The Thesis committee for Alan Sluder certifies that this is the approved version of the following thesis:

Models of RNA Folding in Planetary Environments

APPROVED BY

SUPERVISING COMMITTEE:

Supervisor: _____

John Scalo

_____

Milos Milosavljevic

Models of RNA Folding in Planetary Environments

by

Alan Sluder, B.S.

Thesis

Presented to the Faculty of the Graduate School

of the University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Arts

The University of Texas at Austin

August 2011

Models of RNA Folding in Planetary Environments

by

Alan Sluder, M.A.

The University of Texas at Austin, 2011

SUPERVISOR: John Scalo

Abstract

Multiple lines of evidence suggest that RNA performed all of the biological functions in the first life forms on earth. These functions included cleavage, ligation, polymerization, recognition, binding, and replication. In order to perform these functions, populations of RNA molecules with unevolved sequences must have been able to fold into compact three dimensional shapes, in unregulated environments, and without the help of proteins. Folding into compact tertiary structures is difficult because of the high charge density of RNA. Consequently, the ranges of temperature, salinity, pH, and pressure that allow RNA to fold into functional shapes is very restricted. We use thermodynamic arguments and Brownian dynamics simulations to compute the range of these environmental parameters that will allow RNA to fold. This is a non-

trivial calculation due to the formation of an ion atmosphere around RNA that reduces its electric field. The results can be used to clarify the environments in which the transition to life is possible. Our preliminary calculations suggest that environments with low temperatures ($0 - 50°C$) and high salt concentrations (greater than 100mM) are the most favorable for unassisted RNA folding and thus the transition to RNA-based life. Applications of our results include determining the environments on early earth where life formed, assesing the habitability of Europa, Titan, and (using modeled parameters) extrasolar planets.

Contents

1 Introduction

Although planets in the traditional habitable zone can maintain stable bodies of liquid water (Kasting and Catling D. 2003), life will not be able to originate on these planets if biopolymers cannot form on them . Further, these biopolymers must be able to spontaneously fold into compact three dimensional conformations in order to carry out their prebiological functions. They must be able to do this without proteins, in unregulated environments, and with unevolved sequences. If none of the environments on a planet allow for the formation and folding of biopolymers, the transition to life cannot take place on that planet. Significant work has been done on determining the conditions in which biomolecules can form, but less work has been done on determining the conditions in which biopolymers can fold (Orgel 2004; Moulton 2000).

Evidence is mounting that the first biopolymer was RNA. Unlike DNA and proteins, RNA can replicate itself (Lincoln and Joyce 2009). Further, RNA can store information in its nucleotide sequence like DNA and can catalyze chemical reactions like proteins (Joyce 2002). Ribozymes (RNAs that can catalyze chemical reactions) have been evolved in the lab from random sequences (Joyce 2004). The structure of the ribozyme, which combines amino acids into proteins, suggests that the primordial ribozyme was made entirely of RNA, with proteins added later for

structural support only (Bokov and Steinburg 2009). RNA can also bind with many different types of molecules, allowing it to act as a sensor and thus regulate gene expression (Butler et al. 2011). Ribozymes also exist in nature, especially in plants, bacteria, and viruses (Pena and Garcia 2010). All of this strongly suggests that the current DNA-RNA-protein world emerged from a primodial RNA world. For these reasons, we will assume that RNA was the first biopolymer not only on earth, but on any planet that develops a DNA-RNA-protein based biosphere.

As with any polymer, RNA must be able to fold into compact conformations in order to have biological or prebiological functions. Because RNA is highly negatively charged, there is a strong electrostatic barrier to folding. Calculating the folding time of RNA requires knowing the electric field around RNA, which is complicated by the fact that RNA attracts positive ions, which have their own electric field that must be included in the analysis. We develop a model of RNA and its ion atmosphere, and use it to compute the folding times for RNAs in various environments.

We proceed as follows. In section 2 we discuss the stages of RNA folding, the electrostatic barrier to folding, and the formation of the ion atmosphere. In section 3 we quantify the electrostatic barrier to RNA folding and use the results to compute the melting temperature of RNA as a function of salinity. In section 4 we

develop a coarse-grained model of RNA that allows us to compute the properties of the ion atmosphere around RNA as well as the RNA folding times. We conclude in section 5.

2 RNA Folding

2.1 Primary Structure

The building block of RNA is the ribonucleotide, which is made by connecting a ribose sugar with a base (either adenine, uracil, guanine, or cytosine) to make a nucleoside, and then connecting the nucleoside to a phosphate. The ribonucleotides are connected to make an RNA strand, called an oligonucleotide. The sequence of bases along the chain (which stick out of the ribose-phosphate backbone) is called the primary structure of the oligonucleotide (see Figure 1).

2.2 Secondary Structure

Thermal fluctuations cause the conformation of the chain to continuously change. When bases distant on the chain approach each other due to these conformational changes, they can form hydrogen bonds (called base pairing). Adenine forms 2 hydrogen bonds with uracil, and guanine forms 3 hydrogen bonds with cytosine. Base pairs are flat, and adjacent base pairs can stack one on top of the other. These pairs are held in place by dispersion forces (caused by correlated electron fluctuations) and $\pi - \pi$ interactions (caused by overlap of p orbitals). Base pairs will form and break until the lowest free energy configuration is reached, called the

secondary structure. At this point the RNA consists of stiff base paired regions called stems (which are shaped like a double helix) and flexible single stranded regions called loops. Loops that begin and end on the tip of a helix are called terminal or hairpin loops, loops that connect two helices are called junction loops, and loops that form within helices are called bulge loops (lone single stranded regions) or internal loops (two single stranded regions). RNAs made of n helices connected by a junction loop are called n-helix junctions. We will focus on three helix junctions because they are the basic functional units of RNA structure.

2.3 Tertiary Structure

The building blocks of secondary structure (helices and loops) are themselves subject to thermal fluctuations. The helices and loops can make contacts with each other (tertiary contacts) as they diffuse around. Coaxial stacking occurs when the end base pairs of one helix stacks with the end base pairs of another helix forming a long psuedohelix. Docking occurs when two helices are parallel and side by side, and this configuration could be stabalized by base pairing between terminal loops (kissing hairpin), hydrogen bonding between the ribose-phosphate backbones (ribose zipper), or inserting an adenine from a loop into the minor groove of a helix. A 3HJ is in its functional conformation when two of the helices are coaxially

stacked and two are docked.

2.4 Electric Charge

The phosphate group present on each ribonucleotide contains four oxygen atoms, and is thus very electronegative. The result of this high electronegativity is that electrons that are being shared between the oxygen atoms and other atoms will be drawn toward the oxygen atoms, thus making the bond polar. One of the oxygen atoms on the phosphate group is bonded to a hydogen atom. Because of the electronegativity effect, the two electrons in the bond spend most of their time near the oxygen. The result is that the hydrogen is barely bound to the oxygen at all. In fact, when a water molecule collides with the O-H group it will steal the proton and leave the two bonding electrons. The result is that the phosphate group becomes negatively charged (it lost one proton but no electrons) and a water molecule has been converted into a hydronium H_3O^+. Hence all ribonucleotides have a charge of $-e$, where e is the proton charge.

The charge on a ribonucleotide depends on the pH (the pH increases as the concentration of H_3O^+ decreases). At low pH there are so many hydronium molecules in the surrounding medium that the phosphate group does not become negatively charged. At high pH there are so few hydronium molecules that other

groups on the RNA become negatively charged.

2.5 The Ion Atmosphere

Because RNA is negatively charged, it attracts positive ions (cations) and repels negative ions (anions). The cations form an ion atmosphere around the RNA, similar to the earth's atmosphere but with an osmotic force gradient balancing electrostatic attraction instead of a pressure gradient balancing gravity. The electric field from the cations partially cancels the electric field from the phosphates, reducing the electrostatic repulsion of the phosphates.

We can derive an approximate formula for the reduced ("screened") electrostatic potential of a phosphate by assuming that the ions form a continuous fluid and are in thermodynamic equilibrium. If the number density at some point of cations of charge z_+e is n_+ and the number density of anions of charge $-z_-e$ is n_-e then the charge density at that point is $\rho = (z_+n_+ - z_-n_-)e$. The charge density and electrostatic potential ϕ is related by Poisson's law:

$\nabla^2\phi = -\rho/\epsilon\epsilon_0 = -(z_+n_+ - z_-n_-)e/\epsilon\epsilon_0$, where ϵ is the dielectric constant and ϵ_0 is the permittivity of free space. If the ions are in thermodynamic equilibrium, their number densities are $n_+ = n_{+0}e^{-z_+e\phi/kT}$ and $n_- = n_{-0}e^{+z_-e\phi/kT}$, where k is Boltzmann's constant, T is the temperature, and the subscript 0 refers to the bulk

concentration (concentration far away from the RNA). The result is the nonlinear Poisson-Boltzmann equation for the screened electrostatic potential:

$$\nabla^2\phi = -\frac{(z_+ n_{+0} e^{-z_+ e\phi/kT} - z_- n_{-0} e^{+z_- e\phi/kT})e}{\epsilon\epsilon_0} \tag{1}$$

If we assume that the energy of each ion is much less than the thermal energy we can expand the exponentials: $e^{-z_+ e\phi/kT} = 1 - z_+ e\phi/kT$ and $e^{+z_- e\phi/kT} = 1 + z_- e\phi/kT$. If we further assume that the salt is monovalent (each ion has charge $\pm e$) then $z_- = z_+ = 1$ and $n_{+0} = n_{-0} = n_0$. If we further assume that the electrostatic potential is spherically symmetric (because we want the electrostatic potential of a point charge) we can replace $\nabla\phi$ with $r^{-2}(\partial/\partial r)(r^2\partial/\partial r)$. The result is the linear Poisson Boltzmann equation:

$$\frac{2}{r}\frac{\partial\phi}{\partial r} + \frac{\partial^2\phi}{\partial r^2} = \lambda_D^{-2}\phi \tag{2}$$

$$\lambda_D = \sqrt{\frac{\epsilon\epsilon_0 kT}{2e^2 n_0}} \tag{3}$$

where λ_D is the Debye length. The solution, subject to the conditions that $\phi(r = \infty) = 0$ and $\phi(\lambda = \infty) = q/(4\pi\epsilon\epsilon_0)$ (where q is the charge of the particle whose potential we want) is:

$$\phi(r) = \frac{q}{4\pi\epsilon\epsilon_0 r} e^{-r/\lambda_D} \tag{4}$$

This is the Debye-Huckel limiting law. Remeber that two important assumptions went into this formula: the ions form a continuum and the electrostatic energy of

an ion is less than the thermal energy. Monovalent ions do behave as a continuum, but multivalent ions have correlated thermal fluctuations in their positions that reduce the electrostatic potential below the Debye-Huckel predictions. While the electrostatic energy of a unit charge around RNA and DNA is greater than kT, the Debye-Huckel approximation is accurate enough that researches use it for DNA simulations in monovalent salt (Doi 2010). To get the electostatic potential around RNA in a multivalent salt, one must use simulations that follow the motion of each ion (molecular dynamics, Brownian dynamics, or Monte Carlo).

Ion concentrations are typically measured in milimoles (mM), where 1mM is 6.02×10^{23} ions per cubic meter or 55,000 water molecules per ion.

3 Barriers to RNA Folding

3.1 Introduction

We will focus on 3HJs, and assume that the 3HJs can be in one of four "macrostates," or configurations: coil (no base pairing), open (secondary structure only), stacked (two helices coaxially stacked), or docked (two helices coaxially stacked and a third parallel to the stack). See Figure 2. Each state has a free energy $G = H - TS$, where H is the enthalpy and S is the entropy. H is the energy from hydrogen bonds (contributes negative H), dispersion forces (contributes negative H), and electrostatic interactions (contributes negative H for opposite charges and positive H for like charges). S depends on Ω, the number of conformations that correspond to a given macrostate (multiplicity): $S = k\ln\Omega$. The coil state has the highest Ω, followed by the open state, the coaxially stacked state, and the docked state. Both H and S are assumed independent of temperature. The most stable state is the state with the lowest free energy.

3.2 Temperature

Because multiplicity increases as RNA unfolds, the unfolded state is favored at high temperatures. This can be seen by taking the high temperature limit of

$G = H - TS$, which gives $G_{HOT} = -TS$. At high temperatures, entropy determines the free energy, and the state with the highest entropy (coil) will have the lowest free energy (most stable). In the low temperature limit $G_{COLD} = H$, so the state with the lowest enthalpy (the docked state because it has the most hydrogen bonds and base stacks) is the most stable. This suggests that if you started a with a population of docked 3HJs at a low temperature, and increased the temperature, the 3HJs would transform from docked to stacked to open to coil. Indeed this is the case and this process is called RNA melting. The temperature that makes the free energy of two states equal is called the melting temperature for the transition between the two states.

Melting temperatures for various transitions are measured in experiments. For helix melting experiments, we start with an ensemble of helices at low temperatures and heat the ensemble until all of the helices have dissociated into two separate strands. The temperature at which half of the helices dissociates is the melting temperature. We can represent this symbolically by HELIX$\rightleftharpoons$COIL+COIL. However, in a real RNA the process HELIX$\rightleftharpoons$COIL occurs, because the helix was made by a single strand not two separate strands. Similarly, in coaxial stacking experiments a coaxial stack of two helices is heated until they separate into two free helices, but in a real RNA the two helices would still be part of the same molecule.

Because these experiments dissociate one object into two, the melting temperature is concentration dependent. By measuring the melting temperature at various concentrations, we can determine the enthalpy and entropy change of the reaction. Let the degree symbol denote a quantity measured at one mole per liter concentration. Then for the process A+B$\rightleftharpoons$AB we have $\Delta G = \Delta G^\circ + RT\ln[\mathrm{AB}]/[\mathrm{A}][\mathrm{B}] = \Delta H^\circ - T\Delta S^\circ + RT\ln[\mathrm{AB}]/[\mathrm{A}][\mathrm{B}]$. At the melting temperature we have $\Delta G = 0$ and [AB]=[A]=[B] so

$$\frac{1}{T_m} = \frac{\Delta S^\circ}{\Delta H^\circ} + \frac{R}{\Delta H^\circ}\ln[\mathrm{A}] \tag{5}$$

By plotting $1/T_m$ versus ln[A], we can find ΔH from the slope and ΔS from the y-intercept. These experiments are done in 1M Na^+. We can construct a two dimensional space whose coordinates are $(-\Delta H, -\Delta S)$. Every transition corresponds to a point in this plane. In Figure 2 we have plotted many different experimental transitions as points in the enthalpy-entropy plane.

We need to convert the measured ΔH and ΔS to a melting temperature that is relevant to a 3HJ. To do this we note that the components of a 3HJ are free to move in a volume L^3, where L is the length of a helix. Since typical helix lengths are around 3nm, typical volumes are around 113nm^3. In the formula for melting temperature we need the concentration of one component when half of the components have melted. There is 1/2 such components per volume, so the

concentration is $(1/2)/113\text{nm}^3$=.00442nm^{-3} or about 7mM. Thus the melting temperature is

$$T_m = \frac{\Delta H^\circ}{\Delta S^\circ - 0.00986\,\text{kcal/mol/K}} \tag{6}$$

3.3 Coulomb Barrier

3.3.1 Introduction

Because RNA is negatively charged, folding into a compact state is equivalent to diffusing across a Coulomb barrier. We can calculate the electrostatic potential and electric field around RNA as follows. Assume the charge distribution of the helical regions is that of an ideal double A-form helix (both strands are coaxial and have radius 0.88nm and pitch 0.281nm, one strand begins 0.188nm above the other). Neglect the charges in the single stranded regions. Then orient the helices as seen in Figure 1. The electrostatic potential at a given point is

$$\phi(r) = \sum_i \frac{e}{4\pi\epsilon\epsilon_0 r_i} e^{-r_i/\lambda_D} \tag{7}$$

where r_i is the distance between charge i and point r and the sum is over all charges. We have assumed that only monovalent ions are present so that we can use the Debye-Huckel limiting law. The electric field is $\mathbf{E} = -\nabla\phi$. We plot in Figure 3 the electrostatic potential and the electric field around 3HJs with 12bp

helices in the open, stacked, and docked states with no salt ($\lambda_D = \infty$).

3.3.2 Calculations

It is informative to compute the change in electrostatic energy that occurs when the RNA changes configurations. In particular, for the open to stacked and stacked to docked transitions. The electrostatic energy of a given state is

$$E(r) = \sum_i \sum_j \frac{e^2}{4\pi\epsilon\epsilon_0 r_{ij}} e^{-r_{ij}/\lambda_D} \tag{8}$$

where r_i is the distance between charges i and j and the sum is over all pairs of charges. We plot the change in electrostatic energy as a function of helix length and [Na^+] for the open to stacked and stacked to docked transitions in Figure 4. The melting temperature is a function of [Na^+]. We have already computed an expression for the melting temperature, but we need to correct it by adding the electrostatic energy of the transition to the enthalpy:

$$T_m = \frac{\Delta H^\circ + \Delta H_{el} - \Delta H_{el}([\mathrm{Na}^+] = 1M)}{\Delta S^\circ - 0.00986\,\mathrm{kcal/mol/K}} \tag{9}$$

Since ΔH° is measured at 1M NaCl and thus includes the electrostatic energy at 1M, we have subtracted our calculated electrostatic energy at 1M Na^+ so that our calculations will agree with the experiments at that concentration. In Figure 5 we plot the melting temperature as a function of [Na^+] for the data points in the Figure 2.

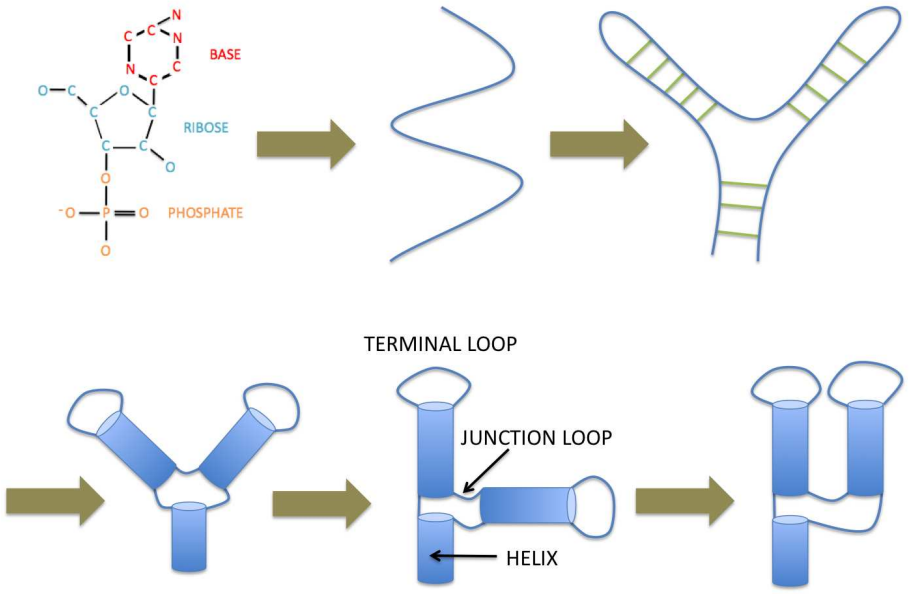

Figure 1: Folding of a three-helix junction (3HJ). *Top Left* Structure of a ribonucleotide. *Top Middle* Ribonucleotides polymerize into a linear molecule known as a coil. *Top Right* Regions of the RNA base pair and base stack with other regions of the same molecule. *Bottom Left* Open configuration of a 3HJ. The regions of the RNA that have base paired are stiff and represented by cylinders, while the unpaired regions are flexible and represented by tethers. *Bottom Middle* The stacked configuration of a 3HJ. *Bottom Right* The docked configuration of a 3HJ.

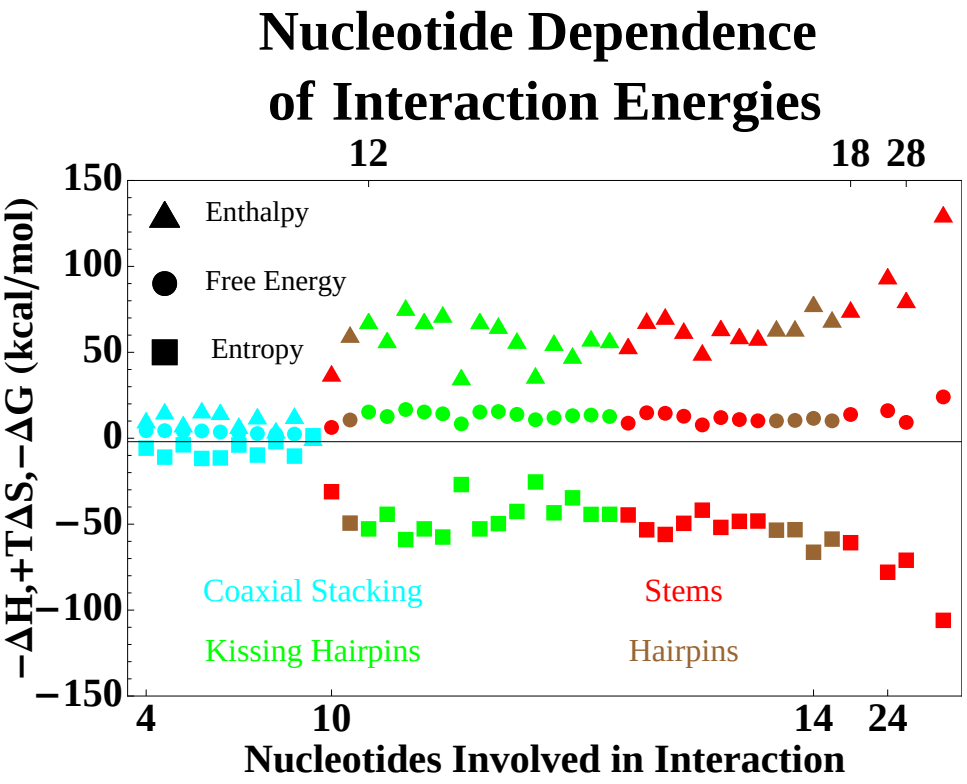

Figure 2: Change in entropy (squares), enthalpy (triangles), and free energy (circles) during RNA transitions. Cyan symbols represent experimental values for coaxial stacking (Walter and Turner 1994). Green symbols represent experimental values for kissing hairpins, which is a type of docking in which the terminal loops of two helices form base pairs with each other (Weixlbaumer et al. 2004; Lorenz et al. 2006; Tan and Chen 2010). Brown symbols represent the melting of a hairpin (helix with a loop on top) and are based on Turner's rules (values taken from Tan and Chen 2008). Red symbols represent the melting of a helix without a terminal loop and are also based on Turner's rules (Tan and Chen 2007).

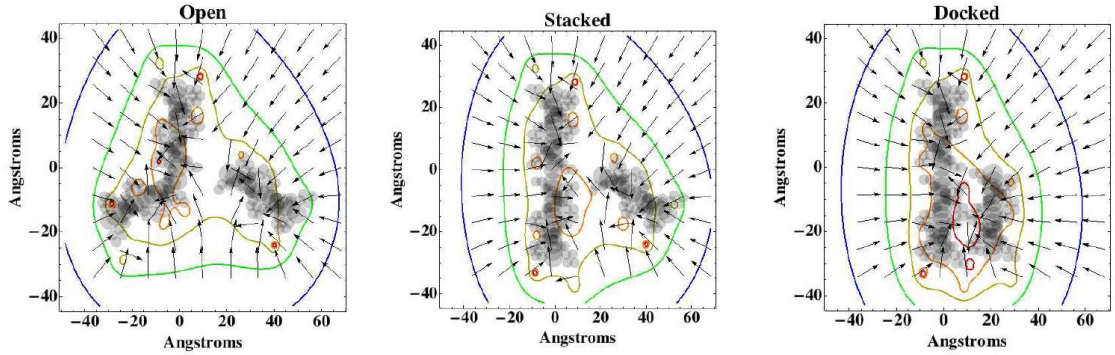

Figure 3: Electrostatic potential and electric field of a 3HJ. The direction of the electric field is shown by the arrows. The colored lines are lines of conatant electrostatic potential (in units of -e/kT, red → 30, orange → 25, yellow → 20, green → 15, blue → 10). The location of the helices is shown in transparent black.

3.3.3 Discussion

We see in Figure 3 that RNA is surrounded by a strong electric field that points toward it (because it is negatively charged). The magnitude of the electrostatic potential increases as we approach the molecule, to values as high as $30e/kT$. This means that bringing a single negative charge (equal to the electron charge) from far away to near the RNA would require doing $30kT$ worth of work. Since this is so much higher than kT, we see why it will be a problem for thermal fluctuations to bring the negatively charged components of RNA together.

We can see that the magnitude of the electrostatic potential, and thus the magnitude of the electrostatic potential energy, increases as the RNA progresses from open to stacked to docked. In Figure 4 we can see how much the electrostatic potential energy actually changes when we transition between the states. It takes more energy to dock than stack since stacking only reqires that the tips of the helices touch, but docking requires that a whole side of one helix is adjacent to another.

The folding energies are very much dependent on the sodium concentration and helix length. Folding energies incrase with increasing helix length because larger helices have more charges coming together. Folding energies decrease with increasing sodium concentration because the higher the sodium concentration the more effectively the negative charges are screened. If the sodium concentration exceeds a critical value, increasing the helix length does not change the folding energy because the additional charges are far enough apart to be screened by the salt.

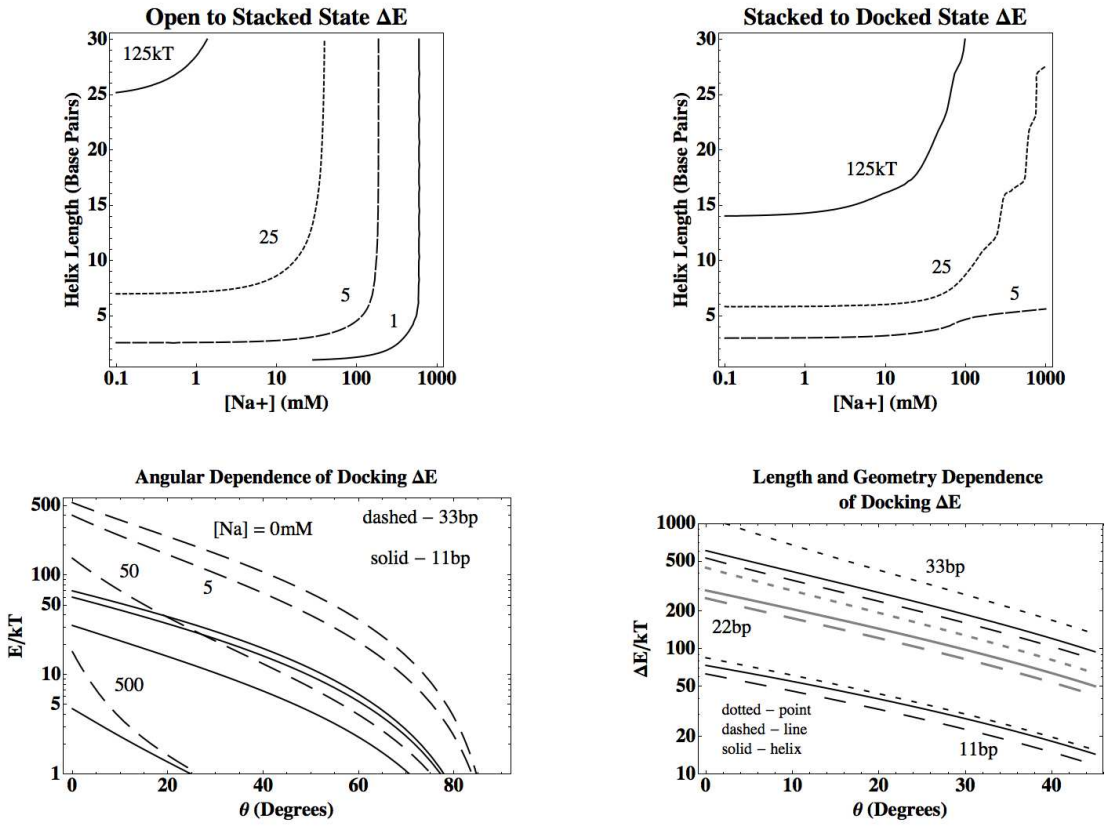

Figure 4: Electrostatic Folding Energies of Docking and Stacking. *Top Left* Change in electrostatic energy in going from the open state to the stacked state as a function of sodium concentration and helix length. *Top Left* Change in electrostatic energy in going from the stacked state to the docked state as a function of sodium concentration and helix length. *Bottom Left* Electrostatic energy as a function of angle between helix A and C (relative to 90°). Dashed line for 33bp helices and solid line for 11bp helices. Computed for 4 values of sodium concentration (from top to bottom: 0mM, 5mM, 50mM, 500mM). *Bottom Right* We replace the helix (solid line) with a point charge (dotted line) and a uniform charged line (dashed). Helix length is varied, and sodium the concentration is zero.

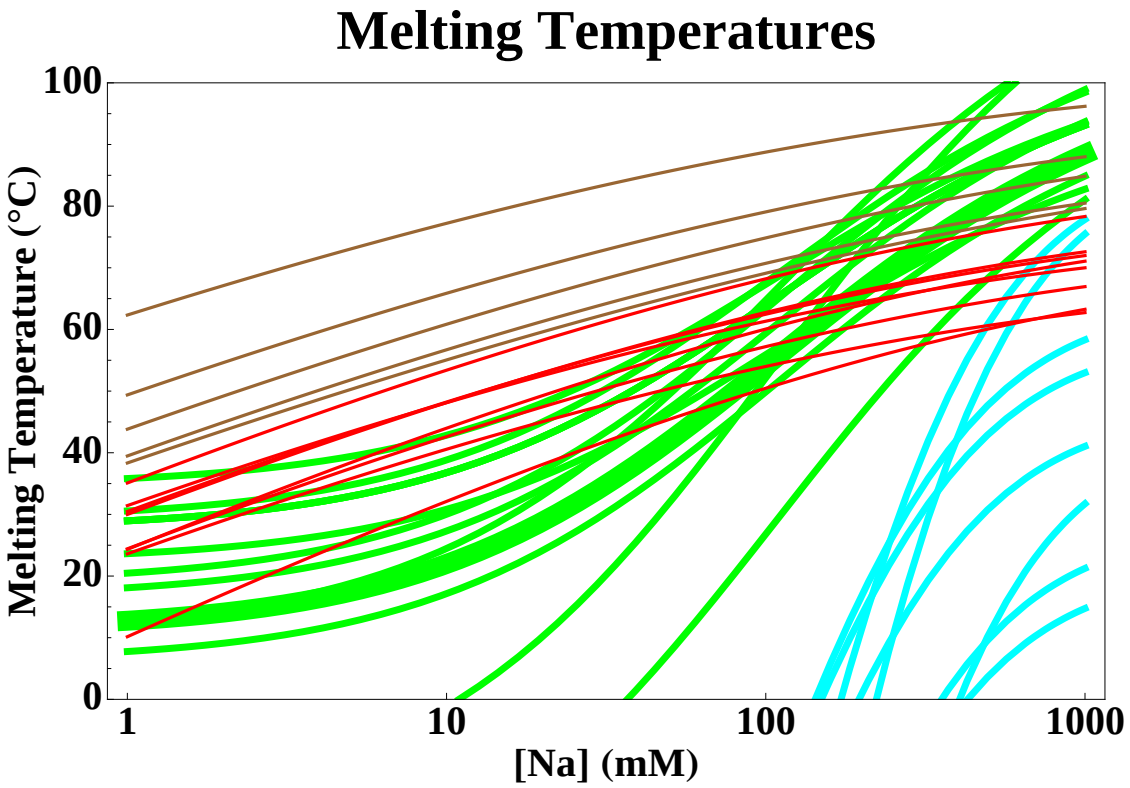

Figure 5: The melting temperatures for each of the tranitions in Figure 2 were computed as a function of sodium concentration with the assumptions outlined in the text. The color code is the same as Figure 2 (brown - hairpin, red - stem, green - kissing hairpin, cyan - stacking). Hairpin and stem melting temeratures are calculated using the TBI model.

The bottom left panel shows how the energy of a stacked 3HJ increases as the non-stacked helix changes its angle from 90° (perpendicular to the stack) to 0° (parallel to the stack). As the angle decreases, the energy increases because negative charges are approaching each other. For low salt concentrations the energy steadily increases, but for higher salt concentrations the energy increases slowly at first (because all charges are separated by more than the Debye length) then increases more rapidly as the charges become closer than the Debye length.

In the above calculations the charge on the helices was distributed in the shape of an A-form helix. In order to determine how important the geometry of the charge distribution is to the electrostatic energy, we repeated the calculation of energy versus angle for a stacked three helix junction, but with three different charge distributions: a single point charge at the center of mass of each helix (dotted line), a line charge with the same length as the helix and coxial with the helix (dashed line), and the usual A-form helix (dotted line). The results are plotted in the bottom right panel of Figure 4 (for no salt). Short helices are well represented as point charges because they are roughly spherical. Long helices are well represented by line charges because they are roughly linear. The fact that replacing the helical charge distribution with a point charge gives similar electrostatic energies for short helices suggests that we can use a single point charge per helix in our most coarse

grained dynamical simulations of RNA folding.

Using the results from the top panels of Figure 4 (gives us ΔH_{el} for each helix length and sodium concentration) and Equation 9, we computed the melting temperature as a function of sodium concentration for each of the data points in Figure 2 (we assumed that stacking and docking melting experiments were done with 6bp helices, and the length of the hairpin and stems in their melting experiments are given in the references). The results are plotted in the right panel of Figure 2. The slope of the melting temperature curve in the $T_m - [\mathrm{Na}^+]$ plane is different for each type of transition. In general, the larger the difference between the enthalpy due to the tertiary contact and the enthalpy due to the electrostatic force, the shallower the curve. Coaxial stacking has very small binding energy, and thus it is very sensitive to the salt concentration. Kissing hairpins have more tertiary contacts, and thus are a little less sensitive to salt concentration. Stems (helices) and hairpins (helices with terminal loops) have many hydrogen bonds and base stacks, and are thus the least sensitive to changes in salt concentration.

At low salt concentrations even the most stable tertiary contacts melt below 40°C, strongly supporting the idea that just because water is a liquid doesn't mean that biomolecules can function. Coaxial stacking is not stable at any temperature unless the sodium concentration exceeds 100mM. Since coaxial stacking is essential to the

formation of RNA tertiary structure, the cyan curves put the strongest constraint on habitability.

4 Modeling The Dynamics of RNA Folding

4.1 Introduction

Thermodynamics does not tell us how long an RNA molecule will take to fold. To do this we must simulate the folding process on a computer. Dynamical simulations can be very detailed (all atom quantum mechanical) or very coarse grained. We cannot use all atom simulations because they can only be run for a milisecond at most, and we need to run a simulation long enough to model one of the steps in tertiary folding. Instead, we will model the helices as rigid objects (rigid cylinders with a helical charge distribution), the ions as solid spheres, and the loop as a rigid circle. We only need to use classical physics, even for the smallest component of our system (the individual ions). For example, our sodium ions are around 0.5nm in diameter, have a mass of 3.81×10^{-26}kg, and travel with a thermal velocity of $\sqrt{kT/m} = 330$m/s. Using the uncertainty principle $\Delta x \propto \hbar/\Delta p = 0.008$nm. In other words, the uncertainty in the position of an ion is 1000 times less than its size, and can safely be ignored.

4.2 Water

RNA molecules exist in aqueous solution, and thus are surrounded by water molecules. These water molecules have three effects on the components of RNA: (1) water molecules are polar and orient themselves in such a way that they reduce the electric field of a charge, (2) water exerts a friction force as the components move through it, and (3) water molecules collide with the components causing random displacemnets in their positions and orientations. We will not include water molecules explicitly, but instead implicitly take the above three effects into account. The reduction in the electric field is taken into account by replacing ϵ_0 with $\epsilon\epsilon_0$ in all of the electrostatic formulas. Friction is taken into account by adding a friction force to the equation of motion of each component. The thermal motions induced by collisions with the water molecules is taken into account by adding a random force to the equation of motion (or equivalently randomly displacing the positions and orientations of each component each timestep).

4.3 Model Helices

Our model helix is constructed by arranging spheres into an ideal A-form helix, complete with a major groove, minor groove, and terminal loop. There are two charged spheres representing the phosphate groups (charge $-e$ each) and three

neutral spheres (a core sphere and two mantle spheres) per base pair. In addition 10 spheres are arranged in a circle at the top of the helix to represent the terminal loop. If an N base pair helix is coaxial with the z-axis and resting on the xy-plane (body frame), the coordinates of the charged spheres are

$$x(i,j) = r_0 \cos(di + f(j)) \tag{10}$$

$$y(i,j) = r_0 \sin(di + f(j)) \tag{11}$$

$$z(i,j) = c * i + g(j) \tag{12}$$

where $i = 1, 2, ..., N$, $j = 1, 2$, $r_0 = 0.88$nm, $d = 0.5712$, $f(1) = 0$, $f(2) = 2.681$, $c = 0.281$nm, $g(1) = 0$, and $g(2) = 0.188$nm (Tan and Chen 2007). The coordinates of the mantle spheres are

$$x(i,j) = r_1 \cos(di + f(j)) \tag{13}$$

$$y(i,j) = r_1 \sin(di + f(j)) \tag{14}$$

$$z(i,j) = c * i + g(j) \tag{15}$$

where $r_1 = 0.58$nm. The core spheres have coordinates

$$x(i) = 0 \tag{16}$$

$$y(i) = 0 \tag{17}$$

$$z(i) = c * i \tag{18}$$

The terminal loop spheres have coordinates

$$x(i) = r_2 * \cos(\pi i/5) \tag{19}$$

$$y(i) = 0 \tag{20}$$

$$z(i) = r_2 * \sin(\pi i/5) + Nc + r_2 \tag{21}$$

where $i = 1, ..., 10$ and $r_2 = 1\text{nm}$. See Figure 6.

In our simulations the center of the bottom core is fixed in space at a point on the rigid junction loop. The helix is allowed to rotate freely in three dimensions around that point. Each sphere can have two forces on it: an electrostatic force and a steric force both from the ions and other helices. The steric force is a repuslive force that kicks in when the center of two spheres are closer than the sum of their radii. For simplicity we assume the force is a constant equal to 10^{-10}N directed radially away from their center of mass. The helix itself has an angular velocity ω, and thus has a rotational friction torque acting on it $\boldsymbol{\tau}_f = -\gamma_R \boldsymbol{\omega}$, where γ_R is the rotational friction coefficient. We assume the rotational friction coefficient is that of a cylinder of length $0.281\text{nm}N$ and radius 0.88nm, and use the friction coefficient from Tirado et. al. 1984.

The orientation of a helix can be described by a rotation matrix that converts the coordinates of the spheres from the space frame (the frame the simulation is run in)

to the body frame:

$$\mathbf{x}' = \mathbf{R}\mathbf{x} = \begin{pmatrix} \mathbf{a}\cdot\mathbf{i} & \mathbf{a}\cdot\mathbf{j} & \mathbf{a}\cdot\mathbf{k} \\ \mathbf{b}\cdot\mathbf{i} & \mathbf{b}\cdot\mathbf{j} & \mathbf{b}\cdot\mathbf{k} \\ \mathbf{c}\cdot\mathbf{i} & \mathbf{c}\cdot\mathbf{j} & \mathbf{c}\cdot\mathbf{k} \end{pmatrix} \mathbf{x} \tag{22}$$

where $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ are the basis vectors of the space frame, $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ are the basis vectors of the body frame and primes refer to coordinates in the body frame.

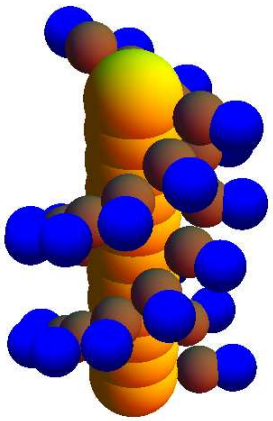

Figure 6: The model helix we use in our calculations and simulations. The blue spheres represent the phosphates with charge $-e$. The brown spheres are the neutral mantle spheres, while the yellow spheres are the neutral core spheres.

4.4 Model Ions

The ions are treated as spheres whose radius includes the hydration shell (6 water molecules that are coordinately covalently bonded to them). Sodium is treated as a sphere of radius 3.5nm and charge $+e$, magnesium as a sphere of radius 4.5nm and charge $+2e$, and chlorine as a sphere of radius 3.5nm and charge $-e$. There are two forces that can act on the ions: electrostatic and steric. The steric force kicks in when the center-to-center distance between the ions and any other component of RNA is less than the sum of their radii and is equal to 10^{-10}N and directed radially away from their center of mass. We take the diffusion constant of sodium ($1.33 \times 10^{-9} m^2 s^{-1}$) and chlorine ($2.03 \times 10^{-9} m^2 s^{-1}$) from Cheung et al. 1998. We assume the diffusion constant of magnesium is $10^{-9} m^2 s^{-1}$.

4.5 Model Loops

We currently do not have a model for the junction and terminal loops. We simply treat them as solid rings, forcing the end of the helices to be fixed onto the loop. Including their internal degrees of freedom will be difficult because the timescales for loop motion is much less than the time scale for helix motion.

4.6 Brownian Dynamics

The state of the system in our simulation is given by the position of the ions and the orientation of the helices. The equation of motion for the ions is:

$$\mathbf{v} = \frac{d\mathbf{x}}{dt} \tag{23}$$

$$m\frac{d\mathbf{v}}{dt} = \mathbf{F}_{el} + \mathbf{F}_{st} - \gamma\mathbf{v} + \xi \tag{24}$$

where $\mathbf{F}_{el}$ is the electrostatic force, $\mathbf{F}_{st}$ is the steric force, $\gamma = kT/D$ is the friction coefficient, and ξ is the random force due to thermal fluctuations. The random force is drawn from a Gaussian with mean 0 and variance $2kT\gamma/\delta(0)$. Since the terminal velocity is reached on time scales much shorter than scales of interest we can set $d\mathbf{v}/dt = 0$:

$$\mathbf{v} = \frac{d\mathbf{x}}{dt} \tag{25}$$

$$\mathbf{v} = \frac{\mathbf{F}_{st} + \mathbf{F}_{el}}{\gamma} + \frac{\mathbf{xi}(t)}{\gamma} \tag{26}$$

The equation of motion for the helix is:

$$\frac{d\mathbf{R}}{dt} = \begin{pmatrix} 0 & \omega'_z & -\omega'_y \\ -\omega'_z & 0 & \omega'_x \\ \omega'_y & -\omega'_x & 0 \end{pmatrix} \mathbf{R} \tag{27}$$

Assuming that terminal angular velocity is reached on time scales shorter than the scales of interest we set $\boldsymbol{\omega} = \boldsymbol{\tau}/\gamma_R$. The torque comes from electrostatic and steric

forces and is found by summing $\mathbf{x}_i'' \times (\mathbf{F}'_{el} + \mathbf{F}'_{st})_i$ over all components i, where $\mathbf{x}_i''$ is the coordinates of the components in the frame where the pivot point (bottom core particle) is at the origin. The full equation of motion also has a random thermal torque $\boldsymbol{\xi}_R$ drawn from a Gaussian with mean zero and variance $2kT\gamma_R/\delta(0)$:

$$\frac{d\mathbf{R}}{dt} = \frac{1}{\gamma_R} \begin{pmatrix} 0 & \tau'_z + \xi'_{R,z} & -\tau'_y - \xi'_{R,y} \\ -\tau'_z - \xi'_{R,z} & 0 & \tau'_x + \xi'_{R,x} \\ \tau'_y + \xi'_{R,y} & -\tau'_x - \xi'_{R,x} & 0 \end{pmatrix} \mathbf{R} \tag{28}$$

A simulation based on Equations 25, 26, and 28 is called a Brownian dynamics simulation. The continous time variable t is divided into discrete time steps of duration Δt. At the beginning of each time step the total electrostatic and steric force is computed for each ion and each helix component, then the torque is computed for each helix component summed to find the total torque on each helix. The random forces and torques $\boldsymbol{\xi}$ and $\boldsymbol{\xi_R}$ are then drawn from a Gaussian distribution with mean 0 and variance $2kT\gamma/\Delta t$ and $2kT\gamma_R/\Delta t$, respectively, and added to the total. The positions of the ions and the rotation matrices of the helices are then changed according to an integration algorithm.

4.7 Choosing an Integrator

The simplest algorithm to change the position of the ions and orientation of the helices is the Euler method, which moves the ions by an amount $\Delta x = F_{ext}\Delta t/\gamma + R$, where R is a random number drawn from a Gaussian with mean 0 and variance $2kT/\gamma$. A more sophisticated algorithm is the second order stochastic Runge-Kutta method (Honeycutt 1991). In this algorithm we first compute the change in position Δx_{Elr} as if we were using the Euler method. We evaluate the force at the new position, and move the particle another Euler step. Since we have moved two Euler steps in one time step, we move to the midpoint of our initial (before moving any Euler steps) and final (after moving two Euler steps) positions.

Changing the rotation matrix is similar. Using the Euler method we have $\Delta R_{ij} = (d\mathbf{R}_{ij}/dt)\Delta t$, where in the expression for $d\mathbf{R}_{ij}/dt$ we draw the random torque from a Gaussian with mean zero and variance $2kT\gamma_R/\Delta t$. To use the second-order stochastic Runge-Kutta model, we follow the same steps as for the ions, but with torque instead of force.

4.8 Choosing a Random Number Generator

The above integration algorithms require drawing numbers from a Gaussian. Computers usually can only draw from a uniform distribution, often from 0 to 1. We need an algorithm that converts numbers drawn from a uniform distribution to numbers drawn from a Gaussian distribution. An efficient algorithm that we use is the Box-Mueller method (Box-Mueller 1958). If U_1 and U_2 are numbers drawn from a uniform distribution, then $\sqrt{-2\sigma^2 \ln U_1}\cos(2\pi U_2)$ and $\sqrt{-2\sigma^2 \ln U_1}\sin(2\pi U_2)$ are equivalent to numbers drawn from a Gaussian distribution with mean 0 and variance σ^2.

4.9 Using Quaternions to Represent the Orientation of a Helix

Throughout the simulation, the rotation matrix of each helix changes as the helix changes its orientation. The rotation matrix must transform the coordinantes of the components of the helices from the lab frame to the space frame each time step. Since the helices are rigid, the distance between any two components should not change during the rotation. This requires that the rotation matrix be orthogonal. However, numerical errors that occur when integrating Equation 28 will cause the rotation matrix to become non-orthogonal, and thus distort the helices as it

transforms the helix from the body frame to the space frame. In order to prevent this it would be necessary to correct the rotation matrix at each time step to ensure it is orthogonal, a difficult task.

There are two popular methods to get around this problem. One is to use Euler angles (ϕ, θ, ψ) to represent the orientation of a helix. Instead of solving the equation of motion for the rotation matrix, one would solve the equation of motion for the Euler angles. Unfortunately if the z-axis of the space and body frames are aligned then the equation of motion for ϕ and ψ become $\dot{\phi} = \dot{\psi} = \pm\infty$. As a result, we will not use Euler angles.

To go from the lab frame to the body frame one can rotate the helix by an angle ϕ about an axis $\mathbf{n}$. This suggests using the four numbers (ϕ, n_x, n_y, n_z) to represent the orientation of the helices. Indeed, if one forms the object $q = [q_0, q_1, q_2, q_3] = [\cos(\phi/2), \mathbf{n}\sin(\phi/2)]$, and demands that $q_0^2 + q_1^2 + q_2^2 + q_3^2 = 1$, then this object (called a quaternion) obeys the equation of motion

$$\frac{dq}{dt} = \frac{1}{2} \begin{pmatrix} -q_1 & -q_2 & -q_3 \\ q_0 & -q_3 & q_2 \\ q_3 & q_0 & -q_1 \\ -q_2 & q_1 & q_0 \end{pmatrix} \begin{pmatrix} \omega_x' \\ \omega_y' \\ \omega_z' \end{pmatrix} \tag{29}$$

One still needs the rotation matrix to get the coordinates of the helix components

in the space frame, and one can get it from:

$$\mathbf{R} = \begin{pmatrix} q_0^2 + q_1^2 - q_2^2 - q_3^2 & 2(q_1 q_2 + q_0 q_3) & 2(q_1 q_3 - q_0 q_2) \\ 2(q_1 q_2 - q_0 q_3) & q_0^2 - q_1^2 + q_2^2 - q_3^2 & 2(q_2 q_3 - q_0 q_1) \\ 2(q_3 q_1 + q_0 q_2) & 2(q_3 q_2 - q_0 q_1) & q_0^2 - q_1^2 - q_2^2 + q_3^2 \end{pmatrix} \tag{30}$$

There are no singularities in the equation of motion and the normalization constraint is easy to implement (just divide each quaternion by its magnitude). Thus we will use quaternions to keep track of the orientation of the helices, solving Equation 29 instead of 28.

4.10 A Two Dimensional Coarse-Grained Simulation

4.10.1 Calculation

Before discussing the detailed model we here use a corase grained version to illustrate how we can compute the time it takes RNA to transition from stacked to docked. Docking of one helix to another in a 3HJ essentially consists of a charged helix diffusing across the Coulomb barrier of the other 2 helices (which are stacked). Call the stacked helices A and B and the docking helix C (see Figure 6). As a first approximation we can replace the helices with point charges at their centers of mass. We place point A at $\mathbf{r}_A = (0, L/2)$, point B at $\mathbf{r}_B = (0, -L/2)$, and point C at $\mathbf{r}_C = (2nm, 0)$, where $L = 0.1405N\text{nm} - .0465\text{nm}$. Points A and B are

fixed in space while point C is allowed to move on a circle centered at the origin. At each time step ($\Delta t = 1$ps) the torque by A and B on C is computed (the friction coeficient is that of a cylinder), and the position of the particle is changed using the second order stochastic Runge-Kutta method. This continues until C becomes parallel to A or B (reaches the end of the semicircle), and the time elapsed (docking time) is then recorded. This is repeated 1000 times and the results averaged to get the mean docking time. This is done for various values of helix length and salt concentration. The results of such a simulation is shown in Figure 7. We plot docking time as a function of sodium concentration for helices of length 5bp, 10bp, and 15bp.

4.10.2 Discussion

In Figure 7 we see that longer helices take longer to fold and that 3HJs fold faster the higher the sodium concentration. Longer helices take longer to fold because (1) they have a larger friction coefficient and so take a longer time to rotationally diffuse and (2) they have more charges and thus a higher Coulomb barrier to cross. Salt screens the charges and reduces the Coulomb barrier, explaining why 3HJs fold faster at higher salt concentrations. Increasing the salt concentration above 1M will not decrease the folding times because all of the charges are already screened at 1M

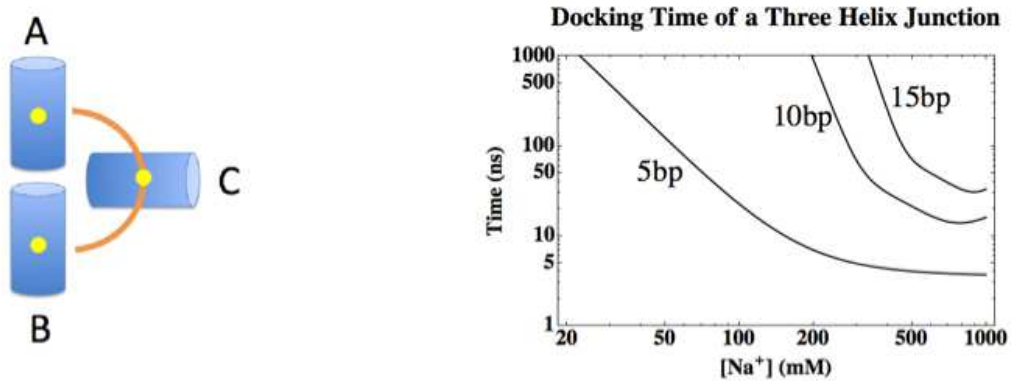

Figure 7: *Left* The point model treats helices (blue cylinders) as point charges (yellow points). Points A and B are fixed in space, while C is free to move on the orange semicircle. *Right* The mean time it takes C to diffuse to the end of the semicircle from the middle in the presence of the electric field of A and B based on the simulation described in the text. Each curve is for a different helix length.

and cannot be screened more by increasing the salt concentration.

4.11 Calculating the Ion Atmosphere

4.11.1 Calculation

Up to this point we have been treating the ions as a continuous fluid obeying Boltzmann statistics and having electrostatic energy less than kT. While this is a good approximation for sodium and other monovalent ions, the high charge of

multivalent ions makes their discrete nature important. The thermal fluctuations in the position of the ions are correlated (they move away from each other and toward ions of opposite charge) in such a way that the electrostatic energy of two charges in a multivalent salt solution is less than that predicted by the Poisson-Boltzmann equation. In order to take into account the discrete nature of ions, we must include them explicitly in the Brownian dynamics simulations. It will be instructive to do this for a single helix first, and then later for the full 3HJ.

We place a single helix in the body frame in a simulation box of size 10nm $\times$ 10nm $\times$ $2L + 4$nm in the x, y, and z directions, respectively. Ions are randomly placed throught the box (the number of ions added is the concentration times the volume). In order to make the simulation box neutral, we add an additional $6N$ sodium atoms to compensate for the helices negative charges. At each time step the force is calculated on each ion (steric and electrostatic) and the ion is moved using the Euler method. Periodic boundary conditions are implemented (if the ion leaves the simulation box, it enters the opposite side). A time step of 1ps is used. We run the simulation for 100ns to allow the ions to reach an equilibrium distribution. Then we divide the region around the helix into 0.1nm thick cocentric shells centered on the helix axis, and run the simulation for another 100ns. At each timestep we count how many ions of each type are in each shell, and sum the result over all timesteps.

We divide this number by the number of timesteps to get the average number of ions in each shell. We divide by the volume of the shell to get the number density of ions in each shell, and convert that to milimoles per liter. The results are shown in the top panels of Figure 8. In addition, we chose 10 ions of each type at random and kept track of its trajectory through the first 100ns of the simulation (the equilibration period). The results are shown in the bottom panels of Figure 8.

4.11.2 Discussion

The top panels of Figure 8 show several key properties of the ion atmosphere. First, the sodium and magnesium concentration near the helix is many times greater than the asymptotic concentration (concentration far away). This is because positive ions are attracted to the helix due to the large number of negative charges present. For 10mM NaCl, the concentration near the helix is an astounding 30 times higher than the asymptotic concentration. It is only 5.5 times higher for the 100mM NaCl case, showing that higher salt concentrations produce less of a difference between the asymptotic concentration and the concentration near the RNA. This is because the higher the salt concentration, the more the negative charges are screened, and thus the less gradient is needed in ion concentration to balance the electrostatic attraction. Also note that chlorine ions are depleted near

the helix (they are negatively charged and thus repeled) and no ions are closer than 0.5nm because of steric constraints.

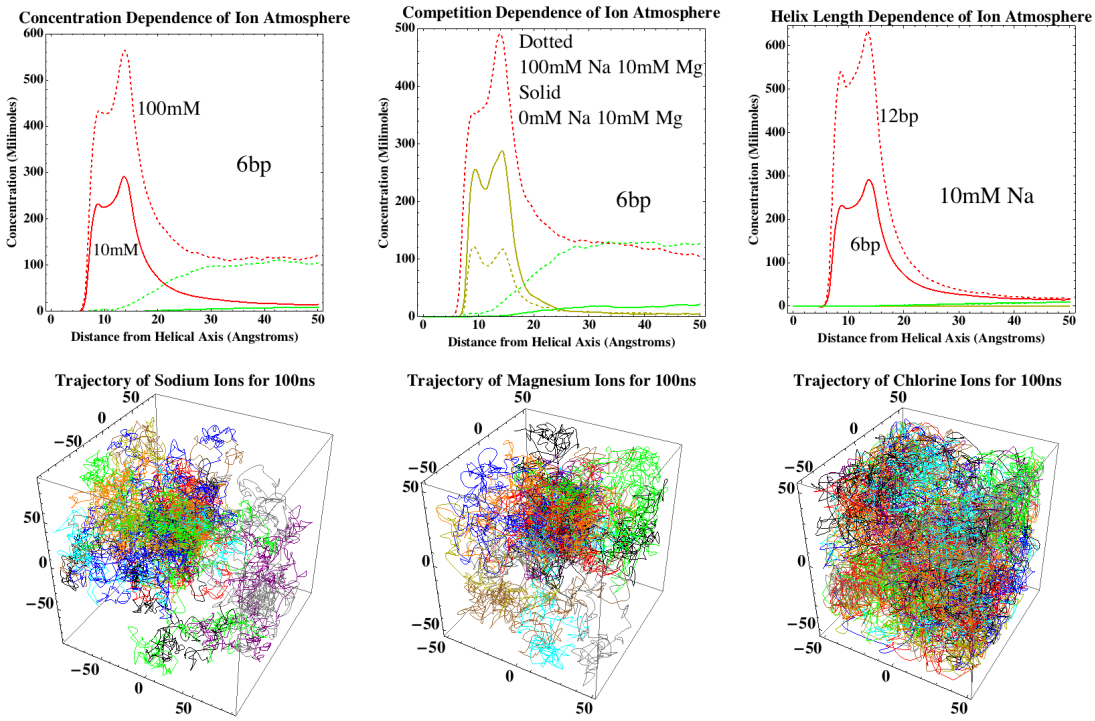

Figure 8: Properties of the ion atmosphere around a single RNA helix. *Top Left* Concentration of sodium (red) and chlorine (green) around a 6bp helix. Dashed lines represent 100mM of NaCl, solid lines represent 10mM NaCl. *Top Middle* Concentration of sodium, magnesium (yellow), and chlorine around a 6bp junction. Solid lines 10mM MgCl and no sodium, Dashed lines 100mM NaCl and 10mM MgCl. *Top Right* Concentration of sodium and chlorine around a 6bp (solid) and 12bp (dashed) helices in a 10mM NaCl solution. Concentrations refer to values far away from the helix. *Bottom Left* The trajectory of 10 sodium ions (each ion a different color) for the first 100ns of a simulation. A 12 bp helix is located at the center of each box. 10mM NaCl was in the box. *Bottom Middle* Trajectories of 10 magnesium ions for the first 100ns of a simulation (20mM MgCl). *Bottom Right* Trajectories of 10 chlorine ions for the first 100ns of a simulation (20mM NaCl).

Second, ions compete for locations near the helix. The top middle panel shows two cases, both of which have the same amount of magnesium (10mM) but different amounts of sodium (100mM for the dotted line and 0mM for the solid line). Notice that there is significantly less magnesium near the helix when sodium is present than when it is absent. The physics behind the fraction of the ion atmosphere that is magnesium versus sodium is the following. For an ion to enter the ion atmopshere, it loses an entropy ΔS_{ion} and loses an enthalpy ΔH_{ion}. Assuming the entropy loss is the same for each ion, but the enthalpy loss of magnesium is twice that of sodium, we get the free energy change of an ion entering the ion atmosphere as

$$\Delta G_{Na} = \Delta H_{Na} - T\Delta S_{ion} \tag{31}$$

$$\Delta G_{Mg} = 2\Delta H_{Na} - T\Delta S_{ion} \tag{32}$$

The ratio of sodium ions to magnesium ions (N_{Na}/N_{Mg}) in the ion atmosphere is the ratio of their asymptotic concentrations times the ratio of their Boltzmann factors for entering the ion atmosphere:

$$\frac{N_{Na}}{N_{Mg}} = \frac{N_{0,Na}}{N_{0,Mg}} \frac{e^{-\Delta G_{Na}/kT}}{e^{-\Delta G_{Mg}/kT}} = \frac{N_{0,Na}}{N_{0,Mg}} e^{\Delta H_{Na}/kT} \tag{33}$$

If we assume the electrostatic potential around a helix is kT/e (screening will bring

it down to that level) then $\Delta H_{Na} = -kT$ and

$$\frac{N_{Na}}{N_{Mg}} = \frac{N_{0,Na}}{N_{0,Mg}} e^{-1} = .368 \frac{N_{0,Na}}{N_{0,Mg}} \tag{34}$$

In our case $N_{0,Na}/N_{0,Mg} = 100/10 = 10$ so $N_{Na}/N_{Mg} = .368 * 10 = 3.68$. Our calculation shows that $N_{Na}/N_{Mg} = 490/120 = 4.08$, showing that our thermodynamic argument gives a reasonable estimation for the relative amount of sodium and magnesium in the ion atmosphere.

Third, longer helices have higher concentrations in their ion atmosphere than shorter ones do. Longer helices have stronger electric fields which must be balanced by larger gradients in the osmotic pressure (and thus larger gradients in the ion density). Demanding a larger gradient in ion concentration requires a larger difference between the asymptotic concentration to the concentration in the ion atmosphere itself, and thus a larger concentration in the ion atmosphere (see top right panel).

It is insightful to focus on an individual ion and follow it for a given period of time. An individual cation will enter and leave the ion atmopshere repeatedly throught the simulation. It will spend most of its time near the helix because of its attraction to the negative charges. The bottom panel of Figure 8 shows the trajectories of ions folowed for 100ns. Each ion's trajectory has been colored a different color, and we have followed the trajectory of 10 of each type of ion. On

the left we see the trajectories of sodium ions. A 12bp helix is located at the center of each box (numbers are in Angstroms). Most of the trajectories cluster near the center of the box where the helix is. The middle panel shows the trajectory of magnesium ions. Because they have twice the charge as sodium ions, their trajectories cluster even closer to the helix. The right panel shows the trajectory of chlorine ions. The trajectories of chlorine ions fill the box because they are negatively charged and repelled by the helix.

4.12 A Three Dimensional Model

4.12.1 Setup

We can combine all of our previous work into a single model, one of the models that can be used to compute the folding times for RNA 3HJs. We proceed as follows to set up the simulation.

We first choose the radius of the junction loop, the concentration of sodium and magnesium, the length of the helices and the temperature (the dielectric constant and viscosity of water are determined by the temperature). We place the 3 helices at 120°, with the bottom core sphere placed on the junction loop (the points where the helices contact the loop are equally spaced). The number of ions that are needed to fill the simulation box are calculated, and the ions are placed at random

in the box. An additional $6N$ sodium ions are added at random locations to ensure charge neutrality.

The following operations are performed each timestep. The quaternion representing the orientation of each helix is converted to a rotation matrix, and the rotation matrix is used to convert the coordinates of the helix components from the body frame to the space frame. Using these coordinates, the total force (steric and electrostatic) on each ion and helix component is calculated. From this the total torque on each helix is found, and the Euler method or second-order Runge-Kutta method is used to find the new ion position and new quaternion corresponding to each helix.

4.12.2 Discussion

The previously described model has ions included explicitly and helices with the correct charge distribution that can rotate in three dimensions. As a result it is more computationally intensive than a model with no ions and helices confined to two dimensions or represented as point charges. Consequently we have not been able to run it and obtain results on RNA folding. When we do we will compute the average time it takes a 3HJ, starting at the open configuration, to reach a docking configuration where the distance between two spheres on the terminal loop is less

than a certain value (probably around 0.5nm). We will run the simulation 1000 times, record the time it took to dock for each run, and average the results to get the mean docking times for a given helix length and salt concentration.

To give an idea of how the simulation works, we have shown four screenshots from different simulations in Figure 9. Each simulation had 12bp helices, a junction loop with radius 1.5nm, a temperature of 298K, a dielectric constant of 77, and a viscosity of water of 9×10^{-4} Ns/m^2. The top panel shows the initial state of the system for 10mM NaCl and 10mM MgCl. Sodium ions are red, magnesium ions are yellow, chlorine ions are green, loops are orange, phosphates are blue, mantle spheres are brown, and core spheres are white. The top right panel shows the same system 100ns into the simulation. Note how the helices are connected to the junction loop at their bottom core sphere. Also note that two helices have terminal loops. The magnesium ions start out randomly distributed but 100ns into the simulation they are clustered near the helices, as expected. The bottom left panel shows the initial configuration of a simulation in the earth's ocean (500mM NaCl and 50mM MgCl). The bottom right panel shows a simulation in 1mM MgCl and 0mM NaCl. Note that even though the concentration of NaCl is zero, there is still sodium ions which were added to keep the simulation box electrically neutral.

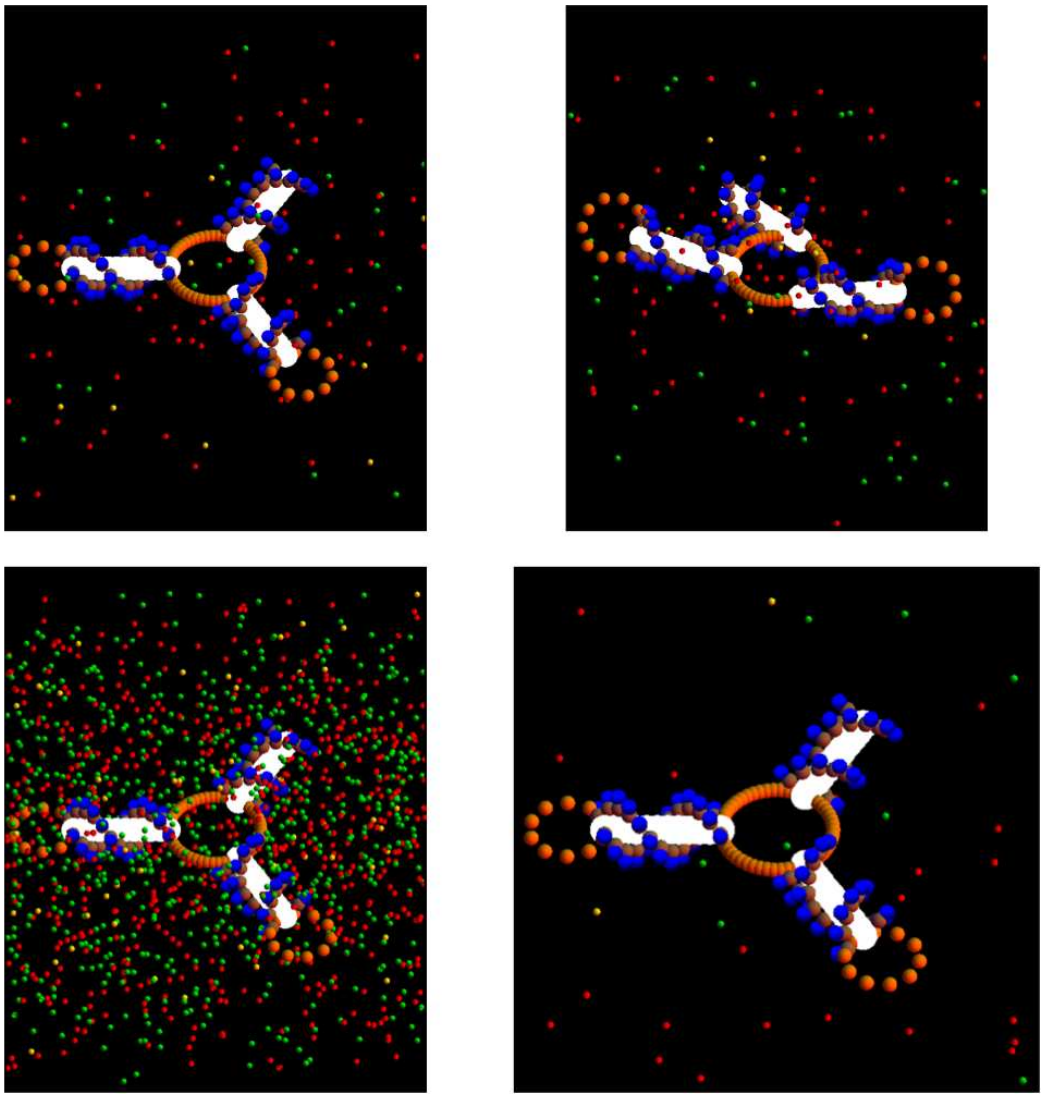

Figure 9: Frames from a fully dynamical simulation of the folding of a 3HJ. Each helix is 12bp long. *Top Left* Initial setup for 10mM NaCl and 10mM MgCl. *Top Right* A snapshot 100ns into a simulation of 10mM NaCl and 10mM MgCl. *Bottom Left* Initial setup for the earth's ocean, 500mM NaCl and 50mM MgCl. *Bottom Right* Initial setup for 1mM MgCl.

5 Future Prospects

When we get time on a supercomputer we will be able to run our model and compute RNA 3HJ folding times in a wide variety of salt conditions and for different helix lengths. Environments in which RNA takes longer to fold than a critical value will be deemed uninhabitable, or at least less likely to allow an RNA world to arise than environments that allow quicker RNA folding times. Because computer time scales with folding time, we can only model the fastest folding 3HJs. We can calculate the folding time as a function of salt concentration for high salt concentrations, and then fit an exponential to that function and extrapolate it to lower salt concentrations.

We would also like to include an attractive potential between the spheres on the terminal loops. This is supposed to take into account the effect of hydrogen bonding. Then when we run the simulation, we can compute the fraction of time that the helices remain in tertiary contact. We can run the simulations at different temperatures, and the temperature at which helices are bound half of the time is the melting temperature.

We would like to include the effects of pressure and pH. pH is straightforward to include, because its effect is to change the charge per nucleotide, and one can derive a formula that gives the charge per nucleotide as a function of temperature and pH.

Pressure is harder to include because it affects many different things. Increasing the pressure tends to favor processes that decrease the volume of the system. This includes the volume of the RNA itself, the ions, and the water molecules.

At the ocean floor the pressure is so high that water can remain liquid past 100°C. This allows hydrothermal vents to emit liquid water at temperatures up to 405°C (Martin et al. 2008). At such high temperatures the dielectric constant of water may be as low as 10, which is much less than it is at room temperature (about 77) (Pitzer 1983). The Debye length at a given salt concentration would be half of what it is at room temperature, but the electrostatic energy between two unscreened charges would be about eight times stronger due to the decreased dielectric screening. Because it is widely believed that life started near hydrothermal vents, it is important that we simulate RNA folding in these extreme conditions.

6 Conclusion

We have constructed a coarsed grained model of RNA 3HJs and used Brownian dynamics simulations to compute the folding time of these 3HJs. Our results will have astrobiological implications since they will restrict the environments in which RNA can fold and thus function during the transition to life. We already have determined that environments with high salt concentrations and low temperatures are the most favorable for RNA folding. This is because thermal fluctuations unfold RNA at high temperatures and electrostatic repulsions unfold RNA at low salt concentrations. We expect environments with high pH to be unfavorable due to the increased negative charge per nucleotide that occurs at high pH. The effect of high pressure is unknown, except that it will reduce the volume of the system. This project draws many suprising connections between different fields of science. It connects habitability with the physics and chemistry of polymers. It emphasizes the importance of the origin of life in determining the habitability of a planet. It introduces RNA folding into origin of life studies. It gives geochemists reasons to model the geochemical cycles on extrasolar planets, the early earth, and moons in our solar system. It constrains the habitable zone, possibly severely. By elucidating the mechanisms by which a world built out of lifeless atoms was transformed into the animated world we see today, we will be one step closer to deciphering the

riddle of our own existence.

7 References

Berendsen HJC. 2007. *Simulating the Physical World.* Cambridge University Press.

Bokov K, Steinberg SV. 2009. A Hierarchial Model for Evolution of 23S Ribosomal RNA. *Nature* 457: 977-980

Box GEP, Muller ME. 1958. A Note on the Generation of Random Normal Deviates. *Ann. Math. Statist.* 29: 610-611

Brion P, Westhof E. 1997. Hierarchy and Dynamics of RNA Folding. *Annu. Rev. Biophys. Biomol. Struct.* 26: 113-137

Butler E. et al. 2011. Structural Basis of Cooperative Ligand binding by the Glycine Riboswitch. *Chemistry and Biology* 18: 293-298

Cheung et al. 1998. Brownian Dynamics Study of Ion Transport in the Vestibule of Membrane Channels. *Biophysical Journal* 74: 37-47

Doi K et al. 2010. Development of Coarse-Graining DNA Models for Single-Nucleotide Resolution Analysis. *Phil. Trans. R. Soc. A* 368: 2615-2628

Downey et al. 2007. Hydrostatic Pressure and Cosolutes on RNA Tertiary Structure. *J. Am. Chem. Soc.* 129: 9290-9291

Dubins DN, Lee A, Macgregor RB, Chalikian TV. 2001. On the Stability of Double Stranded Nucleic Acids. *J. Am. Chem. Soc.* 123: 9254-9259

Honeycutt LH. 1991. Stochastic Runge-Kutta Algorithms. I. White Noise. *Physical*

Review A 45:600-603

Joyce GF. 2002. The Antquity of RNA Based Evolution. *Nature* 418: 214-221

Joyce GF. 2004. Directed Evolution of Nucleic Acid Enzymes. *Annu. Rev. Biochem.* 73: 791-836

Kasting JF, Catling D. 2003. Evolution of a Habitable Planet. *Annu. Rev. Astron. Astrophys.* 41: 429-463

Lincoln TA, Joyce GF. 2009. Self-Sustained Replication of an RNA Enzyme. *Science* 323: 1229-1232

Lorenz C et al. 2006. Stabilities of HIV-I DIS Type RNA Loop-Loop Interactions in vitro and in vivo. *Nucleic Acid Res.* 34:334-342

Martin W. et al. 2008. Hydrothermal Vents and the Origin of Life. *Microbiology* 6:805-814

Moulton V. et al. 2000. RNA Folding Argues against a Hot-Start Origin of Life. *Journal of Molecular Evolution* 51:416-421

Nakano S. et al. 1999. Nucleic Acid Duplex Stability: Influence of base Composition on Cation Effects. *Nucleic Acid Res.* 27: 2957-2965

Orgel LE. 2004. Prebiotic Chemistry and the Origin of the RNA World. *Critical Reviews in Molecular Biology and Biochemistry* 39:99-123

Pabit L et al. 2011. Double Stranded RNA Resists Condensation. *Physical review*

Letters 106:18101(4)

Pastor RW, Karplus M. 1987. Parametrization of the Friction Constant for Stochastic Simulations of Polymers. *American Chemical Society* 88: 2636-2641

Pena M, Dufour D, Gallego J. 2009. Three-way RNA junctions with remote tertiary contacts: A recurrent and highly versatile fold. *RNA*

Pena M, Garcia I. 2010. Ubiquitous Presence of the Hammerhead Ribozyme Motif on the Tree of Life. *RNA* 16: 1943-1950

Pitzer K. 1983. Dielectric Constant of Water at Very High Temperature and Pressure. *PNAS* 80: 4575-4576

Qu X et al. 2008. Single Molecule Nonequilibrium Periodic Mg^{2+}-Concentration Jump Experiments Reveal Details of the Early Folding pathways of a Large RNA. *PNAS* 105:6602-6607 Tan ZJ, Chen SJ. 2005. Electrostatic Correlations and Fluctuations for Ion Binding to a Finite Length Polyelectrolyte. *J. Chem. Phys.* 122:044903

Tan ZJ, Chen SJ. 2007. RNA Helix Stability in Mixed Na+/Mg2+ Solution. *Biophysical Journal* 92:3615- 3632

Tan ZJ, Chen SJ. 2008. RNA Helix Stability in Mixed Na+/Mg2+ Solution. *Biophysical Journal* 95:738- 752

Tan ZJ, Chen SJ. 2010. Salt Dependence of Nucleic Acid Hairpin Stability.

Biophysical Journal 99:1565- 1576

Tirado MM, Martinez CL, Garcia de la torre J. 1984. Comparison of theories for the translational and rotational diffusion coefficients of rod-like macromolecules. Application to short DNA fragments. *J. Chem. Phys.* 81:2047-2052

Walter AE, Turner DH. 1994. Sequence Dependence of Stability for Coaxial Stacking of RNA Helixes with Watson-Crick Base Paired Interfaces. *Biochemistry* 33:12715-12719

Weixlbaumer A. et al. 2004. Determination of Thermodynamic Parameters for HIV DIS Type Loop-Loop Kissing Complexes. *Nucleic Acid Res* 32:5126-5133