computational exploration of mobile ion distributions around rna duplex," *The journal of physical chemistry.B*, vol. 114, no. 24, pp. 8207–8220, 2010.

[47] J. L. Fiore, E. D. Holmstrom, and D. J. Nesbitt, "Entropic origin of mg2+-facilitated rna folding," *Proceedings of the National Academy of Sciences*, vol. 109, no. 8, pp. 2902–2907, 2012.

[48] K. J. Travers, N. Boyd, and D. Herschlag, "Low specificity of metal ion binding in the metal ion core of a folded rna," *RNA (New York, N.Y.)*, vol. 13, no. 8, pp. 1205–1213, 2007.

[49] E. Koculi, C. Hyeon, D. Thirumalai, and S. A. Woodson, "Charge density of divalent metal cations determines rna stability," *Journal of the American Chemical Society*, vol. 129, no. 9, pp. 2676–2682, 2007.

[50] A. A. Andersen and R. A. Collins, "Rearrangement of a stable rna secondary structure during vs ribozyme catalysis," *Molecular cell*, vol. 5, no. 3, pp. 469–478, 2000.

CHAPTER 6

CONCLUSIONS

In this thesis we presented three works bringing us towards an understanding of the dependence of RNA structure on sequence, topology, and ions. In Chapter 3 we discussed the effects of helix length and junction sequence on a model RNA junction. We found that these structural factors influence the overall conformation of the junction and may be used biologically to tune structures in solution. In Chapter 4 we extended this work to include double-stranded junctions and found that junction topology can restrict helix-helix orientations and that Mg^{2+} ions play a sequence dependent role in RNA structure. In Chapter 5 we presented a study of a 3-helix junction which folds in the presence of Mg^{2+} ions and found that the ions influence the rates of collapse when specific binding sites are present.

A complete picture of RNA structure is still a work in progress, and there are several directions which we can take to expand on the work in this thesis. For example, while our experiments in Chapter 5 elucidated the timescales of electrostatic collapse, there are many questions remaining about the folding mechanism of tP5abc. The confinement of the fluorescent dyes precluded the determination of intramolecular distances. With an improved labeling scheme, we can gain more insight into the specific structural changes observed. Furthermore, using our single-molecule capabilities for both equilibrium and time resolved experiments would further elucidate structural information and conformational heterogeneity.

The base-stacking propensity of single-stranded RNA is still poorly understood. We have made some preliminary progress towards using the SHAPE-

seq technique to study the base-stacking propensity of RNA using the poly(A) and poly(U) junctions described in Chapter 3. SHAPE-seq is a technique that combines the ability to chemically modify flexible regions of RNA in solution with the high resolution of single-molecule sequencing. With the help of David Loughrey and Julius Lucks in Chemical and Biomolecular Engineering, we set out to see if this technique could be used to study the base-stacking propensity of un-paired RNA. Unfortunately, preliminary results were unclear, likely due to either sample degradation or sequence-specific interference with the reverse-transcriptase enzyme. Future experiments are needed to investigate the causes of the poor data quality. Moving forward, the combination of SHAPE-seq and smFRET will provide a useful avenue for studying conformations of RNA solution.

Finally, a significant outcome of this thesis is to provide experimental data for validating molecular dynamics (MD) simulations of RNA. MD has become an increasingly useful tool for modeling protein dynamics and interactions. However, many of the force fields have been tuned for proteins, and lead to unphysical results for nucleic acids. We have shown that a complete model of RNA must include accurate descriptions of 1) base-stacking, 2) helix length, 3) junction topology and 4) ion binding sites. An ongoing collaboration with Tongsik Lee and Ron Elber from the University of Texas at Austin and Serdal Kirmizialtin from NYU Abu Dhabi is working towards modeling the folding of tP5abc. The availability of both folded and unfolded-state structure of tP5abc makes it an ideal candidate for simulating the entire time course of folding. Additionally, the smFRET data provided for poly(U) and poly(A) junctions may be useful for improving base-stacking parameters in MD force fields. Combining and comparing the experimental data with MD simulations promises to provide

additional insights into RNA folding at the molecular level.

APPENDIX A

PROTOCOL FOR SINGLE-MOLECULE FRET MEASUREMENTS

Microscope basics

1. Detectors

   - Single-photon counting detectors are extremely sensitive to light. Exposing them to more than 1MHz can be harmful to the detector. Therefore, it is important to always TURN OFF THE ROOM LIGHTS before turning on the APD detectors. Furthermore, ensure that fluorescent samples are sufficiently dilute so as not to exceed the threshold. I will typically try to keep the intensity below ~ 600KHz.

2. Software

   - There are two programs associated with the Flex2kx2-12D correlator card: Flex2kd and Photon.exe. Flex2kd is the FCS mode, which measures both the intensity (sampled every second) and the correlation (either auto or cross). This program is also useful when aligning the pinholes as it shows you the photon counts over time for both channels. When diagnosing FRET signals, use the 'Dual Auto' setting to take the autocorrelation (use 'Dual Cross' when you want to measure the cross correlation between channels) and set the duration to whatever you like.

   - Photon.exe records the arrival time of each individual photon and is used for recording smFRET data. The 16 bit mode is best for the low intensities typical in smFRET measurements. It is important to type

in the '.dat' extension after your file name or else the program will not record the name properly. For example, you must type in 'example.dat', for Photon.exe to produce two files named: 'example1A.dat', and 'example1B.dat' containing the data from each channel. Because the file sizes get large, don't take data for more than 60s at a time. Use the script called 'photon-annotate.au3' to collect multiple .dat files from a given sample.

- Be warned: The correlator software can be buggy and may require you to restart the computer (e.g. Flex2kd and Photon.exe cannot be open at the same time without crashing).

3. Focusing

    - To focus the sample, use the Flex2kd program to monitor the intensity. The best indication for the position of the focal volume is to look for the a change in signal while changing the z-position of the sample stage. When the focus moves from your sample into the coverglass you will either see an increase or decrease in signal depending on the sample you are using. (Laser scatter from the cover glass can increase the background relative to pure water.) Once you find this intersection between sample and glass, increase the z-position by about 35μm. The fluorescence signal should not be very sensitive to the exact z-position.

Taking smFRET measurements

1. Incubate 8-well coverglass chambers with 0.05% BSA in water for at least one hour or overnight.
2. Wash out BSA using the Nanopure water and keep chambers full of water until use.
3. Thaw and prepare FRET-labeled samples as needed. While single-molecule samples require ~100pM concentrations, RNA should not be stored below ~100nM in microcentrifuge tubes for long periods of time.
4. Turn on laser and check alignment of optical fibers (pinholes): To do this you can use free dye (e.g. Alexa 488 for the donor channel or Alexa 594 for the acceptor channel) which is concentrated enough to give a good amount of signal (e.g. > 100KHz). Maximize the counts using the x and y translation knobs. Typically x and y alignment should be checked at the start of the day, but z alignment is less sensitive and does not need to be checked every day.
5. For each sample:
   (a) Take a buffer background: Pipette out the water from one of the chambers in your 8-well sample holder. Get as much water out of the corners as possible, but try not to disturb the BSA from the glass in the center of the well. Add at least 250μl of the desired measurement buffer to the well. Focus the laser into the sample using Flex2kd and use Photon.exe to take a 30s measurement. The average of the photon counts will be your background for the next measurement.
   (b) Prepare the sample: Dilute 0.25μl sample into the buffer that you just used to take the background. Mix with the larger pipette. Check that

the FCS intercept is large enough to signify single-molecule detection. I typically see $G(0)$ of 7-10 for the donor and 10-20 for the acceptor channel. If FCS signal is too noisy, chances are you aren't going to get good FRET data. If that is the case, add another 0.1-0.25μl of sample, mix, and see if signal has improved.

(c) Take data: Open the photon-annotate.au3 script and change the file name to be used for the sample. Open Photon.exe and enter the desired time for each file (usually 60s). Run the photon-annotate.au3 script and follow the instructions to start collecting data. I usually take 30 measurements of 60s each.

6. When you are done taking measurements ensure that detectors are turned off (as they always should be when the lights are on!) and turn off the laser. Transfer the data to another computer so the hard drive doesn't fill up.

Aligning the microscope

A more thorough check of the microscope's alignment should be done periodically. I will typically do this before starting a new set of experiments, or when the microscope has not been used for several weeks. The main goals are to align the laser beam so that it is centered and vertical through the objective and to align the pinholes (fibers) to maximize the signal intensity.

1. Start by aligning the laser beam through the beam expander: Center the beam on an iris at the input, and then use the large iris on the output of the beam expander to estimate how centered it is at the output.

2. Align the laser into the objective: Remove the objective and replace it with a long lens tube with irises at either end (use the RMS to SM1 adaptor to connect the lens tube to the objective turret). Use the two mirrors closest to the input of the microscope to iterate the alignment until the laser beam is centered through both ends of the lens tube. Closing the iris at the output of the beam expander can help with this process by reducing the beam size, but the beam must be centered through the beam expander if it is going to be centered on the objective.

3. Align the pinholes: Replace the objective into the turret, add some immersion water and use a sample of free dye to align the pinholes. Maximize the counts by tweaking x and y translation knobs. Note the maximum counts reached and measure the position of the xy translation stage along the cage rails using a caliper. Loosen the set screws of the translation stage and slide it a couple millimeters to a different z-position. Re-tweak the x and y position to maximize the counts. Iterate until you have found the z-position with the maximum counts. You should notice the maximum counts increase and then decrease as you pass through the ideal pinhole position. I find that the z- position is less sensitive than x and y, and can usually be within a ~2 mm range without affecting the total counts.

APPENDIX B

PROTOCOL FOR LABELING AMINO-MODIFIED RNA WITH NHS ESTER FLUOROPHORES

Overview of labeling process

The protocol for labeling RNA consists of mixing NHS ester fluorophores with amino-modified RNA, waiting for the reaction to go to completion and then purifying the RNA from the remaining free dye. Typical labeling efficiencies can be 70-90%. One potential pitfall is that the NHS ester reacts with water, so it can lose its reactivity over time once dissolved. To circumvent this issue, the dye should either be dissolved right before carrying out the reaction, or dissolved in anhydrous DMSO (or DMSO stored with molecular sieves) and stored at -80°C. Another important consideration for labeling RNA is the solution pH. If the pH is too low the NHS ester will take too long to react. If the pH is too high hydrolysis of the ester is more likely to occur than the reaction between the dye and the modified RNA. The ideal pH range is 8.3-8.5. I have used 0.1M phosphate, but other buffers may be better suited for this pH range. Some resources claim that TRIS buffers will react with the dyes, but there appears to be disagreement in the literature on this issue. Best to avoid TRIS if possible.

Helpful resources

- http://www.lumiprobe.com/protocols/nhs-ester-labeling
- Labeling DNA (or RNA) for Single-Molecule FRET, Chirlmin Joo and

Taekjip Ha Adapted from Single-Molecule Techniques (ed. Selvin and Ha). CSHL Press, Cold Spring Harbor, NY, USA, 2008.

- Methods of site-specific labeling of RNA with fluorescent dyes, Sergey Solomatin and Daniel Herschlag. Methods in Enzymology, volume 469, 2009.

Reagents

- Dimethyl sulfoxide (DMSO; Fisher Scientific), degassed
- RNA sample with amino-modification
- Fluorescent dyes in the NHS ester form. I typically buy them in 0.1mg packs and then dissolve a fresh tube for each labeling reaction. Examples: Cy3 mono-reactive dye pack (manufactured by GE Healthcare, PA23001), CY5 Mono NHS Ester (PA25001), or Alexa Fluor 488 TFP ester (from Life Technologies, A37570).
- Reaction buffer with pH 8.3-8.5, e.g. 0.1M Phosphate buffer

Method

1. Decide how much RNA sample you want to label. Then determine the amount of dye you need. *I typically use 10x moles of dye per mole of RNA.*
2. Determine the final reaction volume. Optimal final RNA concentration is 1-10 mg/ml. *I typically use about 20uL for 0.02mg of RNA. However if using Alexa 488 TFP, I have found that it crashed out of aqueous solution at 2.3mg/ml, so adjust the volume to keep the final Alexa 488 concentration at or*

below ~2.0mg/ml. Cy3 or Cy5 are more soluble in water and can be kept at higher concentrations.

3. Prepare reaction buffer. The pH needs to be between 8.3-8.5 for optimum labeling efficiency.
4. Degas buffer and DMSO for about 1 hour.
5. Dissolve RNA in reaction buffer to a concentration 10/9 times more concentrated than final desired concentration determined in step 2.
6. Add 9/10 final reaction volume RNA to a clean reaction tube *(e.g. 18uL for a 20uL rx. volume)*.
7. The next steps are time sensitive. Since the NHS ester will hydrolyze in aqueous environments, you want to mix the dye with your RNA quickly after dissolving it.
   (a) Dissolve dye in DMSO to be 10x more concentrated than desired.
   (b) Proceeding quickly, add 1/10th reaction volume of dye to reaction tube containing RNA. *(i.e. Add 2ul dye in DMSO to the 18uL RNA)*.
8. Vortex well. Store in the dark on a rocker *(i.e. wrap in Al foil and tape to vortexer set to lowest setting)* at 4°C overnight.
9. To separate the free dye from the labeled molecules, do an ethanol precipitation, use an amicon concentrator, or Micro Bio-Spin chromatography column. *Free Alexa 488 seems to be less efficiently removed by ethanol precipitation than Cy3 or Cy5. I normally will do one ethanol precipitation and then buffer exchange the RNA until the flow-through no longer has any visible absorption.*

APPENDIX C

SUPPLEMENTARY INFORMATION FOR CHAPTER 3

E_{FRET} Simulations

To assist in interpreting the E_{FRET} for our RNA constructs we used tools developed elsewhere [1] to model the sterically accessible volume (AV) of the dyes attached to RNA as in reference [2]. Briefly, .pdb files of RNA structures were input to the FRET_nps program developed in [1] and the C5 atom of the desired base was chosen as attachment point for the dye linker. We used ref. [2]'s values for the linker length (L_{link}) and width (w_{link}) and dye radii (R_{dye}). These values are provided in Table C.1. As in ref [2], the accessible volume of three different dye radii were combined to account for the different dye dimensions, and the AV of Cy5 was rotated by an extra 20 degrees about the helical axis to account for the dye's asymmetric nature.

To compare simulations to experiment, we must account for experimental correction factors as well as the correct averaging regime. We chose to average simulated FRET efficiencies assuming fast orientational dye dynamics, but slow inter-dye distance fluctuations [2-4] so that:

$$\mathcal{E}_\gamma = \frac{1}{mn} \sum_{j=1}^{m} \sum_{i=1}^{n} \mathcal{E}_{\gamma,ij} \tag{C.1}$$

Where $\mathcal{E}_\gamma$ is the gamma-corrected FRET efficiency given by:

$$\mathcal{E}_{\gamma,ij} = \frac{1 + \epsilon(R_{ij}/R_0)^6}{1 + (1+\epsilon)(R_{ij}/R_0)^6} \tag{C.2}$$

assuming no crosstalk of acceptor signal into donor channel. R_{ij} is the distance between points R_i within the acceptor AV and R_j within the donor AV, R_0 is the Förster radius (52Å for this dye pair [2]), and ϵ is a correction factor to account for donor cross talk into the acceptor channel [3]. Finally, $\mathcal{E}_\gamma$ is related back to the measured E_{FRET} using:

$$\mathcal{E}_\gamma = \frac{E_{FRET}}{E_{FRET} + \gamma(1 - E_{FRET})} \tag{C.3}$$

where $\gamma = \eta_A\phi_A/\eta_D\phi_D$, and $\phi_{A,D}$ and $\eta_{A,D}$ are the quantum yields and detection efficiencies of the acceptor and donor, respectively. Values for $\phi_{A,D}$ were assumed to be the same as in [2], while the ratio of η_A/η_D was measured using the method described in the Supporting Appendix of reference [5]. Finally, the cross-talk correction factor, β, was measured as the ratio of photon counts recorded in the acceptor channel to those in the donor channel from sample containing only Alexa 488. The factor ϵ, defined in [3] equals β/γ. For our microscope, we found $\beta = 0.027$, and $\gamma = 1.2$.

Differences in fluorophore accessible volume for 12 and 24 bp helices

Label sites were designed to maintain a fixed distance (8bp) from the junction for both 12 and 24bp helix constructs in order to keep the dye environments comparable between different constructs. Since the above method of modeling dye positions requires a known RNA structure, we used the Nucleic Acid Builder (NAB) web server [6] to generate structures where the junction and he-

lices are all A-form. This structure is most likely found for the construct containing poly(A) junctions, at high $MgCl_2$ concentration. We cannot easily compare data to simulations at other salt concentrations since junction conformations are unknown and the RNA conformations are dynamic. We find that the simulated AVs lead to an E_{FRET} difference of 0.067 for the 12 and 24bp helices since the 24bp helix restricts the AV of the dyes relative to the 12bp helix (Figure C.1, Table C.2), but this difference is not observed experimentally at high [$MgCl_2$]. Because previous experiments validated the assumption of freely rotating dyes [2], it is likely that the actual dye environments are more similar than the calculation suggests. Others report that this modeling method has some discrepancies with molecular dynamics simulations [2] and does not account for interactions between the RNA and the fluorophore. Both Alexa 488 and Cy5 contain negative charges at pH7, which may further restrict their accessible volume. In either case, we expect that any differences between dye environments should be exaggerated at extreme salt concentrations. However, our FRET data only shows differences between constructs of different helix lengths at intermediate salt concentrations. Thus it is unlikely that differences between dye environments alone can explain the observed differences in E_{FRET} between 12 and 24bp helices.

Comparison between simulated and experimental E_{FRET}

Despite the lack of agreement between the 12bp AV simulations and the data, the simulations for the 24bp poly(A) construct agree very well with our measurement at high [$MgCl_2$] (E_{FRET} = 0.396 vs. 0.406, Table C.2). However, due to the lack of structural information on the RNA junctions, we can only make qual-

itative comparisons between data and simulation at other salt concentrations. If we assume a salt-independent R_0 and freely rotating dyes, we expect an overall increase in E_{FRET} for the constructs containing poly(U) junctions compared to the A-form junction at low salt (because helices are skewed) and at high salt (since helices can approach one another).

Although the generated A-form structure reflects a coaxial helix conformation, it does not necessarily reflect the minimum possible E_{FRET}. Thus measurements at other salt concentrations that yield E_{FRET} values as low as ~ 0.34 are still consistent with our model. For example, because dyes are attached internally via a flexible C6 carbon linker, they are offset from the helical axis. As a consequence, the inter-dye distance (and thus the E_{FRET}) is sensitive to both the relative angular orientation of the helices as well as overall RNA conformational relaxation. Therefore it is possible for dyes to be farther apart than in the construct containing the fully stacked junction if one helix is rotated with respect to the other, and the dyes end up on opposite sides of the molecule. For these reasons, caution should be used in interpreting the data; it is not possible to decouple helix approach and rotation through E_{FRET} alone.

Possibility of intramolecular end-to-end stacking

Although intermolecular stacking is not a problem (smFRET measurements are performed on very dilute samples of RNA), the possibility remains that the two tethered RNA helices may stack due to their close proximity. For this stacking to occur, the junction would have to loop out of the way, and thus is most likely to happen when the construct contains the flexible poly(U) junction. To address

the question of end- to-end stacking we simulated the E_{FRET} values obtained from a molecule in which the two helices are fully stacked. We again focus on the 24bp duplex, and we used NAB to generate a structure of a 48bp duplex to mimic two 24bp helices stack end-to-end. This structure yields an E_{FRET} of 0.682 (Table C.2, 24bp - stacked), which is larger than the range of E_{FRET} values measured for poly(U) junctions, even in 300mM $MgCl_2$ (the most likely condition to exhibit end-to-end stacking). Thus, the most likely cause of the increase in E_{FRET} is the continuous approach of the two helices, rather than end-to-end stacking.

RNA sequences

RNA sequences were chosen to reduce undesired hairpins and heterodimers as described in the main text. Sequences are provided below. Amino-modified dT bases were incorporated instead of uridine at blue color locations in strands 1 and 2, and were used to attach either Alexa Fluor 488 TFP (strand 1) or Cy5 NHS ester (strand 2). The placeholder X is used to represent the junction, which is either U or A.

12bp construct:

1. 5′ GGGAGUAUAGGG 3′
2. 5′ GCGAUUAGGAGG 3′
3. 5′ CCCUAUACUCCCXXXXXCCUCCUAAUCGC 3′

24bp construct:

1. 5′ GGGAGUAUAGGGAAAAGGGAGUCG 3′
2. 5′ GGAACAGGGAUAGCGAUUAGGAGG 3′
3. 5′ CGACUCCCUUUUCCCUAUACUCCCXXXXXCCUCCUAAUCGCUA-

Dye	w_{link}(Å)	L_{link}(Å)	$R_{dye,1}$(Å)	$R_{dye,2}$(Å)	$R_{dye,3}$(Å)
Alexa 488	4.5	20	5	4.5	1.5
Cy5	4.5	22	11	3	1.5

Table C.1: parameters from [2] used in AV simulations

	$\mathcal{E}_\gamma, calc$	$E_{FRET,AV}$	$E_{FRET,experimental}$ For poly(A) in 300mM $MgCl_2$
12bp	0.290	0.329	0.407 ± 0.007
24bp	0.353	0.396	0.406 ± 0.003
24bp - stacked	0.641	0.682	-

Table C.2: comparison between experimental E_{FRET} and AV simulation for RNA constructs.

UCCCUGUUCC 3′

Poisson-Boltzmann simulations

We used the Adaptive Poisson Boltzmann Solver (APBS) [7] to simulate the electrostatic potential around isolated RNA duplexes. PDB coordinates of A-form RNA duplexes were created using the Nucleic Acid Builder web server [6] for a 12bp and a 24bp helix, created using sequence 1 above for each construct and its complement. Simulations were done at four different ion concentrations: 20, 50, 100, and 200mM KCl. The input parameters used for APBS were: ion radius = 2Å (K^+ and Cl^-), and the minimum boundary distance, was set to 60Å for 20 and 50mM KCl, and 40Å for the remaining salt concentrations. APBS output was analyzed using MATLAB (Figures C.11 and C.12).

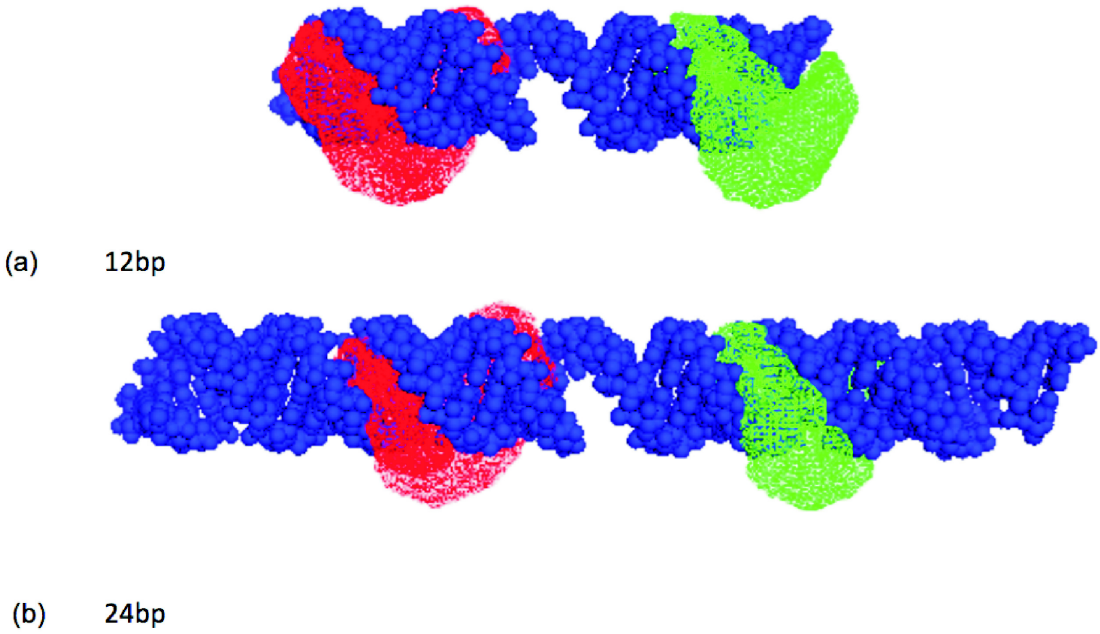

Figure C.1: Surface representation of accessible dye positions, for Alexa 488 (green) and Cy5 (red) around A-form RNA (blue) for (a) a 12bp and (b) a 24bp RNA helix. Simulations were performed using software developed by [1]. Image was generated using PYMOL.

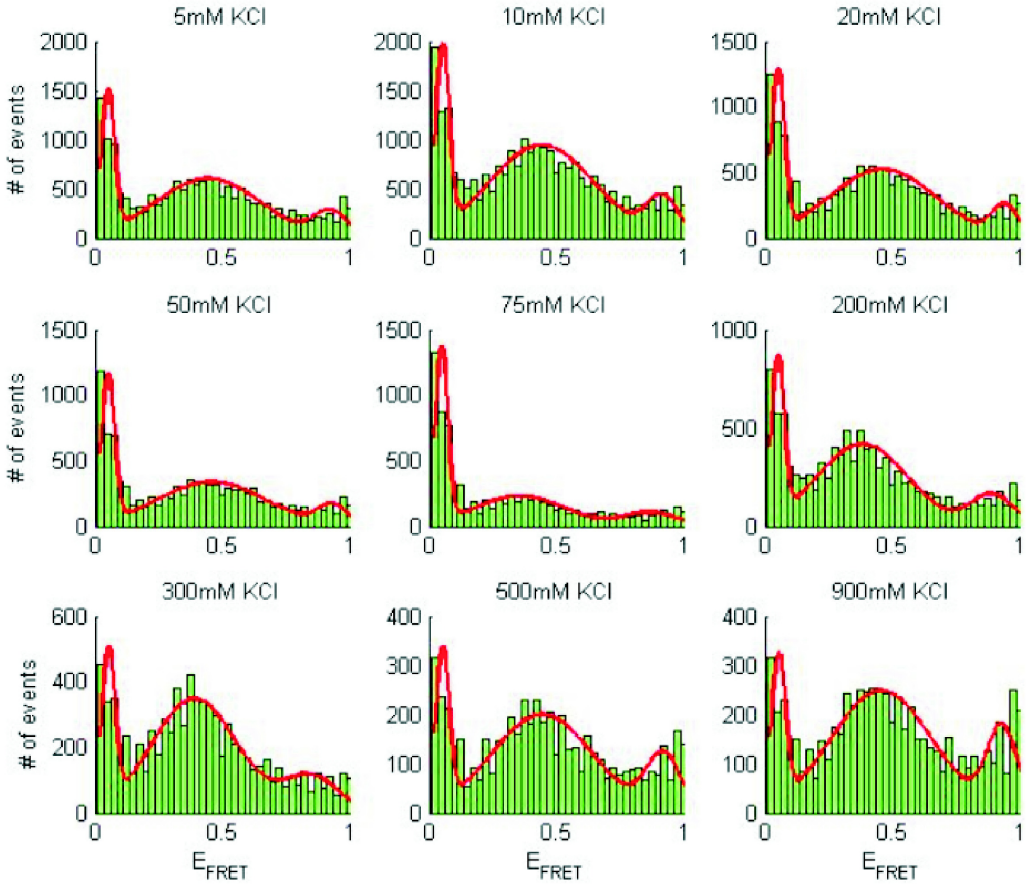

Figure C.2: Example histograms (green) fit to 3 gaussians (red) for RNA constructs with 12bp helices and poly(A) junctions in varying [KCl]. Data was truncated for $E_{FRET} \leq 0.02$.

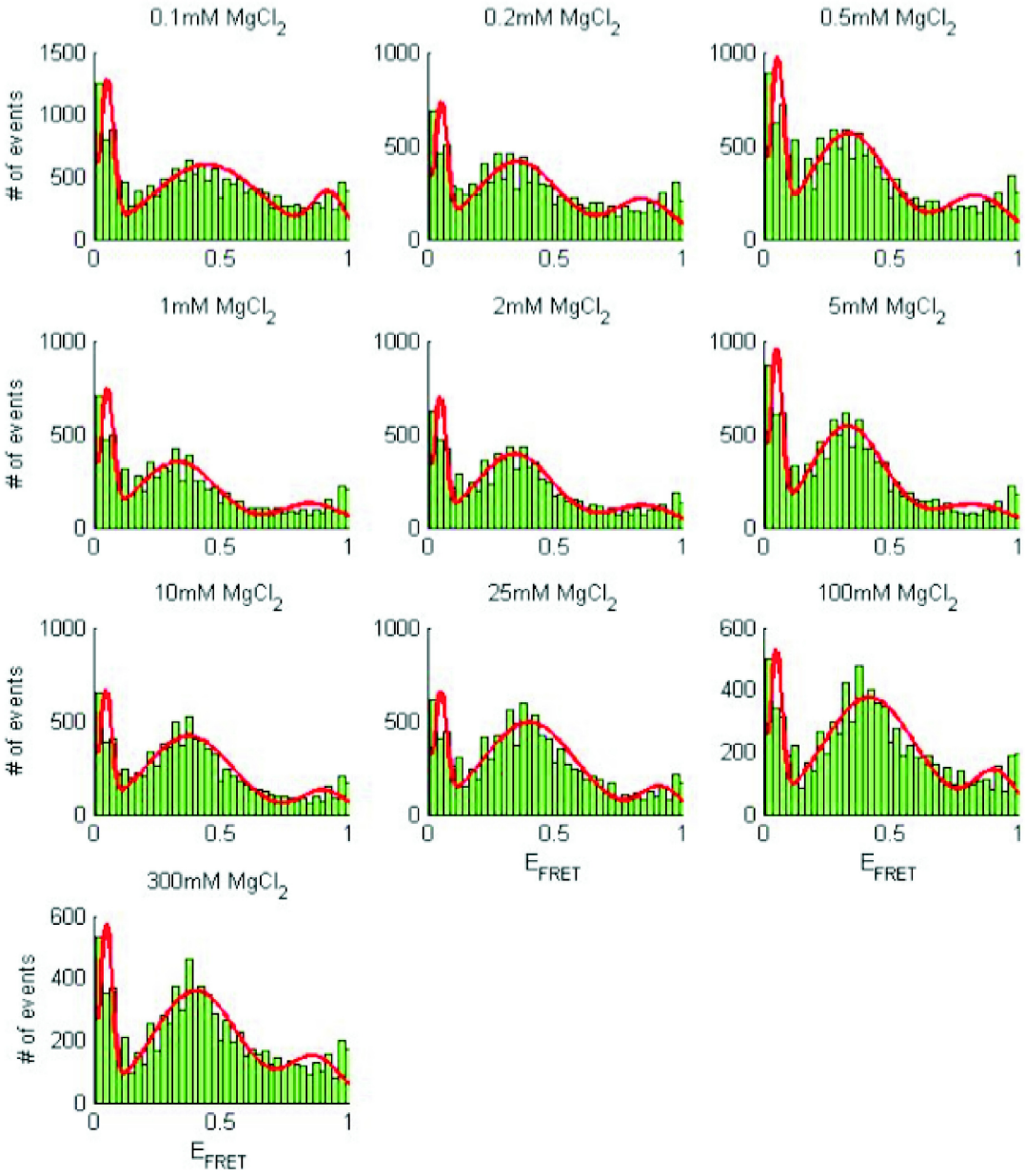

Figure C.3: Example histograms (green) fit to 3 gaussians (red) for RNA constructs with 12bp helices and poly(A) junctions in varying [$MgCl_2$]. Data was truncated for $E_{FRET} \leq 0.02$.

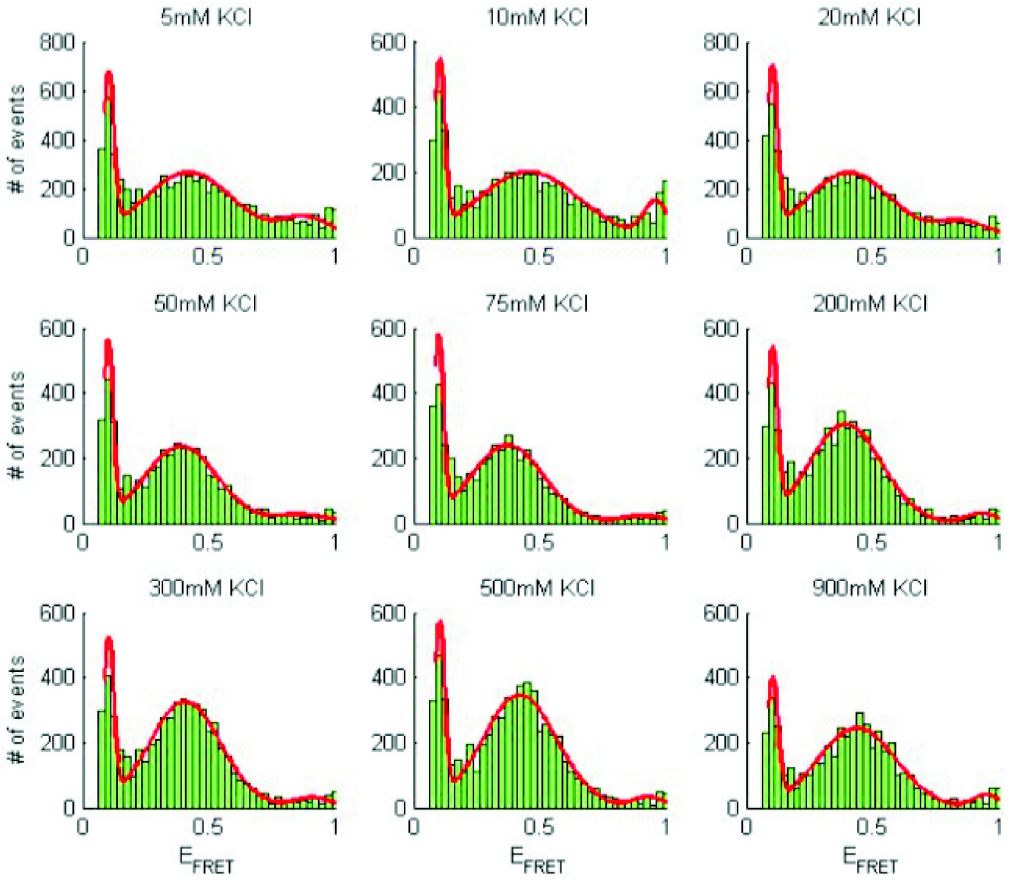

Figure C.4: Example histograms (green) fit to 3 gaussians (red) for RNA constructs with 24bp helices and poly(A) junctions in varying [KCl]. Data was truncated for $E_{FRET} \leq 0.08$.

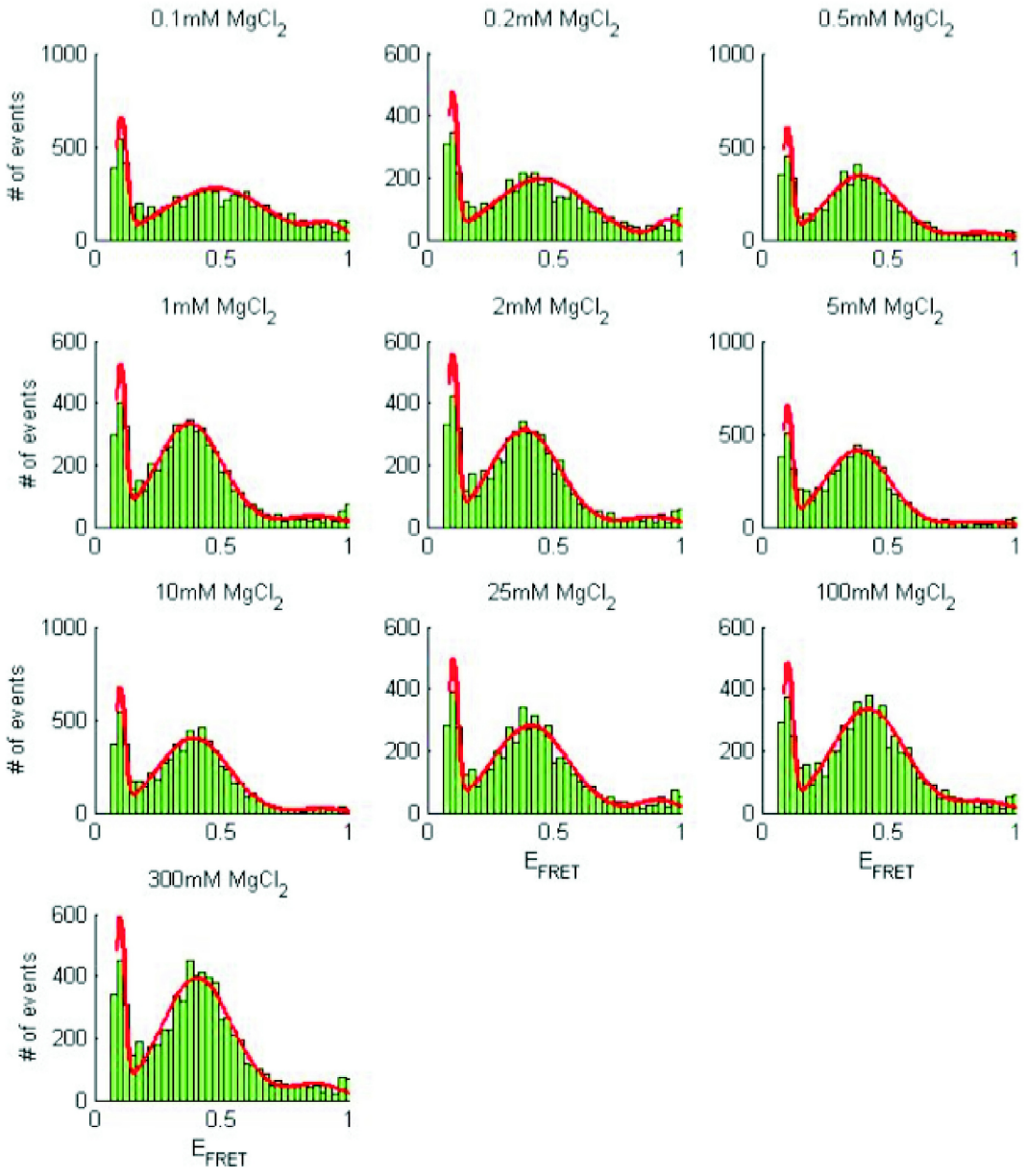

Figure C.5: Example histograms (green) fit to 3 gaussians (red) for RNA constructs with 24bp helices and poly(A) junctions in varying [$MgCl_2$]. Data was truncated for $E_{FRET} \leq 0.08$.

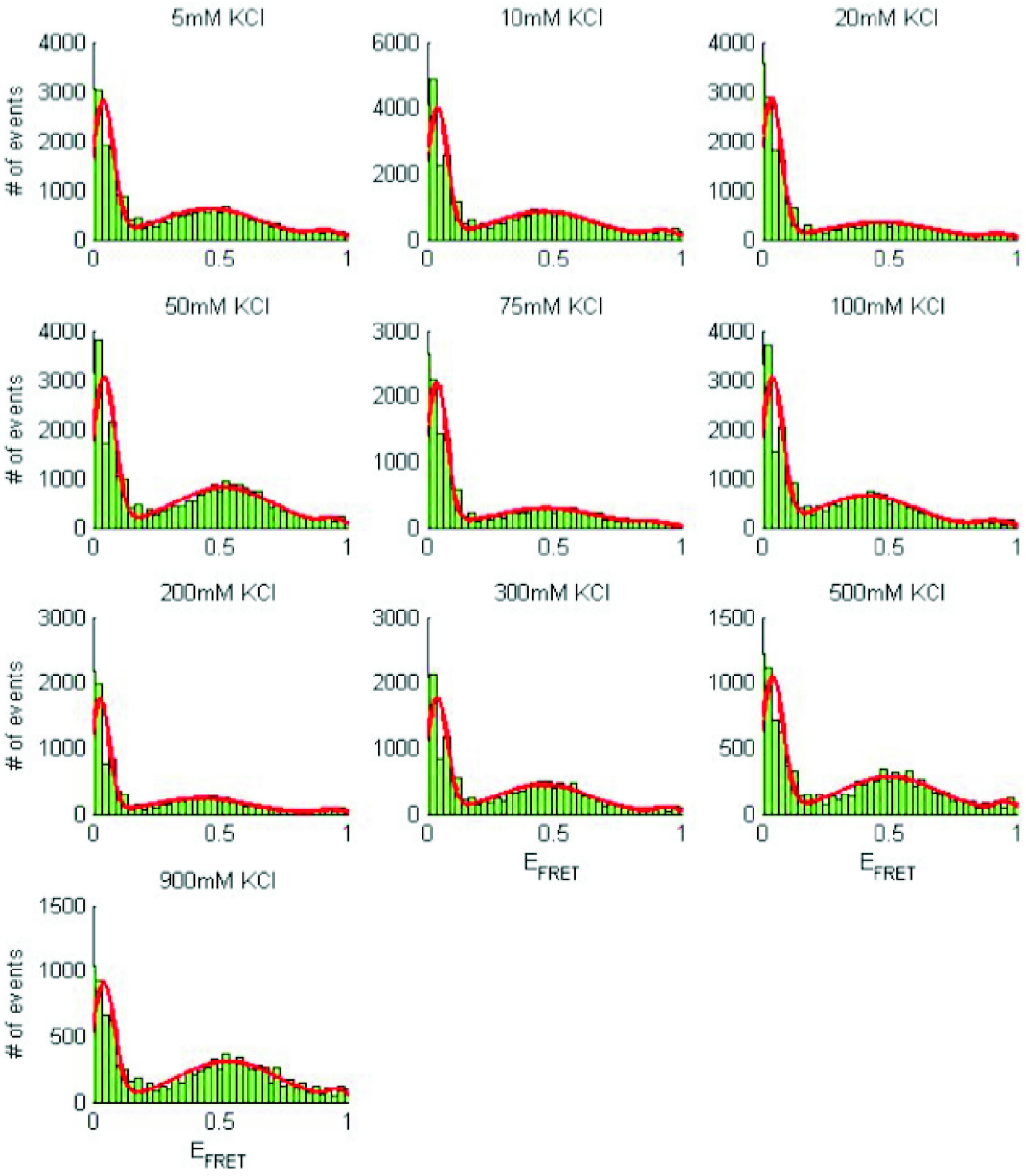

Figure C.6: Example histograms (green) fit to 3 gaussians (red) for RNA constructs with 12bp helices and poly(U) junctions in varying [KCl].

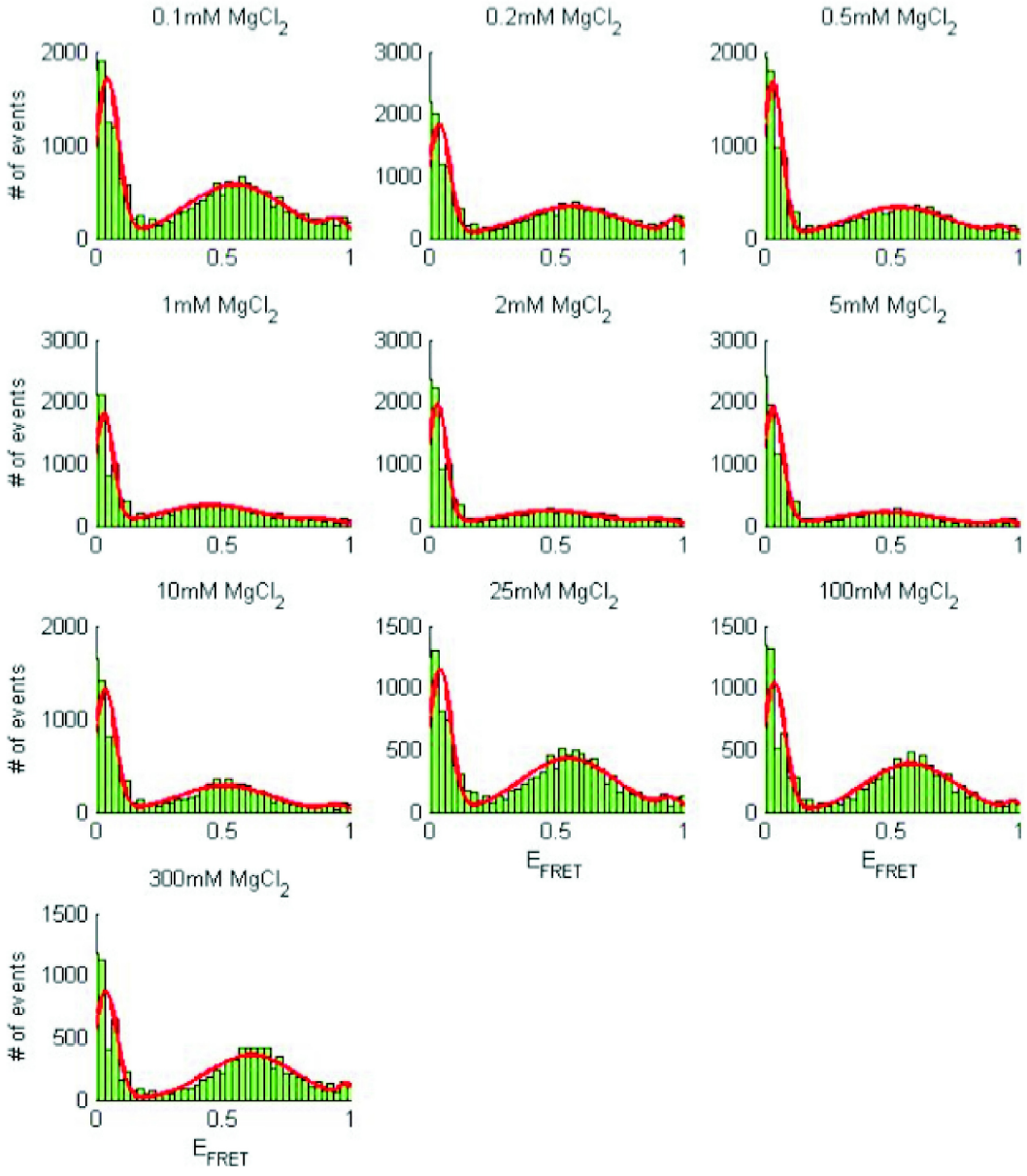

Figure C.7: Example histograms (green) fit to 3 gaussians (red) for RNA constructs with 12bp helices and poly(U) junctions in varying [$MgCl_2$].

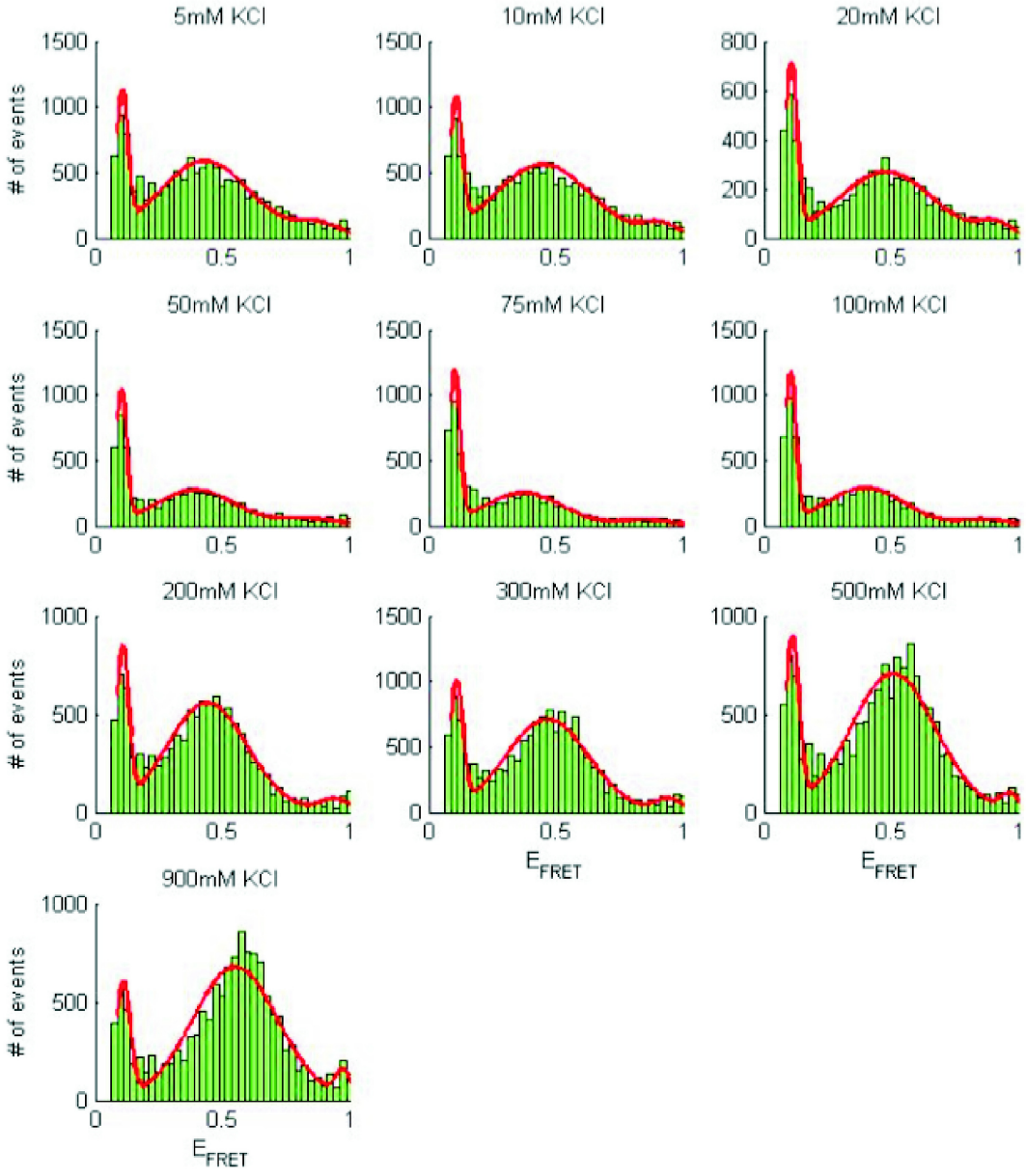

Figure C.8: Example histograms (green) fit to 3 gaussians (red) for RNA constructs with 24bp helices and poly(U) junctions in varying [KCl]. Data was truncated for $E_{FRET} \leq 0.08$.

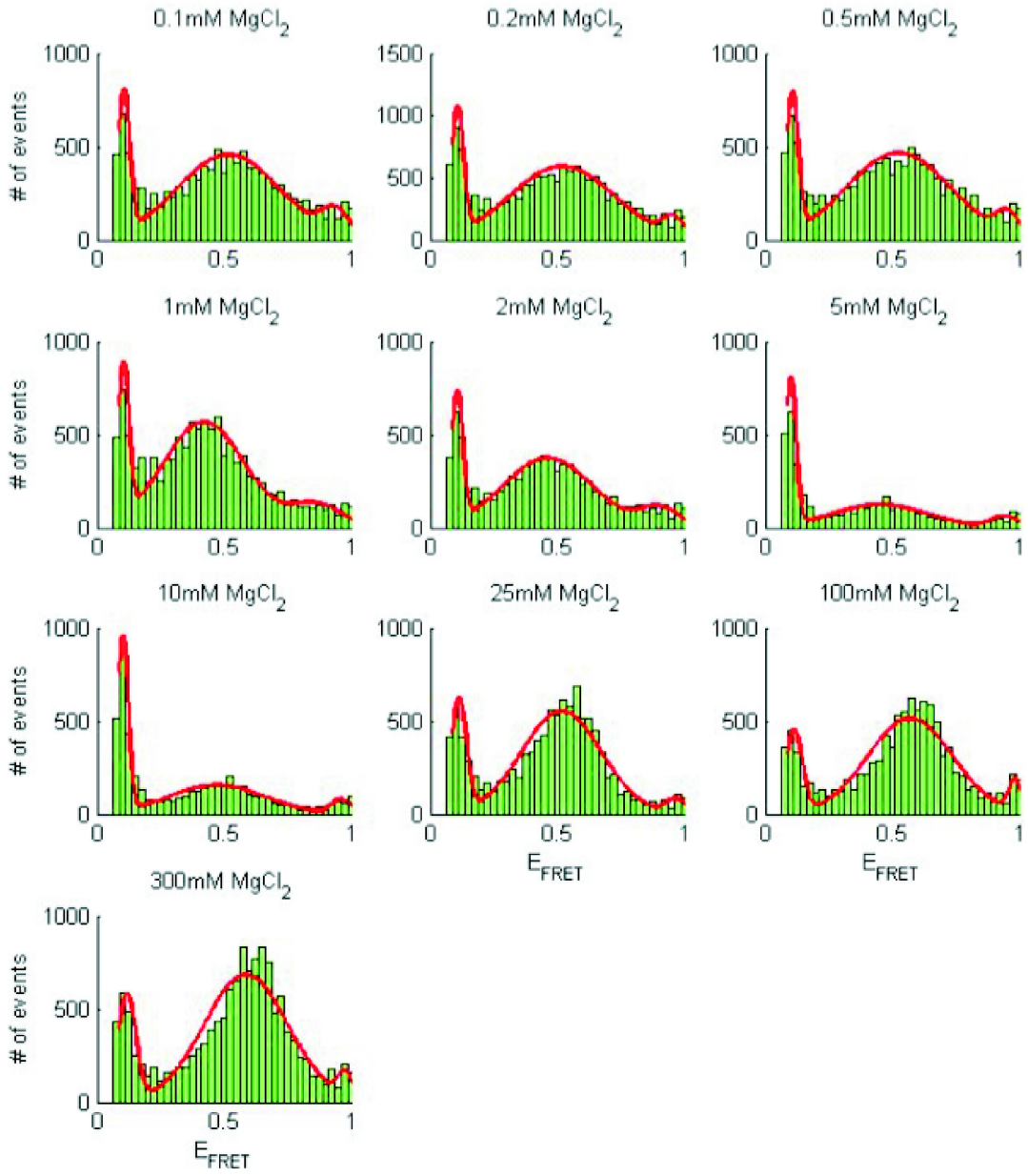

Figure C.9: Example histograms (green) fit to 3 gaussians (red) for RNA constructs with 24bp helices and poly(U) junctions in varying [$MgCl_2$]. Data was truncated for $E_{FRET} \leq 0.08$.

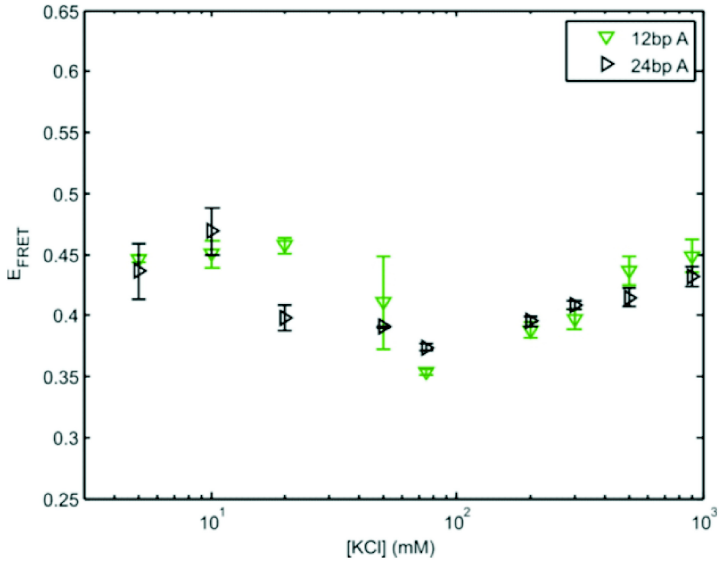

Figure C.10: Helix length comparison of poly(A) constructs show similar behavior as constructs with poly(U) junctions. Helix length effects are only observed at intermediate salt concentrations (20-100mM KCl).

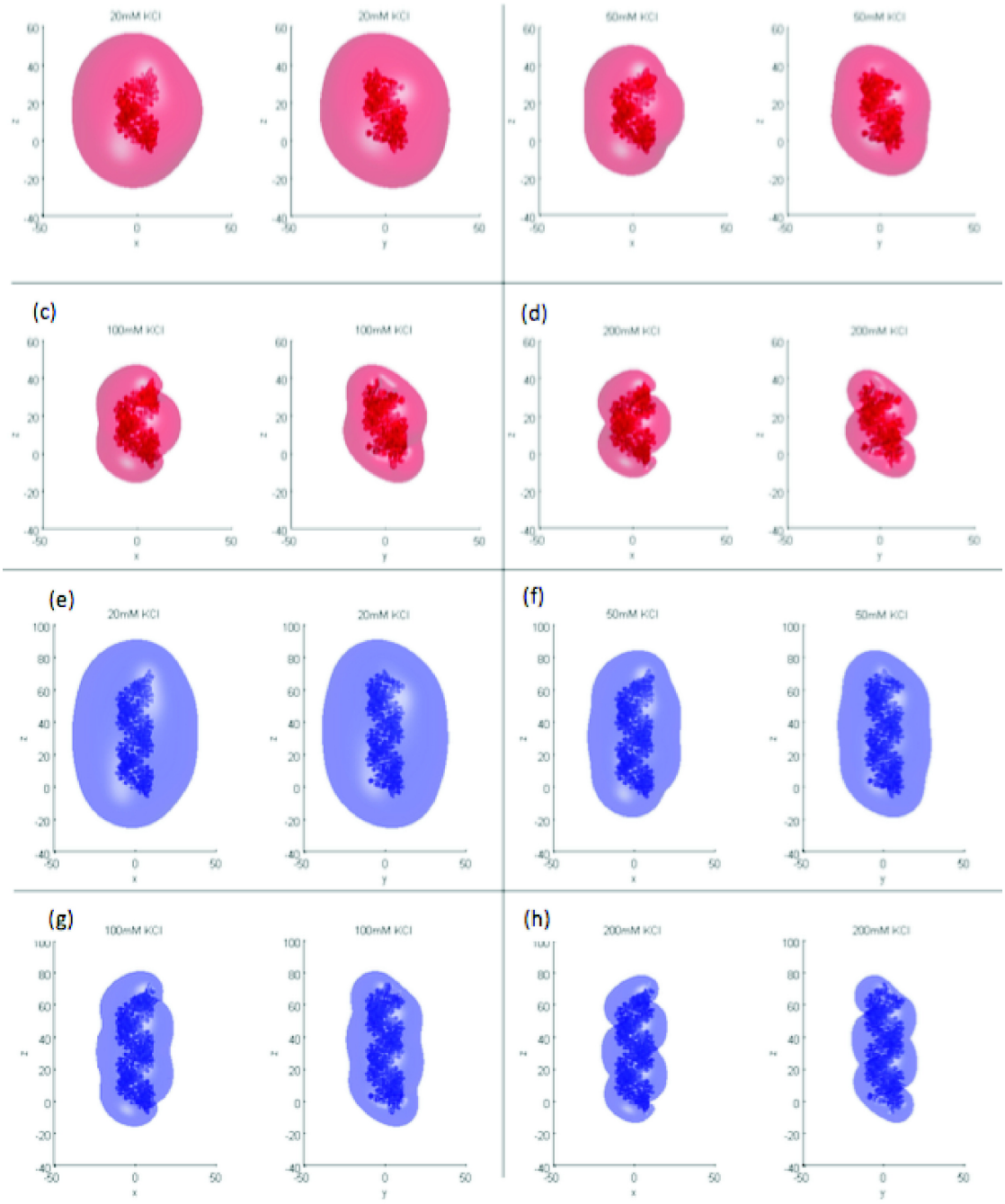

Figure C.11: Potentials are visualized using $-0.5k_BT/e$ isosurfaces. Each panel shows the same isosurface from two angles. Panels (a)-(d) show the 12bp helix in increasing [KCl] (20, 50, 100, and 200 mM) and panels (e)-(h) show the 24bp helix in increasing [KCl].

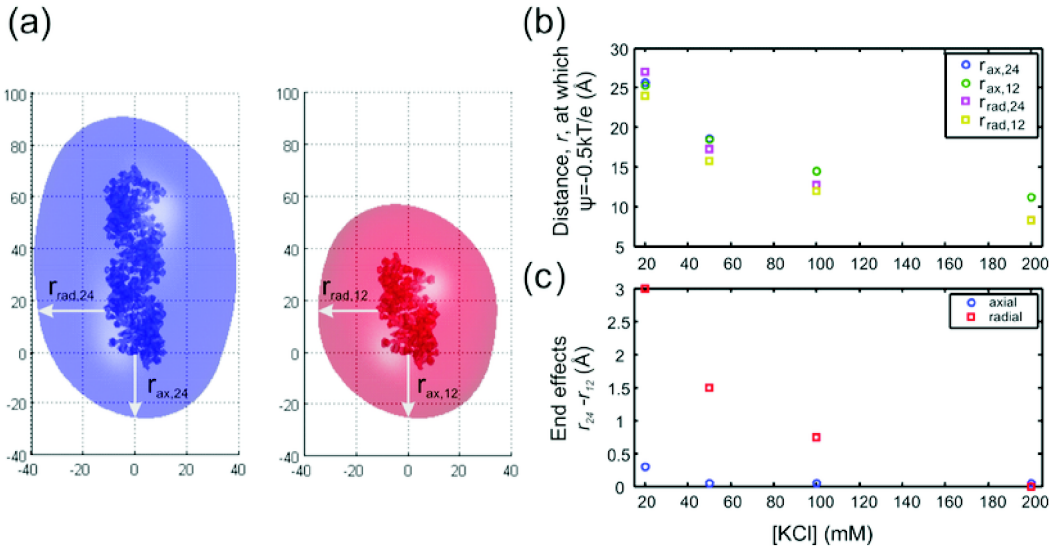

Figure C.12: Comparison of end effects for 12 and 24bp helices. (a) Schematic showing relevant lengthscales. To compare the two helix lengths, we cylindrically average the electrostatic potential to get $\psi(r, z)$. We then define a characteristic distances, r_{ax} and r_{rad} as the distance from the RNA surface to where $\psi(r, z) = -0.5 k_B T/e$. (b) Comparison between r_{ax} and r_{rad} for different helix lengths. r_{rad} was measured at $z = 16$Å which corresponds to halfway along the 12bp helix. (c) End effects as illustrated by the non-zero difference between r_{ax} and r_{rad} for 12 and 24bp helices. Notice how the magnitude of the end effects decreases as [KCl] is increased and how differences in the axial potential are negligible compared to the radial direction.

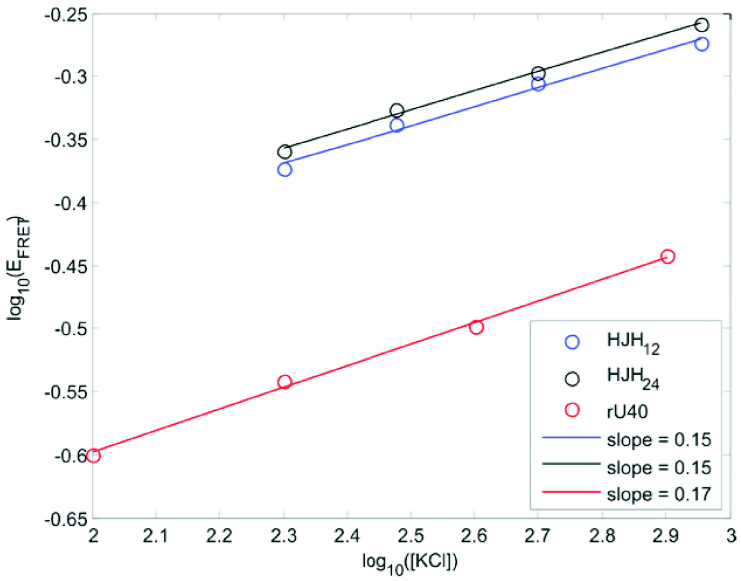

Figure C.13: High salt scaling of poly(U)$_{40}$ compared to single-stranded poly(U)$_5$ junction flanked by two helices. Data from Chen et al. [8] was converted into E_{FRET} using their published R_0. On a log-log plot, the scaling of E_{FRET} in KCl is the same (within ~10%).

References

[1] A. Muschielok, J. Andrecka, A. Jawhari, F. Brückner, P. Cramer, and J. Michaelis, "A nano-positioning system for macromolecular structural analysis," *Nature methods*, vol. 5, no. 11, pp. 965–971, 2008.

[2] S. Sindbert, S. Kalinin, H. Nguyen, A. Kienzler, L. Clima, W. Bannwarth, B. Appel, S. Muller, and C. A. Seidel, "Accurate distance determination of nucleic acids via forster resonance energy transfer: implications of dye linker length and rigidity," *Journal of the American Chemical Society*, vol. 133, no. 8, pp. 2463–2480, 2011.

[3] I. V. Gopich and A. Szabo, "Theory of single-molecule fret efficiency histograms," in *Single-Molecule Biophysics: Experiment and Theory*, ser. Single-Molecule Biophysics, T. Komatsuzaki, M. Kawakami, S. Takahashi, H. Yang, and R. J. Silbey, Eds., vol. 146, John Wiley & Sons, Inc., 2011, pp. 245–297.

[4] A. K. Wozniak, G. F. Schröder, H. Grubmüller, C. A. M. Seidel, and F. Oesterhelt, "Single-molecule fret measures bends and kinks in dna.," *Proceedings of the National Academy of Sciences*, vol. 105, no. 47, pp. 18 337–18 342, 2008.

[5] A. C. M. Ferreon, Y. Gambin, E. a. Lemke, and A. A. Deniz, "Interplay of alpha-synuclein binding and conformational switching probed by single-molecule fluorescence.," *Proceedings of the National Academy of Sciences*, vol. 106, no. 14, pp. 5645–5650, 2009.

[6] T. J. Macke and D. A. Case, "Modeling unusual nucleic acid structures," in, vol. 682, American Chemical Society, 1997, pp. 379–393.

[7] N. A. Baker, D. Sept, S. Joseph, M. J. Holst, and J. A. McCammon, "Electrostatics of nanosystems: application to microtubules and the ribosome," *Proceedings of the National Academy of Sciences*, vol. 98, no. 18, pp. 10 037–10 041, 2001.

[8] H. Chen, S. P. Meisburger, S. A. Pabit, J. L. Sutton, W. W. Webb, and L. Pollack, "Ionic strength-dependent persistence lengths of single-stranded rna and dna," *Proceedings of the National Academy of Sciences*, vol. 109, no. 3, pp. 799–804, 2012.

APPENDIX D

SUPPLEMENTARY INFORMATION FOR CHAPTER 4

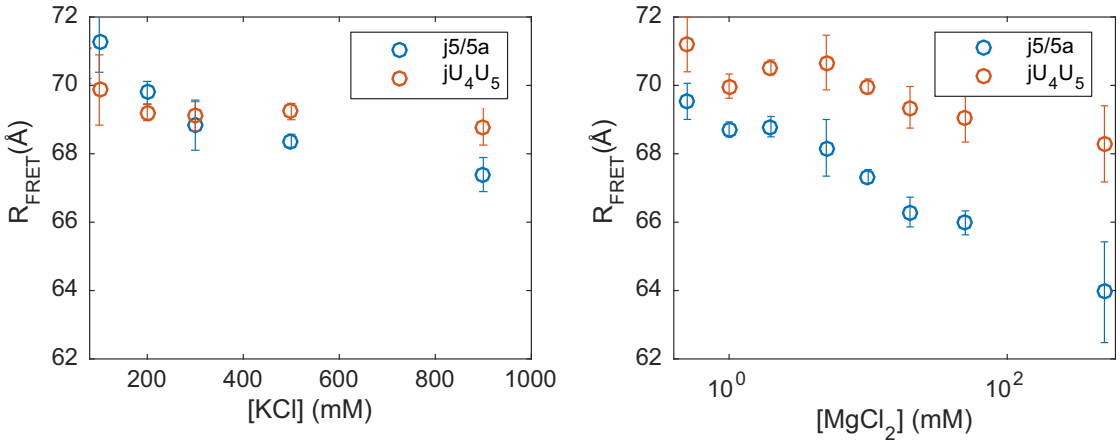

Figure D.1: Comparison between R_{FRET} for RNA junctions in (a) KCl and (b) $MgCl_2$.

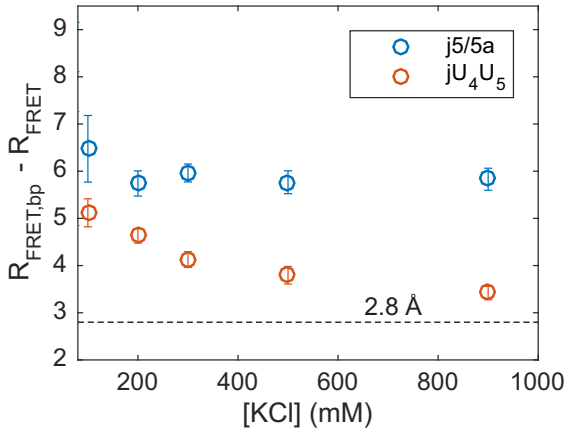

Figure D.2: Difference between R_{FRET} for the BP control and RNA junctions j5/5a and jU_4U_5 in KCl.

APPENDIX E

SUPPLEMENTARY INFORMATION FOR CHAPTER 5

Fluorescence anisotropy measurements show that Dy547 is not freely-rotating

In order to determine the collapse kinetics of tP5abc, we monitor the efficiency of Förster Resonance Energy Transfer (E_{FRET}) between a donor fluorophore (Fluorescein) and an acceptor fluorophore (Dy547, a dye that is spectrally similar to Cy3), attached to the tp5abc molecules of interest. E_{FRET} is related to the distance (R) between the donor and acceptor via:

$$E_{FRET} = \frac{1}{1 + \left(\frac{R}{R_0}\right)^6} \tag{E.1}$$

where R_0 is the distance at which $E_{FRET} = 1/2$, known as the Förster radius. R_0 depends on κ^2, a factor representing the relative orientation of donor and acceptor transition dipoles. It is often assumed that the donor and acceptor are freely-rotating, giving rise to the case where $\kappa^2 = 2/3$, and one can obtain a value for R from a measurement of E_{FRET}. However, the presence of interactions between the fluorophores and the nucleic acid can make this assumption invalid, which precludes an accurate determination of R.

To determine the validity of the $\kappa^2 = 2/3$ assumption, we performed time-resolved fluorescence anisotropy measurements of the attached fluorescein and Dy547 using the time-correlated single photon counting (TCSPC) technique. The time dependence of the fluorescence anisotropy is a function of the rotational correlation time of the fluorophore which relates to its orientational

freedom and hence κ^2. The anisotropy decay ($r(t)$) of a fluorophore linked to a macromolecule can be modeled by a two-component exponential [1]:

$$r(t) = r_0 \left(\alpha e^{-\frac{t}{\tau_F}} + (1 - \alpha) \right) e^{-\frac{-t}{\tau_M}} \tag{E.2}$$

where r_0 is the fundamental anisotropy of the fluorophore, α is the is the fraction of depolarization due to fluorophore's internal motion, τ_F and τ_M are the rotational correlation times of the fluorophore and macromolecule respectively. For a freely rotating label $\alpha \approx 1$ and $r(t)$ is dominated by the fast motion of the fluorophore while for a fixed label $\alpha = 0$ and $r(t)$ is dominated by the slow rotation of the macromolecule.

Measurements were performed on a custom fluorescence lifetime setup. Briefly, a 1060nm 80MHz laser (Fianium, Beverly, MA) was reduced to a repetition rate of 20MHz using a pulse picker (ConOptics, Danbury, CT). It was then frequency doubled to 530nm for Dy547 excitation. A 440nm 20MHz Picosecond Diode Laser (Becker & Hickl, Berlin, Germany) was used to excite the fluorescein. A standard L-format arrangement [1] was used to measure fluorescence emitted perpendicular to the excitation path. Fluorescence emission polarized parallel and perpendicular to the excitation polarization was measured separately. Emission from the sample cuvette was collimated and filtered (495/20 and 635/50 band pass filters for fluorescein and Dy547, respectively), then focused onto a microchannel plate PMT (Hamamatsu, Bridgewater, NJ) for detection. A Becker & Hickl SPC-830 photon counting card was used for data acquisition. The anisotropy ($r(t)$) was calculated using:

$$r(t) = \frac{I_{VV}(t) - G I_{VH}(t)}{I_{VV}(t) + 2G I_{VH}(t)} \tag{E.3}$$

where $I_{VV}(t)$ and $I_{VH}(t)$ are the vertically and horizontally polarized emission,

respectively, of the fluorophore with vertically polarized excitation. The G factor, a correction for the sensitivity of the detector at different polarizations, was calculated using the "tail matching" method [2] on free fluorescent dyes which have anisotropy decays much shorter than their lifetimes (Alexa fluor 488 for 440nm excitation and Rhodamine B for 530nm excitation).

The anisotropy decay of the A186U mutant in 25mM Mg^{2+} is shown in Figure E.1. Similar trends were observed with no added salt and in 160mM K^+ (data not shown). From the fit to Equation E.2, the fluorescein decay is dominated by a fast component ($\alpha = 0.68 \pm 0.03$, $\tau_F = 0.535 \pm 0.004$ns, $\tau_M = 6.9 \pm 0.3$ns) indicating some rotational freedom. We also see rotational freedom of fluorescein in no added salt and in 160mM K^+ where we obtain $\alpha = 0.73 \pm 0.02$ and $\alpha = 0.78 \pm 0.02$, respectively. In contrast, the Dy547 decay is dominated by a slow component ($\alpha = 0.287 \pm 0.007$, $\tau_F = 0.0614 \pm 0.0005$ns, $\tau_M = 19.0 \pm 0.2$ns) indicating restriction of the fluorophore. Similarly, we measure constrained rotation of Dy547 with no added salt ($\alpha = 0.178 \pm 0.006$) and in 160mM K^+ ($\alpha = 0.242 \pm 0.006$).

This result is consistent with a scenario whereby Dy547 is rotationally constrained, likely due to stacking interactions with the nearby nucleotides in the P5c stem-loop. In fact, equilibrium fluorescence emission scans show the calculated E_{FRET} for both the tP5abc A186U mutant and wild-type decrease as a function of increasing ionic strength (Figure E.5). Assuming a constant R_0, the decrease in calculated E_{FRET} would imply that the molecule becomes more extended as ionic strength is increased, contrary to the compaction observed by fluorescence correlation spectroscopy (FCS) and by comparison of the structures of the extended and folded states of tP5abc [3, 4]. Due to the restricted motion of the Dy547, the actual value of κ^2 is unknown, hence E_{FRET} cannot be used to

quantitatively measure distances between the donor and acceptor fluorophores. However, we do observe a change in E_{FRET} values between extended and compact conformational states. A set of control experiments (below) demonstrate that measured changes in fluorescence are the result of conformational changes and thus we can still use the time-dependent change in E_{FRET} to measure the collapse timescales of tP5abc.

Equilibrium fluorescence controls

To confirm that changes in E_{FRET} report the overall collapse of tP5abc, we examined the effect of salt concentration on the emission of free donor dye, free acceptor dye and directly excited labeled acceptor. Control experiments were performed by measuring the salt-dependence of fluorescence spectra using a Cary Eclipse Fluorescence Spectrophotometer. We measured the intensity dependence on salt concentration of free fluorescein and Dy547 under the range used for kinetic measurements (0, 30mM and 160mM K^+, and 0, 1mM and 25mM Mg^{2+}). The results show minimal intensity changes for free donor dye (Figure E.6 A, B) and negligible changes in free acceptor dye (Figure E.6 C, D).

From anisotropy data, we know that the donor fluorophore is freely rotating under all salt conditions ($\alpha \geq 0.68$), and thus we do not expect any significant changes in dye photophysics to influence the tP5abc donor intensity. Therefore, we focus our attention on how the RNA environment affects the acceptor dye. We measured the fluorescence resulting from the direct excitation of Dy547 attached to the A186U mutant (Figure E.6 E, F). The decrease of Dy547 intensity when directly excited may indicate some quenching of the dye due to the

change in dye environment after addition of salt. However, even after we correct the fluorescence intensity using the above results, we still observe a change in E_{FRET} upon the addition of salt. Taken together with FCS data indicating large scale collapse, these control experiments demonstrate that our measured changes in E_{FRET} are a consequence of overall collapse of tP5abc.

Microfluidic mixer experimental details

We used a microfluidic mixer design described in reference [5] to study the collapse of tP5abc RNA. This mixer was previously used to study fast protein folding kinetics [6]. A schematic of the device is shown in Figure E.2A. The RNA starts in a low-salt buffer and mixes with a higher ionic strength solution. The mixer uses hydrodynamic focusing [7, 8]; faster flow velocities in the side channels containing the salt solution focus the slower RNA-containing center channel flow. The narrow sample jet at the mixing region facilitates fast diffusion of ions into the RNA sample. Flow in the two diagonal channels prevents pre-mixing of the side and center channels before the thin jet is formed. To demonstrate hydrodynamic focusing of the center channel flow, we show the fluorescence image of the 5-port mixing device in Figure E.2B with Alexa 488-Dextran in the center channel, Alexa 405 dye in the diagonal channels and a non-fluorescent buffer solution in the side channels. The mixing region and upper part of the observation region is shown in the figure (boxed in white).

For our RNA collapse experiments, we use RNA concentrations of 4μM and flow rates of 10 μL/min in the side channels, 0.42 μL/min in the diagonal channels and 0.4 μL/min in the center channel. Flow is driven by syringe pumps

(Harvard Apparatus, Holliston, MA). These flow rates were chosen to maximize mixer stability. Prior to every measurement set, each device is characterized using KI quenching of Alexa 488-Dextran to determine the focused jet width and the mixing efficiency [5]. The mixer performance is dependent on the focused jet width and the rate at which ions diffuse into the jet from the side channels (diffusion coefficients for the salt solutions used in this work are shown in Table E.1). We define the mixing dead time as the time it takes for the ion concentration in the center channel to reach $1 - (1/e)$ of its final value. For the above flow rates, the jet width was 1 μm, the mixing dead time for K^+ and Rb^+ ions was 140 μs, and that for Mg^{2+} and Sr^{2+} ions was 235 μs. Collapse of tP5abc is monitored by changes in the effective efficiency of energy transfer (E_{FRET}) between the two attached fluorescent labels (fluorescein and Dy547). As we have determined by fluorescence anisotropy measurements that Dy547 is not freely-rotating when attached to tP5abc (see above discussion), we are unable to use FRET as a "spectroscopic ruler" to measure of the actual size of the molecule. However, we are still able to use the changes in the effective E_{FRET} as an indication of molecular compaction. Fluorescent samples were imaged in the mixer using an Olympus Fluoview FV1000 BX61WI confocal microscope (Center Valley, PA) with an Olympus UAPO 20X water immersion objective. An excitation dichroic mirror (DM405/488/559) was used to separate laser scatter (at 488 nm) from the fluorescence signal while a SDM560 dichroic mirror was used to separate the donor and acceptor signals. Two additional emission filters BA505-525 and BA575-675, were used for the donor and acceptor channels, respectively. We simultaneously collected donor channel (D_{ch}) and acceptor channel (A_{ch}) images using the Olympus Fluoview software and used Matlab (Natick, MA) for image analysis and fitting. We calculated $E_{FRET} = A_{ch}/(D_{ch} + A_{ch})$ from fluorescence

Salt	Diffusion Coefficient (x10^{-5} cm^2 s^{-1})
KCl	1.994
RbCl	2.052
$MgCl_2$	1.250
$SrCl_2$	1.334

Table E.1: Diffusion coefficients of salt solutions used in microfluidic mixing experiments [9]

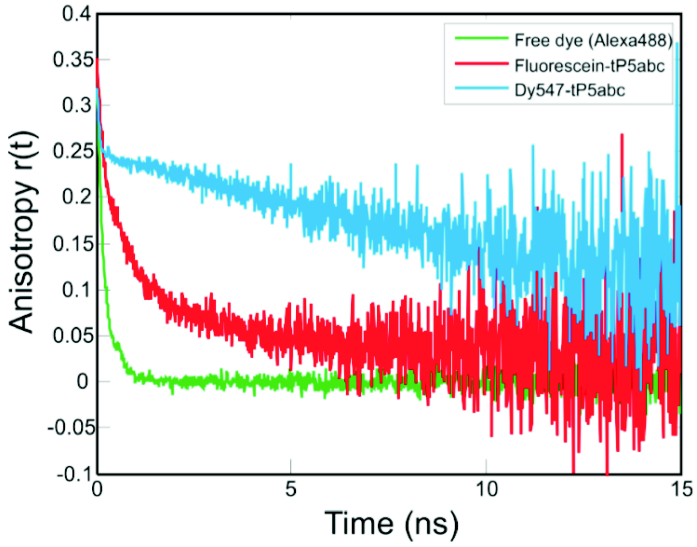

Figure E.1: Time-resolved fluorescence anisotropy decay of the fluorophore-labeled tP5abc A186U mutant in 25mM Mg^{2+}.

intensities measured per laser scan. E_{FRET} from a set of at least 100 fluorescence images of the mixer were averaged for each buffer condition. Numerical simulations of flow speed in the mixer using COMSOL Multiphysics (Burlington, MA) were used to convert the distance along the outlet channel to the time after mixing [5].

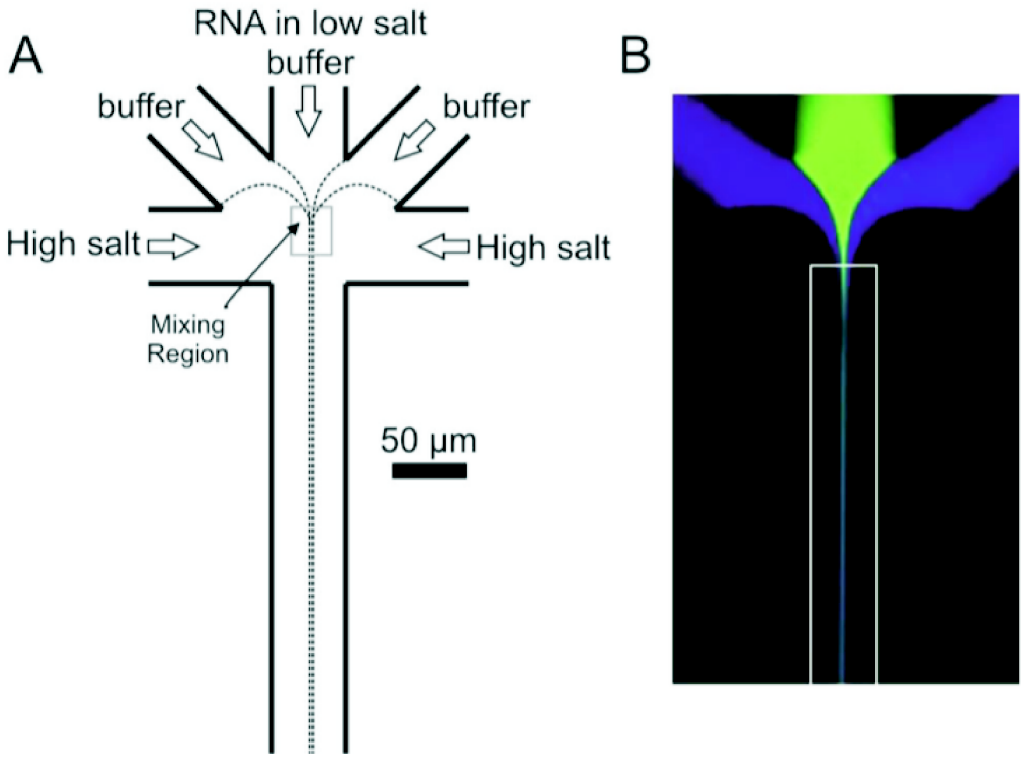

Figure E.2: We used a microfluidic mixer to change the ionic conditions and trigger RNA collapse. (A) A schematic of the mixing device. RNA in a low salt buffer (in the center channel) is rapidly mixed with a solution that contains the same buffer with added salt (in the side channels). Hydrodynamic focusing due to the fast-flowing side channels causes the narrowing of the center channel flow, facilitating fast diffusion of ions. Mixing uniformity is further improved by preventing premixing using flow of low salt buffer in the diagonal channels. Dashed lines represent a typical outline of flow patterns. All channels have a width of 50 μm and a depth of 100 μm. (B) A composite fluorescence image of flow in the mixer with Alexa 488- Dextran in the center channel, Alexa 405 in the diagonal channels and buffer (non-fluorescent) in the side channels. The white box shows the beginning of the observation region where kinetic traces are acquired.

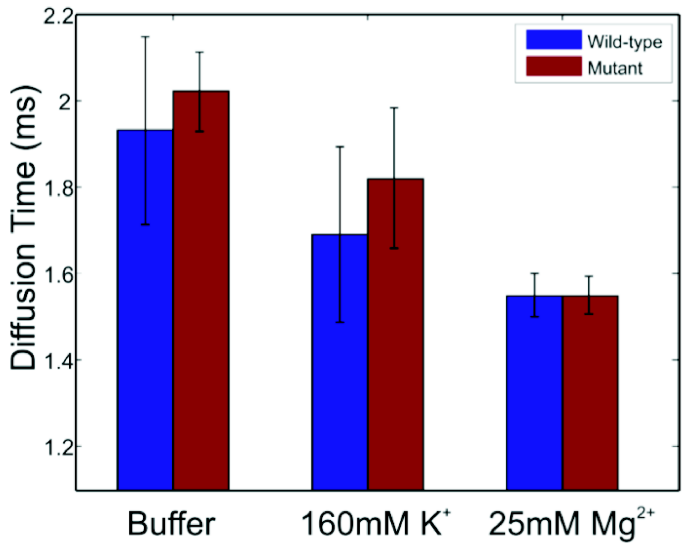

Figure E.3: Diffusion times (τ_D) of both the A186U mutant (blue) and the wild type tP5abc (red) as measured by FCS. Wild-type tP5abc and A186U mutant diffusion times are indistinguishable. Similar to the discussion in the text, we see a decrease in τ_D with the addition of salt (160mM K^+ or 25mM Mg^{2+}) in the wild type, which demonstrates the electrostatic relaxation of the molecule. Within the limitations of FCS, we cannot resolve differences between the folded state of the wild type when Mg^{2+} is present and the collapsed state of the mutant. The data in Figure E.3 were taken with a different focal volume as the data shown in Figure E.4, and in Figure 5.2B and C of the main text, and so differ slightly.

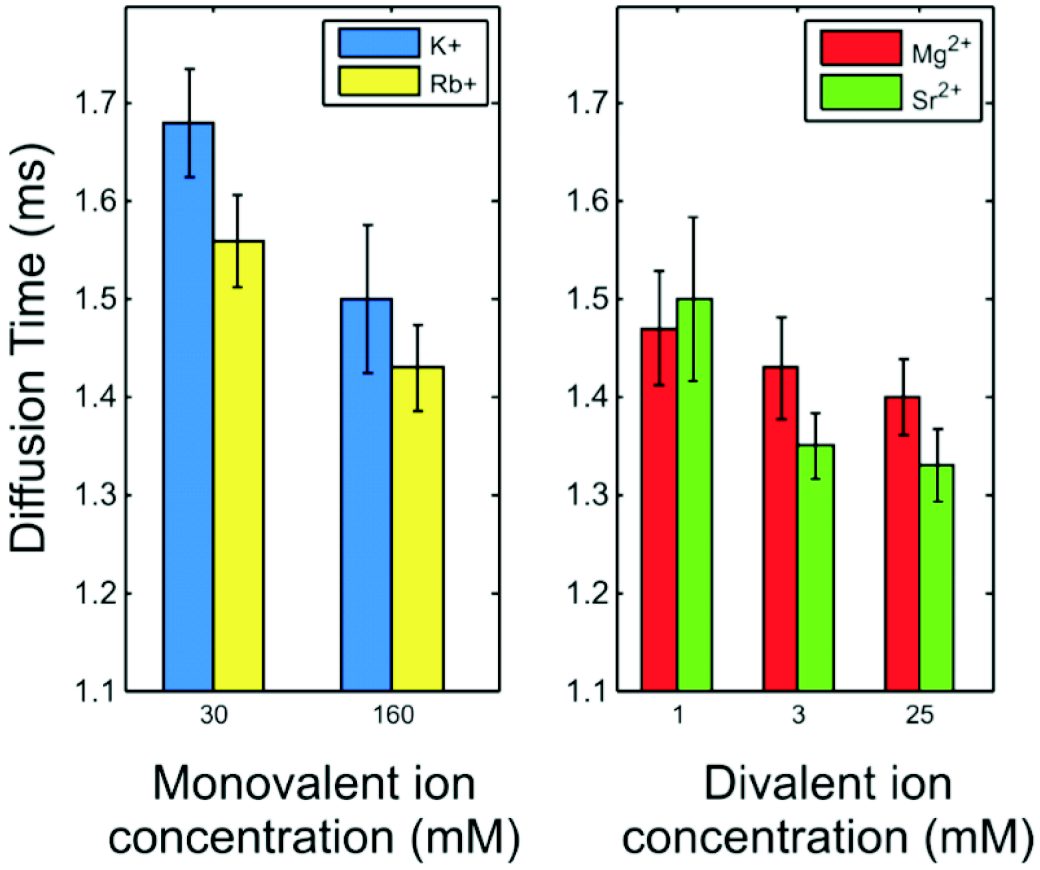

Figure E.4: Dependence of A186U mutant diffusion time on ion type. We used FCS to measure τ_D for the A186U mutant under conditions with different ion types. As before, we see a decrease in τ_D for increasing salt concentration. Similar to the results of second virial coefficients (Figure 5.1A), we cannot distinguish between K^+ and Rb^+ or between Mg^{2+} and Sr^{2+}. This result is consistent with the picture that the ion type does not affect the electrostatic collapse of tP5abc for a given ion valence.

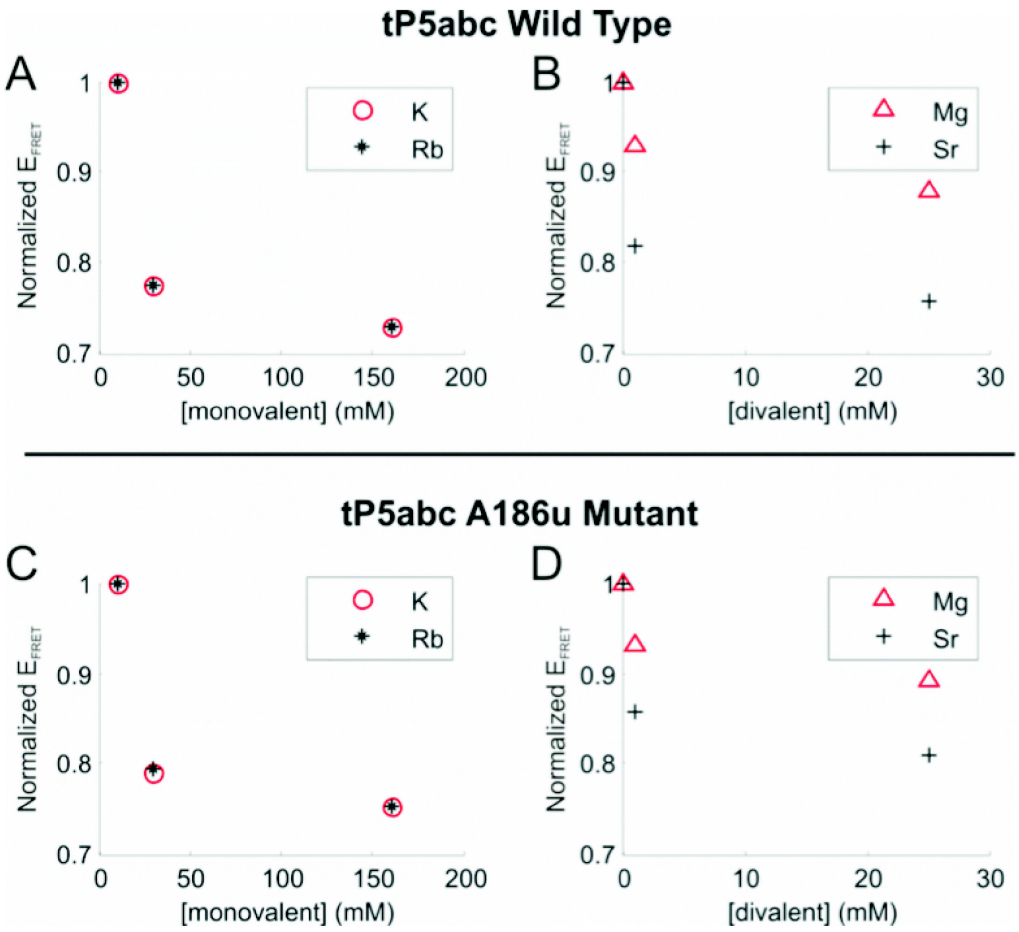

Figure E.5: Equilibrium FRET measurements for the tP5abc wild type (A,B) and A186U mutant (C,D). E_{FRET} was calculated as $A/(A+D)$ were A and D are the peak vales of acceptor and donor fluorescence emission, respectively. It was then normalized by the E_{FRET} measured with no added salt analogous to data taken in the mixer. Since the acceptor dye is not freely rotating, E_{FRET} cannot be interpreted as an intramolecular distance.

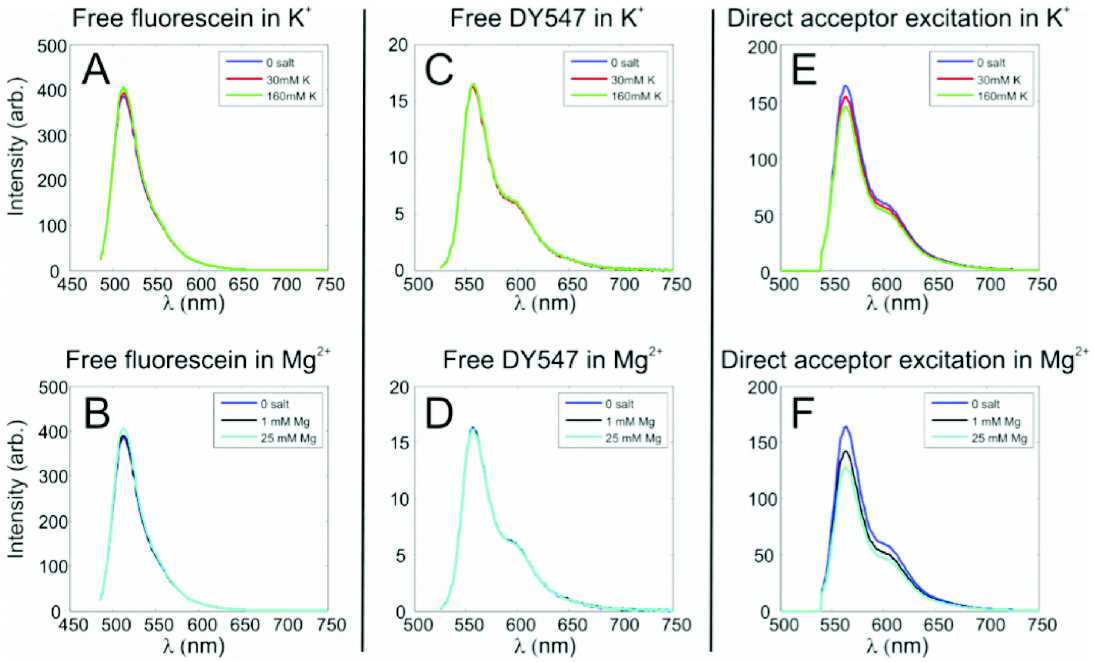

Figure E.6: Change in fluorescence intensity as a function of salt concentration for free donor dye (A, B), free acceptor dye (C, D) and directly excited acceptor dye in the A186U RNA mutant (E, F). The top panels represent monovalent ions while the lower panels use divalent ions.

References

[1] J. R. Lakowicz, *Principles of fluorescence spectroscopy*, 3rd. New York: Springer, 2006, p. 954.

[2] D. V. O'Connor and D. Phillips, *Time-correlated single photon counting*. London: Academic Press, 1984.

[3] M. Zheng, M. Wu, I. T. Jr, and I. Tinoco, "Formation of a gnra tetraloop in p5abc can disrupt an interdomain interaction in the tetrahymena group i ribozyme," *Proceedings of the National Academy of Sciences*, vol. 98, no. 7, pp. 3695–3700, 2001.

[4] J. H. Cate, R. L. Hanna, and J. A. Doudna, "A magnesium ion core at the heart of a ribozyme domain," *Nature structural biology*, vol. 4, no. 7, pp. 553–558, 1997.

[5] H. Y. Park, X. Qiu, E. Rhoades, J. Korlach, L. W. Kwok, W. R. Zipfel, W. W. Webb, and L. Pollack, "Achieving uniform mixing in a microfluidic device: hydrodynamic focusing prior to mixing," *Analytical Chemistry*, vol. 78, no. 13, pp. 4465–4473, 2006.

[6] H. Y. Park, S. A. Kim, J. Korlach, E. Rhoades, L. W. Kwok, W. R. Zipfel, M. N. Waxham, W. W. Webb, and L. Pollack, "Conformational changes of calmodulin upon ca2+ binding studied with a microfluidic mixer.," *Proceedings of the National Academy of Sciences*, vol. 105, no. 2, pp. 542–547, 2008.

[7] J. B. Knight, A. Vishwanath, J. P. Brody, and R. H. Austin, "Hydrodynamic focusing on a silicon chip: mixing nanoliters in microseconds," *Physical Review Letters*, vol. 80, no. 17, pp. 3863–3866, 1998.

[8] L. Pollack, M. W. Tate, N. C. Darnton, J. B. Knight, S. M. Gruner, W. A. Eaton, and R. H. Austin, "Compactness of the denatured state of a fast-folding protein measured by submillisecond small-angle x-ray scattering," *Proceedings of the National Academy of Sciences*, vol. 96, no. 18, pp. 10 115–10 117, 1999.

[9] W. M. Haynes and D. R. Lide, *CRC handbook of chemistry and physics: a ready-reference book of chemical and physical data*, 92nd ed. Boca Raton, FL: CRC Press, 2011.