PROBING NON-CODING RNA STRUCTURAL DYNAMICS WITH 2-AMINOPURINE

by

Michael Samir Dahabieh

B.Sc. (Molecular Biology and Biochemistry), Simon Fraser University, 2008

THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE

in the
Department of Molecular Biology and Biochemistry
Faculty of Science

SIMON FRASER UNIVERSITY

Spring 2012

Approval

Name: **Michael Samir Dahabieh**

Degree: **Master of Science**

Title of Thesis: ***Probing Non-Coding RNA Structural Dynamics with 2-Aminopurine***

Examining Committee:

Chair: **Dr. Mark Paetzel,**
Associate Professor, Department of Molecular Biology and Biochemistry

---

Dr. B. Mario Pinto for Dr. Melanie A. O'Neill
Senior Supervisor,
Professor, Department of Chemistry, Vice-President Research

Dr. Dipankar Sen
Supervisor
Professor, Department of Molecular Biology and Chemistry

Dr. Peter J. Unrau
Supervisor
Associate Professor, Department of Molecular Biology and Biochemistry

---

Dr. Gratien Prefontaine
Internal Examiner
Assistant Professor, Department of Health Sciences

Date Defended/Approved: February 24, 2012

Declaration of
Partial Copyright Licence

Abstract

This thesis investigates sequence-dependent structural dynamics of non-coding (nc) RNAs utilizing the fluorescent base 2-Aminopurine (2Ap). We conclude that the highly homologous adenine (ARNA) and guanine (GRNA) riboswitches exhibit distinct structural dynamics in the ligand-bound state. Relative to ARNA, GRNA is more preorganized towards the closed/native conformer (C_C), displays enhanced thermostability, and higher magnesium (Mg^{2+}) binding affinity. We then focus on the role of nucleoporin 50-kilodalton (*Nup50) Alu* double-stranded (ds) RNAs in adenosine-to-inosine (A-to-I) editing. Here, we deduce the folding pathways of three constructs containing an internally substituted 2Ap, which are representative of *Nup50-Alu* dsRNAs in human, chimp, and rhesus species, respectively. Our fluorescence-based data do not fit well to a two state (folded/unfolded) model, but are well modeled by a four-state (two intermediate) model. Despite the high interspecies homology, the human sequence is most heavily edited *in vivo*. Interestingly, we find that the 2Ap at position five in the human *Nup50-Alu* dsRNA has the most thermodynamically stable intermediate, which is an optimal substrate for the A-to-I editing enzyme. The investigations of ncRNA folding dynamics underscore how Nature utilizes subtle sequence variations to achieve remarkable diversity.

Keywords: purine riboswitches; 2-Aminopurine; adenosine-to-inosine editing; sequence-dependent; folding dynamics; non-coding RNA; primate evolution.

Dedication

This thesis is dedicated to Dr. Melanie A. O'Neill, my family, and friends:

In memory of my supervisor, friend and mentor, Dr. Melanie A. O'Neill. Without your encouragement I may not have realized my potential as a researcher, and human being. Your hard work, sheer intelligence, zest for life, and perseverance to uncover Nature's mysteries are an inspiration to many others and myself. Forever in my heart, and imprinted in my mind, you are a gentle reminder to simply embrace life.

To Mom, Dad, and Melissa. You've helped me through so many difficult times, and have never hesitated to support me in any way possible. My love for you is eternal and unconditional. Who ever knew that a home could magically turn into a bank and restaurant?

My friends, many from kindergarten and high school. Your scientist friend is not done yet…Thanks so much for all the laughs, support and encouragement through happy and sad times. You've kept me in tune with the equally important 'outside world.' I love you all.

Acknowledgements

1. Dr. B. Mario Pinto for warm-heartedly supporting me through the remainder of my thesis.

2. Committee members Dr. Dipankar Sen, and Dr. Peter Unrau for their much-appreciated assistance. Dr. Jean-Claude Brodovitch for his help and insights, Dr. Kathleen Fitzpatrick for her advice.

3. Lab members who have helped me along the way: Sam Eskandari, Oksana Prychyna, Neahlanna Mcleod, Farnaz Changizi, Amy Sung, Matthew Brolich, Dhrubajyoti Samanta, Roberto Trasolini, Stephanie Luongo, and Mike Damiani.

4. Fellow researchers in chemistry and MBB: Gurpreet Sekhon, Paul Cernak, Mariana Ovideo-Ovando, Christian Frech, and Krzysztof Lubieniecki.

Table of Contents

List of Tables

List of Schemes

List of Figures

Glossary

2'OH	2'-hydroxyl group of ribose in a nucleoside
2Ap/Ap	2-aminopurine
5-HT_{2C}	serotonin 2C receptor
5'UTRs	5'-untransated region(s)
7SLRNA	signal recognition particle RNA
A	adenine
Å	Ångstrom = 0.1 nanometre
A-T	adenine- thymine base pair
A-to-I	adenosine-to-inosine
ADAR	adenosine deaminase acting on double-stranded RNA
add ARNA	adenine riboswitch
Alu element	approximately 300 nt primate-specific region recognized by *Alu* enzyme
ARNA	adenine riboswitch
ATP	adenosine triphosphate
B. subtilis	*Bacilus subtilis*
b15 RNA	terminal intron of cytochrome b mRNA from yeast
bp(s)	basepair(s)
C	cytosine
c-fos	cellular protooncogene homologous to Finkel-Biskis-Jinkins (FBJ) murine osteosarcoma virus oncogene

c-myc	transcription factor gene homologous to myelocytomatosis viral oncogene (*v-Myc*)
C-to-U	cytosine to uracil
C74U GRNA	adenine specificity nucleotide mutant guanine riboswitch
CBP2	mitochondrial protein involved in splicing of group I intron αI5 of the COB pre-mRNA
C_c/C_3	closed conformer
CD	circular dichroism
ChAp5/13	chimp *Nup50 Alu* construct with Ap at position 5/13
C_i/C_2	intermediate conformer
ClustalW	nucleic acid sequence alignment program
cm	centimetre
C_o/C_1	open conformer
CTP	cytidine triphosphate
DNA	deoxyribonucleic acid
dsRNA	double-stranded RNA
DTT	dithiothreitol
E.coli	*Eschericia coli*
EDTA	ethylenediaminetetraacetic acid
EMSA	electromobility shift assay
f_1, f_2, f_3	fractions of short, intermediate, and long lifetimes, respectively

f_b	fraction bound
f_{Cn}	fraction of a given conformer
f_f/f_{fold}	fraction folded
f_i	fraction intermediate
f_m, f_j	fractions of intermediates m and j, respectively
FRET	Förster resonance energy transfer
f_{SQ}	fraction static quenching
f_{TQ}	fraction total quenching
f_u	fraction unfolded
FWHM	full width at half maximum
G	guanine
G-C	guanine- cytosine base pair
G13C	purine riboswitch stem mutant
GARNA	GRNA mutant with A binding site
glmS	glucosamine-6-phosphate gene
GNRA	guanosine, any nucleoside, purine, adenosine
GRNA	guanine riboswitch
GTP	guanosine triphosphate
h	hour
H	hypoxanthine
HIV-1	human immunodeficiency virus-1
HuAp5/13	human *Nup50 Alu* construct with Ap at position 5/13

in vitro	experiments with organisms or their components outside of their natural biological setting
in vivo	experiments with organisms or their components inside of their natural biological setting
IRE	iron responsive element
K^{-1}	inverse Kelvin
kcal/mol	kilocalorie per mole
kDa	kiloDalton
k_{fold}	rate constant for folding
K_L	ligand binding affinity
K_{Mg}	magnesium binding affinity
k_{off}	off-rate constant
k_{unfold}	rate constant for unfolding
L(1,2,3)	purine riboswitch loops
MD	molecular dynamics
Mfold	algorithm used to predict RNA secondary structure
Mg^{2+}	magnesium divalent cation
$MgCl_2$	magnesium chloride
miRNA	micro RNA
mM	millimolar
mRNA	messenger RNA
n	Hill coefficent

N48	nucleotide 48 (purine riboswitches)
N48	nucleotide 48 of A/GRNA
N72	nucleotide 72 of A/GRNA
N74	nucleotide 74 (purine riboswitches)
NaCl	sodium chloride
ncRNA	non-coding RNA
nm	nanometre
nM	nanomolar
NMR	nuclear magnetic resonance
NMT1	N-myristoyltransferase 1
ns	nanosecond
nt(s)	nucleotide(s)
Nup50	nucleoporin 50 kiloDalton gene
P(1,2,3)	purine riboswitch stems
pbuE	purine base efflux pump gene
PCR	polymerase chain reaction
pri-RNA	primary micro RNA
ps	picosecond
RBD	RNA binding domain
RhAp5/13	rhesus *Nup50 Alu* construct with Ap at position 5/13
Rho (ρ)	prokaryotic transcription factor
RNA	ribonucleic acid

RNAP	RNA polymerase
rRNA	ribosomal RNA
s^{-1}	per second
SAM	S-adenosylmethionine
SRL	sarcin-ricin loop
SSF	steady-state fluorescence
ssRNA	single-stranded RNA
T	thymine
TAR	transactivating response element
TAT	transactivator protein
TCSPC	time correlated single photon counting
TE	tris ethylenediaminetetraacetic acid
TEMED	Tetramethylethylenediamine
TM	melting temperature (half population in single-strand state)
TPP	thiamine pyrophosphate
TRF	time-resolved fluorescence
TRIS-HCl	tris(hydroxymethyl)aminomethane hydrochloride
tRNA	transfer RNA
U	uracil
UTP	uridine triphosphate
UV	ultraviolet
V. vulnificus	*Vibrio vulnificus*

vide infra	refers reader to below part of text
WT	wildtype
xpt-pbuX	xanthine phosphoribosyltransferase and xanthine-specific purine permease operon
α15γ	group II catalytic intron
α_n	relative fluorescent amplitude
β-interferon	immune system signaling protein
ΔG_{DISS}	free energy of dissociation
ΔH_n	change in enthalpy
ΔS_n	change in entropy
μL	microlitre
μM	micromolar
τ	fluorescence lifetime

CHAPTER 1: Introduction

1.1 History and Utility of 2-Aminopurine

1.1.1. 2-Aminopurine Fluorescent Properties

2-Aminopurine (2Ap/Ap) is an analogue of the base adenine (A/6-Aminopurine) and has long been an invaluable tool that has deepened our understanding of nucleic acid functions from biological and technological perspectives (Figure 1-1). Currently, 2Ap is primarily utilized for its sensitive fluorescence properties, which are reflective of its local environment.[1-3]

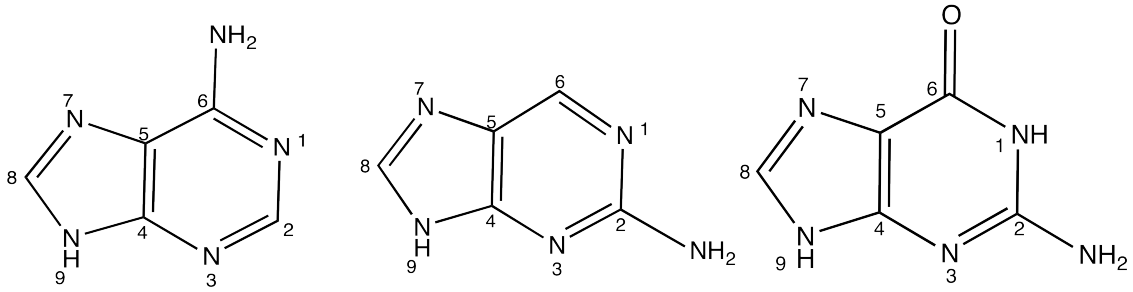


Figure 1-1. Chemical structures of 6-aminopurine (adenine), 2-aminopurine (2Ap), and guanine (G).

Investigations of why 2Ap exhibits a highly detectable fluorescence decay, compared to adenine, have been conducted by Mishra *et al*.[4] Here, *ab initio*

studies of free 2Ap in aqueous solution suggest two highly populated states: protonated nitrogen 7 (N7H), and protonated nitrogen 9 (N9H). Interestingly, these states existed in significantly lower quantities in adenine, when experiments were performed under the same conditions. It is believed that the relative abundance of N7H and N9H in 2Ap contributes to its high fluorescence quantum yield, as well as the red-shifted absorbance at 320 nm, in comparison to that of cellular nucleobases (average 260 nm).

The fluorescence decay of a molecule is well modeled by a Jablonski diagram (Figure 1-2). Here, a nucleotide such as 2Ap, may be excited at 320 nm from the singlet ground state, S_O, towards the higher energy S_1 or S_2 states. If the energy of excitation causes a 'jump' to the S_2 state, an internal conversion may exist, relaxing the system back to S_1 . From the S1 state, the molecule may return directly to the ground state S_0; for 2Ap this results in the emission of a photon at 350 nm (Stokes shift), and fluorescence decay. Alternatively, in certain systems containing heavy atoms (bromine, iodine), intersystem crossing may occur, whereby photons enter the first triplet state T_1. The transition from T_1 to S_0, also known as phosphorescence, is extremely slow compared to fluorescence decay, and is generally forbidden.

1.1.2. Early Cellular Investigations

Initial studies were unaware of its optic properties, primarily due to the lack of fluorescence measurement technology, and its established antibacterial properties.[5-11] G.H. Hitchings first reported the function of 2Ap in 1948, where he

demonstrated that it could outcompete other purines to hamper cellular replication.[5] A follow up study by Elion and colleagues in 1950 supported Hitchings' work by showing that 2Ap compromised bacterial growth with folic acid as a nutrient, yet the effects could be reversed by the addition of adenine, guanine (G), or folic acid.[6] These results expanded our understanding of the role of 2Ap within the bacterial cell. Specifically, the recovery upon addition of guanine suggested that there exists a crosstalk and conversion between cellular purines.

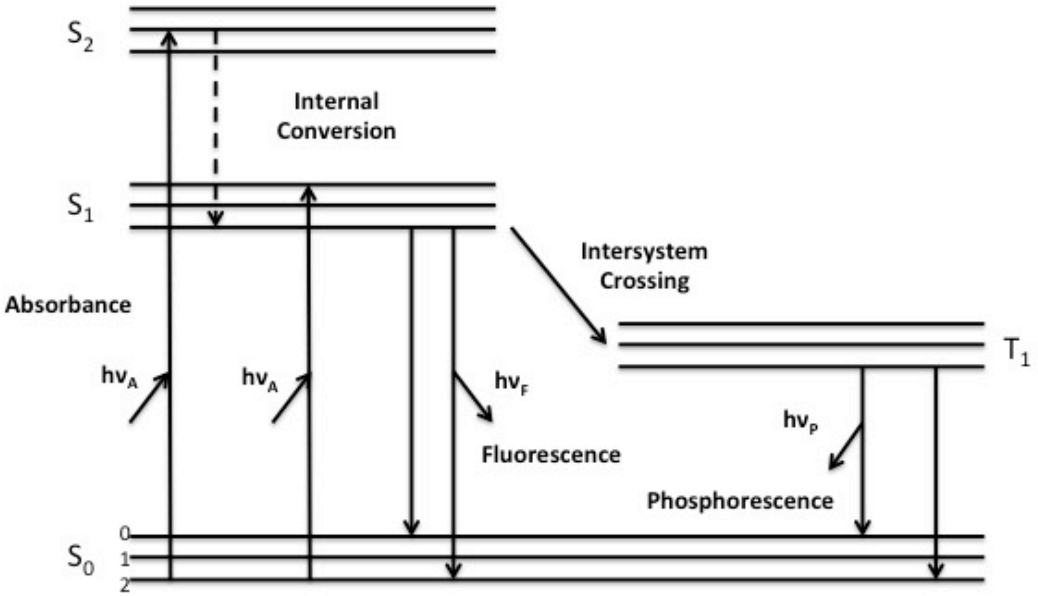


Figure 1-2. Modified form of a Jablonski diagram. S_0, S_1, and S_2 represent ground, first singlet, and second singlet states, respectively. Numbers 0,1, and 2, adjacent to S_0 represent vibrational levels; they exist in other states, but are excluded for clarity. The dashed line from S_2 to S_1 represents an internal conversion. hv_A, hv_F and hv_P represent energy of absorbance, fluorescence, and phosphorescence emission, respectively.

The role of 2Ap in eukaryotic cells differs however; In 1958 Bergmann et al. demonstrated that 2Ap is oxidized at position six by mammalian xanthine oxidase to form guanine.[7] Although this demonstrated the diversity of purine synthetic pathways across organisms, it also deflated the idea that 2Ap could be administered to mammals to thwart bacterial infections. Nonetheless, the results from Bergmann's study were serendipitous, as they facilitated the unveiling of the mammalian purine biosynthetic pathways, contributing to our current knowledge of intermediary metabolism.

1.1.3. Birth of Fluorescence-Based Investigations

With advances in fluorescence-based techniques in the late 1960s, highly sensitive spectroscopic investigations with 2Ap began to surface.[11,12] In solution, 2Ap has a high quantum yield and exhibits a single exponential decay of approximately 10 nanoseconds (ns), (absorption 320 nm, emission 350 nm). Conversely, when 2Ap is incorporated into a nucleic acid strand, its neighbouring nucleotides, whether adjacent or Watson-Crick paired, partially quench the fluorescence emission, resulting in a multi-exponential decay (Figure 1-3).[13] Such a profile suggests that 2Ap resides in multiple environments when incorporated or bound to nucleic acid. Thus, due to its environment-dependent decay, 2Ap has become a widely used probe of nucleic acid local/global conformational dynamics.[4,14,15]

In 1969, Ward and Reich published a seminal paper describing the fluorescent properties of 2Ap.[11] Under physiological conditions they performed

fluorescent intensity measurements with free 2Ap, and 2Ap internally placed in a RNA strand.

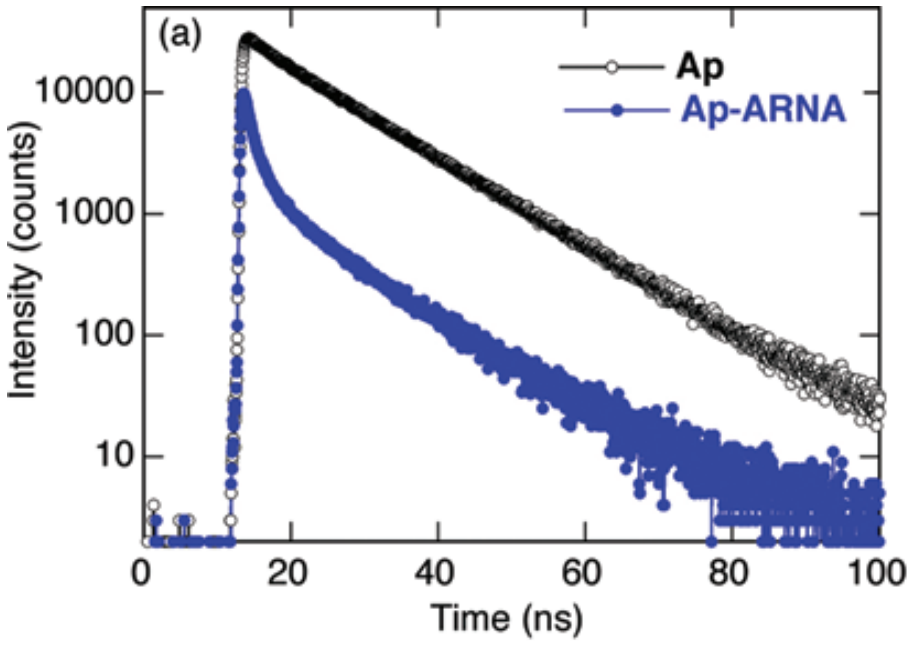


Figure 1-3. Log of fluorescence intensity vs. time (ns) for free 2-aminopurine (Ap), and 2-aminopurine bound to ARNA. Note the single exponential and multi-exponential decay for free and bound Ap respectively. Reproduced with permission from *Journal of the American Chemical Society* **2007**,*129*, 11308.

One of the key observations was that 2Ap could be selectively excited in the presence of normal bases due to its differential absorption compared to the average of cellular nucleobases (average 260 nm). Since the fluorescence intensity is directly related to the environment of the fluorophore, this allowed for site-specific, or local, monitoring under a variety of conditions that altered the conformation of a nucleic acid polymer (pH, temperature, cation concentration). This also laid the groundwork for 2Ap fluorescent studies, and emphasized its

potential as an extremely powerful tool to investigate protein-nucleic acid interactions,[16] nucleic acid conformational transitions,[17] and nuclease behaviours.[18]

In order to monitor the fluorescence of 2Ap, investigators utilize two broad techniques: steady-state fluorescence (SSF) and time-resolved fluorescence (TRF).[13] In the former, the fluorescence intensity is monitored over a constant wavelength and time scale (minutes) that exceeds the fluorescence lifetime of 2Ap. The observed fluorescence is an average of the intensities within a population, and is not necessarily indicative of the variable local/global conformers that may exist in solution. SSF is still a widely used technique however, as it can be used to determine site-specific folding/unfolding from marked fluorescence changes, and is significantly less costly to acquire and maintain than TRF machinery.

Although TRF requires expensive additional components, it is able to resolve fluorescence intensity in the nanosecond regime, and rather than exposure to a constant excitation wavelength, samples are exposed to pulses of light.[13] Unlike SSF, TRF provides a high-resolution view of multiple conformers in solution, by the multi-exponential fluorescence decay of 2Ap. Advanced TRF techniques involving lasers are able to resolve conformers in the femtosecond and picosecond regime, providing an even finer view of the conformational dynamics within a system.[19-21]

1.1.3 Continued Cellular Investigations

With the burgeoning field of 2Ap fluorescence studies, investigations of the molecule from a non-fluorescent perspective were not discarded. A variety of analyses on cellular systems paralleled the explosion of fluorescent studies, and unveiled cellular events including mutagenesis,[22] mismatch repair,[23] and protein kinase activity.[24]Although 2Ap is an intermediate in the mammalian purine biosynthetic pathway,[7] it is not necessarily restricted to such a role within the cell. A study by Zinn et al. in 1988 demonstrated that 2Ap is a protein kinase inhibitor, blocking the induction of β-interferon at the transcriptional level. Furthermore, addition of 2Ap to cellular media blocked the induction of *c-fos* and *c-myc* oncogenes, further suggesting that a protein kinase plays a role in regulation of the cell cycle.[24] These investigations not only broadened our understanding of kinase-mediated gene regulation, but also demonstrated that 2Ap is a diverse molecule that can shuttle to different environments within the cell.

In 1977, Goodman and colleagues hypothesized a specific explanation as to why 2Ap inhibited bacterial growth. They suggested that 2Ap converted Watson-Crick adenine-thymine (A-T) basepairs (bps) to guanine-cytosine (G-C) bps. [25] Assuming that 2Ap could base pair with T or C throughout the genome of an organism, mismatches upon replication in non-coding or coding regions, could have detrimental effects. Sequencing analyses confirmed that such conversions did occur, while nuclear magnetic resonance (NMR) investigations supported the

idea that 2Ap could alternatively pair with cytosine,[26] with the latter leading to a G-C mutagenesis (Figure 1-4).[27]

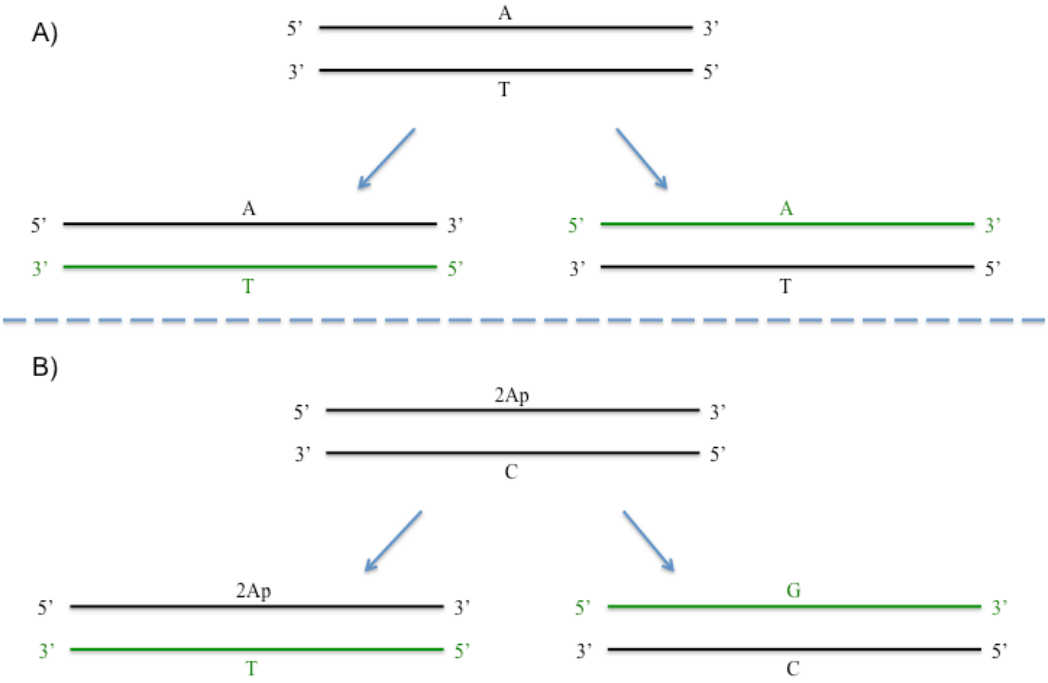


Figure 1-4. Normal vs. 2Ap-mutagenic replication. A) Under normal conditions, adenine (A) and thymine (T) pair during DNA replication. B) 2Ap is able to pair with C, causing G-C mutagenesis. Note strands are coloured differently to indicate semi-conservative nature of replication.

1.2 RNA Hairpins

A generic structure that has been heavily investigated over the last thirty years is the DNA/RNA hairpin. From an anatomical perspective hairpins are straightforward entities, consisting of a minimal, two basepair (bp) stem, capped by a loop composed of at least three nucleotides. *In vivo*, hairpin stems can be hundreds of bps, with intervening bulges, and capping loops averaging fifteen

nucleotides (Figure 1-5).[28] Although they serve imperative functions at the DNA level, such as telomere organization,[29] and antibody generation in higher eukaryotes,[30] hairpins are most abundant in non-coding (nc) RNA.

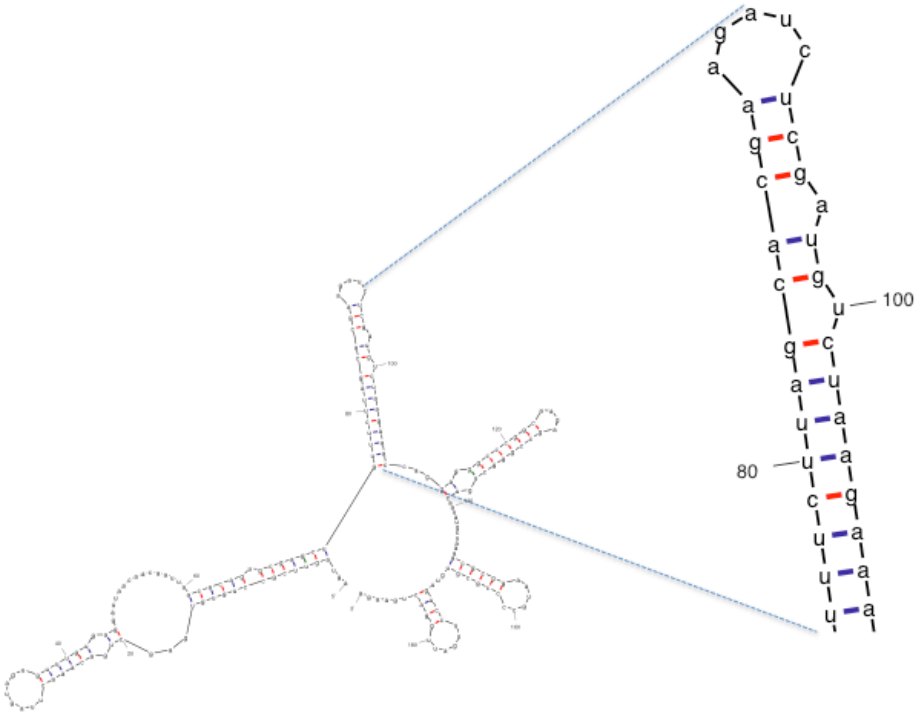


Figure 1-5. Mfold predicted structure at 37 degrees C of the 5'UTR of West Nile Virus (left). Zoom-in of hairpin, showing stem, bulge (position 100), and capping loop at top (right).

Across all kingdoms of life, hairpins are the heart of ncRNAs, and are believed to nucleate the assembly of the larger more complex RNAs in which they collectively reside.[31,32] Eukaryotic hairpins are involved in processes such as RNA interference,[33,34] and spliceosome assembly.[35] In prokaryotes hairpins may function independently to regulate Rho-independent transcription,[36] and temperature sensitive genes (e.g. RNA thermometers).[37] Furthermore, in viruses

including human immunodeficiency virus-1 (HIV-1), West Nile and influenza, hairpins serve as protein binding sites,[38,39] and are key elements in orchestrating viral replication.[40]

Hairpins perform functions both as independent entities, and substructures of large ncRNA complexes. Although they typically exist within large entities, investigating hairpins independently is a worthwhile endeavour, as they provide a glimpse of the RNA folding problem that exists *in vitro* and *in vivo*.

1.2.1 Using Hairpins to Study the RNA Folding Problem

First discussed in Tinoco's seminal paper, 'How RNA Folds,' the RNA folding problem is described as the kinetic and thermodynamic barriers a negatively charged RNA must overcome to fold into its native state.[41] Although ncRNAs are functional in their native states, the folding landscapes they traverse are of great interest, as conformers along such pathways are believed to be in dynamic equilibrium,[42,43] and may be 'signatures' of ncRNA function. For instance, nearly identical crystal structures have been presented for adenine and guanine riboswitches (Figure 1-6),[44,45] which are composed of hairpins, yet variability in sequence accounts for different folding pathways (see Chapter 2). Thus, investigations with small RNA hairpins are an effective approach to understanding the folding events that occur in larger, more complex structures found *in vivo*.

1.2.1.1 Probing RNA Hairpin Folding with 2-Aminopurine

2Ap serves as an excellent probe of RNA folding events due to its short, conformationally-sensitive fluorescent lifetimes,[1,13,46] which are about or less than the timescale by which RNA folds (ps-min).[47-49] Investigations of small RNA hairpins using 2Ap as a probe have provided fascinating results, demonstrating the amazing capability of ncRNAs.

Figure 1-6. Homologous crystal structures of ligand-bound ARNA and GRNA aptamer domains. Left, *Vibrio vulnificus* ARNA aptamer (PDB: 1Y26) with adenine bound (green). Right, *Bacillus subtilis* GRNA aptamer domain (PDB: 1Y27) with guanine bound (green).

A recurring theme exists in 2Ap probing of RNA hairpins; in the smallest, and seemingly elementary structures, a wide range of conformational dynamics

typically exists. Conversely, DNA hairpins of similar size and sequence generally obey a two-state folding model, whereby DNA hairpins appear to exist in folded and unfolded states[50,51]. This highlights the profound effects of the 2'OH, sugar pucker, and alternative helical structures found in RNA. Investigating the conformations of the intermediate(s) in equilibrium with unfolded and native states is of great biological and medical relevance, as many hairpins require proteins to chaperone folding into the native state,[52] and interrogating non-native conformers provides alternative avenues for drug design.[31,53]

A study by Hall and colleagues utilized a four-bp, six-nucleotide (nt) hairpin, representative of the iron –responsive element (IRE) RNA. This RNA interacts with the IRE-binding protein *in vivo* to regulate iron metabolism in metazoans. Here, 2Ap was substituted at variable sites within the loop to interrogate the level of structural dynamics using both SSF and TRF.[14] Since fluorescent lifetimes are correlated with nucleotide environment, short, intermediate, and long lifetimes were assigned to stacked, unstacked, and solvent-exposed 2Ap conformers, respectively. These observations were supported by subsequent molecular dynamics (MD) simulations, and together suggest that loop nucleotide dynamics contribute to the docking of the IRE binding protein. Similar results were observed by Sarkar and colleagues with 2Ap placed within the stem and loop of a three-bp tetraloop hairpin. Using thermal melts, and fast laser temperature jumps (TRF-based), a dynamic

equilibrium between native, frayed, unfolded, and unstacked conformers was proposed, and also complemented by MD simulations.[54]

An investigation in 2007 by Ballin et al. utilized SSF and TRF to characterize the fluorescence lifetime of the three states which a nucleotide could exist in within a hairpin: bulged, looped, and helical. 2Ap-substituted hairpins were subjected to thermal denaturation, urea titration, and cation-dependent folding. Fluorescence intensity, as a function of each variable, was plotted, but most importantly each plot was compared and normalized to a truncated, linear RNA with the same surrounding nucleotides.[55] The purpose of this approach was to emphasize the significant effect that neighbouring nucleotides have on 2Ap fluorescence intensity, acting as either antennae or quenchers of fluorescence. Aside from the proposed need for normalization, fluorescence profiles of the three general hairpin nucleotide states were presented, which serve as signatures of nucleotide environmental transitions that can be applied to subsequent investigations.

Viral RNA hairpin dynamics upon ligand binding have been recently investigated via ultra-fast TRF on a 2Ap-substituted hairpin segment of human immunodeficiency virus-1 trans activating response region (HIV-1 TAR) RNA. As anticipated, a wide range of conformational fluctuations was observed with unbound TAR RNA. Conversely, a significant fraction of fluorescence intensity was attributed to an intermediate or 'native-like' state upon binding of transactivator protein (TAT), TAR's endogenous ligand. Such findings support

the notion that proteins do not exclusively bind the native state. Furthermore, conformational dynamics were also observed upon interaction of the aminoglycoside antibiotic neomycin, indicating that the drug may bind more than one structure, and guide the RNA into a dysfunctional state.[56]

1.3 Riboswitches

A well-characterized element that involves hairpins is the riboswitch, which is a metabolite-binding RNA that regulates downstream gene expression.[57] Riboswitches typically reside in 5' untranslated regions (5'UTRs) of messenger RNAs (mRNAs), and are most abundant in prokaryotes,[58] emphasizing their role in a RNA-based world.

Structurally, they are composed of two domains: first, is the aptamer, which harnesses allosteric structural changes brought about by metabolite binding. Second, is the downstream expression platform, which alters its structure to regulate corresponding genes.[57] Aptamer domains are highly conserved, leading to the classification of seven different families that bind eight different molecules. Conversely, a higher degree of structural variability exists in the expression platforms. Since riboswitches control gene expression at the level of transcription, translation, and splicing, variability in the expression platform may allow a riboswitch to tune its mode of gene regulation.[58]

1.3.1 Modes of Riboswitch Regulation: Transcription, Translation, and Splicing

At the level of transcription, riboswitches regulate RNA synthesis in the coding region of mRNA by collapse or formation of intrinsic terminator stems, which are extended stem-loop structures typically followed by a poly-uracil (U) run. Depending on the riboswitch type, an intrinsic terminator stem is formed either by the presence or absence of metabolite.[58] Consequently, the RNA polymerase is derailed, causing a halt to transcription and subsequent gene expression. A well-documented example of terminator disruption upon metabolite binding involves the adenine riboswitch (ARNA), which is upstream of the purine efflux pump-encoding *pbuE* gene in *Bacillus subtilis*. Upon Watson-Crick interaction of adenine with the specificity nucleotide U-74 in the aptamer binding pocket, the entire domain undergoes a conformational change, leading to the disruption of a downstream terminator in the expression platform.[59] Additionally, in *B. subtilis*, the guanine riboswitch (GRNA) is found upstream of *xpt-pbuX*, which encodes for a protein involved in purine salvage pathways and transport. The binding pockets of purine riboswitches in *B.subtilis* are identical to that of ARNA, with the exception of the specificity nucleotide, where the ligand interacts with the guanine riboswitch containing a cytosine, as opposed to uracil at position 74. Interestingly, the GRNA regulates genes at the transcriptional level but, as opposed to ARNA, a terminator is stabilized upon metabolite binding.[44]

Regarding the translational level of regulation, a transcriptional terminator stem ceases to form, but structural changes to the ribosome binding-site can be induced in the full length mRNA.[57] This mode is well exemplified in an investigation of the coenzyme B12 riboswitch, whereby metabolite (coenzyme B12) binding to the aptamer domain induces structural changes, preventing the formation of stable ribosome complexes *in vitro*.[60] In support, mutant aptamer domains that were unable to bind coenzyme B12, produced transcripts that bound the ribosome with significantly higher affinity.[61,62]

Examples of riboswitch regulation at the level of transcription and translation are found through all kingdoms of life, but recently an alternative splicing mode of gene regulation in the thiamine pyrophosphate (TPP) riboswitch has been exclusively uncovered in eukaryotes, namely yeast and fungi. In a study by Chea et al. of the TPP riboswitch in *Neurospora crassa*, reverse transcription analysis and protein purification determined that an alternatively spliced transcript governs whether the downstream N-myristoyltransferase 1 (NMT1) mRNA will be translated. When the TPP concentration was low, nascently transcribed mRNA adopted a structure that occluded the second 5' splice site, resulting in expression of the NMT1 protein. Conversely, when TPP concentration was high, and bound to the aptamer domain, an allosteric change was induced that altered the splice product, preventing proper translation of *NMT1* mRNA.[63] Regulation by alternative splicing may not be exclusive to the TPP riboswitch; sequence and expression pattern comparisons of transcriptomes

and proteomes under variable metabolite conditions would help determine if other riboswitches modulate genes by this unconventional mode.

1.3.2 Probing Riboswitches with 2-Aminopurine

Similar to hairpins, the folding dynamics of riboswitches are of great interest, as distinct pathways may be signatures of function. In addition, identity of conformers along folding pathways can aid in drug design, as the majority of riboswitches are found in bacteria, viruses and lower eukaryotes.[58] A common theme surrounding riboswitches, and of this thesis, is the presence of homologous secondary and tertiary/native structures across families, yet with divergent function. Several studies of purine riboswitches have used 2Ap as a probe because it is a cognate ligand mimic of ARNA,[59,64-67] and a C74U conversion of GRNA (G_ARNA) in *B. subtilis* with 2Ap or adenine as the binding partner does not alter gene regulation *in vivo*.[68] Here RT-qPCR assays demonstrated that the levels of elongated transcripts in GRNA and G_ARNA were approximately equal.

The aptamer domains found in the purine riboswitches are the smallest in Nature. They consist of stems P1, P2 and P3, which are joined by a three-way junction at the centre. P1 forms the base of the aptamer domain, while P2 and P3 are capped by loops L2 and L3, respectively. Both magnesium and ligand binding facilitate 'kissing loop' interactions between L2 and L3, inducing a conformational change in the aptamer domain, causing the expression platform

to form anti-terminator and terminator stems in ARNA and GRNA, respectively (Figure 1-7).[69]

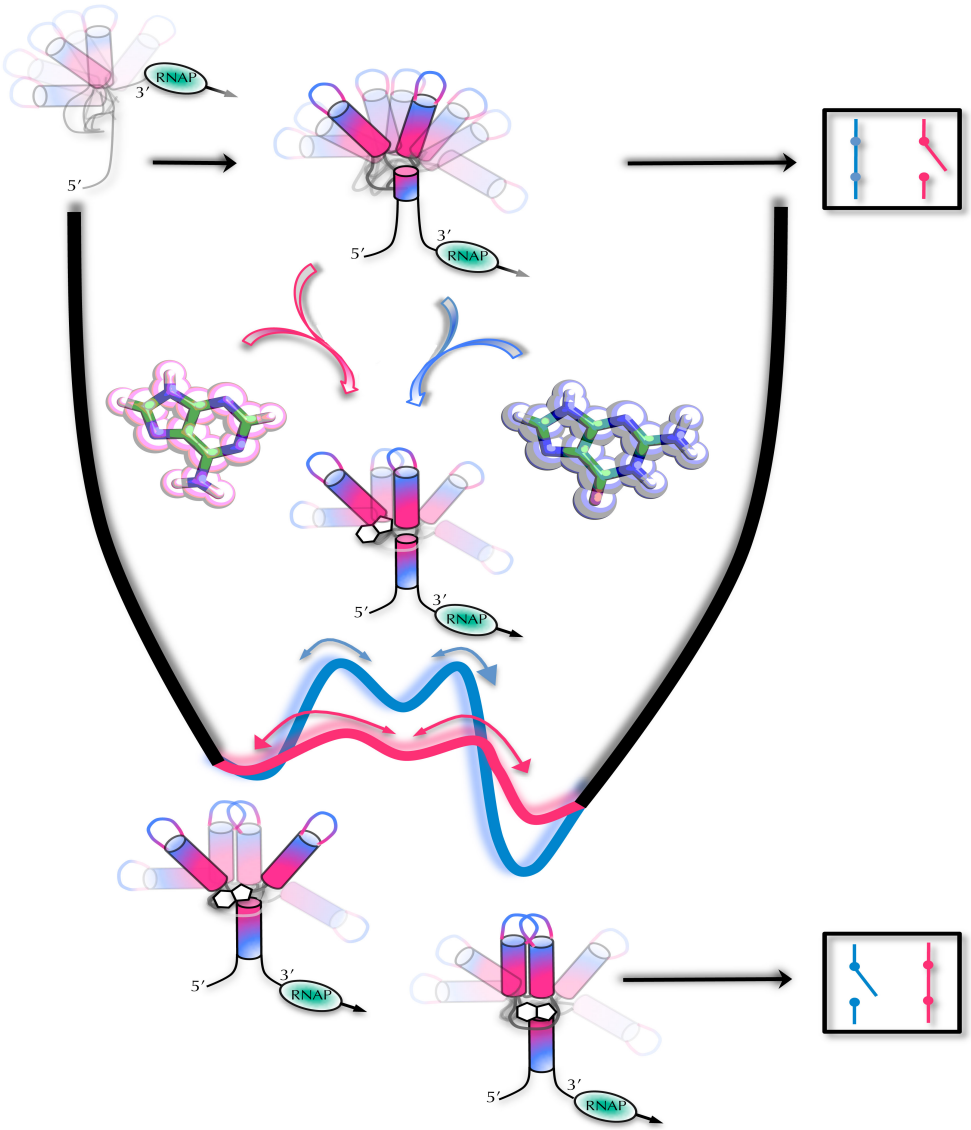


Figure 1-7. Cartoon of ARNA (pink) and G_ARNA (blue) aptamer domain folding with and without ligand binding. When no cognate ligand is present, the *pbuE* gene downstream of ARNA is turned off, while *xpt-pbuX* is turned on by G_ARNA (top), by formation and disruption of terminator stems, respectively. When ligand is present, both riboswitches are 'funelled' into their folding pathways, whereby *pbuE* is turned on and *xpt-pbuX* is turned off. Note: RNAP = RNA polymerase, double arrows highlight conformational equilibrium. Front cover reproduced with permission from Prychyna, O.; Dahabieh, M. S.; Chao, J.; O'Neill, M. A. *Biopolymers* **2009**, *91*, 953. Copyright © 2009, John Wiley & Sons, Inc. All rights reserved.

1.3.2.1 Probing ARNA Structure and Function with 2-Aminopurine

In an investigation of the ARNA, SSF was used to monitor the 2Ap fluorescence intensity upon titration of the aptamer domain, and full-length (aptamer and expression platform), with a constant amount of 2Ap. Results displayed a decrease in fluorescence intensity upon an increase in ARNA aptamer concentration, which validated the interaction of 2Ap and RNA. Most importantly, the decrease in fluorescence intensity suggests that 2Ap mediates RNA folding, as surrounding nucleotides sequester the emission.[59]

To further demonstrate that 2Ap mediates ARNA folding, L2 and L3 mutant aptamer constructs were prepared, resulting in compromised 'kissing loop' interactions. Since L2 and L3 Watson-Crick pairings have been suggested to aid folding of the three-way junction, or ligand-binding site, titration with L2 or L3 mutants quenched 2Ap fluorescence significantly less than the wildtype ARNA aptamer domain.[59]

Additional experiments of full-length ARNA indicate that a conformational competition exists between the aptamer domain and the expression platform. Using SSF, Lemay et al. titrated ARNA constructs ranging in size, with the minimum and maximum corresponding to aptamer and full-length ARNA respectively. Upon increasing the concentration of the aptamer domain, with 2Ap held constant, the fluorescence intensity was heavily quenched, as demonstrated above. Interestingly, as the length of the constructs approached that of the *entire* riboswitch, a more prominent terminator structure began to form, resulting in less

quenching of 2Ap fluorescence. Thermodynamically, this was rationalized by the significantly decreased free energy (as predicted by Mfold) of full-length terminator formation, compared to ligand-mediated aptamer domain folding.[59]

1.3.2.2 Probing GRNA Structure and Function with 2-Aminopurine

Investigations have also been conducted on GRNA in an effort to uncover how residues within the aptamer binding pocket tune the level of transcriptional regulation.[44,70,71] Studies were motivated by the observation that guanine riboswitches demonstrate a high degree of variability at nucleotide 48 (N48, ARNA numbering), which is across from nucleotide (N74) of the binding pocket.[68] As previously shown with ARNA, if N48 is complementary to N74, a Watson-Crick pairing may occur, inducing a self-collapse of the binding pocket, preventing the recognition of the cognate ligand.[64] If this same behaviour exists with GRNA, depending on the identity of N48, guanine may fail to associate with N74, causing the aptamer domain to alter its conformation. This would transmit a 'signal' to the downstream expression platform to secure an anti-terminator structure. Thus, the level of transcriptional regulation may be governed by the identity of N48, and purine riboswitches may have developed a range of ligand binding affinities to respond to environmental pressures.

To determine whether this behaviour exists with guanine riboswitches, G_ARNA was utilized due to its ability to bind 2Ap, yet retain all other GRNA folding and functional characteristics. In addition, a variety of N48 mutants were prepared, which included hypoxanthine (H), guanine, and adenine. SSF was

used to determine the level of 2Ap binding to U74 by the degree of fluorescence quenching, with the A48/U74 combination exhibiting the highest degree of fluorescence due to self-induced collapse of the binding pocket. The authors of this article make no attempt to analyze the specific nature of the N48/U74 interaction, but use in-line chemical probing assays to demonstrate that different aptamer conformers are formed with respect to N48, as H, G, A, and U at this position displayed variable aptamer domain cleavage patterns.[68]

1.3.2.3 Probing Ligand-Bound Dynamics of Purine Riboswitches with 2-Aminopurine

The aforementioned studies have greatly contributed to our understanding of purine riboswitches and riboswitches in general, but do not address dynamics of the ligand-bound state. Crystal structures of guanine and adenine riboswitches clearly broaden our understanding of RNA-ligand interactions, but provide a static view.[44,45] Furthermore, slow biphasic ligand-binding kinetics have stimulated the development of induced-fit models whereby a static, ordered structure is formed from ligand interacting with heteregenous RNA conformers.[72,73] In free A/GRNA, the stems are organized, while the junction and loop display a high degree of conformational dynamics. Such motions may contribute to the specificity of the binding pocket, as there appears to be a relationship between binding pocket closure and 'kissing loop' interactions.[59,64,67]

An examination of recent NMR experiments of purine riboswitches provides an alternative view. Ligand-bound GRNA exhibited local junction

dynamics within a static global conformer, yet both free and ligand-bound ARNA exhibited similar dynamics, questioning an induced fit model.[48,49] Given the observation that ARNA and GRNA exhibit nearly identical secondary and tertiary structures, alongside 90% sequence conservation in binding pocket residues, why is ARNA more dynamic than GRNA? If sequence variation is the contributing factor, can dynamics of the ligand-bound state provide clues to purine riboswitch gene function?

To address these questions, Eskandari et al., utilized TRF to study fluorescent lifetimes and fractions of 2Ap bound to ARNA. In a plot of fluorescence intensity (counts) vs. time (ns), free 2Ap exhibited a single exponential decay, whereas 2Ap bound to ARNA displayed a multi-exponential decay, suggestive of at least three environments in solution (Figure 1-3). Consequently, long, medium, and short lifetimes were assigned to open (C_1), intermediate (C_2) and closed (C_3) conformers, respectively. Fractions of each conformer were subsequently monitored upon varying temperature, concentrations of Mg^{2+}, and RNA. As anticipated, the highest melting temperature corresponded to C_C due to its presumably compact structure consisting of an array of hydrogen bonds and Van Der Waal's interactions. C_C also displayed the highest Mg^{2+} binding affinity, due to the conformer's relatively high demand to counter the charge-charge repulsion in such a compact structure. Additionally, when the amount of 2Ap was held constant, and ARNA

was titrated into solution, the equilibrium between conformers was driven towards C_C, indicated by its high fraction at low RNA concentrations (Figure 1-8).[66]

The presence of distinct lifetime populations attributed to open, intermediate, and closed conformers in the ligand-bound state, lends support to the aforementioned NMR studies observing dynamics in free and ligand-bound purine riboswitches.[48,49] A comparison of WT and mutant forms of ARNA and G_ARNA in chapter two further illustrates how sequence-dependent conformational dynamics may govern riboswitch function.

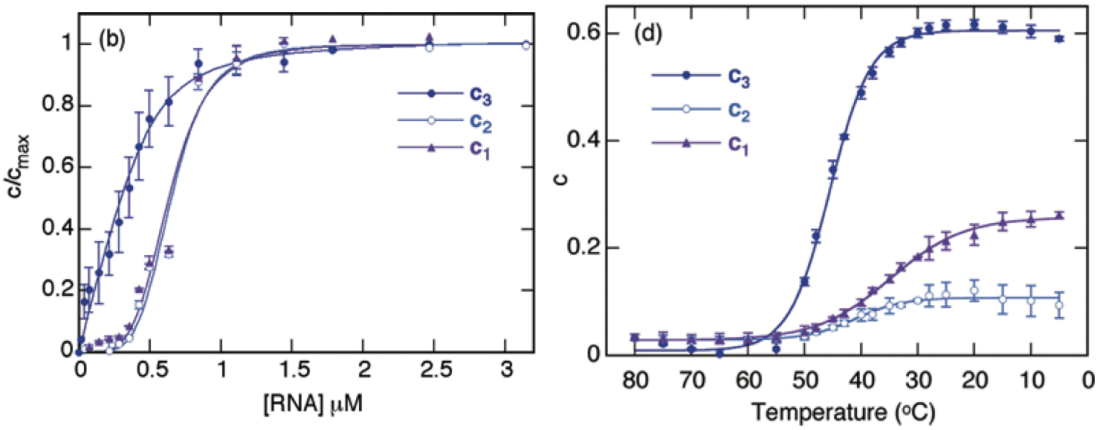


Figure 1-8. RNA titration and temperature variation curves of 2Ap fluorescent lifetime fractions (C_1 = open conformer, C_2 = intermediate, C_3 = closed) bound to *Bacillus subtilis* ARNA aptamer domain. C_3 is the most predominant conformer at low RNA concentrations, indicating that Ap (fixed concentration) drives the equilibrium to the closed state (left). The closed conformer is also the most thermostable, as its T_M is highest of the three (temperature shown in reverse) (right). Reproduced with permission from *Journal of the American Chemical Society* **2007**,*129*, 11308. Copyright © 2007 American Chemical Society. All rights reserved

Ligand-bound ARNA and GRNA regulate gene expression by disruption and formation of terminator structures. Thus, it is likely that the signal transmitted to

the expression platform is rooted in the subtle sequence differences between the purine riboswtiches.

1.4 RNA Editing

Sequence variability rooted at the genomic level, with riboswitches as a quintessential example, is not the only means by which an organism can achieve functional diversity. Oftentimes, the repertoire of RNA can be co- synthetically, or post-synthetically altered. With the exception of 5'-capping, polyadenylation and splicing, this process is coined 'RNA editing', whereby ribonucleotides are deleted, inserted, or chemically modified.[74] Benne and colleagues discovered RNA editing in the mid-1980s, where the insertion of uridines into the cytochrome oxidase subunit II mRNA in kinetoplasts of *Trypanosoma brucei* and *Crithidia fasciculata* was observed.[75] Examples of RNA editing now extend from unicellular protozoa to man, with substrates including transfer (t)RNAs,[76] ribosomal (r)RNAs,[77] signal recognition particle (7SL) RNAs,[78] and UTRs.[79]

The majority of RNA edits target mRNA sequences, where the creation of novel proteins, or truncation of proteins may ensue. In the mRNA of trypanosomatids, plants and man, nucleotide insertions, and cytosine to uracil (C-to-U) conversions create start and stop codons that are not represented by the genome.[74] tRNAs are also subject to high levels of RNA editing, resulting in remodelling at the primary, secondary and tertiary levels. Consequently, edited

tRNAs are able to change their identity by altering aminoacyl synthase specificity.[80]

Editing of rRNAs is less frequent than in tRNAs and mRNAs, yet has profound implications.[74] Single C-to-U conversions have been documented in the strictly conserved 530 loop of the small subunit rRNA of Dictyostelium mitochondria, resulting in modifications to tRNA docking and proofreading ability.[81] Furthermore, tandem arrays of C-to-U conversions have been observed in both the small and large rRNAs of *E.coli*, tuning the affinity of the ribosomal acceptor site binding.[82,83] Although infrequent, C-to-U conversions in rRNAs may be an alternative mode by which an organism alters its transcriptome in response to environmental pressures.

1.4.1 Adenosine-to-Inosine Editing of Primate Alu *RNA Duplexes*

Another type of RNA co/post transcriptional modifications that has garnered recent attention is adenosine-to-inosine (A-to-I) editing.[84] Here, adenosine deaminase acting on double-stranded RNA (ADAR) docks onto a stable RNA duplex, where it flips out adenosine, and hydrolytically deaminates the base to form inosine (Figure 1-9).[85] A-to-I edits are present in a variety of organisms,[86] but seem to be most abundant in primates due to the increased presence of stable dsRNA ADAR substrates (discussed in detail in Chapter 3).[87]

Recent studies of A-to-I edits in primates have sparked some debate, namely, whether the abundance of edits in humans that are surfacing in the literature are a result of innate sequencing errors, or whether the A-to-I edits previously documented were merely a glimpse of the edit sites in our transcriptome.[84,88-90] Interestingly, the Rechavi group performed a genome vs. transcriptome analysis of several known *Alu* elements across primates, corresponding to the 'same' genes, and observed that a much higher editing level exists in humans, specifically in the brain tissue.[91] If humans possess a higher level of A-to-I editing in the brain, then what is the cause?

One possibility could be the enzyme itself. In terms of amino acid sequence, an array of ADAR isoforms exists across species,[92] and this diversity may be compounded by post-translational modifications *in vivo*.[93] Another valid

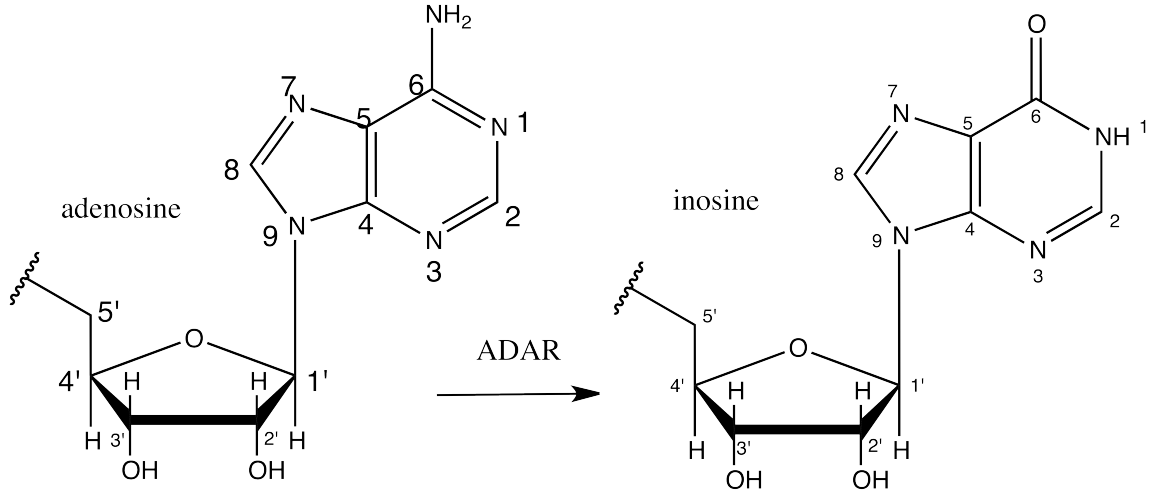


Figure 1-9. ADAR-mediated adenosine-to-inosine conversion. Note: This event occurs within a dsRNA duplex, but 3' terminal nucleotides are shown here for clarity.

explanation would be the RNA itself. The analysis of the *Alu* elements corresponding to the *Nup50* gene from the Rechavi group demonstrated that the human sequences had the highest A-to-I edit levels compared to chimp and rhesus.[91] However, an in depth analysis of sequence differences across primates was not pursued further.

1.4.1.1 Probing A-to-I Editing with 2-Aminopurine

An extensive discussion of the sequence comparison between primate *Nup50 Alu* elements is presented in Chapter three. To note briefly, our investigations of primate *Nup50 Alu* model constructs with an internal 2Ap have discovered that subtle sequence variations exist amongst nucleotides surrounding edit sites. Such variations have profound effects on duplex stability of our model constructs. Since a stable RNA duplex is a requirement for an A-to-I edit, we believe that subtleties in sequence found across primates contribute to differential editing levels.

We are the first to embark on 2Ap-based conformational dynamics studies of ADAR substrates. Investigations by others have utilized 2Ap within RNA duplexes to study A-to-I editing from an enzymatic and mechanistic perspective.[94,95] In a study by the Beal group, 2Ap was substituted into a RNA duplex, corresponding to positions of edited and unedited sites. An approximately four-fold increase in fluorescence intensity was observed when ADAR was incubated with an RNA substrate containing 2Ap at a previously established edit site, versus 2Ap positioned at an unedited site.[94] These

observations are consistent with a mechanism whereby the modified base is flipped out from the duplex,[85] resulting in less quenching of 2Ap fluorescence.

A follow up study by the same group treated 2Ap-substituted A-to-I substrates with full-length ADAR (ADAR-FL), or its RNA binding domain (RBD) to further understand the base-flipping mechanism. Binding assays demonstrated high-affinity interactions of both ADAR-FL and RBD, however only the ADAR-FL caused an increase in 2Ap fluorescence at edit sites. This indicated that the RBD alone is sufficient for docking, but insufficient for base-flipping, and hence editing.[95] Investigations similar to those above are scant, yet highly valuable, as they validate the A-to-I editing mechanism at the enzymatic and substrate level. Investigating changes in fluorescence of 2Ap within a substrate can complement sequencing evidence of edit sites and support proposed, edit site positions gathered from bioinformatics surveys.

1.5 Thesis Overview

The aim of this thesis is to investigate the effects of sequence-variation in folding dynamics of ncRNAs using 2Ap. Our studies emphasize structural homology, at both the two and three-dimensional levels, yet divergent function, rooted in subtle sequence variations.

Chapter 1 provides a history of 2Ap, from its initial use as an antibacterial agent, to its wide use as a fluorescent probe. A general view of RNA hairpins,

purine riboswitches, and mammalian A-to-I editing substrates is presented, followed by a summary of 2Ap-based conformational dynamics investigations for each.

Chapter 2 presents a manuscript (Prychyna, O.; Dahabieh, M. S.; Chao, J.; O'Neill, M. A. *Biopolymers* **2009,** *91,* 953) that describes the sequence-dependent conformational dynamics of adenine and guanine riboswitches in their ligand-bound state. We attribute unique folding pathways to divergent riboswitch function, despite nearly identical ligand-bound crystal structures.

Chapter 3 presents conformational dynamics analyses of RNA hairpin constructs, representative of *Nup50 Alu* dsRNA, which is differentially edited (A-to-I) across primates. Here, SSF, and TRF are used to interrogate the environment of an internally substituted 2Ap. We conclude that the construct corresponding to the human *Nup50 Alu* dsRNA is more thermostable overall, and has a more prevalent, and thermostable intermediate at physiological temperature. The prevalence of such a conformation likely accomodates the A-to-I editing enzyme (ADAR) most effectively *in vivo*, relative to nearly identical sequences found in chimp and rhesus. Enhanced editing levels found in humans are believed to be associated with our enhanced cognition, as well as glioblastomas and psychiatric disorders. We believe that editing levels are partially governed by the sequence-dependent conformational dynamics of dsRNA substrates.

Chapter 4 presents future experiments and perspectives. Specifically, recent investigations of functional RNA intermediates are discussed, and ideas for future 2Ap-directed experiments of riboswitches and A-to-I editing substrates are presented.

1.6 References

1. Jean, J. M.; Hall, K. B. *Proc Natl Acad Sci U S A* **2001**, *98*, 37.
2. Jean, J. M.; Hall, K. B. *Biochemistry* **2002**, *41*, 13152.
3. Serrano-Andres, L.; Merchan, M.; Borin, A. C. *Proc Natl Acad Sci U S A* **2006**, *103*, 8691.
4. Mishra, S. K.; Shukla, M.K.; Mishra, P.C. *Spectrochim Acta* **2000,** *56,* 1355.
5. Hitchings, G. H.; Elion, G. B.; Vanderwerff, H. *Fed Proc* **1948**, *7*, 160.
6. Elion, G. B.; Hitchings, G. H. *J Biol Chem* **1950**, *187*, 511.
7. Bergmann, F.; Levin, G.; Kwietny, H. *Biochim Biophys Acta* **1958**, *30*, 509.
8. Rudner, R. *Z Vererbungsl* **1961**, *92*, 361.
9. Frederiksen, S. *Biochem Pharmacol* **1965**, *14*, 651.
10. Danilov, V. I.; Krugliak Iu, A.; Kuprievich, V. A.; Shramko, O. V. *Biofizika* **1967**, *12*, 726.
11. Ward, D. C.; Reich, E.; Stryer, L. *J Biol Chem* **1969**, *244*, 1228.
12. McClure, W. R.; Scheit, K. H. *FEBS Lett* **1973**, *32*, 267.
13. Lakowicz, J. R. **1999**.
14. Hall, K. B.; Williams, D. J. *Rna* **2004**, *10*, 34.
15. Rachofsky, E. L.; Seibert, E.; Stivers, J. T.; Osman, R.; Ross, J. B. *Biochemistry* **2001**, *40*, 957.
16. Hariharan, C.; Reha-Krantz, L. J. *Biochemistry* **2005**, *44*, 15674.

17. Ballin, J. D.; Prevas, J. P.; Bharill, S.; Gryczynski, I.; Gryczynski, Z.; Wilson, G. M. *Biochemistry* **2008**, *47*, 7043.

18. Mitsis, P. G.; Kwagh, J. G. *Nucleic Acids Res* **1999**, *27*, 3057.

19. Wang, Q.; Raytchev, M.; Fiebig, T. *Photochem Photobiol* **2007**, *83*, 637.

20. Wan, C.; Fiebig, T.; Schiemann, O.; Barton, J. K.; Zewail, A. H. *Proc Natl Acad Sci U S A* **2000**, *97*, 14052.

21. Xia, T. *Curr Opin Chem Biol* **2008**, *12*, 604.

22. Fidalgo da Silva, E.; Mandal, S. S.; Reha-Krantz, L. J. *J Biol Chem* **2002**, *277*, 40640.

23. Diver, W. P.; Woodcock, D. M. *Mutagenesis* **1989**, *4*, 302.

24. Zinn, K.; Keller, A.; Whittemore, L. A.; Maniatis, T. *Science* **1988**, *240*, 210.

25. Goodman, M. F.; Hopkins, R.; Gore, W. C. *Proc Natl Acad Sci U S A* **1977**, *74*, 4806.

26. Fazakerley, G. V.; Sowers, L. C.; Eritja, R.; Kaplan, B. E.; Goodman, M. F. *Biochemistry* **1987**, *26*, 5641.

27. Holliday, R. *Genet Res* **2007**, *89*, 285.

28. Morse, D. P.; Bass, B. L. *Proc Natl Acad Sci U S A* **1999**, *96*, 6048.

29. Sundquist, W. I.; Klug, A. *Nature* **1989**, *342*, 825.

30. Ou, Z.; Bottoms, C. A.; Henzl, M. T.; Tanner, J. J. *J Mol Biol* **2007**, *374*, 1029.

31. Bowman, G. R.; Huang, X.; Yao, Y.; Sun, J.; Carlsson, G.; Guibas, L. J.; Pande, V. S. *J Am Chem Soc* **2008**, *130*, 9676.

32. Zhang, W.; Chen, S. J. *Proc Natl Acad Sci U S A* **2002**, *99*, 1931.

33. Zhou, H.; Xia, X. G.; Xu, Z. *Nucleic Acids Res* **2005**, *33*, e62.

34. Howard-Till, R. A.; Yao, M. C. *Mol Cell Biol* **2006**, *26*, 8731.

35. Oubridge, C.; Ito, N.; Evans, P. R.; Teo, C. H.; Nagai, K. *Nature* **1994**, *372*, 432.

36. Lesnik, E. A.; Sampath, R.; Levene, H. B.; Henderson, T. J.; McNeil, J. A.; Ecker, D. J. *Nucleic Acids Res* **2001**, *29*, 3583.

37. Rinnenthal, J.; Klinkert, B.; Narberhaus, F.; Schwalbe, H. *Nucleic Acids Res* **2010**, *38*, 3834.

38. Chang, K. Y.; Tinoco, I., Jr. *J Mol Biol* **1997**, *269*, 52.

39. Akarsu, H.; Burmeister, W. P.; Petosa, C.; Petit, I.; Muller, C. W.; Ruigrok, R. W.; Baudin, F. *Embo J* **2003**, *22*, 4646.

40. Li, X. F.; Jiang, T.; Yu, X. D.; Deng, Y. Q.; Zhao, H.; Zhu, Q. Y.; Qin, E. D.; Qin, C. F. *J Gen Virol* **2010**, *91*, 1218.

41. Tinoco, I., Jr.; Bustamante, C. *J Mol Biol* **1999**, *293*, 271.

42. Brion, P.; Westhof, E. *Annu Rev Biophys Biomol Struct* **1997**, *26*, 113.

43. Zhuang, X.; Bartley, L. E.; Babcock, H. P.; Russell, R.; Ha, T.; Herschlag, D.; Chu, S. *Science* **2000**, *288*, 2048.

44. Batey, R. T.; Gilbert, S. D.; Montange, R. K. *Nature* **2004**, *432*, 411.

45. Serganov, A.; Yuan, Y. R.; Pikovskaya, O.; Polonskaia, A.; Malinina, L.; Phan, A. T.; Hobartner, C.; Micura, R.; Breaker, R. R.; Patel, D. J. *Chem Biol* **2004**, *11*, 1729.

46. Hall, K. B. *Methods Enzymol* **2009**, *469*, 269.

47. Furtig, B.; Wenter, P.; Reymond, L.; Richter, C.; Pitsch, S.; Schwalbe, H. *J Am Chem Soc* **2007**, *129*, 16222.

48. Buck, J.; Furtig, B.; Noeske, J.; Wohnert, J.; Schwalbe, H. *Proc Natl Acad Sci U S A* **2007**, *104*, 15699.

49. Furtig, B.; Buck, J.; Manoharan, V.; Bermel, W.; Jaschke, A.; Wenter, P.; Pitsch, S.; Schwalbe, H. *Biopolymers* **2007**, *86*, 360.

50. Kim, J.; Doose, S.; Neuweiler, H.; Sauer, M. *Nucleic Acids Res* **2006**, *34*, 2516.

51. Woodside, M. T.; Anthony, P. C.; Behnke-Parks, W. M.; Larizadeh, K.; Herschlag, D.; Block, S. M. *Science* **2006**, *314*, 1001.

52. Woodson, S. A. *RNA Biol* **2010**, *7*, 677.

53. Verhelst, S. H.; Michiels, P. J.; van der Marel, G. A.; van Boeckel, C. A.; van Boom, J. H. *Chembiochem* **2004**, *5*, 937.

54. Sarkar, K.; Nguyen, D. A.; Gruebele, M. *Rna* **2010**, *16*, 2427.

55. Ballin, J. D.; Bharill, S.; Fialcowitz-White, E. J.; Gryczynski, I.; Gryczynski, Z.; Wilson, G. M. *Biochemistry* **2007**, *46*, 13948.

56. Lu, J.; Kadakkuzha, B. M.; Zhao, L.; Fan, M.; Qi, X.; Xia, T. *Biochemistry* **2011**, *50*, 5042.

57. Winkler, W. C.; Breaker, R. R. *Chembiochem* **2003**, *4*, 1024.

58. Mandal, M.; Breaker, R. R. *Nat Rev Mol Cell Biol* **2004**, *5*, 451.

59. Lemay, J. F.; Penedo, J. C.; Tremblay, R.; Lilley, D. M.; Lafontaine, D. A. *Chem Biol* **2006**, *13*, 857.

60. Oltean, S.; Banerjee, R. *J Biol Chem* **2005**, *280*, 32662.

61. Ravnum, S.; Andersson, D. I. *Mol Microbiol* **2001**, *39*, 1585.

62. Ravnum, S.; Andersson, D. I. *Mol Microbiol* **1997**, *23*, 35.

63. Cheah, M. T.; Wachter, A.; Sudarsan, N.; Breaker, R. R. *Nature* **2007**, *447*, 497.

64. Lemay, J. F.; Lafontaine, D. A. *Rna* **2007**, *13*, 339.

65. Mandal, M.; Breaker, R. R. *Nat Struct Mol Biol* **2004**, *11*, 29.

66. Eskandari, S.; Prychyna, O.; Leung, J.; Avdic, D.; O'Neill, M. A. *J Am Chem Soc* **2007**, *129*, 11308.

67. Prychyna, O.; Dahabieh, M. S.; Chao, J.; O'Neill, M. A. *Biopolymers* **2009**, *91*, 953.

68. Mulhbacher, J.; Lafontaine, D. A. *Nucleic Acids Res* **2007**, *35*, 5568.

69. Kim, J. N.; Breaker, R. R. *Biol Cell* **2008**, *100*, 1.

70. Ottink, O. M.; Rampersad, S. M.; Tessari, M.; Zaman, G. J.; Heus, H. A.; Wijmenga, S. S. *Rna* **2007**, *13*, 2202.

71. Kim, J. N.; Roth, A.; Breaker, R. R. *Proc Natl Acad Sci U S A* **2007**, *104*, 16092.

72. Gilbert, S. D.; Stoddard, C. D.; Wise, S. J.; Batey, R. T. *J Mol Biol* **2006**, *359*, 754.

73. Wickiser, J. K.; Cheah, M. T.; Breaker, R. R.; Crothers, D. M. *Biochemistry* **2005**, *44*, 13404.

74. Gott, J. M.; Emeson, R. B. *Annu Rev Genet* **2000**, *34*, 499.

75. Benne, R.; Van den Burg, J.; Brakenhoff, J. P.; Sloof, P.; Van Boom, J. H.; Tromp, M. C. *Cell* **1986**, *46*, 819.

76. Gott, J. M.; Somerlot, B. H.; Gray, M. W. *Rna* **2010**, *16*, 482.

77. Fernandes, A. P.; Nelson, K.; Beverley, S. M. *Proc Natl Acad Sci U S A* **1993**, *90*, 11608.

78. Ben-Shlomo, H.; Levitan, A.; Shay, N. E.; Goncharov, I.; Michaeli, S. *J Biol Chem* **1999**, *274*, 25642.

79. Rosenberg, B. R.; Hamilton, C. E.; Mwangi, M. M.; Dewell, S.; Papavasiliou, F. N. *Nat Struct Mol Biol* **2011**, *18*, 230.

80. Borner, G. V.; Morl, M.; Janke, A.; Paabo, S. *Embo J* **1996**, *15*, 5949.

81. Barth, C.; Greferath, U.; Kotsifas, M.; Fisher, P. R. *Curr Genet* **1999**, *36*, 55.

82 Mahendran, R.; Spottswood, M. S.; Ghate, A.; Ling, M. L.; Jeng, K.; Miller, D. L. *Embo J* **1994**, *13*, 232.

83. Miller, D.; Mahendran, R.; Spottswood, M.; Costandy, H.; Wang, S.; Ling, M. L.; Yang, N. *Semin Cell Biol* **1993**, *4*, 261.

84. Li, M.; Wang, I. X.; Li, Y.; Bruzel, A.; Richards, A. L.; Toung, J. M.; Cheung, V. G. *Science* **2011**, *333*, 53.

85. Goodman, R. A.; Macbeth, M. R.; Beal, P. A. *Curr Top Microbiol Immunol* **2011**.

86. Zamyatnin, A. A., Jr.; Lyamzaev, K. G.; Zinovkin, R. A. *Biochemistry (Mosc)* **2010**, *75*, 1316.

87. Eisenberg, E.; Nemzer, S.; Kinar, Y.; Sorek, R.; Rechavi, G.; Levanon, E. Y. *Trends Genet* **2005**, *21*, 77.

88. Hayden, E. C. *Nature* **2011**, *473*, 432.

89. Barak, M.; Levanon, E. Y.; Eisenberg, E.; Paz, N.; Rechavi, G.; Church, G. M.; Mehr, R. *Nucleic Acids Res* **2009**, *37*, 6905.

90. Greenberger, S.; Levanon, E. Y.; Paz-Yaacov, N.; Barzilai, A.; Safran, M.; Osenberg, S.; Amariglio, N.; Rechavi, G.; Eisenberg, E. *BMC Genomics* **2010**, *11*, 608.

91. Paz-Yaacov, N.; Levanon, E. Y.; Nevo, E.; Kinar, Y.; Harmelin, A.; Jacob-Hirsch, J.; Amariglio, N.; Eisenberg, E.; Rechavi, G. *Proc Natl Acad Sci U S A* **2010**, *107*, 12174.

92. Keegan, L. P.; Brindle, J.; Gallo, A.; Leroy, A.; Reenan, R. A.; O'Connell, M. A. *Embo J* **2005**, *24*, 2183.

93. Glisovic, T.; Bachorik, J. L.; Yong, J.; Dreyfuss, G. *FEBS Lett* **2008**, *582*, 1977.

94. Stephens, O. M.; Yi-Brunozzi, H. Y.; Beal, P. A. *Biochemistry* **2000**, *39*, 12243.

95. Yi-Brunozzi, H. Y.; Stephens, O. M.; Beal, P. A. *J Biol Chem* **2001**, *276*, 37827.

CHAPTER 2: Sequence-Dependent Structural Dynamics of Purine Riboswitches

This chapter comprises the manuscript "Sequence-Dependent Structural Dynamics of Purine Riboswitches," which has been featured on the front cover and published in *Biopolymers* (**2009**, 91, 953-965).

Oksana Prychyna, Michael S. Dahabieh, Jay Chao, and Melanie A. O'Neill

The adenine (ARNA) and guanine (GRNA) riboswitches contain aptamer domains that are nearly identical in their secondary and tertiary structures, yet they perform opposing functions. In *Bacillus subtilis*, upon binding of its cognate ligand (adenine), the ARNA aptamer domain undergoes a conformational change, relaying a signal to the downstream expression platform, resulting in the formation of an anti-terminator stem, activating the *pbuE* gene. Conversely, in GRNA, and upon binding of guanine, the aptamer domain relays a signal to the

expression platform to promote the formation of a terminator hairpin, and turn off the *xpt-pbuX* gene.

Using 2Ap as a cognate ligand mimic for ARNA and GRNA with a C74U mutations (G_ARNA), which does not affect riboswitch function *in vivo,* we have probed the conformational dynamics of the ligand-bound purine riboswitch aptamer domains from *B. subtilis.* Here we demonstrate that each aptamer domain undergoes a unique folding landscape en route to a structurally conserved native state. We believe that despite the high similarities in secondary and tertiary structures, the conformational dynamics, and effectively the function of the purine riboswitches is rooted at the sequence, or primary level.

Using Mg^{2+}, temperature variations, and mutagenesis, we demonstrate that GRNA is more pre-organized towards a folded state, and less dependent on 'kissing loop' interactions between L2/L3. Furthermore, in both riboswitches, we show that the P1 stem formation is imperative to ligand binding, and hence folding of the aptamer domain. Our results are the first to compare ARNA and GRNA as a dynamic equilibrium upon ligand binding, clearly demonstrating that dynamics are not exclusive to the unbound state.

The thesis author's contribution accounted for 40% of the work, which included, PCR-based mutagenesis, transcription, RNA purification, TRF fluorescence on mutant constructs, and data analysis for ARNA, and G_ARNA. Oksana Prychyna's contribution accounted for 50% of the work, which was identical to the thesis author's contribution, with the exception of TRF

fluorescence analysis on all wild type constructs. Jay Chao performed the remaining 10% of the work, which involved replicate Mg^{2+} titrations on ARNA and G_ARNA constructs.

2.1 Keywords

riboswitches; RNA folding; mutagenesis; 2- aminopurine; time-resolved fluorescence

2.2 Abstract

Riboswitch regulation of gene expression requires ligand- mediated RNA folding. From the fluorescence lifetime distribution of bound 2Ap, we resolve three RNA conformers (C_o, C_i, C_c) of the liganded G- and A-sensing riboswitches from *Bacillus subtilis*. The ligand binding affinities, and sensitivity to Mg^{2+}, together with results from mutagenesis, suggest that C_o and C_i are partially unfolded species compromised in key loop-loop interactions present in the fully folded C_c. These data verify that the ligand-bound riboswitches may dynamically fold and unfold in solution, and reveal differences in the distribution of folded states between two structurally homologous purine riboswitches: Ligand-mediated folding of the G-sensing riboswitch is more effective, less dependent on Mg^{2+}, and less debilitated by mutation, than the A-sensing riboswitch, which remains more unfolded in its liganded state. We propose that these sequence-dependent RNA dynamics, which adjust the balance of ligand-mediated folding and unfolding, enable different degrees of kinetic discrimination in ligand binding,

and fine-tuning of gene regulatory mechanisms.

2.3 Introduction

The discovery of functionally-diverse, non-coding RNA continues to challenge our understanding of the chemistry of this molecule, and its biological roles.[1] Riboswitches are a quintessential example: These RNA molecules, widespread in prokaryotes,[2-5] and being uncovered in eukaryotes,[6-9] directly bind a small molecule ligand in order to independently control gene expression at the level of transcription or translation. Riboswitches that regulate transcription typically occupy a few hundred nucleotides of the 5'UTR immediately upstream of a gene(s). Ligand binding to the riboswitch aptamer domain is thought to influence the conformation of a downstream expression platform, by, e.g., promoting or disrupting the formation of an intrinsic terminator hairpin. Unlike protein transcription factors, the efficacy of these RNA regulators requires ligand binding to rapidly adapt to the structure of the nascent RNA, as it is synthesized by RNA polymerase.[10-12] Molecules that most effectively target riboswitch regulation of gene expression might not be those of the highest affinity, but those that most effectively funnel RNA dynamics into regulatory-active conformations. Design of such molecules[13,14] requires an understanding of these ligand-mediated folding pathways.

Guanine (G) riboswitches are involved in widespread G-mediated regulation of genes in the purine biosynthetic pathway.[15] The conserved RNA

motif comprising the aptamer domain of these riboswitches is distributed in numerous bacteria, and it contributes to the regulation of 16 genes in *Bacillus subtilis* alone. For instance, G binding to the *xpt-pbuX* aptamer domain (GRNA) represses expression of downstream genes encoding xanthine phosphoribosyltransferase and xanthine permease; the mechanism is thought to involve formation of an intrinsic terminator in the expression platform (Fig. 2-1). In several bacteria, modified GRNA sequences with mutations including C74U (*B. subtilis* GRNA numbering) have been discovered in other regions of the genome.[15,16] In *B. subtilis,* one such sequence corresponds to the aptamer domain of an adenine (A) responsive riboswitch (ARNA) that activates expression of the downstream *pbuE* gene encoding a purine efflux pump,[16-18] presumably by disruption of an intrinsic terminator (Fig. 2-1).

The structures of ligand-bound, purine ribowitch aptamer domains have been examined through high-resolution crystallography and NMR,[19-21] and the thermodynamics, and kinetics of their RNA-ligand interactions probed in a series of mechanistic investigations.[12,15,16,22-25] The secondary structures, as well as the tertiary fold of all purine-bound riboswitch aptamer domains known to date are highly homologous (Fig. 2-1). These consist of a three-way junction that carves out the purine binding pocket by joining the paired stems (P1, P2, P3). Stem P1 is linked to the expression platform, while P2 and P3 are capped by loops L2 and L3, respectively. Although *B. subtilis* ARNA has not been crystallized, the high resolution structure of its homolog, the aptamer domain of *Vibrio vulnificus add*-ARNA, closely resembles that of GRNA from *B. subtilis* (Fig. 2-1).[20] Ligand

specificity is encoded by Watson-Crick pairing with N74 which is strictly conserved as C or U in G- and A-sensing riboswitches, respectively. The single base mutation, C/U at N74, exchanges their G/A selectivity, as gauged by ensemble-averaged ligand dissociation constants.[16,22,23] In C74U mutant G riboswitches, all residues that fold directly around the purine are identical to those in the A riboswitch (Fig. 2-1). In both riboswitches, N48 is flipped out into solution where it makes no contacts with ligand or RNA. The identity of this nucleotide is variable, except that in A riboswitches, A is excluded, while in G riboswitches, G is excluded. This is thought to prevent Watson-Crick pairing between N48 and N74, which may fold the unbound riboswitch core, and mimic the ligand-bound state.[24]

Considerable sequence variation amongst A and GRNA exists outside the binding pocket (Fig. 2-1).[15,16,24] These variations have little effect on secondary and ensemble-averaged tertiary structure,[19,20] or the overall thermodynamics of binding.[16,22] Yet, they impact function, since mutation of the specificity nucleotide N74 alone is insufficient to interchange G/A regulation of gene expression *in vivo*.[16] These variations in primary sequence may confer different conformational dynamics on the purine riboswitch aptamer domains. A variety of experiments suggest that free of ligand and/or Mg^{2+} ions, the purine riboswitches are heterogeneous mixtures of conformers in solution.[15,16,22,26-31] These observations have prompted induced-fit models that couple binding with RNA folding, locally or globally, depending on the extent of preorganization in the unbound RNA. They

have also suggested that unbound GRNA may be more ordered, while unbound ARNA may be more conformationally flexible.[28,29]

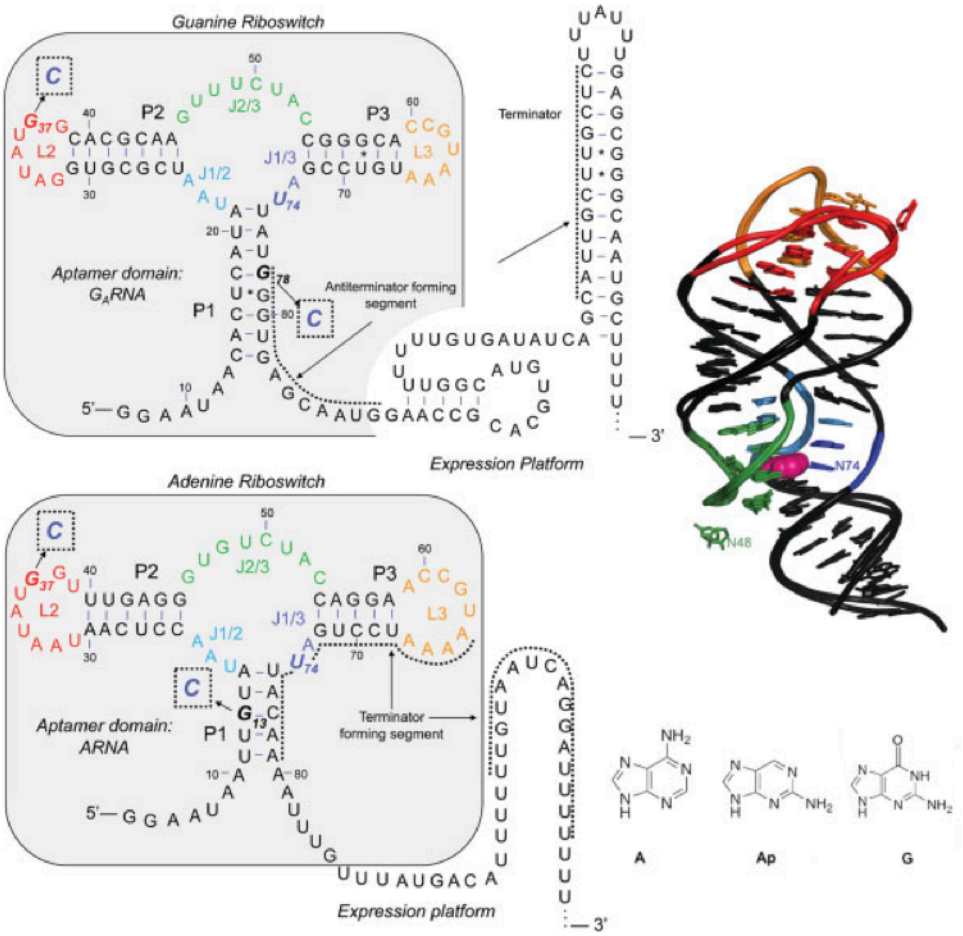


Figure 2-1: Structures of the purine riboswitches and ligands. Shown are the primary sequences and predicted secondary structures of the complete *xpt-pbuX* and *pbuE* purine riboswitches from *Bacillus subtilis*, GRNA, and ARNA, respectively, and the tertiary structure of the aptamer domain of GRNA, and the adenine riboswitch from *Vibrio vulnificus* (PDB: 1Y27, 1Y26, Ref. 20). Sequences of aptamer domains investigated here are highlighted in grey; loop (G37C) and P1-stem (G13C, G78C) mutant sites noted.

Dynamics of the ligand-bound riboswitches have received less attention: The complexes are presumably much more ordered than the free RNA, possibly

existing as a single well-defined tertiary structure in solution. Yet, ligand may not be sufficient to lock down the dynamics, and a conformationally rigid liganded state may not be optimal for riboswitch regulation of transcription, especially if operating under kinetic control. For instance, the ability to partially unfold/refold while ligand is bound may provide opportunities for ligand discrimination, fine tuning of gene regulation by other cellular parameters, and escape from misfolding events.

Conformational heterogeneity within liganded riboswitches was revealed by our recent time-resolved fluorescence investigation of 2-aminopurine (Ap) bound to ARNA from *B. subtilis*.[32] Ap is a natural probe of A-responsive riboswitches as its interactions with these RNA molecules mimic those of A structurally, thermodynamically, and kinetically.[12,16,22,26] From the fluorescence lifetime distribution we resolved three distinct ligand-bound species of ARNA, assigned to three RNA conformers, C_1, C_2, and C_3. The ligand binding affinity, and sensitivity of folding to Mg^{2+}, was substantially higher for C_3, than for C_1 or C_2. The distribution of these conformers, measured in ensemble solution, correlated well with that obtained from rates of folding and partial unfolding of ligand-bound ARNA in single molecule FRET experiments.[26] Consequently, we proposed that C_1, C_2, and C_3 differ in global tertiary structure, and predicted that the folding/unfolding dynamics may be diminished in the liganded state of the structurally homologous GRNA. Here, we provide corroborating evidence for folded (closed, $C_c = C_3$) and partially unfolded (intermediate and open, $C_i = C_2$

and C_o = C_1) conformers through mutagenesis. We test our prediction by comparative investigation of the Ap-bound conformers of ARNA and C74U GRNA; the latter (abbreviated as G_ARNA) possesses an A/Ap binding pocket within a GRNA scaffold (Fig. 2-1). With time-resolved fluorescence we sensitively detect differences in the distribution of folding intermediates between ARNA and G_ARNA. These may impart different degrees of "leakage" into the regulatory mechanisms, and different modes of ligand discrimination.

2.4 Results

2.4.1 Probing Liganded-RNA Conformer Equilibria in Ensemble Solution with Fluorescence Lifetimes

As previously reported for ARNA,[32] in contrast to the single exponential decay of free Ap (t = 11.5 ns, χ^2 ~ 1.2), the time-resolved fluorescence of Ap in the presence of saturating concentrations of G_ARNA, is not fit by a single ($\chi^2 \geq 10$), or a double exponential decay ($\chi^2 \geq 3.5$); it is instead well modeled ($\chi^2 \leq 1.3$) by three exponentials (Fig. 2-2). A significant reduction in the initial fluorescence intensity (loss in peak intensity for a constant run time), or "static" quenching is also detected. This corresponds to an additional component whose lifetime(s) is < 300 ps and, therefore, beyond the resolution of our instrument. Although the lifetime of this species is not measured, its relative population, readily evaluated as the fraction of static fluorescence quenching, dictates our conclusions. The

complete intensity versus time profile of Ap in the presence of either RNA is thus described by eq. 2-1,

$$I(t) = fC_u e(-t/\tau_u) + fC_o e(-t/\tau_o) + fC_i e(-t/\tau_i) + fC_c e(-t/\tau_c) \qquad \textbf{2-1}$$

where fC_c is the statically quenched population with a lifetime, t_C < 300 ps. The lifetimes (t_u, t_o, and t_i) and relative populations (fC_u, fC_o and fC_i) of the other species are obtained from the time-resolved decay, and their populations are adjusted to account for fC_c (Materials and Methods). The fluorescence lifetime distribution described by equation 2-1 reveals that Ap populates at least four distinct environments in the presence of either ARNA or G_ARNA.[33] While a larger number of environments may exist, more than four are not convincingly separated from Ap fluorescence lifetime distribution, irrespective of the instrument resolution.

The longest lifetime, t_u = 11.5 ns, matches that of free Ap, and, in the presence of excess, saturating RNA, $f_u \leq 0.1$. Consequently, t_u is attributed to unbound ligand, and f_u is the fraction of unbound ligand. In our preliminary investigation with ARNA,[32] we proposed that the remaining lifetimes correspond to Ap bound within distinct RNA conformers. Since large scale RNA motion is effectively frozen during the excited-state lifetime (ps-ns), each bound conformer is resolved as a component in the fluorescence decay, with an amplitude (fC_n) equal to its relative population in the ensemble. We assigned the statically

quenched species to Ap bound to a fully folded, "closed" conformer, C_c (previously denoted C_3), resembling that observed at high resolution by crystallography and NMR; C_c may be a single conformer, or a mix of similar structures each having $t_{Ap} < 300$ ps.

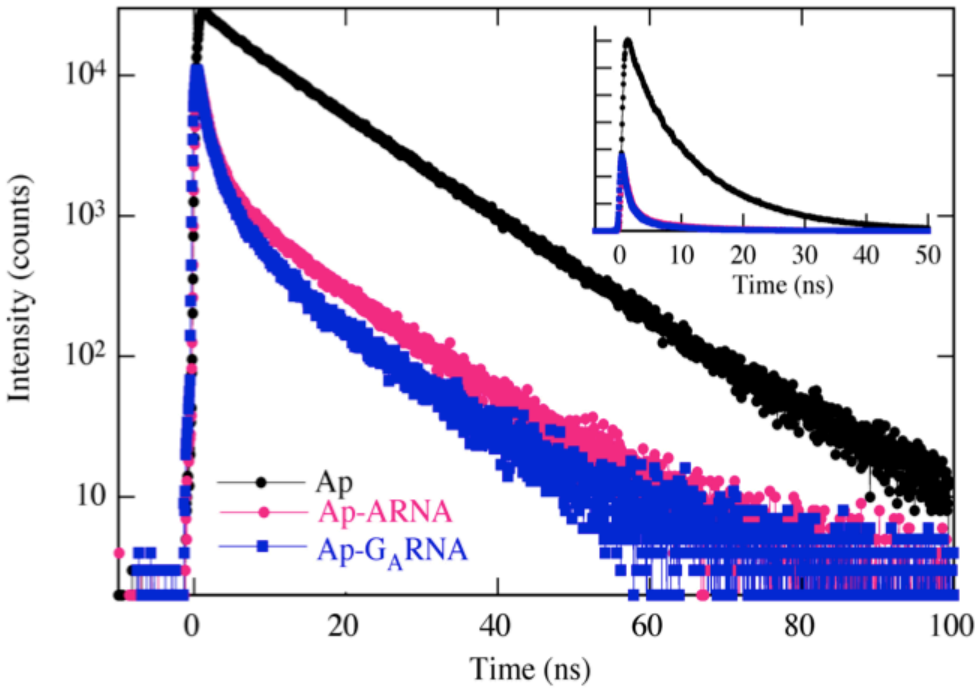


Figure 2-2: Fluorescence decay profile of free Ap, and Ap in the presence of either ARNA or G_ARNA (500 nM Ap, 4 μM RNA, 20°C, 10 mM Mg^{2+}) measured by time correlated single photon counting for a constant run time. The latter are characterized by significantly reduced intensity at zero time ("static" quenching) and a triexponential fluorescence decay. Inset shows decay profiles on linear scale.

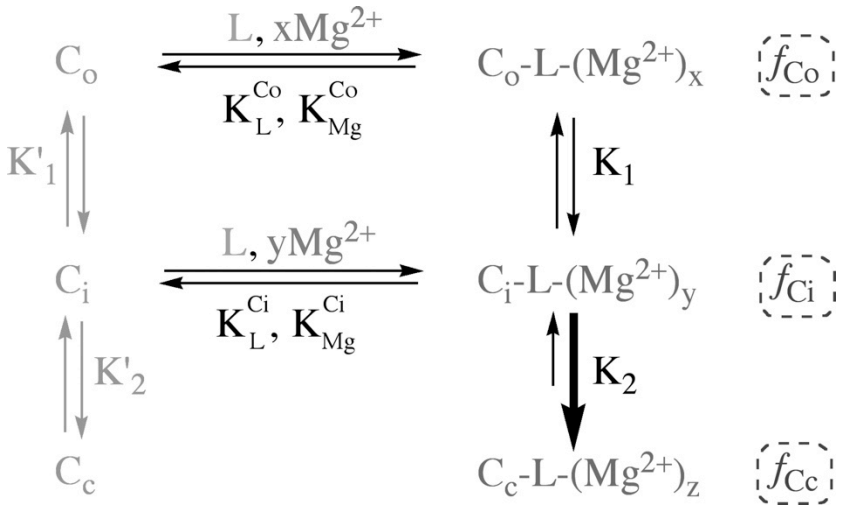
We assigned the partially quenched species as Ap bound to two different, partially unfolded conformers, intermediate, C_i, and open, C_o (previously denoted C_2 and C_1), with lifetimes of (0.5 ± 0.2) and (2.0 ± 0.2) ns, respectively. The

results described below aim to test these assignments and compare the conformer distribution in ARNA and G_ARNA.

Under the conditions represented in Figure 2-2 (10 mM Mg^{2+}, 20°C, excess RNA, Ap:RNA ≤ 1:8), the population of partially quenched species constitutes ~ 25 %: fC_o = 0.14 ± 0.02, fC_i = 0.13 ± 0.02, fC_c = 0.62 ± 0.02, and f_u = 0.10 ± 0.02, the same within experimental error for ARNA and G_ARNA. However, the population of each bound conformer varies with the ratio of Ap:RNA and Mg^{2+}:RNA. If C_o and C_i are folding intermediates, this variation may be understood from the coupled equilibria for ligand binding and folding (Scheme 2-1, see Discussion).

These binding and conformational equilibria may be probed in experiments where the total concentration of either RNA or Mg^{2+} is varied, and the distribution of ligand-bound species is evaluated stepwise. The first experiment provides K_L, the ligand dissociation constant for each of the three conformers (not for the three- just for C_o and C_i now). The second experiment

provides a measure of the sensitivity of ligand-mediated folding of each



Scheme 2-1: Coupled equilibria for Mg^{2+}-dependent ligand binding to three (but not horizontal for Cc) RNA conformers (horizontal arrows) that are folding/unfolding in unbound (left vertical arrows) and bound (right vertical arrows) forms.

conformer (each conformer) to Mg^{2+}, K_{Mg}. Both experiments yield values for K_{-1} = $[C_o\text{-L-}(Mg^{2+})_x]/[\ C_i\text{-L-}(Mg^{2+})_y] = fC_o/fC_i$ and $K_2 = [C_C\text{-L-}(Mg^{2+})^z]/[C_i\text{-L-}(Mg^{2+})_y] = fC_C/fC_i$ measures of the equilibria for exchange between liganded conformers, as a function of Ap:RNA or Mg^{2+}:RNA ratio.

A few points should be noted in reference to Scheme 2-1. First, these fluorescence experiments observe only the ligand-bound forms of C_o, C_i, and C_c. Since we do not probe free RNA, they cannot address the structures of the unbound conformers, nor evaluate K'_{-1} and K'_2, or K_{Mg} in the absence of bound ligand. Existing data suggest that the unbound purine riboswitch aptamer domains adopt multiple conformers,[15,16,22,26-31] possibly even a fully folded

structure, transiently, and if ligand is mediating folding, then K'_2 should be less than K_2. Second, we have not included a pathway for direct binding of ligand to C_c as it is unlikely given that the ligand is completely buried in the fully folded structure.[20] The pathway to liganded C_c is believed to be through binding to C_i and/or C_o. Third, because only bound complexes can be observed, values of K_{Mg} are for the liganded RNA, and because Mg^{2+} is required for ligand binding (*vide infra*), values of K_L presented are those in the presence of Mg^{2+}.

2.4.2 Ligand Binding Affinities of Purine Riboswitch Conformers and Mutants

The binding affinity of each conformer, C_o, C_i, and C_c, for Ap ligand (K_L) may be evaluated in RNA titration experiments. As in our previous report with ARNA, we find that it is the relative population of each Ap-bound conformer, fC_n (eq. 2-1), and not its fluorescence lifetime, that varies most significantly with RNA concentration. Here, we have conducted additional replicates with ARNA in order to compare directly to experiments with G_ARNA, and to increase precision and accuracy of the fC_n values, particularly fC_i and fC_o. We have adopted two approaches to evaluate fC_i and fC_o, and limit their measurement at very low RNA concentration, where C_i and C_o are most weakly populated (Materials and Methods). Consequently, their binding curves are fit to a hyperbolic model like C_c, and the resulting K_L values for C_o and C_i are ~ 4-fold larger than those derived from fitting to a sigmoidal model.[32] The K_L value for C_c is the same,

within experimental error, and the values for C_o and C_i are significantly higher than C_c, as previously reported.

Shown in Figure 2-3 are the ligand binding curves for "wild type" A-sensing, ARNA (a) and G_ARNA (b), measured in the presence of 10 mM Mg^{2+}; the corresponding K_L values (Materials and Methods) are given in Table 2-1. These data present two important observations: 1) For both riboswitches, the Ap binding affinities of C_i and C_o are significantly lower (~25-160-fold) than that of C_C.) The ensemble K_L, evaluated from time-resolved Ap fluorescence, is ~ 2-fold smaller for G_ARNA than for ARNA, predominantly due to tighter binding of G_ARNA C_c (Table 2-1). As a result, the variation in relative binding affinity of the conformers is larger for G_ARNA (~160) than for ARNA (~25). We note that this difference in ensemble K_L values is not detected by steady-state fluorescence, where we find K_L = (300 ± 20) nM for either ARNA or G_ARNA (Fig. 2-4), comparable to previous reports.[16,22,26] This is presumably because the steady-state measurement is skewed towards the species with the longer fluorescence lifetimes, rendering differences in binding affinity due to the short-lived C_c difficult to detect in the ensemble average.

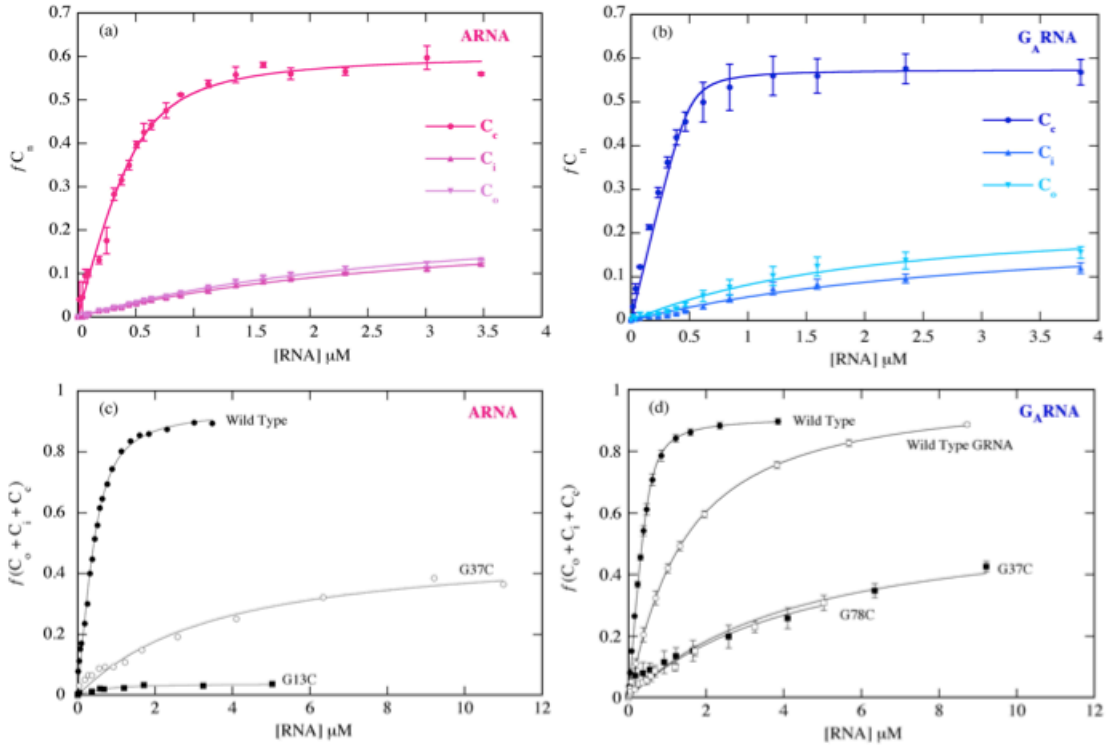


Figure 2-3: Ligand binding to wild-type and mutant purine riboswitches. Top panels show the variation in the fraction of Ap-bound conformers, C_o, C_i, and C_c, of ARNA (a) and G_ARNA (b) as a function of RNA concentration (500 nM Ap, 10 mM Mg^{2+}, 20°C). Bottom panels show the variation in total population bound for wild type, loop- and P1 stem-mutant ARNA (c) and G_ARNA (d) as a function of RNA concentration (500 nM Ap, 10 mM Mg^{2+}, 20°C). Comparison with wild-type GRNA is also shown in panel (d). Plotted are the average data points from three to six experiments with standard errors, and fits (lines) to 1:1 binding model [Eq. (2-7)] which yield the K_L values of the individual conformers, and the ensemble average.

			ARNA		G_ARNA	
			K_L (μM)[b]	K_L rel	K_L (μM)[b]	K_L rel
	Sequence					
Ap	WT	En.[c]	0.20 ± 0.02	1	0.080 ± 0.02	1
		C_c	0.10 ± 0.02	0.5	0.015 ± 0.005	0.2
		C_i	2.4 ± 0.3	12	2.6 ± 0.5	33
		C_o	2.5 ± 0.3	13	1.6 ± 0.4	20
	G37C	En.	3.2 ± 0.5	16	4.0 ± 0.5	50
	G13C	En.	Not detected	>100		
	G78C	En.			3.9 ± 0.5	50
			K_{Mg} (μM)[d]	K_{Mg} rel	K_{Mg} (μM)[d]	K_{Mg} rel
Mg^{2+}	WT	En.	300 ± 20		280 ± 20	
		C_c	200 ± 20	1	150 ± 20	1
		C_i	1500 ± 300	7	1100 ± 200	7
		C_o	2700 ± 400	13	2000 ± 400	13

[a] Averages and standard errors of three to six experiments in Fluorobuffer at 20°C.
[b] Ap ligand binding evaluated at 10 mM Mg^{2+}; K_L values from fit to hyperbolic binding model for 1:1 complex.
[c] Ensemble average.
[d] K_{Mg} values evaluated from fit to Hill model at RNA:Ap ratio ~8 (see text for additional details).

Table 2-1: Ligand Dissociation Constants (K_L) and Mg^{2+}-Dependence of Ligand-Mediated Folding (K_{Mg}) for Individual Conformers and Ensemble Averages of Wild-Type and Mutant ARNA and G_ARNA[a]

Also presented in Figure 2-3 and Table 2-1 are the Ap binding curves and dissociation constants for two site-specific mutants, one in the remote loop, L2, and the other in the P1 stem (Fig. 2-1). The G37C mutation disables one of the Watson-Crick hydrogen bonds between L2 and L3,[20] and is among the most deleterious loop mutations to ligand binding affinity, and presumably loop-loop interactions, in ARNA.[26] The G13C mutation introduces a C-C mismatch in the P1 stem of ARNA; the G-C pair at this position in wild type ARNA is proposed to be a structural keystone essential for zipping up of the P1 stem.[34] In G_ARNA, the corresponding P1 stem C-C mismatch is provided by the G78C mutant. Each mutation substantially reduces (~20 - > 100 fold) the ligand binding affinity of ARNA and G_ARNA (Fig. 2-3c,d). We report only the ensemble averages

because very little C_i and C_o are populated by the mutant riboswitches, and measurement of their very weak binding affinities is not feasible (Materials and Methods). Similar results were observed with the somewhat less deleterious C52G ARNA loop mutant[26] (data not shown). The K_L values of the G37C loop mutants are the same for ARNA and G_ARNA, within experimental error, although the loss in affinity relative to wild type is greater for G_ARNA (Table 2-1). In contrast, while G78C-G_ARNA binds Ap with similar or greater affinity than its loop mutant, no binding of Ap by G13C-ARNA is detected, and the loss in affinity relative to wild type is greater for ARNA. Both the P1 stem and remote loop mutations have a greater impact on ligand binding affinity than mutation of the specificity nucleotide, N74, in the binding pocket. Site-specific binding of hypoxanthine to *B. subtilis* (*pbuE*) ARNA and A to *B. subtilis* (*xpt-pbuX*) GRNA has been detected by NMR,[30] and binding of Ap to GRNA is only ~ 10-fold weaker than to ARNA.[24]

2.4.3 Mg^{2+}-Dependence of Ligand-Mediated Folding of Purine Riboswitch Conformers

The variation in the relative population of liganded C_o, C_i, and C_c with Mg^{2+} concentration (at a constant Ap:RNA ratio) is shown for wild type ARNA and G_ARNA in Figure 2-4. In the absence of Mg^{2+}, addition of saturating concentrations of ARNA (RNA:Ap ~ 8) leaves the Ap fluorescence largely unchanged; the fraction bound is effectively zero. In contrast, addition of G_ARNA

in the absence of Mg^{2+} quenches ~ 20 % of the Ap fluorescence, predominantly through binding to C_c. This difference is seen in Figure 2-4 at the limit near zero Mg^{2+}.

For both RNAs, the population of each conformer increases with Mg^{2+} concentration, exhibiting a sigmoidal dependence, analogous to that observed in ensemble-averaged experiments.[26,35] These data may be fit to a Hill model (Materials and Methods) to extract the midpoint of each folding transition, K_{Mg}.[36] The K_{Mg} value thus reflects the sensitivity of each folding transition to Mg^{2+} concentration. Since Mg^{2+} association is accompanied by both ligand binding and RNA folding, K_{Mg} includes contributions from intramolecular RNA and RNA-ligand interactions, as well as those from Mg^{2+}.[37,38] The latter includes the affinity-weighted average for binding at specific sites, detected by crystallography[19,20] and NMR[28,29] as well as non-specific association by the diffuse ion atmosphere[37] in complexes, RNA:Ap:$(Mg^{2+})_x$. Our analysis makes no attempt to estimate x (or y, or z, Scheme 2-1) or to separate the contributions to binding affinity of site-specific and ion atmosphere interactions.[39] For both riboswitches, K_{Mg} values (Table 2-1) vary as: $C_c << C_i < C_o$, each slightly smaller for G_ARNA than for ARNA. As in ensemble measurements, the near unity Hill coefficients (n = 1-2) suggest that the folding transitions are largely uncooperative with respect to changes in Mg^{2+} concentration.[40]

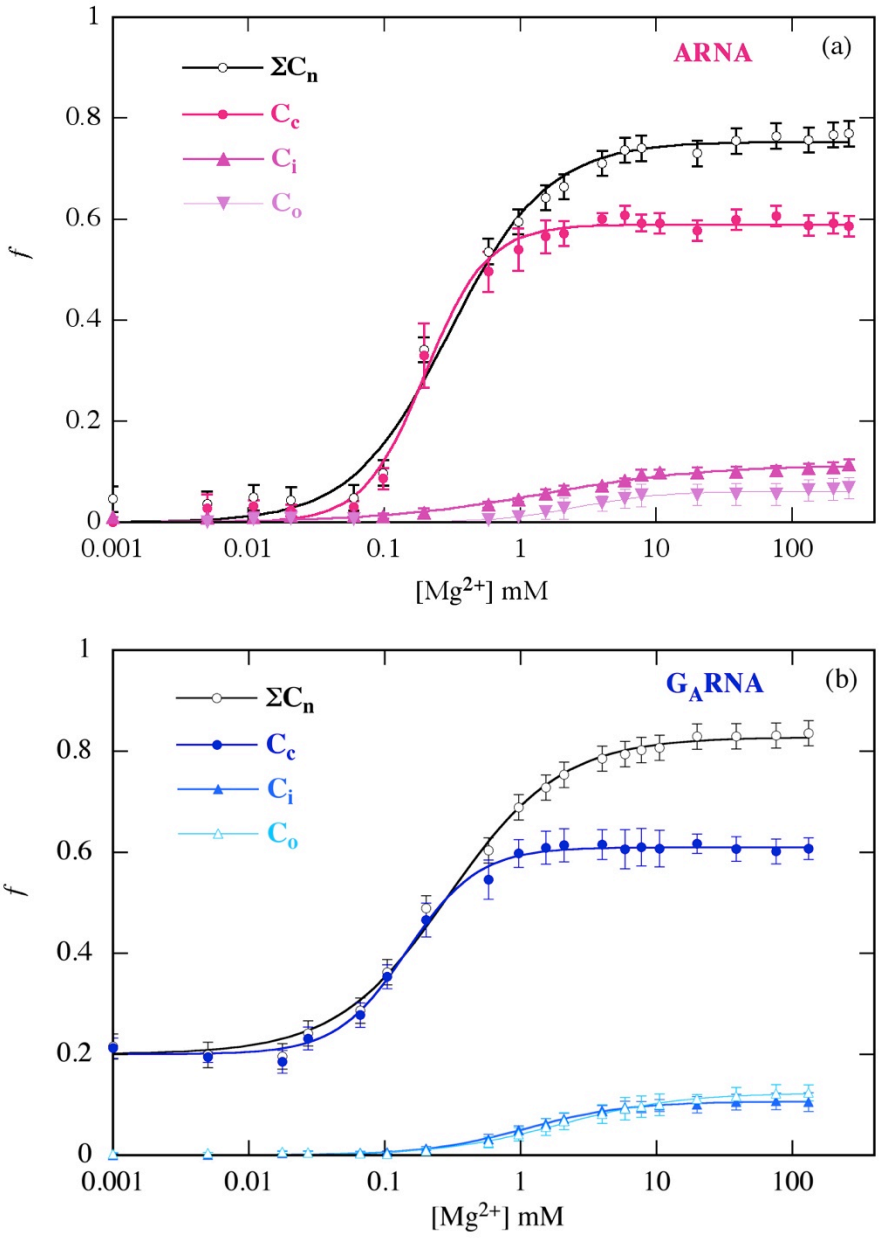


Figure 2-4. Mg^{2+}-dependence of ligand-mediated folding of the three conformers of ARNA (a) and G_ARNA (b). Shown are the variation in the fraction of Ap-bound conformers, C_o, C_i, and C_c, together with the total population bound (ΣC_n) as a function of Mg^{2+} concentration (500 nM Ap, RNA:Ap ~ 8, 20°C). Plotted are the average data points from 3-6 experiments with standard errors, and fits (lines) to Hill model (eq. 2-7) which yield the K_{Mg} values for the individual conformers and the ensemble average. The difference between panels (a) and (b) in the low Mg^{2+} limit is a consequence of the ~ 20 % fluorescence quenching observed upon addition of G_ARNA, but not ARNA to Ap in the absence of Mg^{2+}.

2.5 Discussion

A range of conformational dynamics has been suggested to accompany ligand-mediated folding of the purine riboswitches.[22,26-31,34,41,42] In both time-resolved NMR and single molecule experiments, at least two partially unfolded species, differing significantly in tertiary and/or secondary structure are detected in the presence of ligand.[26,30,34] Although presumably representing intermediates on route to the fully folded structure, the kinetic[34] and thermodynamic[26,32,34] barriers separating these species are small (≤ 1-3 kcal/mol). We propose that the three Ap-bound species of ARNA and G_ARNA observed by time-resolved fluorescence correspond to two such partially unfolded intermediates, C_o and C_i, that are in reversible conformational exchange with the fully folded structure, C_c. The ligand binding affinities and conformer distributions of wild type and mutant riboswitches suggest that the conformers differ with respect to loop-loop interactions, rather than P1 stem structure. Our results also reveal differences in conformer distribution between ARNA and G_ARNA that may play important functional roles in the RNA folding/unfolding that accompanies specific ligand binding and cotranscriptional gene regulation.

2.5.1 Partially Quenched Species as Ap-Bound RNA Folding Intermediates

Since the fluorescence of Ap is exquisitely sensitive to its local environment,[43] several explanations for the multiexponential decay of the RNA-bound Ap need to be considered. For instance, intermolecular RNA interactions

may be promoted at higher RNA concentrations, and the partially quenched species may be those bound to multimers of the form, Ap:$(RNA)_n$:$(Mg^{2+})_x$, rather than unfolded intermediates of the 1:1 complex. This explanation has been eliminated by the close match in the measured hydrated volume of the liganded RNAs with that calculated for the monomeric aptamer domains (unpublished results). Alternatively, it is possible that all RNA molecules adopt a single conformation in solution, that the ensemble is homogeneous, and differences in the fluorescence lifetime reflect minor variations in local interactions at the binding site, or non-specific binding at alternative sites. The former explanation is inconsistent with the substantial differences in binding affinity or RNA-Mg^{2+} interactions (Table 2-1) of the partially and more fully quenched species. The latter explanation is inconsistent with changes in the conformer distribution with mutation in the binding site (wild type GRNA, *vide infra*). Moreover, non-specific binding has not been observed: For instance, no binding of A to the G riboswitch, or to the A riboswitch isolated P2-P3 RNA is detected by NMR.[30]

The coupled equilibria for ligand binding (and Mg^{2+} association) and RNA folding (Scheme 2-1), may also be used to assign the partially quenched species as folding intermediates. If so, than the population of C_o and C_i should decrease, while that of C_c should increase, as the fraction of RNA bound to ligand, f_B, increase. This is true for the total RNA population, since ligand binding to heterogeneous free RNA should direct folding,[44,45] but also within the ligand-bound RNA: Since $K_L(C_o) > K_L(C_i) > K_{L(}C_c)$, the bound population should also

shift towards C_c with increasing f_B. For each Ap:RNA ratio, f_B may be determined from the ensemble K_L values (Table 2-1), and the fraction of each bound conformer (fC_n) from the fluorescence lifetime distribution as described above. This provides the fraction folded within the bound state f_{folded}, (eq. 2-2),

$$f_{folded} = fC_c/(fC_c + fC_i + fC_o) \qquad \textbf{2-2}$$

and, given a simple three-state model, values of $K_2 = fC_c/fC_i$, for folding of C_i to C_c, and $K_{-1} = fC_o/fC_i$, for unfolding of C_i to C_o (Scheme 2-1). While the folding profiles are likely embedded with additional complexities, our experimental data are described by three liganded RNA species. Analogous three state approximations have been applied to, e.g., Mg^{2+}-mediated folding of the *Tetrahymena* ribozyme,[40,46] theophylline-mediated folding of its RNA aptamer,[47] and glycine-mediated riboswitch folding.[48]

For the tight-binding ARNA and G_ARNA, it is possible to measure the f_{folded}, K_{-1} and K_2 over a large binding range (f_B ~ 0.2 - 0.8); these vary as expected for ligand-mediated folding of C_o and C_i to C_c. The f_{folded} decreases as the f_B decreases (Fig. 2-5a). In the presence of 10 mM Mg^{2+}, this dependence is shallow and comparable for the two RNA sequences; Mg^{2+} facilitates folding, and a considerable f_{folded} (> 0.6) is reached at the lowest measurable f_B. We expect the f_{folded} to continue to decrease at lower f_B, however, since we cannot monitor

the unbound RNA, the profile of this dependence is unknown. For instance, it is possible that the f_{folded} approaches zero in the unbound state, or that its dependence on f_B remains essentially linear, and a significant fraction of the unbound RNA may be folded in 10 mM Mg^{2+}.

The Mg^{2+} titrations (Fig. 2-4) also attest the importance of Mg^{2+} in ligand-mediated folding of all three conformers, especially for ARNA where no ligand binding is detected in the absence of the divalent cation. In excess Mg^{2+}, K_{-1} is independent of ligand concentration (Fig. 2-5b). These observations suggest that Mg^{2+} plays a major role in folding of C_o to C_c. However, even in high Mg^{2+} concentration (10 mM), at the lowest f_B, a significant amount of the RNA remains unfolded (Fig. 2-5a), as a consequence of minimal population of C_c (Fig. 2-5b); K_2 approaches zero at the lowest f_B, and increases with ligand binding. Both riboswitches require ligand to advance from the intermediate C_i to the fully folded C_C. This equilibrium is shifted sharply towards the folded form by Ap binding, and K_2 increases to ~ 10 and 25 for ARNA and G_ARNA, respectively, for the same range of Ap:RNA ratios (higher f_B for G_ARNA, due to its tighter ensemble K_D). These findings match well with recent investigations of ligand-induced folding of *add*-ARNA from *V. vulnificus*.[27] While folding could be partially driven by Mg^{2+}, subsequent binding of A facilitates additional conformational changes, detected in the remote L2-L3 region.

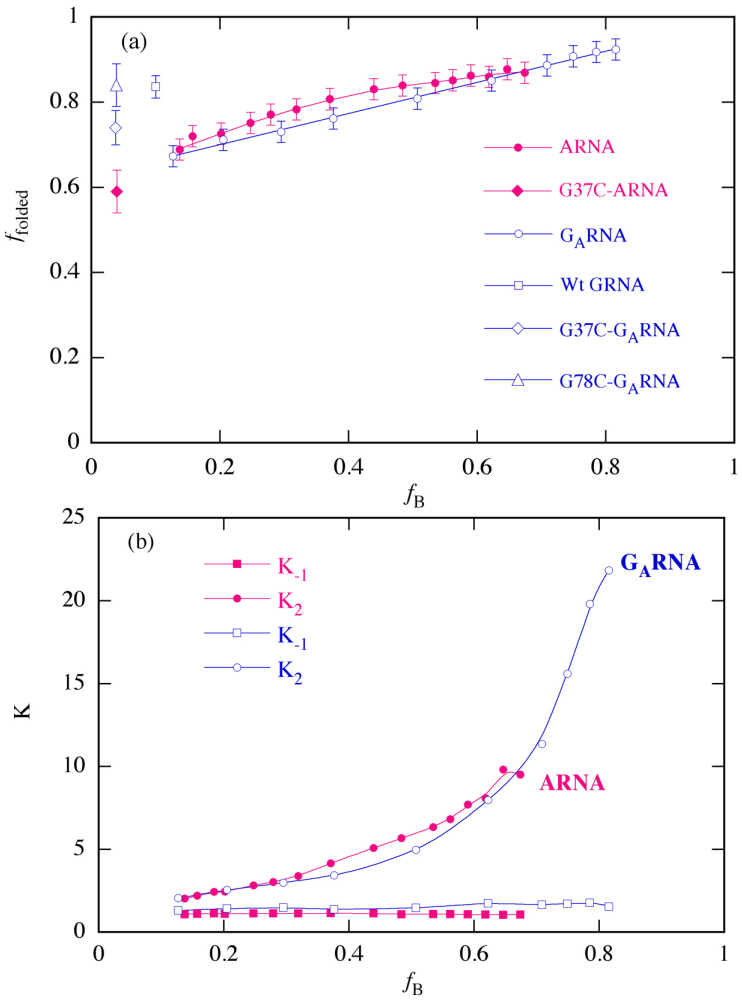


Figure 2-5. Dependence of the fraction of liganded-RNA folded, f_{folded} (a), and ligand-bound conformer equilibrium constants K_{-1}/K_2 (b), on the fraction of RNA bound, f_B (see text for additional details).

For the weak-binding mutant riboswitches, the value of f_B is very small (< ~10 %), even at relatively high concentrations of RNA (> 10 μM). Consequently, the f_{folded} (Materials and Methods) is quantified only at the highest values of f_B, most comparable to wild type (Fig. 2-5a). Significantly, both the binding pocket and the loop mutations shift the conformer distribution relative to wild type. If the partially quenched species were a consequence of non-specific intercalation at

alternative sites, the single point mutation in the binding pocket should have no impact on their distribution. If the partially quenched species reflected minor stacking perturbations of Ap bound within the fully folded RNA, mutations distal from the binding site should not affect their distribution. Thus, these shifts in mutant conformer distribution corroborate our assignment of C_i and C_o as partially folded intermediates of the site-specifically bound riboswitches (Fig. 2-6).

Reversible conformational exchange within the ligand-bound ensemble should be related to the kinetics of conformer exchange in the bound RNA, relative to the kinetics of ligand dissociation. While it is difficult to predict the outcome of this competition for these Ap-bound riboswitches from existing studies, the limited kinetic data suggests that dissociation may be slow relative to conformer exchange: For ARNA, k_{off} ~ 0.01 s^{-1} has been reported for Ap dissociation at 20°C and 2 mM Mg^{2+}.[12] For Ap-bound *add*-ARNA from *Vibrio vulnificus*, $k_{fold}/_{unfold}$ ~ 1 s^{-1} have been measured at 22°C and 0.5 mM Mg^{2+}.[26] Global conformational motion has been observed in liganded complexes of the transactivating response element (TAR) RNA of HIV-1.[49] Binding by TAR to its target, the transactivator protein (Tat), induces significant conformational rearrangement: Two helical domains move from a bent to a coaxial alignment, not unlike that suggested for folding of the P2-P3 stems in the purine riboswitches. Dynamic folding and unfolding of the TAR helices when unbound is presumably functionally significant and facilitates Tat recognition. This motion is arrested by a certain highly charged ligand, possibly due to electrostatic

stabilization of the negatively charged RNA backbone, and deactivation of global motions.[49,50] However, when bound to less charged ligands of lower affinity, even at the helical junction, global motions comparable to those of the free RNA persist.[49] Thus, ligand binding may be insufficient to lock down RNA conformational dynamics, especially for neutral ligands like purines. As in the unbound state, dynamics within riboswitch complexes are presumably of functional significance.

2.5.2 Model for Conformer Structure and Dynamics

The dominant species, C_c, binds ligand with the highest affinity, in a structure that leads to the most fluorescence quenching, and whose folding is very sensitive to Mg^{2+}. These observations strongly suggest that C_c is the most tightly folded, and that it most closely resembles the structure of liganded aptamer domains seen at high resolution.[19,20,23] This structure is indeed "closed" both around ligand at the binding pocket, and with respect to RNA interactions, as all three stems are fully formed, and P2-P3 are aligned tightly parallel through multiple L2-L3 interactions (Fig 2-6).

Partial unfolding in C_o and C_i may be accomplished by disruption of tertiary interactions, e.g. the base pairing between L2-L3, and/or loss of secondary structure, e.g. unzipping of the P1 stem (Fig. 2-6). The data reported here and elsewhere suggests that the P1 stem is formed in C_o and C_i. Single molecule measurements,[34] and structure-based simulations[51] on the A and SAM riboswitch aptamer domains, respectively, suggest that "zipping" up of the P1

stem, mediated by ligand binding, is the ultimate folding step. In unbound RNA, the local P2 and P3 stems form, and L2-L3 interactions exist,[26] providing a platform to initiate ligand binding, while P1 is likely (partially) unfolded.[34] Pairing of P1, stabilized by ligand, is likely maintained in the bound state. Block and coworkers have shown that the liganded ARNA aptamer domain lacking a P1 stem is ~ 4 kcal/mol above the fully folded structure.[34] Due to the ~ 1000-fold loss in affinity, the P1-unfolded state should rarely be bound, and should be undetectable in our assays. Similarly, investigations of "strong" and "weak" P1 stems in ARNA by Crothers and coworkers indicate that ligand dissociation precedes breakdown of major structural elements, likely including the P1 stem,[12] while studies by Lafontaine and coworkers find a stem length of at least 2 base pairs is required for detectable binding to ARNA.[24] Consistent are data presented here on the lack of Ap binding by G13C-ARNA (Fig. 2-3c); in the context of the short, A-U rich P1 stem of ARNA, the single C-C mismatch likely prohibits stem formation, and detectable ligand binding. Binding of Ap to G78C-G_ARNA is likely possible because the C-C mismatch is insufficient to block P1 stem formation in its longer P1 stem which contains 3 G-C pairs. In contrast, ligand does remain bound to aptamers with compromised L2-L3 interactions, e.g. the G37C riboswitches. This mutant has been shown to exist in a structure globally distinct from wild type ARNA, either as a conformer lacking loop-loop interactions[26] or a mix of open and closed conformers in rapid exchange on the electrophoresis time scale.

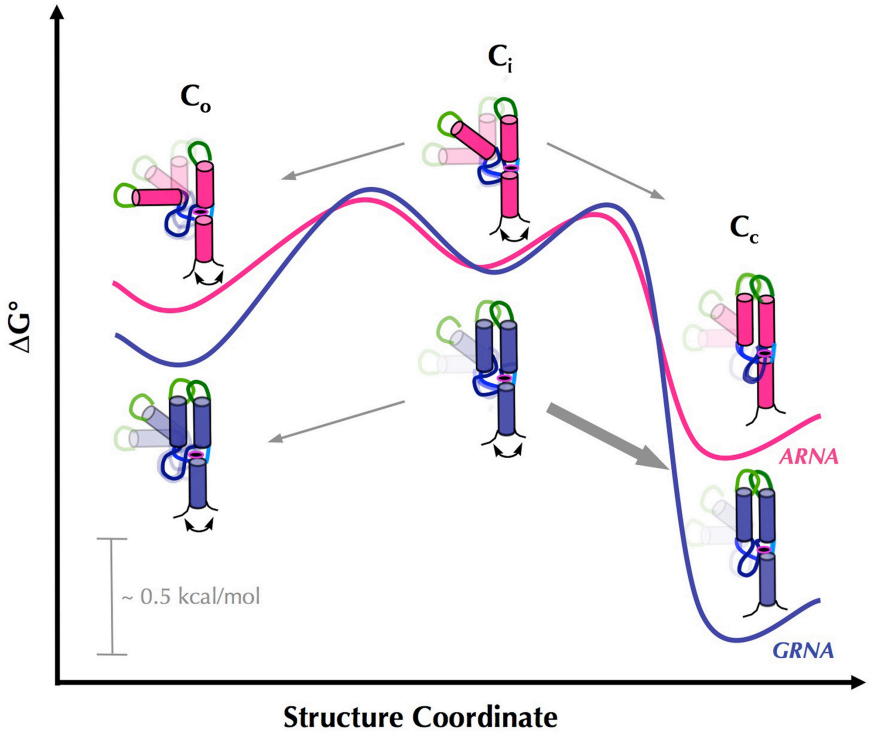


Figure 2-6. Schematized energy landscapes for ligand-bound conformers of the purine riboswitches. In the proposed model, liganded-RNA exchanges between three states (C_O, C_i, and C_C, illustrated here as cartoons) that differ in local and global structure (e.g. P2-P3 interhelical angle and/or P1 stem unzipping (arrows), both decreasing as $C_O > C_i > C_C$). The relative conformer energies measured for ARNA (pink) and G_ARNA (blue) are estimated based on the equilibrium constants, K_{-1} and K_2 measured for the C_i-C_O and C_i-C_C equilibria, respectively, from fC_n values near the maximum f_{folded}. Sequence-dependent partitioning of ligand-bound C_i leads to greater unfolding to C_O in the more dynamic ARNA complex, and drives the more preorganized G_ARNA towards the fully folded C_C in the bound state.

For both A and G_ARNA, we find that the affinity of the G37C loop mutants for Ap are comparable to those of the wild type C_i and C_O conformers (Table 2-1). Partial unfolding via disruption of L2-L3 interactions in C_O and C_i is corroborated by single molecule FRET detection of an “unfolded” ligand-bound conformer of ARNA with an L2-L3 distance ~ 70 Å, as well as an intermediate with L2-L3

distances ~ 50 Å.[26] Similar conclusions may be drawn from time-resolved NMR[30] and stopped-flow fluorescence[27] results. The latter indicate that ligand mediates purine riboswitch folding, both at the binding pocket and at the remote L2-L3. Together, the observations are consistent with a model where, unlike the free RNA, all bound states possess a stabilized P1 stem, and it is instead the extent of L2-L3 interactions that differ. In C_c, tight ligand binding is maintained also through strong L2-L3 interactions,[26] while these interactions are weakened in C_i and C_o. Analogous to the interplay between pseudo-knot and P1 stem formation observed for the SAM riboswitch,[51] a dynamic balance may exist between P1 stem and L2-L3 interactions in the purine riboswitches. In the unbound state, L2-L3 binding may impede P1 stem formation, while in the bound state, P1 stem closure may induce partial unfolding at the loops. This "backtracking"[51] of the folding process may afford a functionally-relevant link between the P1 stem and distal loops.

2.5.3 Sequence-Dependence of Riboswitch Conformer Equilibria

The ensemble K_L values of the G37C loop mutant are the same within experimental error for ARNA or G_ARNA (Table 2-1), and these values are comparable to those for wild type C_i and C_o. Yet, the ensemble K_L values for wild type G_ARNA are smaller than those for ARNA, as a consequence of the 7-fold tighter binding of G_ARNA C_c. The L2-L3 interactions existing in C_c appear to be stronger in G_ARNA than ARNA, and their loss in C_i and C_o, or the G37C mutant,

is accompanied by a larger reduction in binding affinity for G_ARNA (K_{rel}, Table 2-1). The smaller difference in binding affinity between C_c and C_o/C_i in ARNA shifts its conformer distribution towards more unfolding in the liganded state (Fig. 2-6).

This difference in conformer distribution is not evident for the wild type sequences at high Mg^{2+} concentration, where folding is substantial even at low f_B. Differences are revealed by experiments where folding is compromised, e.g. at lower Mg^{2+} concentration, or in mutants. For instance, in the absence of Mg^{2+}, G_ARNA is competent to bind ligand, albeit relatively weakly, while for ARNA, ligand binding is not detected in the absence of Mg^{2+} (Fig. 2-4). Increasing Mg^{2+} concentration facilitates ligand binding and folding of each conformer for both RNAs. However, the maximum f_{folded} is higher, and achieved at a significantly lower Mg^{2+} concentration, for G_ARNA (f_{folded}max ~ 0.95 at ~ 10 μM Mg^{2+}) than for ARNA (f_{folded}max ~ 0.85 at 500 μM Mg^{2+}). The weaker requirement on the divalent cation for ligand-mediated folding may be a consequence of intrinsic preorganization in G_ARNA, as suggested from recent NMR experiments.[21,28] These found that Mg^{2+} is required to stabilize L2-L3 interactions in the more conformationally dynamic free ARNA, while these are more preorganized in free GRNA. The results presented here suggest that this preorganization operates at the level of the C_o-C_i equilibrium: At high concentrations of Mg^{2+}, for both ARNA and G_ARNA, the value of K_{-1} is comparable, and independent of ligand binding,

while the value of K_2 is minimal at low f_B and increases substantially with ligand binding (Fig. 2-5b).

Differences in conformer distribution between ARNA and G_ARNA are also amplified by mutation. For instance, for the same fraction of RNA bound, the liganded state of loop mutant, G37C-G_ARNA is more highly folded than that of G37C-ARNA (Fig. 2-5a). For ARNA we cannot compare the conformer distribution in the loop mutant directly to wild type since the latter is not measurable at such a low f_B where the very weak binding mutant is observable. However, for G_ARNA, the conformer distribution is shifted towards the more folded state in the loop mutant. A similar shift is seen in the P1 stem mutant, and also for Ap bound to wild type GRNA (Fig. 2-5a). These results suggest that all three "mutations", G37C, G78C, and U74C (wild type GRNA), have a greater impact on the Ap binding affinity of C_o and C_i than C_c. For the G78C and U74C mutants, with compromised interactions in the P1 stem and binding pocket, respectively, the bound state may not be as readily maintained by C_o and C_i that lack significant, compensatory loop-loop interactions found in C_c. Some loop-loop interactions must exist, at least transiently, in G_ARNA C_o and C_i; if not, their binding affinities should have been unaltered by the G37C loop mutation, and the distribution shifted towards the more unfolded state.

While the overall driving force for ligand binding and folding is ~ 6-9 kcal/mol,[22,34] folding and unfolding of the ligand-bound riboswitches involves relatively small free energy changes. Values of ~ 1-2 kcal/mol are calculated

from K_2 and K_{-1} (Scheme 2-1) and corroborated by k_{open}/k_{close}[26] and force spectroscopy for single ARNA molecules.[34,42] Consequently, interconversion of ligand-bound C_o, C_i and C_c may exhibit heightened sensitivity[52,53] to subtle structural differences between these purine riboswitches. For instance, the extended loops in ARNA (9 bases) may render L2-L3 interactions more transient, thereby shifting the conformer equilibria towards C_O, while the shorter loops in G_ARNA (7 bases) may be the source of less fluctuating L2-L3 interactions and the higher population of C_C. Indeed, replacement of the L2 proximal A-U base pair in ARNA P2 by a more rigidifying G-C abrogates the need for Mg^{2+} to stabilize L2-L3 interactions.[29] The sequence of both L3 and the P1 stem of the adenine-sensing riboswitch from *V. vulnificus* (*add*-ARNA) more closely resembles the *xpt-pbuX B. subtilis* GRNA than the *pbuE* ARNA investigated here. One might predict that folding/unfolding of *add*-ARNA[27] may be more similar to *xpt-pbuX* GRNA. Unlike *pbuE* ARNA (Fig. 2-1), exchange between expression platform conformations of *xpt-pbuX* GRNA and *add*-ARNA may occur while leaving their aptamer domain folds largely intact. In addition, while we have used the same length overhangs in the A and G aptamer domains studied here, folding of the riboswitch aptamer domains is also expected to be influenced by the nucleotides extending from the 5'- and 3'-termini of the P1 stem.[54]

For some, riboswitches, including the *xpt-pbuX* GRNA and the TPP riboswitch,[55] there appears to exist an intrinsic level of preorganization set at the sequence level that resists further folding of unbound RNA by Mg^{2+} alone, but

also limits unfolding in the bound state. Transition of these riboswitches to the fully folded structure, is "switched" on, and largely maintained by specific ligand binding. For others, such as the catalytic *glmS* riboswitch[56,57] Mg^{2+} drives complete folding of the free RNA, and binding of ligand, glucosamine 6-phosphate, induces no significant RNA structural changes. In this case, changes in RNA conformation alone do not affect gene regulation; because *glmS* acts also as a ribozyme, ligand-dependent cleavage of the downstream mRNA is required. A more delicate balance is seen in *pbuE* ARNA, as well as the glycine riboswitch[48] where folding appears to be mediated by coupled binding of both Mg^{2+} and the specific ligand. As for *glmS*, the conformationally-dynamic free RNA requires Mg^{2+} to bind ligand, and it also maintains more dynamics in the bound state. However, unlike *glmS*, the regulatory responses of these riboswitches depends on the impact of binding on RNA conformational dynamics alone.

2.5.4 Functional Significance of Ligand-Bound Folding Intermediates

Without ligand, L2-L3 tertiary interactions undoubtedly form,[26] and, at least for ARNA, are likely stronger than secondary interactions within the P1 stem.[34] The former may serve to organize the pocket for ligand binding, as *in vivo* the P1 stem is the last to be generated. Strong ligand-induced stabilization of the P1 stem, especially for ARNA, precludes its unzipping once bound, while dynamics within the remote loops persist. These bound state dynamics may be biologically

significant at several levels. For instance, ligand recognition by the purine riboswitches may involve thermodynamic discrimination through hydrogen bonding at the Watson-Crick face with N74, and kinetic discrimination through dynamic sampling. The latter can only be effective if C_c formation is rate limiting ($R = k_2[C_i]$); this may be controlled by limiting build up of liganded C_i through its funnelling also along an unfolding pathway. For the G-sensing riboswitch, thermodynamic discrimination is intrinsically higher, and once bound, GRNA may be less dynamic. The A-sensing riboswitch may exploit its dynamics to achieve a higher level of kinetic discrimination. In addition, dynamics within the bound state provide the opportunity to escape misfolding events, or to fine-tune the regulatory mechanism by requiring other cellular signals to adjust the folding/unfolding equilibrium. These functional roles are rooted in the existence of intermediates along the liganded-folding pathway that afford a layer of control over riboswitches that exploits sequence-dependent RNA dynamics.

2.6 Materials and Methods

2.6.1 RNA Samples

DNA templates of the purine riboswitch aptamer domains were PCR-amplified from the 5'-UTR of the *pbuE* (ARNA) and *xpt-pbuX* (GRNA) genes in *Bacillus subtilis*, cloned into pTZ57R/T vectors, and used to generate coding DNA for all mutants through site-directed mutagenesis (Strategene).

Oligonucleotides for PCR were obtained commercially (Sigma); the 17 nucleotide promoter sequence of T7 RNA polymerase (RNAP) was incorporated at the 5'-end of the forward primers. All DNA templates were sequenced before being transcribed by RNAP. Transcription reactions contained: 40 mM Tris-HCl pH 7.9, 2.5 mM spermidine, 26 mM $MgCl_2$, 0.01% Triton X-100, 10 mM DTT, 8 mM GTP, 4 mM ATP, 4 mM CTP, 2 mM UTP, and 4 U/μL T7 RNAP (prepared in house). Reactions were incubated at 37°C for 2 h and the crude RNA product was purified by 6% denaturing polyacrylamide gel electrophoresis and eluted using 300 mM sodium acetate. Following ethanol precipitation, the purified RNA was quantified using UV-vis spectroscopy, and was stored at -20°C. RNA samples were heat denatured, slowly cooled, and incubated in Fluorobuffer (20 mM Tris-HCl, 5 mM NaCl, pH 7; variable Mg^{2+} concentrations) at ambient temperature prior to experiments.

2.6.2 Time-Resolved Fluorescence Experiments

Fluorescence measurements were made using a Horiba-JobinYvon Fluorolog spectrometer equipped with an IBH detector and electronics for time correlated single photon counting (TCSPC). The excitation source was a 295 nm pulsed nanoLED operating at a 1 MHz repetition rate. Emission was passed through a double monochromator with a 15 nm bandpass centered at 370 nm, or through a 320 nm cut-off filter yielding broad band emission (~340-600 nm). Comparable results (*i.e.* fluorescence lifetimes and amplitudes) were obtained with either configuration, and were conducted under magic angle conditions (54.

7^{o}) to avoid contributions from orientational motion of the ligand. The instrument response function (IRF), measured by scattering the excitation beam off a dilute solution of Ludox was 735 ps FWHM: For our IRF, $t \geq 300$ ps are reliably and reproducibly measured. Temperature was maintained at 20°C with a Peltier-controlled thermostated sample holder.

Titrations involved stepwise addition of either RNA (5 nM – 2-8 μM) to a solution of Ap (500 nM, Sigma) in Fluorobuffer containing 10 mM Mg^{2+}, or addition of Mg^{2+} to a solution of RNA (~ 4 μM) and Ap (500 nM) in Fluorobuffer. Fluorescence decay curves were collected at each step (reverse mode, 100 ns time scale, resolved into 2048 channels), for a constant run time, typically 300 s (≥ 10000 counts in peak channel), in order to evaluate the "static" quenching and $fC_c = 1-(I'/I'_o)$, where I'_o and I' are the counts in the peak channel (intensity at the earliest recorded time) of Ap in the absence and presence of RNA and/or Mg^{2+}, respectively. Decay curves were analyzed using standard iterative reconvolution (DAS software, Horiba-JobinYvon). Fit quality was assessed by reduced chi-squared ($\chi^2 \leq 1.3$ acceptable) and randomness of residuals. The total fluorescence quenching, corresponding to ensemble-averaged binding, is evaluated as $\Sigma fC_n = 1-(I/I_o)$, where I_o and I are the total counts (integrated area under decay curves) of Ap in the absence and presence of RNA and/or Mg^{2+}, respectively.

As the full lifetime decay of Ap bound to the purine riboswitches is not completely resolved in our experiments, fC_c is evaluated from the extent of static

quenching, while fC_i and fC_o are evaluated from their amplitude in the fluorescence lifetime decay, α_n in eq. 2-3 ($\Sigma a_n = 1$).

$$I(t) = \alpha_u e(-t/\tau_u) + \alpha_o e(-t/\tau_o) + \alpha_i e(-t/\tau_i) \quad \textbf{2-3}$$

These relative amplitudes, a_o and a_i, were adjusted to account for fC_c according to equation 2-4:

$$fC_{o,i} = a_{o,i} * (1 - fC_c) \quad \textbf{2-4}$$

The relative amplitude of unbound Ap, α_u, is similarly adjusted to yield the unbound fraction, f_u, and $\Sigma fC_n + f_u = 1$. This analysis accurately (errors ≤ 5%) measures fC_c, but less accurately quantifies fC_i and fC_o, especially in the limit of low ligand binding (*i.e.* low [RNA] or [Mg^{2+}]), where the lifetime decay is dominated by unbound Ap. Another measure of the total population of these conformers comes from eq. 2-5, where fTQ and fSQ are the fractional total and static quenching, respectively.

$$\Sigma(fC_i + fC_o) = fTQ - fSQ \quad \textbf{2-5}$$

At values of $\Sigma(fC_i + fC_o) < 5$ % we do not reliably detect the presence of bound species in our lifetime decay. This precludes independent evaluation of fC_i and

fC_o (and thus their K_L values, and the equilibrium constants, K_{-1} and K_2) for the mutant riboswitches.

The variation in fC_n, or total fluorescence quenching, ΣfC_n, as a function of RNA concentration ([Mg^{2+}] = 10 mM), or Mg^{2+} concentration ([RNA] = 4 mM), was fit to equations 2-6 or 2-7, respectively:

$$fC_n = \frac{(fC_n)_{\max}}{2[Ap]}\{[Ap]+[RNA]+K_L-(([Ap]+[RNA]+K_L)^2-4[Ap][RNA])^{1/2}\} \quad \textbf{2-6}$$

$$fC_n = (fC_n)\max * \frac{[Mg^{2+}]^n}{K_{Mg}+[Mg^{2+}]^n} \quad \textbf{2-7}$$

where n is the Hill coefficient. The former yields values for K_L, the dissociation constant for binding of Ap to each conformer, or the ensemble average. The latter yields values for K_{Mg}, a measure of the dependence of ligand-mediated folding of each conformer, or the ensemble average, on Mg^{2+}.

2.7 Acknowledgements

We thank A. J. Bennet, B. Mario Pinto, and Dipankar Sen for helpful discussions. Our research is generously supported by the Natural Sciences and Engineering Research Council of Canada, and the Michael Smith Foundation for Health Research.

2.8 References

1. Serganov, A.; Patel, D. J. *Nature Rev* **2007**, 8, 776-790.

2. Winkler, W. C.; Breaker, R. R. *Annu Rev Microbio* **2005**, 59, 487-517.

3. Vitreschak, A. G.; Rodionow, D. A.; Mironov, A. A.; Gelfand, M. S. *Rev. Trends in Genetics* **2004**, 20, 44-50.

4. Batey, R. T. *Curr Opin Struct Biol* **2006**, 16, 299-306.

5. Schwalbe, H.; Buck, J.; Fürtig, B.; Noeske, J.; and Wöhnert, J. *Angew Chem Int Ed* **2007**, 46, 1212-1219.

6. Sudarsan, N.; Barrick, J. E.; Breaker, R. R. *RNA* **2003**, 9, 644–647.

7. Cheah, M. T.; Wachter, A.; Sudarsan, N.; Breaker, R. R. *Nature* **2007**, 447, 497-501.

8. Thore, S.; Leibundgut, M.; Ban, N. N. *Science* **2006**, 312, 1208-1211.

9. Croft, M. T.; Moulin, M.; Webb, M. E.; Smith, A. G. *Proc Natl Acad Sci USA* **2007**, 104, 20770-20775.

10. Mironov, A. S.; Gusarov, I.; Rafikov, R.; Lopez, L. E.; Shatalin, K.; Kreneva, R. A.; Perumov, D. A.; Nudler, E. *Cell* **2002**, 111, 747-756.

11. Wickiser, J. K.; Winkler, W. C.; Breaker, R. R.; Crothers, D. M. *Mol Cell* **2005**, 18, 49-60.

12. Wickiser, J. K.; Cheah, M. T.; Breaker, R. R.; Crothers, D. M. *Biochemistry* **2005**, 44, 13404–13414.

13. Sudarsan, N.; Chalamish-Cohen, S.; Nakamura, S.; Emilsson, G. M.; Breaker, R. R.; *Chem Biol* **2005**, 12, 1325-1335.

14. Gilbert, S. D.; Mediatore S. J.; Batey, R. T. *J Am Chem Soc* **2006**, 128, 14214-14215.

15. Mandal, M.; Boese, B.; Barrick, J. E.; Winkler, W. C.; Breaker, R. R. *Cell* **2003**, 113, 577–586.

16. Mandal, M.; Breaker, R. R. *Nat Struct Mol Biol* **2004**, 11, 29–35.

17. Johansen, L. E.; Nygaard, P.; Lassen, C.; Agersø, Y.; Saxlid, H. H. *J Bacteriol* **2003**, 185, 5200-5209.

18. Nygaard, P.; Saxlid, H. H. *J Bacteriol* **2005**, 187, 791-794.

19. Batey, R. T.; Gilbert, S. D.; Montange, R. K. *Nature* **2004**, 432, 411–415.

20. Serganov, A.; Yuan, Y. R.; Pikovskaya, O.; Polonskaia, A.; Malinina, L.; Phan, A.T.; Hobartner, C.; Micura, R.; Breaker, R. R.; Patel, D. *Chem Biol* **2004**, 11, 1729–1741.

21. Noeske, J.; Richter, C.; Grundl, M. A.; Nasiri, H. R.; Schwalbe, H.; Wöhnert, J. *Proc Natl Acad Sci USA* **2005**, 102, 1372–1377.

22. Gilbert, S. D.; Stoddard, C. D.; Wise, S. J.; Batey, R. T. *J Mol Biol* **2006**, 359, 754-768.

23. Gilbert, S. D.; Love, C. E.; Edwards, A. L.; Batey, R. T. *Biochemistry* **2007**, 46, 13297-13309.

24. Lemay, J. F.; Lafontaine, D. A. *RNA* **2007**, 13, 339-350.

25. Mulhbacher, J.; Lafontaine, D. A. *Nucleic Acids Res* **2007**, 35, 5568-5580.

26. Lemay, J. F.; Penedo, J. C.; Tremblay, R.; Lilley, D. M.; Lafontaine, D. A. *Chem Biol* **2006**, 13, 857-868.

27. Rieder, R.; Lang, K.; Graber, D.; Micura, R. *Chem Bio Chem* **2007**, 8, 896–902.

28. Noeske, J.; Buck, J.; Furtig, B.; Nasiri, H. R.; Schwalbe, H.; Wöhnert, J. *Nucleic Acids Res* **2007**, 35, 572-583.

29. Noeske, J.; Schwalbe, H.; Wöhnert, J. *Nucleic Acids Res* **2007**, 35, 5262-5273.

30. Buck, J.; Furtig, B.; Noeske, J.; Wöhnert, J.; Schwalbe, H. *Proc Natl Acad Sci USA* **2007**, 104, 15699-15704.

31. Ottink, O. M.; Rampersad, S. M.; Tessari, M.; Zaman, G. J. R.; Heus, H. A.; Wijmenga, S. S. *RNA* **2007**, 13, 2202-2212.

32. Eskandari, S.; Prychyna, O.; Leung, J.; Avdic, D.; O'Neill, M. A. *J Am Chem Soc* **2007**, 129, 11308-11309.

33. Lakowicz, J. R. Principles of Fluorescence Spectroscopy 2nd Ed.; Springer Science: New York, **2004**; pp. 303-310.

34. Greenleaf, W. J.; Frieda, K. L.; Foster, D. A. N.; Woodside, M. T.; Block, S. M. *Science* **2008**, 319, 630-633.

35. Yamauchi, T.; Miyoshi, D.; Kubodera, T.; Nishimura, A.; Nakai, S.; Sugimoto, N. *FEBS Lett* **2005**, 579, 2583–2588.

36. Heilan-Miller, S. L.; Thirumalai, D.; Woodson, S. A. *J Mol Biol* **2001**, 306, 1157-1166

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

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

39. Das, R.; Travers, K. J.; Bai, Y.; Herschlag, D. *J Am Chem Soc* **2005**, 127, 8272-8273.

40. Pan, J.; Thirumalai, D.; Woodson, S. A. *Proc Natl Acad Sci USA* **1999**, 96, 6149-6154.

41. Stoddard, C. D.; Gilbert, S. D.; Batey, R. T. *RNA* **2008**, 14, 675-684.

42. Lin, J. C.; Thirumalai, D. *J Am Chem Soc* **2008**, 130, 14080-14081.

43. Guest, C. R.; Hochstrasser, R. A.; Sower, L. C.; Millar, D. P. *Biochemistry* **1991**, 30, 3271-3279.

44. Williamson, J. R. *Nat Struct Biol* **2000**, 7, 834-837.

45. Leulliot, N.; Varani, G. *Biochemistry* **2001**, 40, 7947-7956.

46. Celander, D. W.; Cech, T. R. *Science* **1991**, 51, 401-407.

47. Jucker, F. M.; Phillips, R. M.; McCallum, S. A.; Pardi, A. *Biochemistry* **2003**, 42, 2560-2567.

48. Lipfert, J.; Das, R.; Chu, V. B.; Kudaravalli, M.; Boyd, N.; Herschlag, D.; Doniach, S. *J Mol Biol* **2007**, 365, 1393-1406.

49. Pitt, S. W.; Zhang, Q.; Patel, D. T.; Al-Hashimi, H. M. *Angew Chem Int Ed Engl* **2005**, 44, 3412-3415.

50. Casiano-Negroni, A.; Sun, X.; Al-Hashimi, H. M. *Biochemistry* **2007**, 46, 6525-6535.

51. Whitford, P. C.; Schug, A.; Saunders, J.; Hennelly, S. P.; Onuchic, J. N.; Sanbonmatsu, K. Y. Biophys *J: Biophy. Lett* **2009**, 96, L7-9.

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

53. Brion, P.; Westhof, E. *Annu Rev Biophys Biomol Struct* **1997**, 26, 113-37.

54. Rieder, R.; Höbartner, C.; Micura, R. *Meth Mol Biol* **2009**, 540, 15-24.

55. Lang, K.; Rieder, R.; Micura, R. *Nucleic Acids Res* **2007**, *35*, 5730-5738.

56. Hampel, K. J.; Tinsely, M. M. *Biochemistry* **2006**, 45, 7861-7871.

57. Tinsely, R. A.; Furchak, J. R. W.; Walter, N. G. *RNA* **2007**, 13, 468-477.

2.9 Supporting Information

Sequence-Dependent Folding and Unfolding of Ligand-Bound Purine Riboswitches

Oksana Prychyna, Michael S. Dahabieh, Jay Chao, and Melanie A. O'Neill

$[Mg^{2+}]$ (μM)	Total counts	Peak counts	τ_u	α_u	τ_o	α_o	τ_i	α_i
0	4.8734e+06	45935	11.8	0.9253	6.13	0.0427	1.39	0.032
0.0159	4.6445e+06	43972	11.78	0.9241	5.42	0.0479	1.07	0.028
0.0395	4.5062e+06	42879	11.76	0.9371	4.38	0.0331	1.075	0.0298
0.0782	4.1598e+06	40200	11.86	0.9166	4.00	0.0498	1.098	0.0336
0.1562	3.6016e+06	36284	11.88	0.9043	5.00	0.0553	0.962	0.0404
0.2338	3.1455e+06	32941	11.78	0.9126	3.13	0.0499	0.734	0.0375
0.3111	2.7337e+06	30137	11.77	0.8978	2.92	0.058	0.676	0.0441
0.3881	2.3450e+06	27761	11.76	0.8763	2.72	0.0728	0.566	0.051
0.4649	2.0650e+06	27010	11.71	0.847	2.23	0.0945	0.545	0.059
0.6174	1.6028e+06	27010	11.67	0.7566	2.01	0.1601	0.543	0.0833
0.844	1.2166e+06	26145	11.61	0.6504	2.07	0.2162	0.564	0.1334
1.216	8.8569e+05	24303	11.48	0.499	2.22	0.2933	0.609	0.2076
1.5945	7.7288e+05	23828	11.23	0.44	2.05	0.342	0.535	0.218
2.3493	6.6840e+05	22601	11.08	0.372	2.07	0.378	0.524	0.25
3.85	5.9569e+05	22416	10.9	0.2957	2.24	0.3942	0.564	0.3101

Table 2-2. Parameters from a representative experiment monitoring the fluorescence lifetime of 500 nM Ap as a function of G_ARNA concentration (10 mM Mg^{2+}, 20°C). Fluorescence decay fit to: $I(t) = \alpha_u\tau_u + \alpha_o\tau_o + \alpha_i\tau_i$ where α_u, α_o, and α_i are the amplitudes, and τ_u, τ_o, and τ_i are the lifetimes. The change in peak counts provide the static quenching ($\tau_3 \leq$ 300 ps). Together these are used to evaluate the fractions of the 4 species, C_u, C_o, C_i, and C_c present within the ensemble according to eq. 2-1 and 2-2 (see main text for additional details).

$[Mg^{2+}]$ (μM)	Total counts	Peak counts	τ_u	α_u	τ_o	α_o	τ_i	α_i
0	4.9421e+06	21844	11.6	0.97	5.5	0.015	0.55	0.015
0.001	3.9197e+06	17734	11.67	0.955	3.8	0.033	0.38	0.012
0.00498	3.8991e+06	17624	11.69	0.954	3.9	0.034	0.38	0.0131
0.0109	3.7904e+06	17203	11.69	0.954	3.9	0.034	0.48	0.013
0.018	3.6680e+06	16844	11.7	0.955	3.2	0.0318	0.3	0.013
0.027	3.5258e+06	16116	11.68	0.954	3.5	0.032	0.4	0.0134
0.066	3.2255e+06	15140	11.65	0.953	2.9	0.032	0.324	0.015
0.104	2.8865e+06	13886	11.65	0.9497	2.3	0.033	0.25	0.0175
0.201	2.3081e+06	11774	11.61	0.9376	2.11	0.037	0.24	0.0254
0.58	2.2188e+06	11350	11.68	0.9285	3	0.042	0.31	0.0292
0.96	9.8615e+05	8198	11.62	0.8292	2.2	0.079	0.287	0.092
1.53	6.6121e+05	7179	11.53	0.753	1.8	0.1153	0.281	0.133
2.1	5.3991e+05	6932	11.51	0.6889	1.75	0.1509	0.25	0.1602
3.097	4.2329e+05	6944	11.4	0.5825	1.9	0.1853	0.294	0.2322
5.86	3.8451e+05	6727	11.5	0.5377	2	0.2035	0.31	0.259
7.72	3.6656e+05	6757	11.3	0.5046	1.96	0.221	0.292	0.274
10.51	3.6919e+05	6925	11.25	0.4948	1.93	0.2354	0.276	0.2698

Table 2-3. Parameters from a representative experiment monitoring the fluorescence lifetime of 500 nM Ap as a function of Mg^{2+} concentration (~ 3 μM G_ARNA, 20 ^oC). Fluorescence decay fit to: $I(t) = \alpha_u\tau_u + \alpha_o\tau_o + \alpha_i\tau_i$ where α_u, α_o, and α_i are the amplitudes, and τ_u, τ_o, and τ_i are the lifetimes. The change in peak counts provide the static quenching ($\tau_3 \leq$ 300 ps). Together these are used to evaluate the fractions of the 4 species, C_u, C_o, C_i, and C_c present within the ensemble according to eq. 1 and 2 (see main text for additional details).

Chapter 3: Sequence-Dependent Structural Dynamics of Primate A-to-I Editing Substrates

The thesis author's contribution accounted for 50% of the work. This included construct design, gel purification, and initial UV experiments. The thesis author also performed steady-state, and time-resolved fluorescence data analysis. Dhrubajyoti Samanta's contribution accounted for 40% of the work, which included steady-state, and time-resolved fluorescence experiments, with corresponding raw plots. Dr. Jean-Claude Brodovitch developed the multi-state model, which accounted for 10% of the work.

3.1 Introduction

Diversity exists along the DNA, RNA and protein levels, and is essential for the adaptability of an organism. Genomes constantly undergo mutagenesis and translocation of genetic material,[1,2] while proteomes experience post-translational modifications such as methylations, glycosylations, phosphorylations,[3] and even truncations.[4,5] The transcriptome however, experiences a fascinating array of events that are collectively referred to as RNA editing (discussed in Chapter 1), whereby nucleotides are deleted, inserted, or chemically altered. Such modifications have the capacity to alter the properties

of an RNA molecule from structural and functional perspectives.[6]

One mode of RNA editing that has garnered significant attention is A-to-I editing due to recent evidence indicating the abundance of inosines throughout the human transcriptome.[7] Furthermore, aberrant A-to-I editing events are linked to psychiatric disorders,[8] neurodegenerative disease,[9] and cancers.[10] A-to-I edits across all organisms are most prominent in ncRNAs,[11] exemplifying a role for the once envisioned 'junk RNA.'

In a survey of ncRNAs across various organisms and kingdoms of life, the ratio of ncRNA/coding RNA was found to be highest in primates, and especially in humans.[12] Since ncRNAs are the chief edit sites, this enhanced ratio may rationalize our increased A-to-I editing levels relative to other organisms, namely 'lower' primates. Due to the fact that most A-to-I edits occur in brain tissues,[13-16] it is tempting to suggest that enhanced levels may be a cause, or effect, of our relatively higher cognitive abilities. It is possible that increased amounts of ncRNAs, alterations to ncRNAs, or the editing enzyme itself contribute to enhanced A-to-I editing in humans.

3.1.1 Mechanism and Sites of A-to- I Editing

In order for an A–to-I edit to occur, adenosine deaminase acting on double-stranded RNA (ADAR) must clamp on to a stable RNA duplex, whereby it 'flips out' and deaminates the exocyclic amine of adenine.[6,17-19] The ncRNA element that almost singlehandedly facilitates A-to-I editing is the *Alu* element,

which is exclusive to primates, and found at over one million sites, composing 10% of the human genome. *Alu* elements are approximately 300 nt long, highly conserved, both within individuals and across order, and are often oriented in opposite strands in dsDNA. This configuration permits the formation of stable, intramolecular RNA duplexes upon transcription (Figure 3-1). *Alus* capable of forming dsRNAs are found in both coding and ncRNAs, however the majority reside in intronic regions, intergenic sequences, and UTRs.[20]

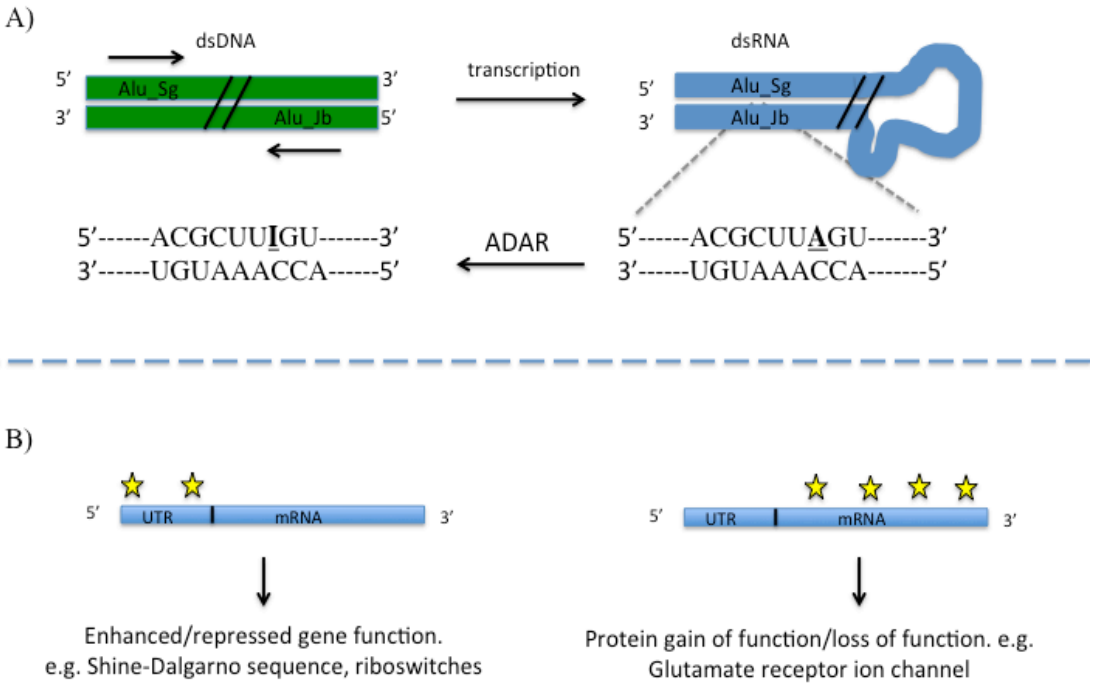


Figure 3-1: Schematic of intramolecular *Alu* duplex formation, followed by A-to-I edit. (A). *Alu* elements that are oriented on opposite DNA strands, both in the 5' -3' direction, may form an intramolecular duplex upon transcription. Depending on the stability of the duplex, ADAR will 'dock' on, facilitating an A-to-I edit. Note: Alu_Sg and Alu_Jb are two *Alu* elements that share approximately 80% sequence identity. Angled parallel lines indicate a potentially large (up to 4000 bp) separation distance between adjacent elements. (B) Effects of editing in coding and ncRNA with stars representing edit sites.

The most documented example of A to I editing in coding RNA is the Q356R conversion in the glutamate receptor β-subunit Ca^{2+} ion channel (*GluB*). Under normal editing, a glutamine (CAG) is edited to CIG, which is read as CGG (arginine) by the translation machinery. If this edit fails to occur, the result is an altering of the permeability and conductance of the ion channel, causing lethality in mice.[21] Another example of editing in protein-coding transcripts involves the serotonin 2C (5-HT_{2C}) receptor, which is a G-protein coupled receptor found in the central and peripheral nervous systems. Here, twenty-four isoforms may be generated as a result of variable A-to-I editing, with many isoforms associated with suicide and depression.[22]

Regarding editing effects in ncRNAs such as UTRs, and introns, A-to-I conversions may tune the thermostability and kinetics of UTR folding, altering regulation of an associated gene. Interestingly, recent bioinformatics analyses of RNA interference (RNAi) target sites have indicated that UTRs are typical binding sites of micro (mi)RNAs,[23] which repress translation when bound. Examples of UTRs that are heavily edited in humans correspond to genes regulating cytochrome P450, GTPases, and junctional adhesion molecules.[24] Since precursors to miRNAs (pri-miRNA) along the RNAi pathway contain duplex regions, they are viable ADAR substrates as well. Thus editing may alter their ability to bind a given UTR, affecting gene regulation. Such a relationship implies that a crosstalk exists between A-to-I editing and RNA interference. The abundance of A-to-I edits in ncRNAs has lead to a race to utilize novel

sequencing approaches to identify 'new' sites,[7,25,26] but this has not been matched by studies interrogating the causes and effects of editing.

3.1.2 Why Study A to I Editing in Primates?

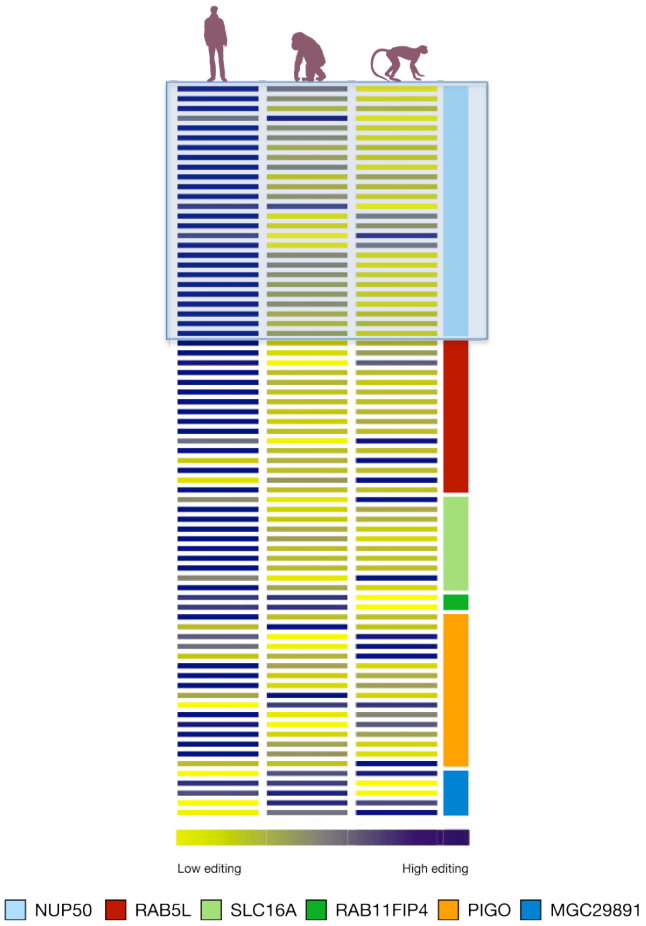


Figure 3-2: Qualitative comparison of A-to-I editing levels in selected primate substrates. The *Alu* element we have directed our studies towards corresponds to *Nup50*, boxed and shown in light blue. Other genes of interest are also shown (below). Reproduced with permission from *Proceedings of the National Academy of Sciences of the United States of America* **2010**, *107*, 12174. Copyright © 2010 National Academy of Sciences, USA. All rights reserved.

A recent investigation by the Rechavi group involved a comparative analysis of A-to-I editing levels in *Alu* elements found in ncRNAs of six evolutionarily conserved genes across human, chimp, and rhesus species. In general, humans displayed a significantly higher level of editing across these sequences, with the largest discrepancies found in the 5'UTR of *Nup50* (Figure 3-2),[24] a 50 kDa nucleoporin gene, which regulates the flow of molecules and ions across the nucleus.[27] *Nup50 Alu* A-to-I conversion in human brain tissues was approximately five times higher compared to chimp/rhesus.[24]

As mentioned above, increased editing levels could be attributed to factors at the substrate (RNA) and/or enzymatic level. To explore the first possibility, Rechavi and colleagues examined the number and length of *Alus*, and alterations to *Alu* sequences between primates, but found no outstanding discrepancies. Here, the numbers and lengths of *Alus* would presumably increase the probability of an editing event, whereas sequence alterations such as mismatches, deletions, and insertions could affect global duplex stability, and hence editing efficiency.

The amount of ADAR present in brain tissues of primates was also explored, with the authors stating that little variations in expression levels were observed. Although exploring A-to-I editing from an enzymatic viewpoint is essential to understanding such a process, almost thirty isoforms exist in humans. Furthermore, post-translational ADAR modifications may exist in humans, whereas most expression systems take advantage of yeast or bacteria,

which may or may not perform the same modifications. Examining A-to-I editing from the substrate level will undoubtedly provide highly valuable information, as a mutual relationship between enzyme and substrate must be satisfied in order for an edit to occur.

We use 2Ap as a probe of conformational dynamics to interrogate the folding pathways of *Nup50 Alu* constructs, representative of the human, chimp, and rhesus sequences *in vivo*. We believe that the observed subtleties in sequences neighbouring the edit sites alter the kinetics and thermodynamics of the RNA folding landscape. These sequenced-dependent structural dynamics have direct effects on the editing efficiency.

3.1.3 Construct Design

We have developed three 16 bp constructs, corresponding to human, chimp, and rhesus *Nup50 AluSg/Jg* dsRNA elements. We focus our efforts on three primate hairpin constructs with 2Ap at position 5 which we refer to as HuAp5, ChAp5, and RhAp5. These constructs were developed because site A5 displayed significant interspecies variability in terms of editing efficiency (Figure 3-3B), yet subtle variations in surrounding sequences. Capping the 16 bp region of each construct is a GNRA tetraloop (GAAA) (Figure 3-3D); we have selected this particular sequence because GNRA tetraloops are abundant in Nature,[28] and will contribute to our general understanding of hairpin folding. Consequently,

these constructs allow us to investigate two intriguing events in the RNA world: A-to-I editing, and hairpin folding.

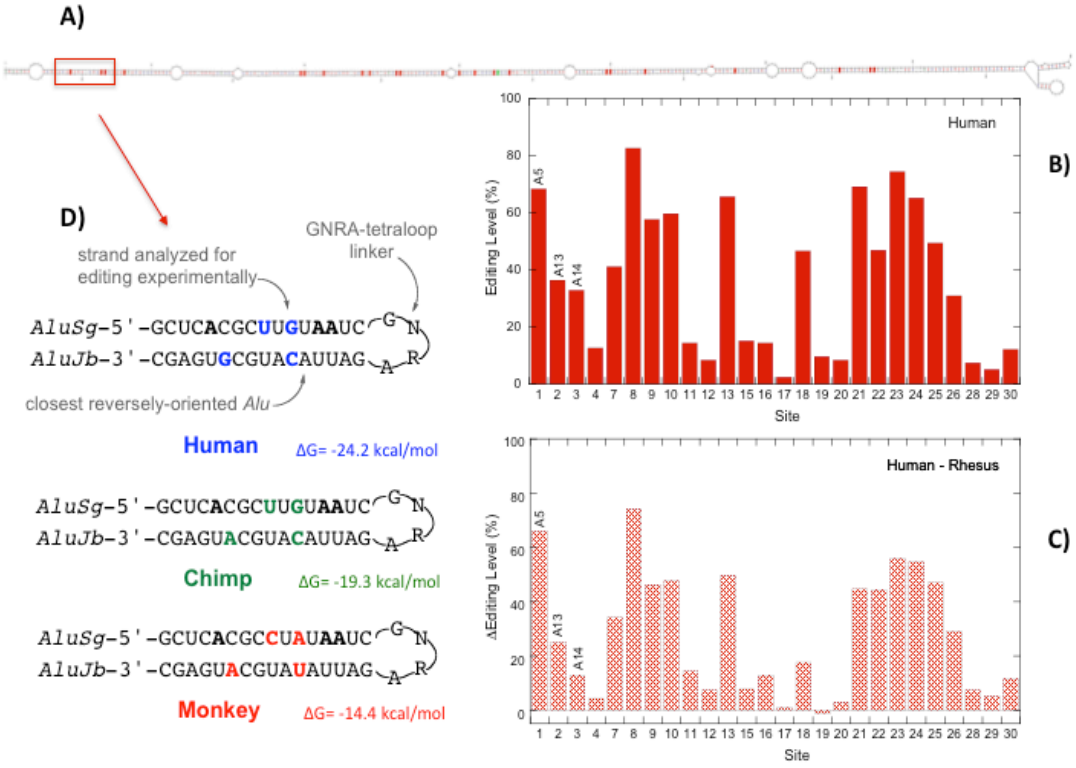


Figure 3-3: Development of *Nup50*-derived primate constructs. A) Mfold predicted secondary structure of approximately 250 bp Human *Nup50 AluSg/Jb* intramolecular hairpin with edit sites highlighted in red. Predicted structures for chimp and rhesus were highly similar. B) Editing level expressed in terms of percent of A-to-I edits, relative to sum of corresponding RNA analyzed from human *AluSg Nup50* RNA. C) Percent of human editing level subtracted by percent rhesus editing level. D) Constructs developed from Mfold predicted structure found in A), consisting of 16 bp stem corresponding to the sequence of each primate species, and GNRA tetraloop to investigate general hairpin folding. Bold sites indicate A5, A13, and A14, while coloured bold sites indicate subtle sequence variations. See Paz-Yaacov, et al. *Proc Natl Acad Sci U S A* **2010**, *107*, 12174, for further details.

3.2 Results

3.2.1 UV Experiments

In order to probe human, chimp, and rhesus hairpin constructs, we have utilized two techniques: UV and fluorescence spectroscopy. Initial UV experiments were performed in order to monitor the global folding/unfolding of each construct by changes in absorbance. Here, we analyzed the change in absorbance at 260 nm (average wavelength of maximum absorbance for nucleic acids) under folding (85°C-10°C) and unfolding (10°C-85°C) conditions. From plots of absorbance (Y) vs. temperature (X), we determined the melting temperatures (T_M) of HuAp5, ChAp5 and RhAp5 to be 69, 56, and 50 °C, respectively (Figure 3-4A); this followed our expected trend, based on G-C content and nearest-neighbour T_m calculations (Mfold).

Furthermore, using the Hill equation (eq 3-1), we determined the cooperativity of the folding/unfolding process, with the highest number indicating the most cooperative transition. With Hill coefficients for human, chimp, and rhesus of 25, 14, and 11, respectively (Figure 3-4A), it appears as though the transition from a folded to unfolded state (or vice versa) is the most cooperative in the human construct, while chimp and rhesus have much broader transitions. This is likely rooted in the sequence of the constructs, as HuAp5 is the most G-C rich of the three.

3.2.2 Steady State Fluorescence Experiments

Although UV experiments provided us with valuable information concerning the global environment of our constructs, we were also interested in probing individual sites to gain nucleotide-specific information. By assessing the change in intensity of 2Ap in our three constructs under the same temperature variations conducted by UV, we were able to probe a site-specific T_M. Furthermore, the use of 2Ap serves multiple benefits: it is non-perturbing, can be substituted at any position, fluoresces under a variety of environments, and supporting examples are abundant in the literature.[29-31]

In a duplex/folded state, surrounding nucleotides quench the emitted light, decreasing the intensity of 2Ap emission. Conversely, at increasing temperatures where a helix-to-coil transition takes place, the internal 2Ap is able to emit light with generally less quenching from surrounding nucleotides. In our site-specific T_M curves (Figure 3-4B), we plotted fraction unfolded (f_u) versus temperature (eq 3-2). Using this information, we calculated the Hill coefficient, which followed the same trends as those calculated by UV (Figure 3-4A).

3.2.3 Steady-State Fluorescence-Based Two-state Model

In addition to site-specific T_M values obtained from steady-state fluorescence, we sought to obtain thermodynamic parameters for each primate construct. Assuming a two-state model, consisting of an equilibrium between

folded and unfolded states, and by using the aforementioned f_u value at a given temperature, we calculated the equilibrium constant for unfolding, K_u, (eq. 3-3).

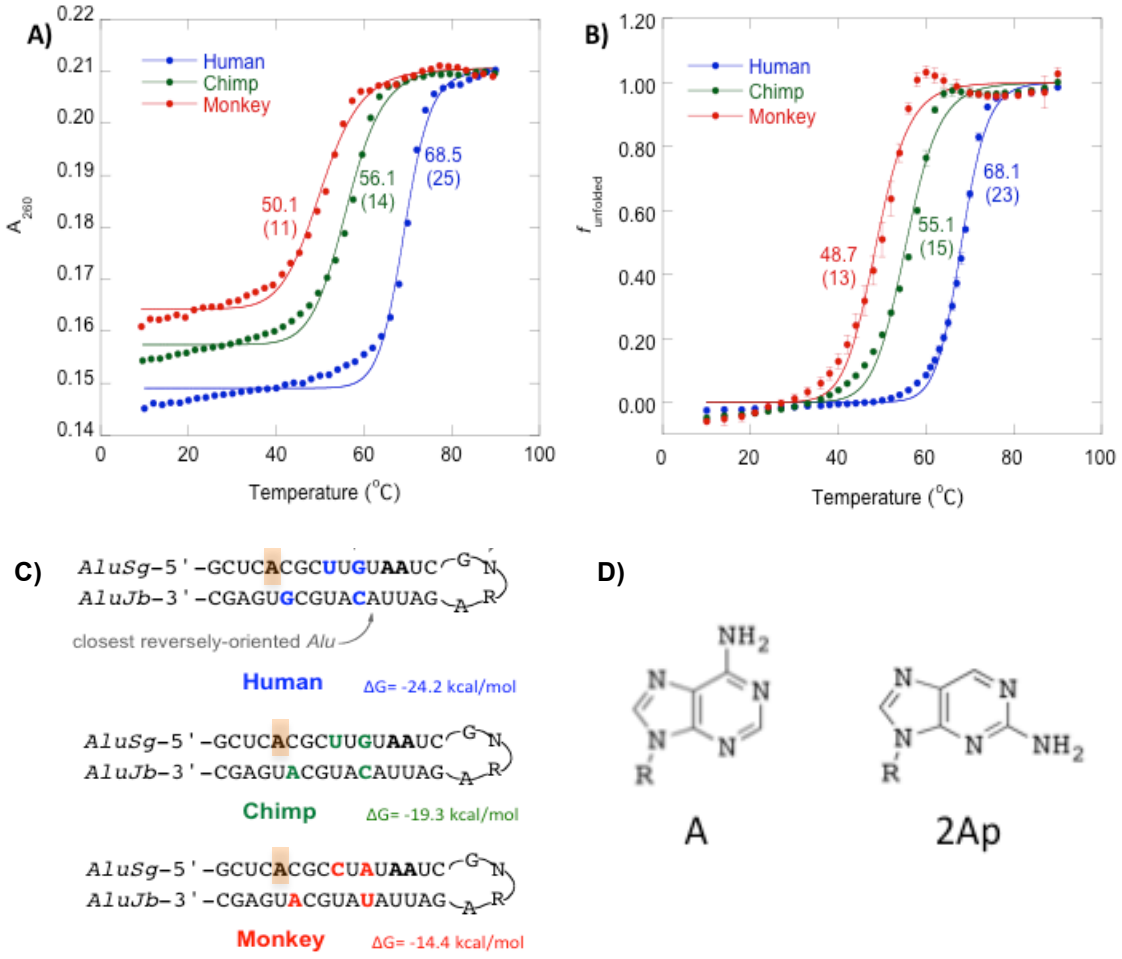


Figure 3-4: Comparison of UV and Ap5 Steady-state fluorescence T_m curves. A) UV and B) steady-state T_m profiles of human, chimp, and rhesus constructs. C) Constructs with Ap5 site highlighted, and D) structures of adenine/A and 2-Aminopurine/2Ap.

Since experiments were conducted over temperatures ranging from 10-85°C, we plotted $\ln(K_{unfold})$ vs. $1/T(K^{-1})$, and used the Van't Hoff equation (eq. 3-4) to extrapolate enthalpy and entropy components. Interestingly, our constructs

did not display a linear relationship across the entire temperature range (Figure 3-5), causing us to discard our straightforward two-state model. Upon observation, certain portions of our Van't Hoff plot display a linear relationship, which lead us to utilize high-resolution TRF data to propose a multi-state model.

3.2.4 Time-Resolved Fluorescence Experiments

In order to determine whether additional conformers are present in solution, we utilized time-resolved fluorescence (TRF), which provides us with a more sensitive detection of fluorescence intensity, rather than the ensemble average intensity detected by steady-state measurements. The multiple fluorescent lifetimes determined from TRF data allowed us to predict the fraction of conformers (folded, intermediate(s) unfolded) in solution, as the fluorescence lifetime of 2Ap is reflective of the surrounding local environment.

This is well demonstrated by our experiments comparing the fluorescent lifetime(s) of free 2Ap, to our chimp hairpin construct, ChAp5. The fluorescence decay profile of the free nucleobase fits well to a single-exponential model (equation 3-5), corresponding to a lifetime of 9.8 ns, which closely matches literature values. This relatively long lifetime is detected due to minimal fluorescence quenching in solution (Figure 3-6).

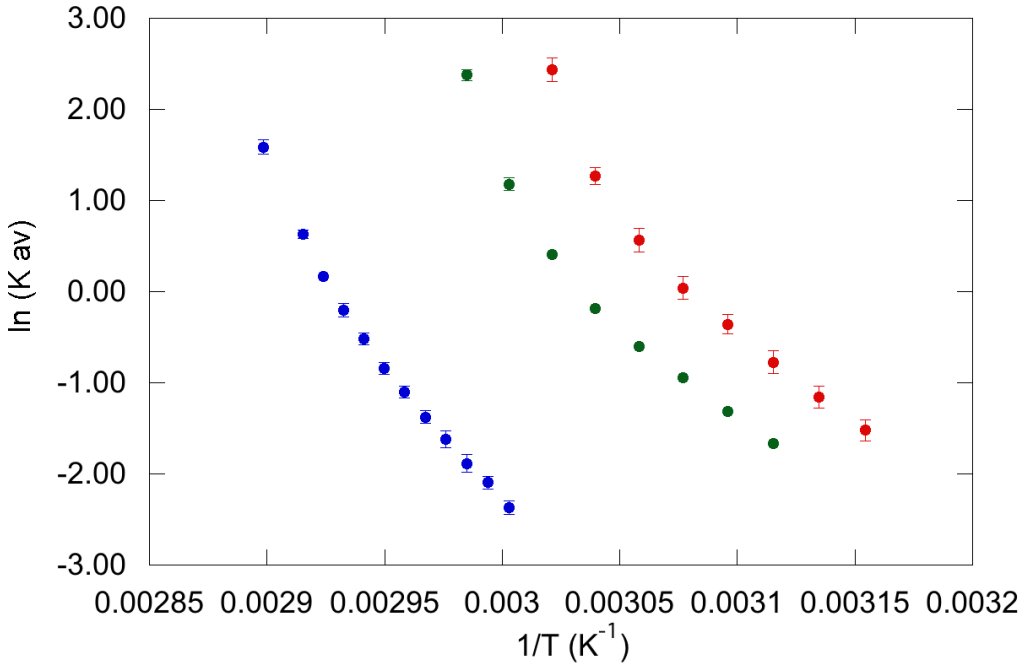


Figure 3-5: Van't Hoff plots of human, chimp, and rhesus Ap5 constructs displayed in blue, green and red, respectively. All Ap5 constructs do not display a linear relationship, thus rejecting a simple two-state folding/unfolding model.

Conversely, the three primate constructs do not fit well to a single-exponential decay, but fit well to a three-exponential decay (eq. 3-6). For example, ChAp5 lifetimes of approximately 0.5(τ_1), 3.0 (τ_2), and 8.5 (τ_3) ns, contribute to 75(f_1), 20(f_2) and 5(f_3) percent of the total fluorescence intensity, respectively (Figure 3-7). As with free 2Ap, these measurements were performed at ambient temperature, thus it is logical to suggest that τ_1, τ_2, and τ_3, alongside their corresponding fractions (data not shown), represent a dynamic equilibrium between folded, intermediate, and base-exposed conformers, respectively.

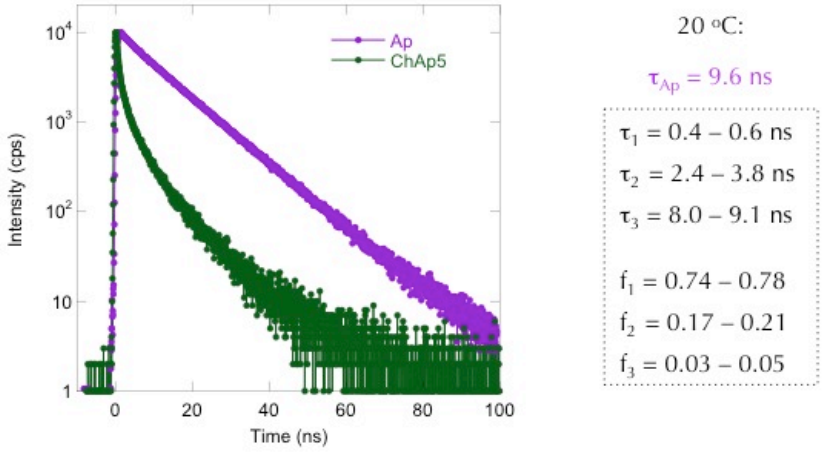


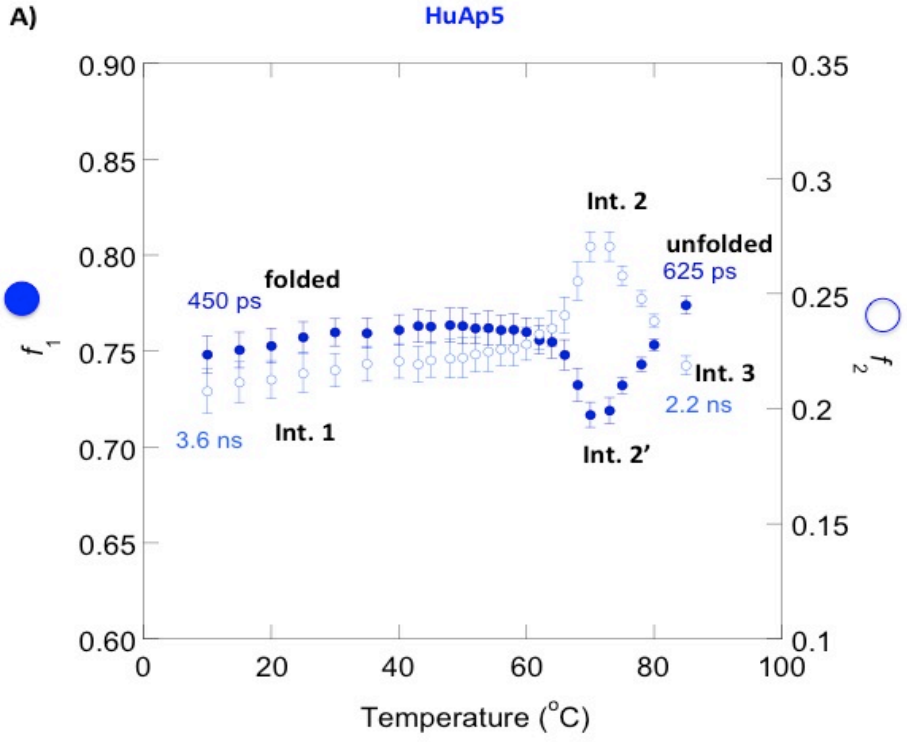
Figure 3-6: Free 2Ap and ChAp5 fluorescent lifetime comparison. Plot of log fluorescence intensity vs. time (ns) for free 2Ap, and 2Ap within ChAp5 construct, with fluorescence lifetimes, τ_n, and fractions of each lifetime, f_n, displayed on the right.

If a dynamic, multi-state equilibrium exists, then variations in temperature should change the fraction of each conformer in solution. To test this hypothesis, we analyzed fluorescence lifetimes and their contributions (f_n) to overall intensity for HuAp5, ChAp5, RhAp5, from 10-85°C. The most profound differences are observed in comparing the lifetime distributions between HuAp5 and Ch/RhAp5 (Figure 3-7). At temperatures ranging from 10-40°C, and notably at 37°C, the contributions of f_1 and f_2 in HuAp5 are 75% and 25%, respectively; while in ChAp5 the contributions are 85 and 15%, respectively. Since we assigned τ_1 and τ_2 to folded and intermediate states, within the 10-40 °C range, it is apparent that more intermediate is present at physiological temperatures in HuAp5. Not only is the amount of intermediate more prevalent at physiological temperatures, but the thermostability of the intermediate is enhanced. The difference in distribution of f_1 and f_2 is more pronounced near the T_M of ChAp5, compared to

the T_M of HuAp5. The relevance of the above observations in relation to A-to-I editing levels *in vivo* is of great interest.

3.2.4.1 Time-Resolved Fluorescence Multi-Intermediate Model

The fractions of fluorescence lifetime regimes with respect to temperature (f_1, f_2, f_3) are suggestive of the local/global conformers in solution (Figure 3-7). f_3, the fraction of the longest lifetime (τ_3), is reflective of 2Ap flipped out into solution with relatively minimal interactions with nearby bases. Since f_3 over all temperatures contributes to <5% of the overall fluorescence intensity (Σf_n =1) we neglect its contribution in further analyses (data not shown).



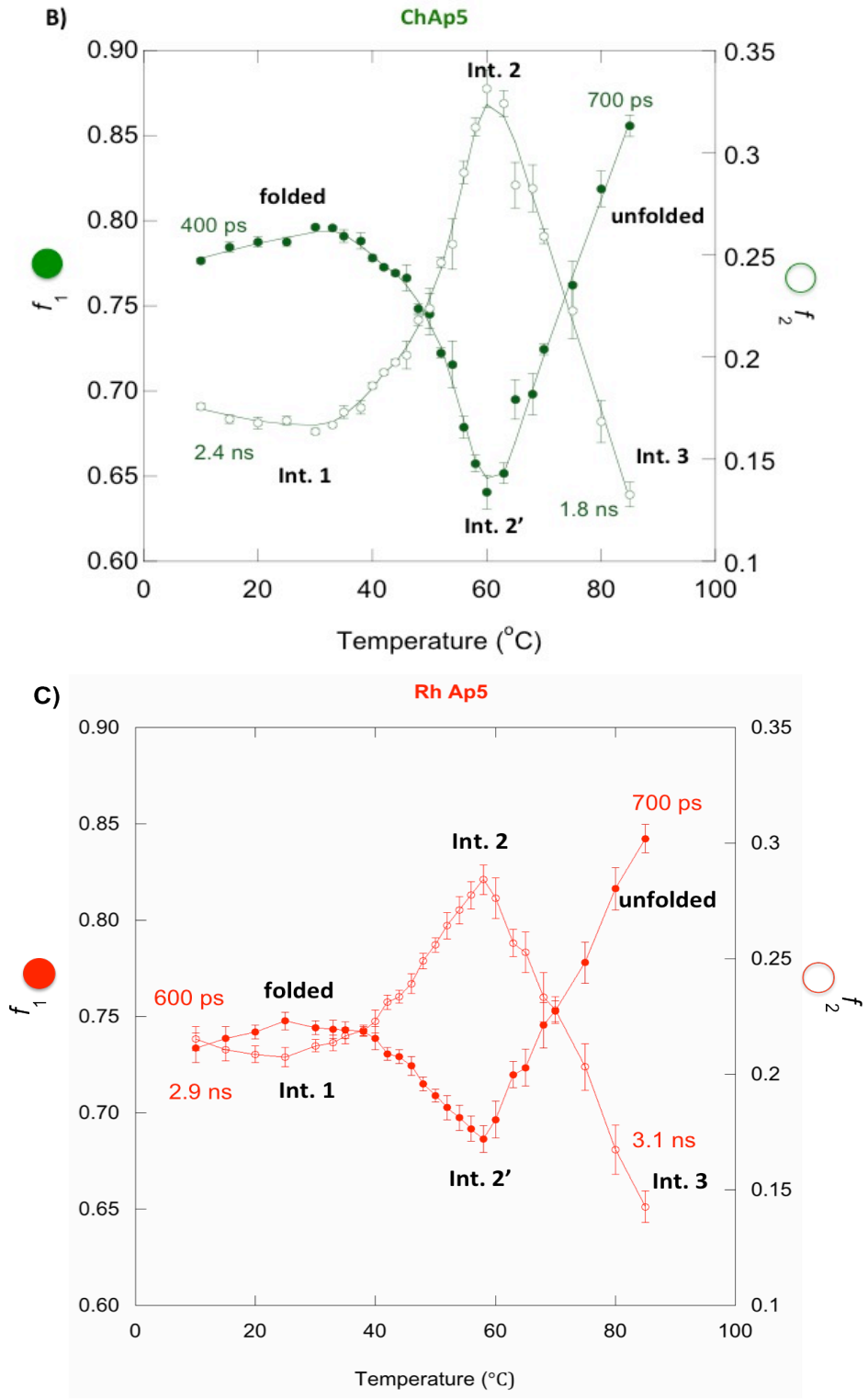
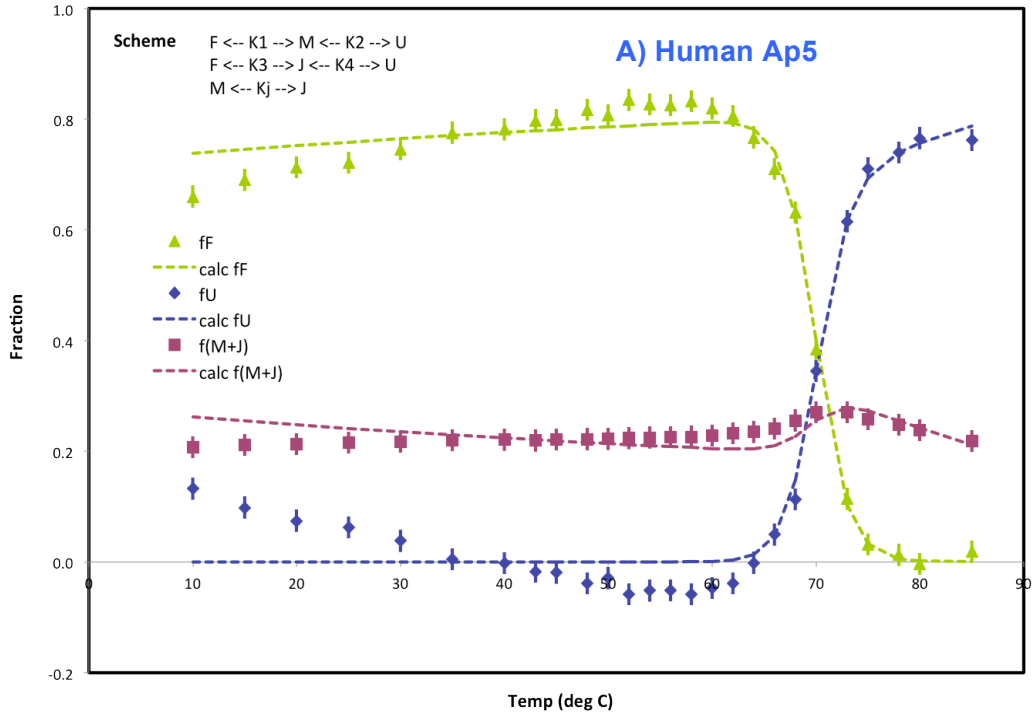


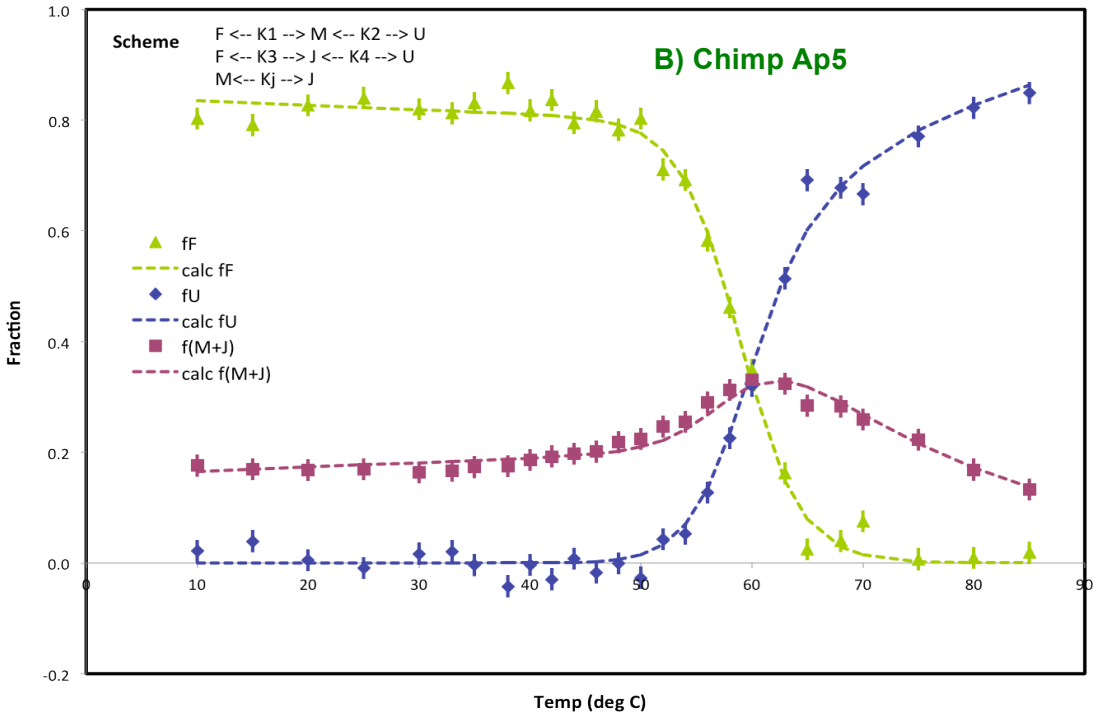
Figure 3-7: Temperature-dependent conformer distribution. Plot of fractions of short lifetime (f_1, closed circles) and intermediate lifetime (f_2, open circles) vs. temperature for HuAp5 (A), ChAp5 (B), and RhAp5 (C) constructs. Proposed conformers (folded, intermediate, unfolded) are noted along curves.

The fraction of the intermediate lifetime, f_2, is used directly in further analyses, and we attribute it to the fraction of intermediate(s) conformers with respect to temperature. For all three constructs, upon increasing temperature, we observe an increase in f_2 as the T_M (roughly) is approached, followed by a decrease in f_2 (Figure 3-7). This is suggestive of the respective development and collapse of the intermediate(s) upon increasing temperature. Such an observation is consistent with a minimal three-state folding/unfolding pathway (folded, intermediate, unfolded).

Analyses of f_1 with respect to temperature provide a unique challenge. Since f_1 is the fraction of the shortest lifetime ($\tau_{1)}$, it is indicative of 2Ap fluorescence quenching by neighbouring nucleotides in a folded conformation. In addition, and specifically at high temperatures when the RNA is in a coiled conformation, adjacent bases may π-stack with one another causing a decrease in fluorescence intensity.[31] Such properties are observed in our plot of f_1 vs. temperature, as we see a decrease in f_1 as the T_M is roughly approached, followed by an increase in f_1 as the T_M is exceeded (Figure 3-7). The specific challenge is assigning folded (f_f) and unfolded (f_u) conformers as components of f_1. To approximate f_u, we used equations 3-7 and 3-8, and to approximate f_f we used equation 3-9 (see experimental section 3.4.6).

Assuming we have folded, intermediate and unfolded conformers in solution, we can extract thermodynamic parameters (K_n, ΔH_n and ΔS_n , see Table 3-1) according to scheme 3-1.


Scheme
F <-- K1 --> M <-- K2 --> U
F <-- K3 --> J <-- K4 --> U
M <-- Kj --> J
A) Human Ap5
fF
calc fF
fU
calc fU
f(M+J)
calc f(M+J)
Fraction
Temp (deg C)


Scheme
F <-- K1 --> M <-- K2 --> U
F <-- K3 --> J <-- K4 --> U
M<-- Kj --> J
B) Chimp Ap5
fF
calc fF
fU
calc fU
f(M+J)
calc f(M+J)
Fraction
Temp (deg C)

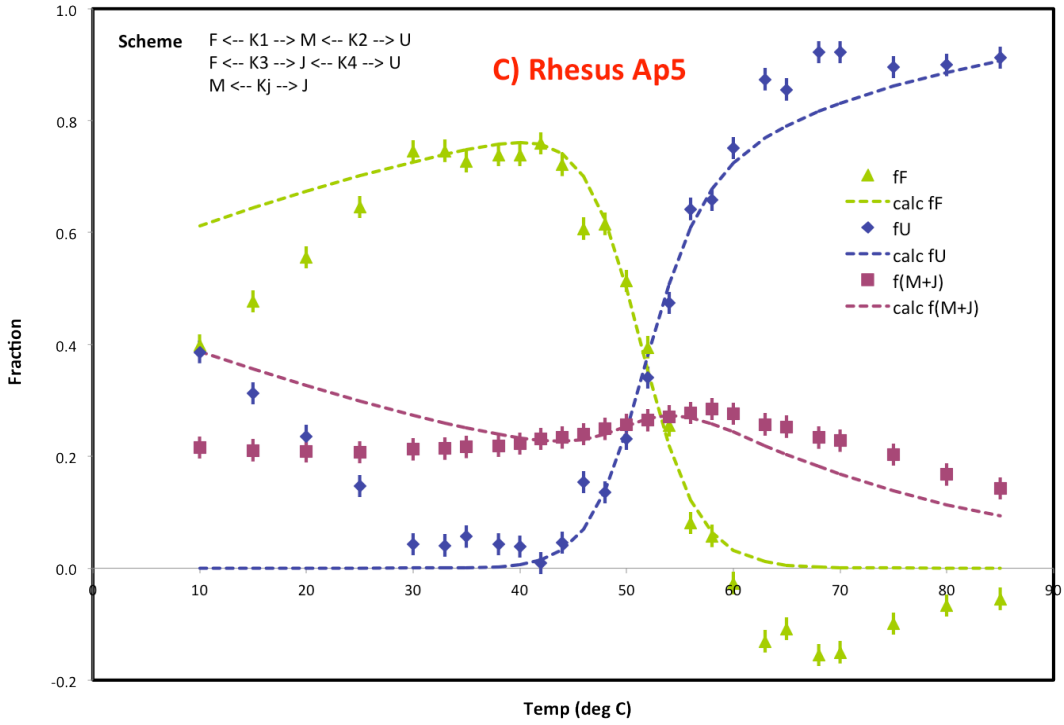
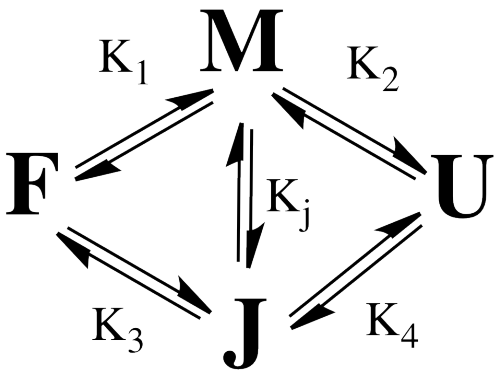


Figure 3-8: Two-intermediate folding/unfolding model. Plotted are fraction of folded, intermediate(s), and unfolded conformers with respect to temperature. HuAp5 (A) generally abides by the mode in Scheme 3-1, with deviations at lower temperatures. The ChAp5 construct is well modelled by Scheme 3-1 (B) at all temperatures, whereas The RhAp5 (C) construct does not fit well to this model at low and high temperatures. Note: calculated values are displayed with dashed lines, and real values are indicated by data points.



Scheme 3-1: Proposed folding/unfolding pathway for primate Ap5 constructs

A folded conformer (F) may transit to intermediates M or J whereby $f_M + f_J = f_i$. Probabilities of such transitions can be approximated by K_1 and K_3,

respectively. Furthermore, this scheme assumes that intermediates can exchange (K_J), and both M and J can independently approach the same unfolded state (U), as estimated by K_2 and K_4, respectively. The behaviour of the chimp construct is modeled quite well (Figure 3-8B), but our human and rhesus data do not abide by this scheme at all temperatures due to negative f_u values obtained via equation 3-7. Nonetheless, human and rhesus constructs generally follow our predicted values, permitting us to reasonably extrapolate thermodynamic parameters (Table 3-1).

3.3 Discussion:

3.3.1 Two-State vs. Multi-State Folding/Unfolding

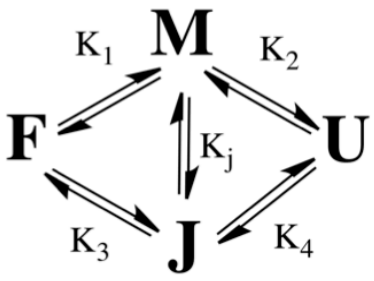
From our UV/temperature variation studies we were able to determine the global T_M of our three constructs with corresponding Hill coefficients. Interestingly, the SSF-based Hill coefficients and site-specific T_Ms coincide with the UV-based values within experimental error. If the T_M for Ap5 *and* the entire construct are roughly the same then is the global folding/unfolding pathway a two-step, or multi-step process?

Our Van't Hoff plots do not obey an overall linear relationship, and allude to a multi-state model. It is important to note that the data used to display the Van't Hoff plots are the same as those used for the site-specific T_Ms. Thus, the imprecise fits, broad transitions, and low Hill coefficients we observe (Fig 3-4B), directly correlate with the Van't Hoff plots. This rationalizes the lack of linearity

and failure to accurately extract thermodynamics parameters using this approach.

Construct	Pathway/Thermodynamic Parameters (ΔH_n kJ/mol, ΔS_n J/mol K, K_n at 37 °C)				
ChAp5	**F→M**	**M→U**	**F→J**	**J→U**	**M→J**
	K_1= 2.16x10^{-4} ± 9 x10^{-6}	K_2= 0.320 ± 1.5 x10^{-2}	K_3 = 0.230 ±1.3 x10^{-2}	K_4 = 3.01 x 10^{-4} ± 3x10^{-5}	K_J = 1065.8±7.71 x10^{1}
	ΔH_1 = 303.85 ± 0.08	ΔH_2 = 57.34 ± 0.09	ΔH_3 = 4.19 ± 0.11	ΔH_4 = 357.0 ± 0.2	ΔH_J = -299.7 ± 0.1
	ΔS_1 = 909.5 ± 0.3	ΔS_2 = 175.4 ±0.3	ΔS_3 = 1.3 ± 0.3	ΔS_4 = 1083.6 ± 0.5	ΔS_J = -908.2 ± 0.4
HuAp5	**F→M**	**M→U**	**F→J**	**J→U**	**M→J**
	K_1= 8.65 x10^{-11} ± 7.2 x10^{-12}	K_2= 0.560 ± 2.53 x10^{-2}	K_3 = 0.290 ± 8.08 x10^{-3}	K_4 = 1.64 x10^{-10} ± 1.62 x10^{-11}	K_J = 3.40 x10^{9} ± 2.98 x10^{8}
	ΔH_1 = 596.63 ± 0.17	ΔH_2 = 36.46 ± 0.09	ΔH_3 = -5.02 ± 0.05	ΔH_4 = 638.1 ± 0.2	ΔH_J = -601.7 ± 0.2
	ΔS_1 = 1731.1 ± 0.4	ΔS_2 = 112.7 ± 0.3	ΔS_3 = -26.4 ± 0.2	ΔS_4 = 1870.2 ± 0.5	ΔS_J = -1757.5 ± 0.5
RhAp5	**F→M**	**M→U**	**F→J**	**J→U**	**M→J**
	K_1= 2.91 x10^{-3} ± 1.69 x10^{-4}	K_2= 0.888± 3.45 x10^{-2}	K_3 = 0.31751 ± 9.58 x10^{-3}	K_4 = 0.00812 ± 6.2 x10^{-4}	K_J = 109 ± 7.2
	ΔH_1 = 291.04 ± 0.11	ΔH_2 = 46.18 ± 0.06	ΔH_3 = -18.74 ± 0.06	ΔH_4 = 356.0 ± 0.1	ΔH_J = -309.8 ± 0.1
	ΔS_1 = 889.9 ± 0.3	ΔS_2 = 147.9 ± 0.3	ΔS_3 = -70.0 ± 0.2	ΔS_4 = 1107.7 ± 0.5	ΔS_J = -959.8 ± 0.4

Table 3-1: Thermodynamic parameters for primate constructs from multi-state model proposed in scheme 3-1 (shown again below).



3.3.3 Nup50 Ch/RhAp5 vs. HuAp5 TRF data

To further investigate the thermodynamics of each construct we utilized TRF. The multiple fluorescence lifetime fractions gathered from TRF measurements indicate variable 2Ap environments, and suggest different local/global conformers. Thus, data gathered from TRF could rationalize the segmented linearity in our Van't Hoff plots, and justify a multi-state folding/unfolding model.

From our TRF measurements on all three primate constructs we noticed the most profound differences between HuAp5 and ChAp5. We have assigned f_1 to 2Ap in a closed/quenched state prior to the T_M, and in an open/π-stacked state exceeding the T_M. We also assign f_2 to 2Ap within an intermediate state throughout the entire unfolding process. Comparing f_2 from HuAp5 and Ch/RhAp5 we notice that the former is less resistant to change with respect to temperature, and is also more prevalent at physiological temperature. Such observations may be indicative of the substrate-level events facilitating A-to-I editing *in vivo*.

Since ADAR is believed to perform a hydrolytic deamination via a docking and base-flip mechanism, a stable duplex must be present.[23,33] From our UV experiments we have established that the HuAp5 construct is the most thermostable. The overall robustness, and proportion of f_2 at 37°C is of profound biological relevance. Since f_2 represents an intermediate lifetime/state, HuAp5 is likely positioned in a fashion whereby it is more solvent-exposed, and less

quenched by neighbouring nucleotides relative to Ch/RhAp5. Thus, the prominence and persistence of f_2 in HuAp5 may be key to explaining the enhanced A-to-I editing found *in vivo*, from the perspective of the substrate.

3.3.4 Nup50 Ap5 Construct Multi-Intermediate model

3.3.4.1 Comparison of Thermodynamics

In addition to our global/local T_M data, our TRF-derived equilibrium constants lend support to the commonly believed model for A-to-I editing, namely, that a thermostable duplex surrounding the edit site is required for docking, alongside the labile/intermediate-like base efficiently flipping out of the duplex. Specifically, our results provide insights into why increased editing levels are found in human *Nup50 AluSg*, relative to other primates.

The multi-intermediate model presented in Scheme 3-1 fits best with the ChAp5 construct's behaviour, and along certain temperature ranges for Hu/RhAp5 constructs (see Fig. 3-8 and Table 3-1). For ChAp5 at physiological temperatures, K_1 (F→M) and K_3 (F→J) are less than 1, indicating that 2Ap is in an overall quenched/folded state. Interestingly, the relatively high K_J (M→J) value of 1065 indicates that if a folded-to-intermediate(s) transition occurred, state J would be heavily favoured. The low K_2 and K_4 values demonstrate that at 37 °C, the unfolded state is not very probable (Table 3-1).

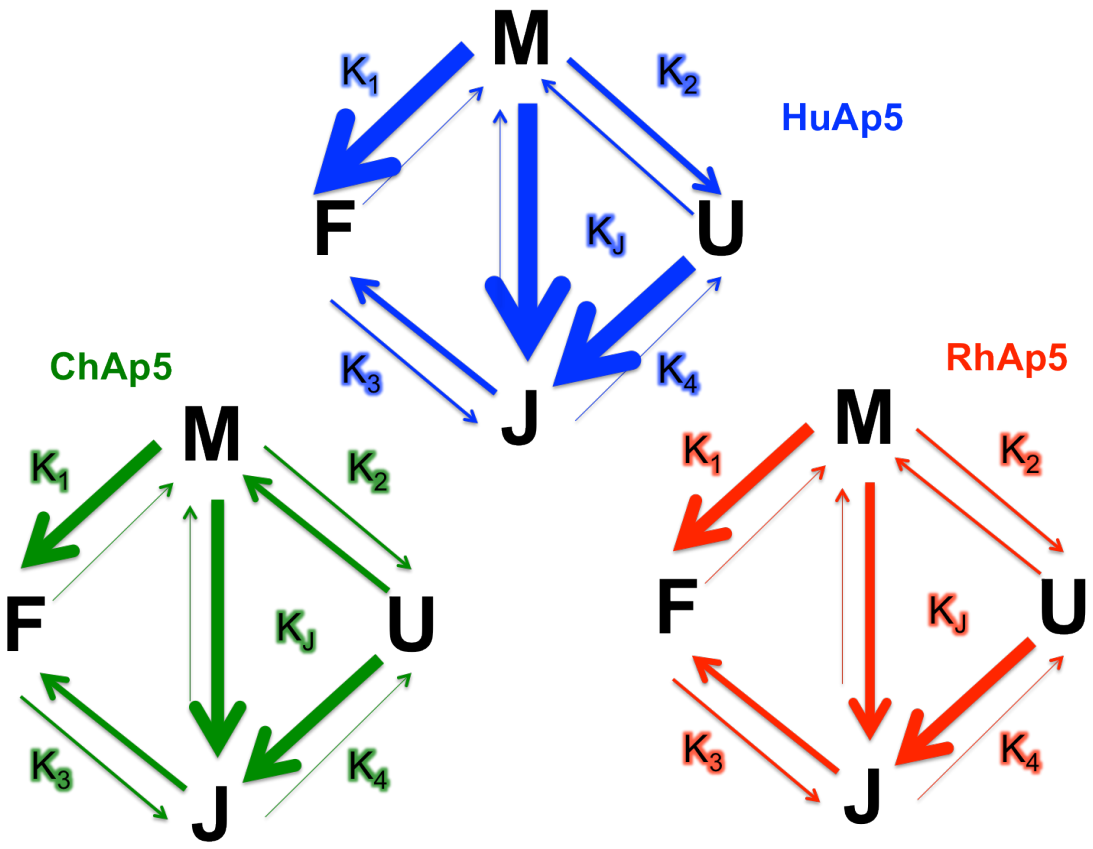
Equilibrium constants (37 °C) from our modeling of the RhAp5 construct were along the same order of magnitude compared to ChAp5. Such values were

accompanied by greater error due to significant deviation from our proposed model, with high levels of negative f_f and f_u values (Figure 3-8C). At low temperatures this is also apparent with the HuAp5 construct and is rooted from the fact that the lowest τ_1 value does not correspond to the lowest temperature (data not shown). Since equation 3-7 assumes that τ_1 is shortest at the lowest temperature, our data sometimes translate to negative f_f and f_u values. Although a negative fraction is not feasible, our analyses take this into account by including error values alongside our thermodynamic parameters (Table 3-1). Negative fractions found in HuAp5 do not deviate substantially from our predicted values (Figure 3-8A).

Despite the slight deviation from our model at low temperatures, examination of HuAp5 equilibrium constants allows us to draw interesting conclusions (Scheme 3-2). At physiological temperature, K_1 (F→M) is extremely low, whereas K_3 (F→J) is similar compared to rhesus and chimp constructs (Table 3-1). The most notable difference is in the value of K_J (M→J), as it is five orders of magnitude higher in HuAp5 relative to Ch/RhAp5. Thus, if any intermediate forms (J or M), the formation of J is heavily favoured in the human construct.

At 37 °C, the K_1 and K_3 values generally favour a folded state. Interestingly, the F→J rate constant (K_3) is approximately 0.3 across the three primate constructs, indicating that Ap5 is typically not in an intermediate state. However, docking of ADAR or other cellular proteins onto a thermostable duplex

could provide the system with the energy required to undergo the F→J transition. Since ADAR is relatively conserved in sequence across primates, it should interact with the human construct with the highest affinity, facilitating the F→J transition.



Scheme 3-2: Proposed pathway of each primate Ap5 construct, with increased arrow size representing (not to scale) increased probability of transition. See Table 3-1 for details.

The M→J rate constant (K_J) is highest in the human construct by five orders of magnitude. This indicates that if the intermediate M were to form from the RNA independently folding, or chaperoned by a protein, nearly the entire population would transit to J. From an RNA perspective, there is less J formation

in Rh/ChAp5 constructs. However, if enzyme docking were to facilitate the F→M transition, it would likely be a less efficient process in Rh/ChAp5 constructs as they are less stable in the folded state than HuAp5.

Another important observation is in the relative K_4 values, with the corresponding HuAp5 being lowest (37°C) by four orders of magnitude. This indicates that the human construct will remain in state J, and not state U, whereas the J→U transition is more probable in Rh/ChAp5 constructs. The prevalence of a thermostable folded state, alongside the greater ability to transfer to the hyperedited (presumably) state J suggests why editing of A5 in *Nup50 AluSg* is highest in humans. The above interpretations are representative of local conformational changes, yet allude to those on a global RNA scale. Further experiments are required from an exclusive RNA perspective, as well as RNA in combination with ADAR.

3.3.5 Conclusions and Perspectives

Our constructs are predictions of the A-to-I editing substrates that exist *in vivo*, as it is currently unknown what the central/edited element actually pairs with. To determine if these constructs are truly representative of substrates *in vivo*, it would be beneficial to treat these substrates with ADAR and analyze editing efficiency via inosine cleavage or sequencing. Results could then be compared with those presented in the foundational Rechavi study presented earlier in this chapter (Figure 3-2). Our investigations are the first to use 2Ap to

probe dynamics of A-to-I editing substrates, but to provide a more representative view additional experiments could be performed.

To determine the global folding/unfolding pathway of our three conformers it would be worthwhile to determine site-specific T_Ms for 2Ap at multiple positions within the hairpin constructs. For all three primates, we currently have Ap13 and Ap18 constructs in our possession. Similar to position 5, position 13 is representative of a frequently edited site in humans, relative to chimp and rhesus. If we observe the same site-specific T_Ms within experimental error compared to our Ap5 constructs, we may be able to draw a conclusion similar to Schwalbe and colleagues,[32] namely, that global unfolding is highly cooperative, but bases may reposition themselves in intermediate states along the helix to coil transition.

Placing 2Ap within the loop of our construct (Ap18) could also provide us with a site-specific T_M that would roughly mirror Ap5 and Ap13 T_M curves. Here, Ap18 would be solvent exposed at low temperatures (10 °C), but quenched via π-stacking at high temperatures (85°C). Consequently, the relative fluorescence at low temperatures will be greater than at higher temperatures.

Despite our thermodynamic limitations, information from TRF regarding the proportion and overall thermostability of the fluorobase at position 13 will be key to our assumptions from our Ap5 constructs. Since Ap13 represents a heavily edited site in humans, we would anticipate that f_2 for the HuAp13

construct would follow a similar trend as f_2 from HuAp5. In addition, we anticipate ChAp13 and RhAp13 constructs to follow similar trends with respect to their corresponding Ap5 constructs. If these observations surface, the ADAR base flipping mechanism would be further supported. Furthermore, we could utilize trends in f_2 as a function of temperature to predict and validate *in vivo* A-to-I edit sites.

The dynamics we observe contribute to the fields of RNA hairpin folding and A-to-I editing. With respect to the former, we have demonstrated that sequence dictates the folding landscape, and that for a seemingly simply structure, we may not have resolved the multitude of conformers in solution. Although many conformers may exist, with respect to A-to-I editing, our data suggest that in particular the abundance and stability of an intermediate conformer may facilitate ADAR docking and editing.

3.4 Experimental

3.4.1 Primate Construct Development

3.4.1.1 *Nup50 Alu* RNA Alignments

Human, Chimp, and Rhesus *Nup50 Alu* DNA sequences were initially located in the UCSC genome browser as described by Rechavi et al. The central *AluSg* was closest to the antiparallel *AluJb* DNA sequence in all three primates, and the two DNA sequences were aligned using CLUSTALW

(http://www.genome.jp/tools/clustalw/). Following this, the *AluSg* sequence was converted to its RNA form (T→ U), while the *AluJb* DNA sequence was converted to its RNA form, followed by a reverse complement conversion. A random 25 nt linker was 'placed' in between the fully converted *AluJb* and *AluSg* elements, which was intended to loosely mimic the approximately 1000 nt that separates these adjacent elements *in vivo*. Consequently, a ssRNA sequence was 'formed' that would promote the formation of an intramolecular duplex (A-to-I editing substrate), according to the Mfold algorithm.

Since positions of edited adenosines in *Nup50 AluSg* elements were already proposed by Paz-Yaacov et al., we identified the corresponding sites within the context of our constructs. Specifically, we analyzed A5 due to its highly variable editing levels across primates, with the highest levels found in human, compared to chimp and rhesus. Furthermore, subtle sequence differences surrounding the edit sites were also found upon Mfold 2-D structural prediction.

3.4.1.2 Construct Design:

We acquired three 36 nt constructs (Dharmacon), corresponding to human, chimp, and rhesus *Nup50 AluSg/Jb* dsRNA elements, with 2Ap substituted at position 5, which we refer to as HuAp5, ChAp5, and RhAp5. Positions 1-16 are reflective of *AluSg*, positions 17-20 contain GAAA (GNRA tetraloop upon folding), and positions 21-36 consist of the RNA-converted reverse complement of *AluJb*. Upon low-temperature folding, nucleotides 1-16

base pair with 17-20, with an intervening tetraloop nucleating the formation of a hairpin construct upon folding conditions.

3.4.2 RNA Processing

3.4.2.1 RNA Deprotection

Approximately 200 nmol of 2'ACE® protected RNAs (pellets) were centrifuged briefly, followed by the addition of 400 uL of deprotection buffer (100 nM acetic acid, pH adjusted to 3.8 with TEMED). RNAs were then dissolved by pipetting up and down, followed by ten seconds (each) of vortexing and centrifugation. Following a 60°C incubation for 30 minutes, samples were vacuum concentrated to dryness (Mini-SpeedVac DDA concentrator, Savant).

3.4.2.2 Gel Purification of RNA

Deprotected RNAs were then resuspended in 100 uL of 0.5x TE and 100 uL of 2X RNA loading dye (95% deionized formamide, 0.025% w/v bromophenol blue, 0.025% w/v xylene cyanol, and 5 mM EDTA, pH 8.0). To purify the RNAs, a 15% acrylamide gel (19:1 Bis: Acrylamide, 6.67 M Urea, (30cm x 60cm x 1cm) was cast, and pre-run to 50°C at 80 W. Resuspended RNAs were heat denatured at 85° C for two minutes and loaded into 10 centimetre-long wells, and run for approximately 2.5 hours.

Via UV shadowing, the gel pieces for HuAp5, ChAp5, and RhAp5, corresponding to the slowest migrating (full-length), most prominent band, were

excised with a sterile blade and subjected to an overnight crush and soak procedure in 300 mM sodium acetate, pH 4.8, at 4°C. The liquid and solid from the resulting slurry were separated by filtration through a 0.25 micron filter (Millipore), and the aqueous solution was precipitated overnight at -20°C in 2.5 volumes of anhydrous ethanol. Samples were then centrifuged at 4°C for 30 minutes, followed by removal of the liquid and three subsequent washes in 200 uL of 70% ethanol. The remaining pellets were briefly vacuum-dried and resuspended in fluorobuffer (10 mM sodium phosphate, 0.1 mM EDTA, ph 7.1).

3.4.3 Temperature-Dependent UV Absorbance Measurements

The absorbance values at 260 nm and 600 nm (for normalization) of 500 nM HuAp5, ChAp5, and RhAp5 in 500 uL fluorobuffer were measured with the Varian Cary 100 spectrophotometer, equipped with a temperature-controlled multi cell holder, and absorbance values at temperatures ranging from 10-85°C were recorded. A graph of absorbance (260 nm) versus temperature (°C) was then plotted in order to determine the 'global' T_M of each construct. Furthermore, using the Hill equation:

$$S(T) = \frac{\Delta S_{max} * T^n}{{T_M}^n + T^n} \qquad \textbf{3-1}$$

where S(T) represents the absorbance as a function of temperature, S_{max} is the absorbance at the maximum temperature, T_m is the melting temperature, and n is the Hill coefficient, we are able to determine the cooperativity of the folding/unfolding process, with the highest number being the most cooperative.

3.4.4 Steady-State Fluorescence Measurements at Variable Temperatures

Hu-Ap5, Ch-Ap5, Rh-Ap5 were individually placed in solution and excited at 320 nm under steady-state conditions, and their emissions (360 nm) were recorded. In our site-specific T_m curves (Figure 3-4B), we plotted fraction unfolded (f_u) (Y) vs. temperature (X),

$$f_u = \frac{\mathrm{I(T)}}{\mathrm{I(T}_{Max}\mathrm{)}} \quad \textbf{3-2}$$

whereby the fraction unfolded is the fluorescence signal (360 nm) at a given temperature divided by the signal at the highest temperature. Using this information, we also calculate the Hill coefficient, which followed the same trends as those calculated by UV (Figure 3-4B/C).

3.4.4.1 Steady-State Fluorescence-Based Two-State Model

Assuming a two-state model, consisting of an equilibrium between folded and unfolded states, we calculated the equilibrium constant for unfolding (K_{unfold}) by the following equation:

$$\mathrm{K_{unfold}} = [\mathrm{U}]\,/[\mathrm{F}] = fu\,/(1 - fu\,) \quad \textbf{3-3}$$

Since experiments were conducted over temperatures ranging from 10-85°C, we plotted ln(K_{unfold}) vs. 1/T(K^{-1}), and used the Van't Hoff equation:

$$\ln\mathrm{K_{unfold}} = \frac{-\Delta\mathrm{H}}{\mathrm{R}} * \frac{1}{\mathrm{T}} + \frac{\Delta\mathrm{S}}{\mathrm{R}} \quad \textbf{3-4}$$

to extrapolate enthalpy (ΔH) and entropy (ΔS) components.

3.4.5 Time-Resolved Fluorescence Measurements

Fluorescence measurements were made using a Horiba-JobinYvon Fluorolog spectrometer equipped with an IBH detector and electronics for time correlated single photon counting (TCSPC). The excitation source was a 295 nm pulsed nanoLED operating at a 1 MHz repetition rate. Emission was passed through a double monochromator with a 15 nm bandpass centered at 370 nm, or through a 320 nm cut-off filter yielding broadband emission (~340-600 nm), under magic angle conditions (54. 7^{o}). The instrument response function (IRF), measured by scattering the excitation beam off a dilute solution of Ludox was 735 ps FWHM: For our IRF, t ≥ 300 ps were reliably and reproducibly measured. Temperatures were maintained at desired temperatures with a Peltier-controlled thermostated sample holder, and involved approximately five-degree steps upon approach of the T_M, two-degree steps within the T_M regime, and five-degree steps at temperatures well exceeding the T_M. Experiments were performed with constant RNA concentration of 500 nM in fluorobuffer. Fluorescence decay curves were collected at each step (reverse mode, 100 ns time scale, resolved into 2048 channels), for a constant run time, typically 300 s (≥ 10000 counts in peak channel). Decay curves were analyzed using standard iterative reconvolution (DAS software, Horiba-JobinYvon). Fit quality was assessed by reduced chi-squared (χ^2 ≤ 1.3 acceptable) and randomness of residuals.

For comparison, the fluorescence decay profile of 500 nM free 2Ap in fluorobuffer was evaluated. The relatively long lifetime detected is due to minimal fluorescence quenching in solution, and was evaluated by a single exponential decay (Figure 3-6).

$$\mathrm{I(t)} = fe(-\frac{t}{\tau}) \quad \textbf{3-5}$$

Here, I(t) is the intensity with respect to time, τ is the fluorescence lifetime, and *f* is the fraction that the corresponding lifetime contributes to the overall intensity (*f* ≅1).

All three constructs do not fit well to a single-exponential decay, rather, they fit well to a three-exponential decay,

$$\mathrm{I(t)} = f_1 e^{-t/\tau 1} + f_2 e^{-t/\tau 2} + f_3 e^{-t/\tau 3} \quad \textbf{3-6}$$

with f_1, f_2 and f_3 (Σf_n =1) representing the fractions of short, intermediate, and long lifetimes (τ_1, τ_2, τ_3, respectively) as a function of temperature, and I(t) representing the total fluorescence signal.

3.4.6 Assigning Conformers from Lifetime Fractions

Since f_1 is representative of both the folded and unfolded states, we estimated the proportion of the unfolded conformer first using equation 3-7

$$f_u(\tau_1) = \frac{\tau_1(T) - \tau_{1folded}}{\Delta\tau_{max}} \quad \textbf{3-7}$$

where f_u (τ_1) is the fraction unfolded as a function of τ_1 determined by the difference between τ_1 at a given temperature, τ_1 (T) and τ_1 at the lowest (presumably most folded temperature) $\tau_{1folded}$, divided by the maximum difference between τ_1 across the given temperature range, $\Delta\tau_{max}$.

$f_u(\tau_1)$ provides us with the proportion of the folded conformer, as a function of τ_1 across our temperature range. However, the fraction of the short lifetime f_1 must also be accounted for. Thus, to determine the fraction unfolded (f_u) we utilized equation 3-8.

$$f_u = f_u(\tau_1) \times f_1 \quad \textbf{eq 3-8}$$

Since $\Sigma f_n = 1$, and we have already determined the proportions of unfolded (f_u) and intermediate ($f_2 = f_i$) conformers, assuming a three-state model, we determined the fraction folded (f_f) by equation 3-9.

$$f_f = 1 - f_u - f_i \quad \textbf{eq 3-9}$$

3.5 References

1. Chalker, D. L. *Biochim Biophys Acta* **2008**, *1783*, 2130.
2. Moore, G. L.; Maranas, C. D. *J Theor Biol* **2000**, *205*, 483.
3. Mann, M.; Jensen, O. N. *Nat Biotechnol* **2003**, *21*, 255.

4. Masterson, J.; O'Dea, S. *Cells Tiss Org* **2007**, *185*, 175.

5. Rashid, M. G.; Sanda, M. G.; Vallorosi, C. J.; Rios-Doria, J.; Rubin, M. A.; Day, M. L. *Cancer Res* **2001**, *61*, 489.

6. Gott, J. M.; Emeson, R. B. *Annu Rev Genet* **2000**, *34*, 499.

7. Li, M.; Wang, I. X.; Li, Y.; Bruzel, A.; Richards, A. L.; Toung, J. M.; Cheung, V. G. *Science* **2011**, *333*, 53.

8. Silberberg, G.; Lundin, D.; Navon, R.; Ohman, M. *Hum Mol Genet* **2011**.

9. Mehler, M. F.; Mattick, J. S. *Physiol Rev* **2007**, *87*, 799.

10. Dominissini, D.; Moshitch-Moshkovitz, S.; Amariglio, N.; Rechavi, G. *Carcinogenesis* **2011**.

11. Osenberg, S.; Paz Yaacov, N.; Safran, M.; Moshkovitz, S.; Shtrichman, R.; Sherf, O.; Jacob-Hirsch, J.; Keshet, G.; Amariglio, N.; Itskovitz-Eldor, J.; Rechavi, G. *PLoS One* **2010**, *5*, e11173.

12. Frith, M. C.; Pheasant, M.; Mattick, J. S. *Eur J Hum Genet* **2005**, *13*, 894.

13. Wahlstedt, H.; Daniel, C.; Enstero, M.; Ohman, M. *Genome Res* **2009**, *19*, 978.

14. Maas, S.; Godfried Sie, C. P.; Stoev, I.; Dupuis, D. E.; Latona, J.; Porman, A. M.; Evans, B.; Rekawek, P.; Kluempers, V.; Mutter, M.; Gommans, W. M.; Lopresti, D. *Biochem Biophys Res Commun* **2011**, *412*, 407.

15. Gommans, W. M.; Tatalias, N. E.; Sie, C. P.; Dupuis, D.; Vendetti, N.; Smith, L.; Kaushal, R.; Maas, S. *RNA* **2008**, *14*, 2074.

16. Blow, M.; Futreal, P. A.; Wooster, R.; Stratton, M. R. *Genome Res* **2004**, *14*, 2379.

17. Keegan, L. P.; Leroy, A.; Sproul, D.; O'Connell, M. A. *Genome Biol* **2004**, *5*, 209.

18. Keegan, L. P.; Brindle, J.; Gallo, A.; Leroy, A.; Reenan, R. A.; O'Connell, M. A. *Embo J* **2005**, *24*, 2183.

19. Goodman, R. A.; Macbeth, M. R.; Beal, P. A. *Curr Top Microbiol Immunol* **2011**.

20. Kim, D. D.; Kim, T. T.; Walsh, T.; Kobayashi, Y.; Matise, T. C.; Buyske, S.; Gabriel, A. *Genome Res* **2004**, *14*, 1719.

21. Brusa, R.; Zimmermann, F.; Koh, D. S.; Feldmeyer, D.; Gass, P.; Seeburg, P. H.; Sprengel, R. *Science* **1995**, *270*, 1677.

22. Morabito, M. V.; Emeson, R. B. *Neuropsychopharmacology* **2009**, *34*, 246.

23. Nishikura, K. *Annu Rev Biochem* **2010**, *79*, 321.

24. Paz-Yaacov, N.; Levanon, E. Y.; Nevo, E.; Kinar, Y.; Harmelin, A.; Jacob-Hirsch, J.; Amariglio, N.; Eisenberg, E.; Rechavi, G. *Proc Natl Acad Sci U S A* **2010**, *107*, 12174.

25. Bahn, J. H.; Lee, J. H.; Li, G.; Greer, C.; Peng, G.; Xiao, X. *Genome Res* **2011**.

26. Greenberger, S.; Levanon, E. Y.; Paz-Yaacov, N.; Barzilai, A.; Safran, M.; Osenberg, S.; Amariglio, N.; Rechavi, G.; Eisenberg, E. *BMC Genomics* **2010**, *11*, 608.

27. Ogawa, Y.; Miyamoto, Y.; Asally, M.; Oka, M.; Yasuda, Y.; Yoneda, Y. *Mol Biol Cell* **2010**, *21*, 630.

28. Costa, M.; Michel, F. *EMBO J* **1997**, *16*, 3289.

29. Rist, M.; Marino, J. *Nucleic Acids Res* **2001**, *29*, 2401.

30. Souliere, M. F.; Haller, A.; Rieder, R.; Micura, R. *J Am Chem Soc* **2011**, *133*, 16161.

31. Ballin, J. D.; Bharill, S.; Fialcowitz-White, E. J.; Gryczynski, I.; Gryczynski, Z.; Wilson, G. M. *Biochemistry* **2007**, *46*, 13948.

32. Rinnenthal, J.; Klinkert, B.; Narberhaus, F.; Schwalbe, H. *Nucleic Acids Res* **2010**, *38*, 3834.

33. Wong, S. K.; Sato, S.; Lazinski, D. W. *RNA* **2001**, *7*, 846.

Appendix: Chapter 3 Thermodynamic Parameters

Chimp Ap5

Fitting Parameters		+/-	to minimize	Include sch 1	1	SSQ (sum of squares)
$-\Delta H_1/R$	-36544	10	108.166282	SSQ (sum of squares)		
$\Delta S_1/R$	109.39	0.03	72	deg freed		
$-\Delta H_2/R$	-6896	10	1.5023095	SSQ/df = Sdfit	<--Used for error calculation	
$\Delta S_2/R$	21.10	0.03	109.66859	SSQ + 1SDfit		
-DH3/R	-504	13	109.66859	Value only (SSQ + 1SDfit)		
DS3/R	0.15	0.04				
$-\Delta H_4/R$	36040	16				
$\Delta S_4/R$	-109.24	0.05				

Define K's as:
$K_1 = [I]/[F]$, $K_2 = [U]/[I]$, $K_3 = [J]/[F]$, $K_4 = [U]/[J]$, $K_J = [J]/[I]$
Assuming that *one cannot distinguish between I and J*,
ie, one sees only $[I_{total}] = [I] + [J] = Sum[Intermediates)$
Assuming a closed system (ie, no other reactions), then only three equilibrium constants, K_1, K_2, K_3 for example, can define the system; with this particular choice, K_4 and K_J are linked to K_1, K_2 and K_3 and can be calculated by:
$K_4 = K_1K_J/K_3$ and $K_J = K_3/K_1 = K_2/K_4$

Enter temperature (to look up K and fraction values)

Temp/C	37	+/-		0.00322435 <--1/T	Calculated fractions @37 C		
K1	2.16E-04	9.43E-06	1/fF =	1 + K1 + K3 + K1K2	fF =	0.81293	
K2	0.320636	1.45E-02	1/fI =	1 + 1/K1 + K3/K1 + K2	fI =	0.00018	
K3	0.229833	1.32E-02	1/fJ =	1 + 1/K3 + K1/K3 + K1K2/K3	fJ =	0.18684	0.1870 <-- f(I+J)
Kj	1.07E-03	7.71E-01	1/fU =	1 + 1/K1K2 + 1/K2 + K3/K1K2	fU =	0.00006	
K4	0.000301	2.56E-05			total =	1.000000	

		+/-	
ΔH_1	303.85	0.08	kJ mol^{-1}
ΔS_1	909.5	0.3	J mol^{-1} K^{-1}
ΔH_2	57.34	0.09	kJ mol^{-1}
ΔS_2	175.4	0.3	J mol^{-1} K^{-1}
ΔH_3	4.19	0.11	kJ mol^{-1}
ΔS_3	1.3	0.3	J mol^{-1} K^{-1}
ΔH_J	-299.7	0.1	kJ mol^{-1}
ΔS_J	-908.2	0.4	J mol^{-1} K^{-1}
ΔH_4	357.0	0.2	kJ mol-1
ΔS_4	1083.6	0.5	J mol-1 K-1

F<-- K1 --> I <-- K2 --> U

F<-- K3 --> J <-- K4 --> U

I <-- Kj --> J

K1 = I/F

K2 = U/I

K3 = J/F

Then

K1 + K3 = Iob/F

K1K2 = U/F

I(observed) = Iob = I + J

Fitting parameters are highlighted purple

1/fU = 1 + 1/(K1K2) + (K1 + K3)/(K1K2)

1/fIob = 1 + 1/(K1 + K3) + (K1K2)/(K1 + K3)

1/fF = 1 + (K1 + K3) + (K1K2)

fIob = 1/(1+K2+(1+K3)/K1) + 1/(1+1/K3+K1(1+K2)/K3)

												+/-								+/-
										$-\Delta H_1/R$	-36544	17					$-\Delta H_2/R$	-6896	11
										ΔH_1	303847	141					ΔH_2	57336	89
										$\Delta S_1/R$	109.4	0.05					$\Delta S_2/R$	21.10	0.03
										ΔS_1	910	0.4					ΔS_2	175.4	0.3
										$-\Delta H_i/R$	-504	9					-(Delh1+Delh2)/R	-43440.33	
										DH3	4190	79					Delh1+DelH2	361182.97	
										DS3/R	0.15	0.03					(dels1+DelS2)/R	130.5	Tinit
										DS3	1.3	0.3					DelS1+DelS2	1085	Tfinal
			0.03		0.036		0.02												
			0.020		0.020		0.020			Not used									
	temp/C	1/T (K^{-1})	fU/fract	+/-	fI/fract	+/-	fF/fract	+/-	K1 + K3	+/-	K1*K2	+/-	K1(calc)	K2(calc)	K3(calc)	[K1 + K3] (calc)	[K1*K2](calc)	fU(calc)	fIob(calc)
	10	0.00353	0.021	0.020	0.176	0.020	0.803	0.020	0.219	0.028	0.017	0.016	0.000	0.038	0.197	0.1968	1.09E-10	9.14E-11	0.1645
	15	0.00347	0.039	0.020	0.170	0.020	0.791	0.020	0.214	0.028	0.031	0.016	0.000	0.059	0.203	0.2030	1.57E-09	1.30E-09	0.1688
	20	0.00341	0.005	0.020	0.168	0.020	0.827	0.020	0.203	0.028	0.004	0.017	0.000	0.088	0.209	0.2092	2.05E-08	1.70E-08	0.1730
	25	0.00335	-0.009	0.020	0.169	0.020	0.840	0.020	0.201	0.028	0.000	0.000	0.000	0.131	0.215	0.2153	2.46E-07	2.03E-07	0.1771
	30	0.00330	0.017	0.020	0.164	0.020	0.820	0.020	0.200	0.028	0.014	0.016	0.000	0.192	0.221	0.2214	2.72E-06	2.23E-06	0.1813
	33	0.00327	0.021	0.020	0.167	0.020	0.812	0.020	0.205	0.028	0.017	0.016	0.000	0.240	0.225	0.2251	1.11E-05	9.05E-06	0.1837
	35	0.00325	-0.004	0.020	0.173	0.020	0.831	0.020	0.208	0.028	0.000	0.000	0.000	0.278	0.227	0.2275	2.79E-05	2.27E-05	0.1853
	38	0.00321	-0.042	0.020	0.175	0.020	0.867	0.020	0.202	0.028	0.000	0.000	0.000	0.344	0.231	0.2314	1.08E-04	8.81E-05	0.1879
	40	0.00319	-0.004	0.020	0.186	0.020	0.818	0.020	0.228	0.028	0.000	0.000	0.001	0.397	0.233	0.2341	2.65E-04	2.14E-04	0.1897
	42	0.00317	-0.029	0.020	0.193	0.020	0.837	0.020	0.230	0.028	0.000	0.000	0.001	0.456	0.236	0.2372	6.38E-04	5.15E-04	0.1916
	44	0.00315	0.008	0.020	0.197	0.020	0.795	0.020	0.248	0.028	0.006	0.016	0.003	0.524	0.238	0.2411	1.52E-03	1.22E-03	0.1940
	46	0.00313	-0.017	0.020	0.201	0.020	0.816	0.020	0.246	0.028	0.000	0.000	0.006	0.600	0.241	0.2466	3.59E-03	2.87E-03	0.1972
	48	0.00311	-0.001	0.020	0.218	0.020	0.782	0.020	0.279	0.028	0.000	0.000	0.012	0.687	0.243	0.2552	0.008	0.0066	0.2020
	50	0.00309	-0.027	0.020	0.224	0.020	0.803	0.020	0.279	0.028	0.000	0.000	0.025	0.784	0.245	0.2700	0.019	0.0150	0.2094
	52	0.00308	0.043	0.020	0.246	0.020	0.711	0.020	0.346	0.028	0.031	0.014	0.049	0.894	0.248	0.2972	0.044	0.0330	0.2216
	54	0.00306	0.053	0.020	0.255	0.020	0.692	0.020	0.369	0.028	0.036	0.014	0.098	1.018	0.250	0.3485	0.100	0.0692	0.2406
	56	0.00304	0.127	0.020	0.290	0.020	0.583	0.020	0.498	0.028	0.074	0.012	0.194	1.157	0.252	0.4464	0.225	0.1344	0.2672
	58	0.00302	0.226	0.020	0.312	0.020	0.462	0.020	0.677	0.028	0.104	0.010	0.379	1.313	0.255	0.6341	0.498	0.2337	0.2974
	60	0.00300	0.320	0.020	0.331	0.020	0.349	0.020	0.951	0.028	0.112	0.009	0.736	1.488	0.257	0.9930	1.095	0.3546	0.3215
	63	0.00297	0.514	0.020	0.324	0.020	0.162	0.020	2.001	0.028	0.083	0.011	1.959	1.790	0.261	2.2194	3.507	0.5214	0.3300
	65	0.00296	0.691	0.020	0.284	0.020	0.025	0.020	11.587	0.028	0.017	0.014	3.726	2.021	0.263	3.9890	7.531	0.6015	0.3186
	68	0.00293	0.678	0.020	0.283	0.020	0.040	0.020	7.115	0.028	0.027	0.014	9.638	2.418	0.266	9.9048	23.309	0.6813	0.2895
	70	0.00291	0.666	0.020	0.259	0.020	0.075	0.020	3.456	0.028	0.050	0.013	17.996	2.721	0.269	18.2644	48.961	0.7176	0.2677
	75	0.00287	0.771	0.020	0.223	0.020	0.007	0.020	32.838	0.028	0.005	0.015	83.070	3.631	0.274	83.3449	301.631	0.7815	0.2159
	80	0.00283	0.822	0.020	0.168	0.020	0.010	0.020	17.609	0.028	0.008	0.016	367.210	4.807	0.280	367.4900	1764.995	0.8273	0.1722
	85	0.00279	0.849	0.020	0.133	0.020	0.019	0.020	7.172	0.028	0.016	0.017	1557.25	6.313	0.286	1557.5392	9830.755	0.8632	0.1368

+/-		SumSQ (all)	SumSQ (fF + fU)						Plot Ln(K)				
11		108.16628	108.1662822	<-- cell to minimize			t/C	1/T	Ln(k1)	Ln(k3)	Ln(K2)	Ln(K1/K3)	
89	deg freed	72					90	0.00275376	8.75564	-1.23323	2.107707	9.988866	
0.03	SSQ/df	1.502309475					5	0.00359531	-21.9984	-1.65736	-3.69557		
0.3	SSQ + 1SDfit	109.66859					37	0.00322	-15				
	Value only	109.66859		<-- used for error estimation			37	0.00322	5				
	Fit over								K1 (calc)	K3(calc)	K2(calc)		DH3-DH1 DS3-Ds1
	Temp range						37	0.00322	0.000216	0.229833	0.320636		-299657 -908.23
Tinit	10	1	1	1		V15	W15	X15	-8.44185	-1.4704	-1.13745	-6.97145	DHj DSj
Tfinal	85	51.55807644	11.85629377	44.751912		V40	W40	X40					-299657 -908.23
		0	0	0									
	whole range	0.00000	0.00000	0.00000									
flob(calc)	fF(calc)	Diff^2(fU)	Diff^2(flob)	Diff^2(fF)	fF+fl+fU	K1 (calc)	K3(calc)	K2(calc)	Ln(K1calc)	Ln(K3calc)	Ln(K2calc)		
0.1645	0.8355	1.15026	0.32225	2.69016	1.00	2.84E-09	1.97E-01	3.85E-02	-19.678	-1.625	-3.258		
0.1688	0.8312	3.87302	0.00205	4.05111	1.00	2.67E-08	2.03E-01	5.87E-02	-17.439	-1.594	-2.835		
0.1730	0.8270	0.06931	0.06739	0.00002	1.00	2.32E-07	2.09E-01	8.83E-02	-15.275	-1.565	-2.427		
0.1771	0.8229	0.19985	0.16552	0.72906	1.00	1.88E-06	2.15E-01	1.31E-01	-13.185	-1.536	-2.032		
0.1813	0.8187	0.69370	0.77754	0.00239	1.00	1.42E-05	2.21E-01	1.92E-01	-11.163	-1.508	-1.651		
0.1837	0.8163	1.11207	0.71537	0.04358	1.00	4.63E-05	2.25E-01	2.40E-01	-9.981	-1.492	-1.428		
0.1853	0.8146	0.03866	0.37434	0.65363	1.00	1.00E-04	2.27E-01	2.78E-01	-9.207	-1.481	-1.282		
0.1879	0.8120	4.45591	0.39551	7.50377	1.00	3.15E-04	2.31E-01	3.44E-01	-8.063	-1.465	-1.066		
0.1897	0.8101	0.03570	0.03234	0.13582	1.00	6.67E-04	2.33E-01	3.97E-01	-7.313	-1.455	-0.924		
0.1916	0.8078	2.24183	0.00186	2.11450	1.00	1.40E-03	2.36E-01	4.56E-01	-6.572	-1.445	-0.785		
0.1940	0.8047	0.10337	0.02913	0.24250	1.00	2.91E-03	2.38E-01	5.24E-01	-5.841	-1.435	-0.647		
0.1972	0.7999	0.98728	0.03486	0.65111	1.00	5.98E-03	2.41E-01	6.00E-01	-5.119	-1.425	-0.510		
0.2020	0.7914	0.13107	0.65181	0.19860	1.00	1.22E-02	2.43E-01	6.87E-01	-4.406	-1.415	-0.376		
0.2094	0.7755	4.38005	0.51511	1.89102	1.00	2.47E-02	2.45E-01	7.84E-01	-3.701	-1.405	-0.243		
0.2216	0.7454	0.24900	1.52237	3.00273	1.00	4.95E-02	2.48E-01	8.94E-01	-3.006	-1.395	-0.112		
0.2406	0.6903	0.67664	0.53970	0.00773	1.00	9.84E-02	2.50E-01	1.02E+00	-2.319	-1.386	0.018		
0.2672	0.5985	0.13718	1.34243	0.62135	1.00	1.94E-01	2.52E-01	1.16E+00	-1.640	-1.377	0.146		
0.2974	0.4690	0.15800	0.56812	0.12655	1.00	3.79E-01	2.55E-01	1.31E+00	-0.969	-1.367	0.273		
0.3215	0.3238	2.97727	0.23996	1.52676	1.00	7.36E-01	2.57E-01	1.49E+00	-0.307	-1.358	0.398		
0.3300	0.1487	0.13410	0.08797	0.43930	1.00	1.96E+00	2.61E-01	1.79E+00	0.672	-1.345	0.582		
0.3186	0.0799	20.17399	2.97223	7.65922	1.00	3.73E+00	2.63E-01	2.02E+00	1.315	-1.336	0.704		
0.2895	0.0292	0.03065	0.12146	0.27467	1.00	9.64E+00	2.66E-01	2.42E+00	2.266	-1.323	0.883		
0.2677	0.0147	6.66813	0.18694	9.08802	1.00	1.80E+01	2.69E-01	2.72E+00	2.890	-1.314	1.001		
0.2159	0.0026	0.29683	0.11247	0.04387	1.00	8.31E+01	2.74E-01	3.63E+00	4.420	-1.293	1.290		
0.1722	0.0005	0.07029	0.03608	0.20692	1.00	3.67E+02	2.80E-01	4.81E+00	5.906	-1.273	1.570		
0.1368	0.0001	0.51391	0.04151	0.84752	1.00	1.56E+03	2.86E-01	6.31E+00	7.351	-1.253	1.843		

Human Ap5

	Fitting Parameters	+/-	to minimize	Include sch 1
				1
$-\Delta H_1/R$	-69492	37	253.022809	SSQ (sum of squares)
$\Delta S_1/R$	204.54	0.11	72	deg freed
$-\Delta H_2/R$	-2655	10	3.5142057	SSQ/df = Sdfit
$\Delta S_2/R$	8.69	0.05	256.53702	SSQ + 15Dfit
-DH3/R	605	11	256.82571	Value only (SSQ + 15Dfit)
DS3/R	-3.19	0.03		
$-\Delta H_4/R$	70097	39		
$\Delta S_4/R$	-207.73	0.11		

<--Used for error calculation

Define K's as:
$K_1 = [I]/[F]$, $K_2 = [U]/[I]$, $K_3 = [J]/[F]$, $K_4 = [U]/[J]$, $K_J = [J]/[I]$
Assuming that *one cannot distinguish between I and J*,
ie. one sees only $[I_{Total}] = [I] + [J] = Sum(Intermediates)$
Assuming a closed system (ie. no other reactions), then only three equilibrium constants, K_1, K_2, K_3 for example, can define the system; with this particular choice, K_4 and K_J are linked to K_1, K_2 and K_3 and can be calculated by:
$K_4 = K_1K_2/K_3$ and $K_J = K_3/K_1 = K_2/K_4$

Enter temperature (to look up K aand fraction values)

Temp/C	37	+/-		0.00322435	<--1/T	Calculated fractions @37 C	
K1	3.31E-09	5.36E-10	1/fF =	1 + K1 + K3 + K1K2		fF =	0.77513
K2	1.136107	6.83E-02	1/fI =	1 + 1/K1 + K3/K1 + K2		fI =	0.00000
K3	0.290114	1.35E-02	1/fJ =	1 + 1/K3 + K1/K3 + K1K2/K3		fJ =	0.22487
KJ	8.78E+07	1.48E+07					0.2249 <-- f(I+J)
K4	1.29E-08	2.32E-09	1/fU =	1 + 1/K1K2 + 1/K2 + K3/K1K2		fU =	0.00000
						total =	1.000000

		+/-	
ΔH_1	577.79	0.31	kJ mol^{-1}
ΔS_1	1700.6	0.9	J mol^{-1} K^{-1}
ΔH_2	22.08	0.09	kJ mol^{-1}
ΔS_2	72.3	0.4	J mol^{-1} K^{-1}
ΔH_3	-5.03	0.09	kJ mol^{-1}
ΔS_3	-26.5	0.2	J mol^{-1} K^{-1}
ΔH_J	-582.8	0.3	kJ mol^{-1}
ΔS_J	-1727.1	0.9	J mol^{-1} K^{-1}
ΔH_4	604.9	0.3	kJ mol-1
ΔS_4	1799.4	1.0	J mol-1 K-1

F<-- K1 --> I <-- K2 --> U
F<-- K3 --> J <-- K4 --> U
I <-- Kj --> J

I(observed) = Iob = I + J

K1 = I/F
K2 = U/I
K3 = J/F

Then K1 + K3 = Iob/F
K1K2 = U/F

Fitting parameters are highlighted purple

		+/-			+/-
$-\Delta H_1/R$	-69492	17	$-\Delta H_2/R$	-2655	11
ΔH_1	577791	141	ΔH_2	22079	89
$\Delta S_1/R$	204.5	0.05	$\Delta S_2/R$	8.69	0.03
ΔS_1	1701	0.4	ΔS_2	72.3	0.3
$-\Delta H_3/R$	605	9	-(Delh1+Delh2)/R	-72147.83	
DH3	-5027	79	Delh1+DelH2	599870.39	
DS3/R	-3.19	0.03	(dels1+DelS2)/R	213.2	Tinit
DS3	-26.5	0.3	DelS1+DelS2	1773	Tfinal

1/fU = 1 + 1/(K1K2) + (K1 + K3)/(K1K2)
1/fIob = 1 + 1/(K1 + K3) + (K1K2)/(K1 + K3)
1/fF = 1 + (K1 + K3) + (K1K2)
fIob = 1/(1+K2+(1+K3)/K1) + 1/(1+1/K3+K1(1+K2)/K3)

			0.03		0.036		0.02	Not used										
			0.020		0.020		0.020											
temp/C	1/T (K^{-1})	fU/fract	+/-	fI/fract	+/-	fF/fract	+/-	K1 + K3	+/-	K1*K2	+/-	K1(calc)	K2(calc)	K3(calc)	[K1 + K3] (calc)	[K1*K2](calc)	fU(calc)	fIob(calc)
10	0.00353	0.132	0.020	0.208	0.020	0.660	0.020	0.315	0.028	0.087	0.013	0.000	0.502	0.349	0.3494	8.72E-19	6.46E-19	0.2589
15	0.00347	0.098	0.020	0.212	0.020	0.690	0.020	0.307	0.028	0.068	0.014	0.000	0.591	0.337	0.3367	7.26E-17	5.43E-17	0.2519
20	0.00341	0.074	0.020	0.213	0.020	0.713	0.020	0.298	0.028	0.053	0.014	0.000	0.691	0.325	0.3248	5.20E-15	3.92E-15	0.2452
25	0.00335	0.063	0.020	0.215	0.020	0.722	0.020	0.299	0.028	0.045	0.014	0.000	0.805	0.314	0.3138	3.22E-13	2.45E-13	0.2388
30	0.00330	0.038	0.020	0.217	0.020	0.745	0.020	0.291	0.028	0.028	0.015	0.000	0.932	0.303	0.3035	1.74E-11	1.34E-11	0.2328
33	0.00327	0.005	0.020	0.220	0.020	0.775	0.020	0.283	0.028	0.004	0.016	0.000	1.016	0.298	0.2976	1.80E-10	1.39E-10	0.2293
35	0.00325	-0.002	0.020	0.221	0.020	0.781	0.020	0.282	0.028	0.000	0.000	0.000	1.075	0.294	0.2938	8.30E-10	6.41E-10	0.2271
38	0.00321	-0.017	0.020	0.219	0.020	0.798	0.020	0.275	0.028	0.000	0.000	0.000	1.168	0.288	0.2883	7.93E-09	6.16E-09	0.2238
40	0.00319	-0.019	0.020	0.221	0.020	0.798	0.020	0.277	0.028	0.000	0.000	0.000	1.233	0.285	0.2847	3.49E-08	2.71E-08	0.2216
42	0.00317	-0.039	0.020	0.222	0.020	0.817	0.020	0.271	0.028	0.000	0.000	0.000	1.301	0.281	0.2813	1.51E-07	1.17E-07	0.2195
44	0.00315	-0.029	0.020	0.222	0.020	0.807	0.020	0.275	0.028	0.000	0.000	0.000	1.372	0.278	0.2779	6.38E-07	4.99E-07	0.2175
46	0.00313	-0.058	0.020	0.224	0.020	0.835	0.020	0.268	0.028	0.000	0.000	0.000	1.446	0.275	0.2746	2.65E-06	2.08E-06	0.2154
48	0.00311	-0.052	0.020	0.225	0.020	0.827	0.020	0.272	0.028	0.000	0.000	0.000	1.523	0.271	0.2714	0.000	0.0000	0.2135
50	0.00309	-0.051	0.020	0.226	0.020	0.825	0.020	0.274	0.028	0.000	0.000	0.000	1.603	0.268	0.2683	0.000	0.0000	0.2115
52	0.00308	-0.058	0.020	0.226	0.020	0.832	0.020	0.272	0.028	0.000	0.000	0.000	1.686	0.265	0.2653	0.000	0.0001	0.2096
54	0.00306	-0.047	0.020	0.228	0.020	0.819	0.020	0.278	0.028	0.000	0.000	0.000	1.773	0.262	0.2625	0.001	0.0005	0.2078
56	0.00304	-0.038	0.020	0.232	0.020	0.806	0.020	0.289	0.028	0.000	0.000	0.001	1.862	0.259	0.2606	0.003	0.0020	0.2063
58	0.00302	-0.002	0.020	0.235	0.020	0.767	0.020	0.306	0.028	0.000	0.000	0.005	1.955	0.256	0.2613	0.010	0.0075	0.2056
60	0.00300	0.049	0.020	0.241	0.020	0.710	0.020	0.339	0.028	0.035	0.014	0.017	2.052	0.254	0.2709	0.035	0.0271	0.2073
63	0.00297	0.113	0.020	0.255	0.020	0.632	0.020	0.404	0.028	0.071	0.013	0.111	2.203	0.250	0.3607	0.245	0.1526	0.2246
65	0.00296	0.345	0.020	0.270	0.020	0.385	0.020	0.703	0.028	0.133	0.010	0.378	2.309	0.247	0.6246	0.872	0.3493	0.2502
68	0.00293	0.615	0.020	0.271	0.020	0.114	0.020	2.370	0.028	0.070	0.013	2.302	2.474	0.243	2.5451	5.694	0.6163	0.2755
70	0.00291	0.711	0.020	0.258	0.020	0.031	0.020	8.222	0.028	0.022	0.014	7.547	2.588	0.241	7.7872	19.534	0.6897	0.2750
75	0.00287	0.739	0.020	0.248	0.020	0.013	0.020	19.408	0.028	0.009	0.015	138.337	2.893	0.235	138.5711	400.166	0.7414	0.2567
80	0.00283	0.765	0.020	0.238	0.020	-0.004	0.020	-66.004	0.028	-0.003	-0.015	2335.335	3.223	0.229	2335.5639	7525.839	0.7631	0.2368
85	0.00279	0.762	0.020	0.219	0.020	0.019	0.020	11.653	0.028	0.014	0.015	36432.54	3.579	0.223	36432.7585	130403.208	0.7816	0.2184

+/-		SumSQ (all)	SumSQ (fF + fU)						Plot Ln(K)					
11		253.02281	253.0228094	<-- cell to minimize			t/C	1/T	Ln(k1)	Ln(k3)	Ln(K2)	Ln(K1/K3)		
89	deg freed	72					90	0.00275376	13.17487	-1.52202	1.37726	14.69689		
0.03	SSQ/df	3.514205686					5	0.00359531	-45.3066	-1.01319	-0.85748			
0.3	SSQ + 1SDfit	256.53702					37	0.00322	-15					
	Value only	109.66859		<-- used for error estimation			37	0.00322	5					
	Fit over								K1 (calc)	K3(calc)	K2(calc)		DH3-DH1	DS3-Ds1
	Temp range						37	0.00322	3.31E-09	0.290114	1.136107		-582819	-1727.14
Tinit	10	1	1	1		V15	W15	X15	-19.5276	-1.23748	0.127607	-18.2901	DHj	DSj
Tfinal	85	150.5196271	29.48979611	73.0133862		V40	W40	X40					-590424	-1749.95
		0	0	0										
	whole range	0.00000	0.00000	0.00000										
flob(calc)	fF(calc)	Diff^2(fU)	Diff^2(flob)	Diff^2(fF)	fF+fl+fU	K1 (calc)	K3(calc)	K2(calc)	Ln(K1calc)	Ln(K3calc)	Ln(K2calc)			
0.2589	0.7411	43.81116	6.58505	16.42572	1.00	1.74E-18	3.49E-01	5.02E-01	-40.895	-1.052	-0.689			
0.2519	0.7481	24.16509	4.06549	8.40823	1.00	1.23E-16	3.37E-01	5.91E-01	-36.636	-1.089	-0.526			
0.2452	0.7548	13.87153	2.64092	4.40542	1.00	7.51E-15	3.25E-01	6.91E-01	-32.522	-1.124	-0.369			
0.2388	0.7612	9.85363	1.36732	3.87963	1.00	4.00E-13	3.14E-01	8.05E-01	-28.546	-1.159	-0.217			
0.2328	0.7672	3.61494	0.64308	1.20907	1.00	1.87E-11	3.03E-01	9.32E-01	-24.702	-1.192	-0.070			
0.2293	0.7707	0.06459	0.23940	0.05536	1.00	1.77E-10	2.98E-01	1.02E+00	-22.455	-1.212	0.016			
0.2271	0.7729	0.01000	0.10233	0.17673	1.00	7.72E-10	2.94E-01	1.07E+00	-20.982	-1.225	0.072			
0.2238	0.7762	0.74106	0.05072	1.17877	1.00	6.79E-09	2.88E-01	1.17E+00	-18.807	-1.244	0.155			
0.2216	0.7784	0.94770	0.00080	1.00357	1.00	2.83E-08	2.85E-01	1.23E+00	-17.381	-1.256	0.210			
0.2195	0.7805	3.73980	0.01222	3.32316	1.00	1.16E-07	2.81E-01	1.30E+00	-15.973	-1.268	0.263			
0.2175	0.7825	2.15877	0.05633	1.51826	1.00	4.65E-07	2.78E-01	1.37E+00	-14.582	-1.281	0.317			
0.2154	0.7846	8.48530	0.16740	6.26984	1.00	1.83E-06	2.75E-01	1.45E+00	-13.209	-1.292	0.369			
0.2135	0.7865	6.67693	0.32197	4.06628	1.00	7.12E-06	2.71E-01	1.52E+00	-11.853	-1.304	0.421			
0.2115	0.7885	6.58753	0.51275	3.42491	1.00	2.72E-05	2.68E-01	1.60E+00	-10.513	-1.316	0.472			
0.2096	0.7902	8.54917	0.68052	4.40520	1.00	1.02E-04	2.65E-01	1.69E+00	-9.190	-1.327	0.523			
0.2078	0.7916	5.72472	1.02968	1.89945	1.00	3.77E-04	2.62E-01	1.77E+00	-7.884	-1.339	0.573			
0.2063	0.7917	4.04547	1.71189	0.49364	1.00	1.37E-03	2.59E-01	1.86E+00	-6.593	-1.350	0.622			
0.2056	0.7869	0.20980	2.13304	1.00456	1.00	4.90E-03	2.56E-01	1.96E+00	-5.318	-1.361	0.671			
0.2073	0.7655	1.24069	2.77106	7.71874	1.00	1.73E-02	2.54E-01	2.05E+00	-4.058	-1.372	0.719			
0.2246	0.6228	3.92094	2.37410	0.19301	1.00	1.11E-01	2.50E-01	2.20E+00	-2.196	-1.388	0.790			
0.2502	0.4005	0.05268	1.02714	0.61460	1.00	3.78E-01	2.47E-01	2.31E+00	-0.973	-1.399	0.837			
0.2755	0.1082	0.00236	0.06119	0.08785	1.00	2.30E+00	2.43E-01	2.47E+00	0.834	-1.415	0.906			
0.2750	0.0353	1.10673	0.73024	0.03892	1.00	7.55E+00	2.41E-01	2.59E+00	2.021	-1.425	0.951			
0.2567	0.0019	0.00959	0.20063	0.29780	1.00	1.38E+02	2.35E-01	2.89E+00	4.930	-1.450	1.062			
0.2368	0.0001	0.01448	0.00419	0.03439	1.00	2.34E+03	2.29E-01	3.22E+00	7.756	-1.475	1.170			
0.2184	0.0000	0.91495	0.00034	0.88031	1.00	3.64E+04	2.23E-01	3.58E+00	10.503	-1.499	1.275			

Rhesus Ap5

	Fitting Parameters	+/-	to minimize	Include sch 1	1
$-\Delta H_1/R$	-35004	10	1730.937696	SSQ (sum of squares)	
$\Delta S_1/R$	107.02	0.03	72	deg freed	
$-\Delta H_2/R$	-5554	10	24.040813	SSQ/df = Sdfit	
$\Delta S_2/R$	17.79	0.03	1754.97850	SSQ + 1SDfit	<--Used for error calculation
-DH3/R	2254	13	1754.97850	Value only (SSQ + 1SDfit)	
DS3/R	-8.41	0.04			
$-\Delta H_4/R$	37258	16			
$\Delta S_4/R$	-115.44	0.05			

Enter temperature (to look up K aand fraction values)

Temp/C	37	+/-		0.00322435	<--1/T	Calculated fractions @37 C	
K1	2.91E-03	1.27E-04	1/fF =	1 + K1 + K3 + K1K2		fF =	0.75586
K2	0.88786	4.01E-02	1/fI =	1 + 1/K1 + K3/K1 + K2		fI =	0.00220
K3	0.317514	1.83E-02	1/fJ =	1 + 1/K3 + K1/K3 + K1K2/K3		fJ =	0.23999
Kj	1.09E+02	7.90E+00		1/fU =	1 + 1/K1K2 + 1/K2 + K3/K1K2	fU =	0.00195
K4	0.00813	6.93E-04				total =	1.000000

		+/-	
ΔH_1	291.04	0.08	kJ mol^{-1}
ΔS_1	889.9	0.3	J mol^{-1} K^{-1}
ΔH_2	46.18	0.09	kJ mol^{-1}
ΔS_2	147.9	0.3	J mol^{-1} K^{-1}
ΔH_3	-18.74	0.11	kJ mol^{-1}
ΔS_3	-70.0	0.3	J mol^{-1} K^{-1}
ΔH_j	-309.8	0.1	kJ mol^{-1}
ΔS_j	-959.8	0.4	J mol^{-1} K^{-1}
ΔH_4	356.0	0.2	kJ mol-1
ΔS_4	1107.7	0.5	J mol-1 K-1

F<-- K1 --> I <-- K2 --> U

F<-- K3 --> J <-- K4 --> U

I <-- Kj --> J

K1 = I/F

K2 = U/I

K3 = J/F

Then

K1 + K3 = Iob/F

K1K2 = U/F

I(observed) = Iob = I + J

Fitting parameters are highlighted purple

1/fU = 1 + 1/(K1K2) + (K1 + K3)/(K1K2)

1/fIob = 1 + 1/(K1 + K3) + (K1K2)/(K1 + K3)

1/fF = 1 + (K1 + K3) + (K1K2)

fIob = 1/(1+K2+(1+K3)/K1) + 1/(1+1/K3+K1(1+K2)/K3)

													+/-						+/-
										$-\Delta H_1/R$	-35004	17					$-\Delta H_2/R$	-5554	11
										ΔH_1	291040	141					ΔH_2	46176	89
										$\Delta S_1/R$	107.0	0.05					$\Delta S_2/R$	17.79	0.03
										ΔS_1	890	0.4					ΔS_2	147.9	0.3
										$-\Delta H_i/R$	2254	9					-(Delh1+Delh2)/R	-40557.74	
										DH3	-18738	79					Delh1+DelH2	337215.76	
										DS3/R	-8.41	0.03					(dels1+DelS2)/R	124.8	Tinit
										DS3	-70.0	0.3					DelS1+DelS2	1038	Tfinal
				0.03		0.036		0.02											
				0.020		0.020		0.020		Not used									
	temp/C	1/T (K^{-1})	fU/fract	+/-	fI/fract	+/-	fF/fract	+/-	K1 + K3	+/-	K1*K2	+/-	K1(calc)	K2(calc)	K3(calc)	[K1 + K3] (calc)	[K1*K2](calc)	fU(calc)	flob(calc)
	10	0.00353	0.386	0.020	0.215	0.020	0.398	0.020	0.541	0.028	0.154	0.011	0.000	0.161	0.635	0.6349	9.91E-09	6.06E-09	0.3883
	15	0.00347	0.313	0.020	0.211	0.020	0.477	0.020	0.442	0.028	0.149	0.011	0.000	0.226	0.553	0.5530	1.19E-07	7.66E-08	0.3561
	20	0.00341	0.236	0.020	0.209	0.020	0.555	0.020	0.376	0.028	0.131	0.012	0.000	0.314	0.484	0.4839	1.31E-06	8.85E-07	0.3261
	25	0.00335	0.147	0.020	0.207	0.020	0.646	0.020	0.321	0.028	0.095	0.013	0.000	0.432	0.425	0.4254	1.34E-05	9.37E-06	0.2984
	30	0.00330	0.043	0.020	0.212	0.020	0.745	0.020	0.285	0.028	0.032	0.015	0.000	0.587	0.376	0.3757	1.26E-04	9.16E-05	0.2731
	33	0.00327	0.041	0.020	0.214	0.020	0.746	0.020	0.286	0.028	0.030	0.015	0.001	0.703	0.349	0.3498	4.67E-04	3.46E-04	0.2591
	35	0.00325	0.057	0.020	0.217	0.020	0.727	0.020	0.298	0.028	0.041	0.015	0.001	0.790	0.333	0.3342	1.10E-03	8.27E-04	0.2503
	38	0.00321	0.043	0.020	0.219	0.020	0.738	0.020	0.297	0.028	0.032	0.015	0.004	0.940	0.310	0.3144	3.93E-03	2.98E-03	0.2385
	40	0.00319	0.039	0.020	0.223	0.020	0.738	0.020	0.302	0.028	0.029	0.015	0.009	1.054	0.296	0.3047	9.04E-03	6.88E-03	0.2320
	42	0.00317	0.009	0.020	0.231	0.020	0.760	0.020	0.304	0.028	0.007	0.015	0.017	1.180	0.283	0.3004	2.06E-02	1.56E-02	0.2274
	44	0.00315	0.046	0.020	0.234	0.020	0.721	0.020	0.324	0.028	0.033	0.014	0.035	1.318	0.270	0.3056	4.63E-02	3.42E-02	0.2260
	46	0.00313	0.154	0.020	0.239	0.020	0.607	0.020	0.394	0.028	0.093	0.013	0.070	1.471	0.259	0.3288	1.03E-01	7.20E-02	0.2296
	48	0.00311	0.136	0.020	0.249	0.020	0.615	0.020	0.405	0.028	0.084	0.013	0.139	1.640	0.248	0.3864	0.228	0.1410	0.2394
	50	0.00309	0.231	0.020	0.256	0.020	0.513	0.020	0.499	0.028	0.118	0.011	0.273	1.825	0.237	0.5096	0.497	0.2478	0.2539
	52	0.00308	0.341	0.020	0.264	0.020	0.395	0.020	0.670	0.028	0.135	0.010	0.531	2.028	0.227	0.7578	1.076	0.3798	0.2674
	54	0.00306	0.474	0.020	0.271	0.020	0.255	0.020	1.063	0.028	0.121	0.011	1.025	2.252	0.218	1.2425	2.308	0.5071	0.2731
	56	0.00304	0.642	0.020	0.277	0.020	0.081	0.020	3.442	0.028	0.052	0.013	1.964	2.496	0.209	2.1723	4.901	0.6071	0.2691
	58	0.00302	0.658	0.020	0.284	0.020	0.058	0.020	4.934	0.028	0.038	0.013	3.733	2.764	0.200	3.9328	10.317	0.6765	0.2579
	60	0.00300	0.751	0.020	0.276	0.020	-0.027	0.020	-10.289	0.028	-0.020	-0.015	7.041	3.057	0.192	7.2329	21.522	0.7233	0.2431
	63	0.00297	0.874	0.020	0.257	0.020	-0.131	0.020	-1.966	0.028	-0.114	-0.018	17.984	3.547	0.181	18.1647	63.791	0.7690	0.2190
	65	0.00296	0.856	0.020	0.253	0.020	-0.108	0.020	-2.333	0.028	-0.093	-0.017	33.295	3.911	0.174	33.4687	130.224	0.7907	0.2032
	68	0.00293	0.922	0.020	0.233	0.020	-0.155	0.020	-1.501	0.028	-0.143	-0.019	82.744	4.519	0.164	82.9083	373.922	0.8167	0.1811
	70	0.00291	0.922	0.020	0.228	0.020	-0.150	0.020	-1.516	0.028	-0.139	-0.019	150.478	4.969	0.158	150.6355	747.694	0.8314	0.1675
	75	0.00287	0.896	0.020	0.203	0.020	-0.099	0.020	-2.058	0.028	-0.088	-0.018	651.255	6.269	0.144	651.3990	4082.755	0.8622	0.1376
	80	0.00283	0.899	0.020	0.167	0.020	-0.067	0.020	-2.507	0.028	-0.060	-0.018	2704.038	7.858	0.131	2704.1691	21247.433	0.8871	0.1129
	85	0.00279	0.912	0.020	0.143	0.020	-0.055	0.020	-2.597	0.028	-0.050	-0.018	10789.75	9.787	0.120	10789.8650	105598.461	0.9073	0.0927

+/-		SumSQ (all)	SumSQ (fF + fU)						Plot Ln(K)						
11		1730.93770	1730.9376962	<-- cell to minimize			t/C	1/T	Ln(k1)	Ln(k3)	Ln(K2)	Ln(K1/K3)			
89	deg freed	72					90	0.00275376	10.63209	-2.20778	2.494559	12.83987			
0.03	SSQ/df	24.04080134					5	0.00359531	-18.8257	-0.31122	-2.17913				
0.3	SSQ + 1SDfit	1754.97850					37	0.00322	-15						
	Value only	109.66859		<-- used for error estimation			37	0.00322	5						
	Fit over								K1 (calc)	K3(calc)	K2(calc)		Kj(calc)	DH3-DH1	DS3-Ds1
	Temp range						37	0.00322	0.002907	0.317514	0.88786			-309778	-959.81
Tinit	10	1	1	1		V15	W15	X15	-5.84053	-1.14723	-0.11894	-4.69329		DHj	DSj
Tfinal	85	953.1292384	256.6283021	521.180156		V40	W40	X40						-303483	-919.68
		0	0	0											
	whole range	0.00000	0.00000	0.00000											
flob(calc)	fF(calc)	Diff^2(fU)	Diff^2(flob)	Diff^2(fF)	fF+fl+fU	K1 (calc)	K3(calc)	K2(calc)	Ln(K1calc)	Ln(K3calc)	Ln(K2calc)				
0.3883	0.6117	373.04990	74.78629	113.77621	1.00	6.16E-08	6.35E-01	1.61E-01	-16.603	-0.454	-1.827				
0.3561	0.6439	244.24989	52.81632	69.90652	1.00	5.26E-07	5.53E-01	2.26E-01	-14.458	-0.592	-1.486				
0.3261	0.6739	139.13278	34.49953	35.06210	1.00	4.18E-06	4.84E-01	3.14E-01	-12.386	-0.726	-1.157				
0.2984	0.7015	53.88340	20.73794	7.76535	1.00	3.09E-05	4.25E-01	4.32E-01	-10.383	-0.855	-0.840				
0.2731	0.7268	4.54743	9.19961	0.81120	1.00	2.15E-04	3.76E-01	5.87E-01	-8.447	-0.979	-0.532				
0.2591	0.7406	4.03885	5.15771	0.06831	1.00	6.65E-04	3.49E-01	7.03E-01	-7.315	-1.052	-0.353				
0.2503	0.7489	7.76787	2.82834	1.22185	1.00	1.40E-03	3.33E-01	7.90E-01	-6.573	-1.100	-0.235				
0.2385	0.7586	3.93805	0.95181	1.01777	1.00	4.18E-03	3.10E-01	9.40E-01	-5.478	-1.171	-0.061				
0.2320	0.7612	2.55738	0.20720	1.30975	1.00	8.57E-03	2.96E-01	1.05E+00	-4.759	-1.217	0.053				
0.2274	0.7570	0.09858	0.03628	0.01527	1.00	1.74E-02	2.83E-01	1.18E+00	-4.050	-1.263	0.165				
0.2260	0.7397	0.32580	0.14295	0.90028	1.00	3.51E-02	2.70E-01	1.32E+00	-3.349	-1.308	0.276				
0.2296	0.6984	16.76745	0.23208	20.94490	1.00	7.01E-02	2.59E-01	1.47E+00	-2.658	-1.352	0.386				
0.2394	0.6196	0.06820	0.23028	0.04785	1.00	1.39E-01	2.48E-01	1.64E+00	-1.975	-1.396	0.494				
0.2539	0.4983	0.71953	0.01059	0.55551	1.00	2.73E-01	2.37E-01	1.82E+00	-1.300	-1.440	0.601				
0.2674	0.3528	3.78796	0.02310	4.40274	1.00	5.31E-01	2.27E-01	2.03E+00	-0.634	-1.482	0.707				
0.2731	0.2198	2.76771	0.01031	3.11592	1.00	1.02E+00	2.18E-01	2.25E+00	0.025	-1.525	0.812				
0.2691	0.1239	3.03071	0.17771	4.67640	1.00	1.96E+00	2.09E-01	2.50E+00	0.675	-1.567	0.915				
0.2579	0.0656	0.83949	1.72933	0.15889	1.00	3.73E+00	2.00E-01	2.76E+00	1.317	-1.608	1.017				
0.2431	0.0336	1.87348	2.73460	9.13528	1.00	7.04E+00	1.92E-01	3.06E+00	1.952	-1.649	1.117				
0.2190	0.0121	27.49170	3.58149	50.91871	1.00	1.80E+01	1.81E-01	3.55E+00	2.889	-1.709	1.266				
0.2032	0.0061	10.51730	6.15068	32.75382	1.00	3.33E+01	1.74E-01	3.91E+00	3.505	-1.749	1.364				
0.1811	0.0022	27.75913	6.83047	62.13713	1.00	8.27E+01	1.64E-01	4.52E+00	4.416	-1.808	1.508				
0.1675	0.0011	20.70230	9.12112	57.30637	1.00	1.50E+02	1.58E-01	4.97E+00	5.014	-1.846	1.603				
0.1376	0.0002	2.77352	10.75952	24.45905	1.00	6.51E+02	1.44E-01	6.27E+00	6.479	-1.940	1.836				
0.1129	0.0000	0.37874	7.42926	11.16353	1.00	2.70E+03	1.31E-01	7.86E+00	7.903	-2.032	2.061				
0.0927	0.0000	0.06209	6.24377	7.54947	1.00	1.08E+04	1.20E-01	9.79E+00	9.286	-2.121	2.281				

Chapter 4: Conclusions and Future Work

4.1 Functional Intermediates in Non-Coding RNAs

Modifications to non-coding RNAs are one of the many avenues an organism can take to achieve functional diversity. Such diversity may not be implemented in the final, or native state of a folding pathway, as homologous structures have been reported for entities with divergent functions.[1,2] Thus, the sequence-dependent folding landscapes traversed by RNAs en route to their native state are likely signatures of function.

The native state is not the sole functional entity of a ncRNA. Several investigations have unveiled functional intermediates and native like conformers, which perform duties in addition to those of the native state. In a study of the sarcin-ricin loop (SRL) of the large subunit rRNA in rat, electromobility shift assays (EMSA) of a 27 nt version of the SRC displayed two distinct species. Fascinatingly, the slower migrating population (30%) was resistant to the endoribonuclease restrictocin,[3] suggesting that alternative functional folds may be an adaptation to environmental pressures.

Intermediates, or native-like conformers may also self-guide, or provide scaffolds for chaperones, en route to the native state. In yeast, the addition of the CBP2 protein to b15 RNA intermediates yields a markedly higher amount of

active complex as compared to when CBP2 is added to unfolded b15RNA.[4] Furthermore, the aI5γ RNA exists in a collapsed state at physiological Mg^{2+} levels, but can be driven toward the native state in the presence of proteins.[5]

These findings further warrant our investigations on ncRNA folding pathways, with an emphasis on characterizing the intermediate(s). Our studies, have, and will not only broaden our understanding of ncRNA folding and function, but can also contribute to the field of drug design, as characterizing a multitude of structures broadens the number of drug targets.

4.2 Purine Riboswitches: Closing Remarks and Future Experiments

2Ap probing of WT and mutant purine riboswitch aptamer domains (GRNA, ARNA), has allowed us to identify unique sequence-dependent folding pathways. The primary sequence of the aptamer domains is the root of the signal transmitted to the downstream expression platform, resulting in downstream gene activation and repression of ligand-bound ARNA and GRNA, respectively.

Several questions remain unanswered in the general riboswitch field, and are worthy of future experimentation. Regarding the purine riboswitches, investigations of the full-length forms (aptamer and expression platform) are scant. Specifically, there is a lack of information regarding the *entire* landscape: from ligand-bound aptamer folding to signalling towards the downstream

expression casette. To address this, rather than using 2Ap as a cognate ligand mimic, it could be internally substituted at different positions along the full-length riboswitch. For instance, binding experiments with GRNA and guanine, with 2Ap substituted at positions of choice along the entire riboswitch would provide an extended, high-resolution folding pathway that would more accurately represent the events *in vivo*.

4.2.1 Prospective Studies of Putative Riboswitches

Bioinformatics-based identification has demonstrated the utility and abundance of riboswitches across prokaryotes and yeast.[6,7] Not only has bioinformatics allowed us to identify riboswitches in organisms by their sequence homology to already classified ones, but novel riboswitches have also been proposed.[8] Consequently, the identity of putative riboswitches provides opportunities for researchers to identify their cognate metabolite. Oftentimes, downstream genes that these riboswitches regulate may be undefined, or may not be directly related to the metabolite. To approach such a problem one can exploit 2Ap fluorescence changes with respect to potential aptamer interactions with candidate ligands. SInce the ligand/metabolite induces a packing of the aptamer domain, change in fluorescence intensity of an internally substituted 2Ap can probe binding affinity, and identify of the cognate binding partners that would exist *in vivo*.

Such an approach can be extended to eukaryotes as well, however discovery of eukaryotic riboswitches has been futile. For instance, the TPP

riboswitch is found across all three domains of life, yet is the only one that exists in eukaryotes.[9] One could argue that riboswitches are vestiges of an RNA-based world, and with the evolution of more complex organisms a more intricate protein-based system evolved to regulate genes. Regarding the TPP riboswitch in yeast, it is well established that it mediates splicing events. However, as eukaryotes evolved into more complex beings, splicing likely became a more tightly regulated function that may have required a more advanced means of regulation, namely the eukaryotic spliceosome.

The search for eukaryotic riboswitches should not be discarded, but rather the perspective could be altered. Firstly, eukaryotic genomes (and transcriptomes) are generally larger than prokaryotes, and consequently less sequence information is available. Thus, riboswitches may exist in eukaryotes, but have not yet been unveiled. Redefining the riboswitch may be worthwhile as well; it is well established that eukaryotes possess an increased amount of ncRNAs, specifically introns, which often house highly structured RNAs. Such entities may not necessarily bind small molecules as in prokaryotes, but perhaps interact with secondary messengers and/or compounds.[10] Consequently, and pending future discoveries, the definition of a riboswitch may be extended to include compounds/secondary messengers. Since riboswitch-based bioinformatics approaches utilize pre-established sequences to search for homologs, this may explain the lack of eukaryotic riboswitches.

4.3 A-to-I editing: Closing Remarks and Future Experiments

Our studies of primate *Nup50 Alu* dsRNA constructs have addressed two important issues regarding ncRNA folding. First, we have validated that RNA hairpin folding is not governed by a simple two-state model, but rather can be described by a multi-state pathway along a rugged free-energy landscape. Secondly, regarding A-to-I editing in primates, our findings have provided a glimpse of the events occurring *in vivo.* The variability in thermostability found across the human, chimp and rhesus constructs demonstrates the potency of a single nucleotide mismatch with respect to T_M. Using TRF, our investigations suggest a more prominent and thermostable intermediate, both requirements for high, ADAR-mediated editing efficiency. Our results do only report on one of many edit sites, however. Thus, studies with 2Ap placed at different positions along the duplexes would complement our findings.

4.3.1 Prospective 2-Aminopurine-Based Studies of A-to-I Editing Constructs

To predict A-to-I edit sites, a software tool (Inosine predict- http://bioserver.hci.utah.edu:8080/Bass/InosinePredict) has been developed by the Bass group. Here, factors such as enzyme (ADAR1, ADAR2) RNA duplex length, neighbour identity, and bp partners are taken into account to predict the probability of an A-to-I edit.[11] To validate this algorithm one can place 2Ap at putative edit sites and monitor its fluorescence intensity, either with the enzyme present, or without the enzyme, but subject to folding/unfolding conditions. In the

presence of ADAR, we would expect to see a relative increase in fluorescence upon base flipping, whereas changes in fluorescence intensity and populations and stability of intermediate states could be determined using TRF.

Another interesting study would involve synthesizing an extremely stable duplex that would hopefully lack an intermediate at the putative edit site. Presumably, this would not accommodate ADAR well as it would challenge the base-flipping step. Treatment of this model duplex with ADAR followed by analysis of editing levels via inosine cleavage could validate the algorithm,[12] as well as a base-flipping model.

The big picture issue of whether A-to-I editing is in fact a cause, or effect of our higher cognition amongst other primates will likely not be answered in the near future. However, the above studies, lay the groundwork for addressing such a question, and are a testament to the importance of ncRNA folding.

4.4 References

1. Prychyna, O.; Dahabieh, M. S.; Chao, J.; O'Neill, M. A. *Biopolymers* **2009**, *91*, 953.

2. Serganov, A.; Yuan, Y. R.; Pikovskaya, O.; Polonskaia, A.; Malinina, L.; Phan, A. T.; Hobartner, C.; Micura, R.; Breaker, R. R.; Patel, D. J. *Chem Biol* **2004**, *11*, 1729.

3. Korennykh, A. V.; Plantinga, M. J.; Correll, C. C.; Piccirilli, J. A. *Biochemistry* **2007**, *46*, 12744.

4. Webb, A. E.; Weeks, K. M. *Nat Struct Biol* **2001**, *8*, 135.

5. Bokinsky, G.; Nivon, L. G.; Liu, S.; Chai, G.; Hong, M.; Weeks, K. M.; Zhuang, X. *J Mol Biol* **2006**, *361*, 771.

6. Chang, T. H.; Huang, H. D.; Wu, L. C.; Yeh, C. T.; Liu, B. J.; Horng, J. T. *RNA* **2009**, *15*, 1426.

7. Abreu-Goodger, C.; Merino, E. *Nucleic Acids Res* **2005**, *33*, W690.

8. Serganov, A.; Patel, D. J. *Nat Rev Genet* **2007**, *8*, 776.

9. Cheah, M. T.; Wachter, A.; Sudarsan, N.; Breaker, R. R. *Nature* **2007**, *447*, 497.

10. Sudarsan, N.; Lee, E. R.; Weinberg, Z.; Moy, R. H.; Kim, J. N.; Link, K. H.; Breaker, R. R. *Science* **2008**, *321*, 411.

11. Eggington, J. M.; Greene, T.; Bass, B. L. *Nat Commun* **2011**, *2*, 319.

12. Sakurai, M.; Yano, T.; Kawabata, H.; Ueda, H.; Suzuki, T. *Nat Chem Biol* **2010**, *6*, 733.