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

© Michael Samir Dahabieh 2012

SIMON FRASER UNIVERSITY

Spring 2012

All rights reserved.
However, in accordance with the Copyright Act of Canada, this work may be reproduced, without authorization, under the conditions for “Fair Dealing.” Therefore, limited reproduction of this work for the purposes of private study, research, criticism, review and news reporting is likely to be in accordance with the law, particularly if cited appropriately.
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

The author, whose copyright is declared on the title page of this work, has granted to Simon Fraser University the right to lend this thesis, project or extended essay to users of the Simon Fraser University Library, and to make partial or single copies only for such users or in response to a request from the library of any other university, or other educational institution, on its own behalf or for one of its users.

The author has further granted permission to Simon Fraser University to keep or make a digital copy for use in its circulating collection (currently available to the public at the “Institutional Repository” link of the SFU Library website <www.lib.sfu.ca> at: <http://ir.lib.sfu.ca/handle/1892/112>) and, without changing the content, to translate the thesis/project or extended essays, if technically possible, to any medium or format for the purpose of preservation of the digital work.

The author has further agreed that permission for multiple copying of this work for scholarly purposes may be granted by either the author or the Dean of Graduate Studies.

It is understood that copying or publication of this work for financial gain shall not be allowed without the author's written permission.

Permission for public performance, or limited permission for private scholarly use, of any multimedia materials forming part of this work, may have been granted by the author. This information may be found on the separately catalogued multimedia material and in the signed Partial Copyright Licence.

While licensing SFU to permit the above uses, the author retains copyright in the thesis, project or extended essays, including the right to change the work for subsequent purposes, including editing and publishing the work in whole or in part, and licensing other parties, as the author may desire.

The original Partial Copyright Licence attesting to these terms, and signed by this author, may be found in the original bound copy of this work, retained in the Simon Fraser University Archive.

Simon Fraser University Library
Burnaby, BC, Canada
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.

# Table of Contents

Approval................................................................................................................................. ii
Abstract................................................................................................................................. iii
Dedication............................................................................................................................... iv
Acknowledgements............................................................................................................... v
Table of Contents....................................................................................................................... vi
List of Tables........................................................................................................................... ix
List of Schemes........................................................................................................................ x
List of Figures.......................................................................................................................... xi
Glossary................................................................................................................................. xv

## CHAPTER 1: Introduction ........................................................................................................ 1
  1.1 History and Utility of 2-Aminopurine............................................................................... 1
    1.1.1 2-Aminopurine Fluorescent Properties...................................................................... 1
    1.1.2 Early Cellular Investigations ..................................................................................... 2
    1.1.3 Birth of Fluorescence-Based Investigations............................................................... 4
  1.2 RNA Hairpins.................................................................................................................... 8
    1.2.1.1 Probing RNA Hairpin Folding with 2-Aminopurine................................................. 11
  1.3 Riboswitches.................................................................................................................... 14
    1.3.1 Modes of Riboswitch Regulation: Transcription, Translation, and Splicing.............. 15
    1.3.2 Probing Riboswitches with 2-Aminopurine................................................................. 17
      1.3.2.1 Probing ARNA Structure and Function with 2-Aminopurine............................ 19
      1.3.2.2 Probing GRNA Structure and Function with 2-Aminopurine............................ 20
      1.3.2.3 Probing Ligand-Bound Dynamics of Purine Riboswitches with 2-Aminopurine.... 21
  1.4 RNA Editing..................................................................................................................... 24
    1.4.1 Adenosine- to-Inosine Editing of Primate Alu RNA Duplexes.................................... 25
      1.4.1.1 Probing A- to-I Editing with 2-Aminopurine......................................................... 27
  1.5 Thesis Overview............................................................................................................... 28
  1.6 References....................................................................................................................... 30

## CHAPTER 2: Sequence-Dependent Structural Dynamics of Purine Riboswitches ............... 35
  2.1 Keywords........................................................................................................................ 37
  2.2 Abstract.......................................................................................................................... 37
  2.3 Introduction.................................................................................................................... 38
  2.4 Results............................................................................................................................ 43
    2.4.1 Probing Liganded-RNA Conformer Equilibria in Ensemble Solution with Fluorescence Lifetimes......................................................................................... 43
    2.4.2 Ligand Binding Affinities of Purine Riboswitch Conformers and Mutants................ 48
    2.4.3 Mg\(^{2+}\)-Dependence of Ligand-Mediated Folding of Purine Riboswitch Conformers......................................................................................................................... 52
  2.5 Discussion....................................................................................................................... 55
    2.5.1 Partially Quenched Species as Ap-Bound RNA Folding Intermediates.................... 55
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chapter 2: Model for Conformer Structure and Dynamics</td>
<td>61</td>
</tr>
<tr>
<td>2.5.2</td>
<td>61</td>
</tr>
<tr>
<td>2.5.3</td>
<td>64</td>
</tr>
<tr>
<td>2.5.4</td>
<td>68</td>
</tr>
<tr>
<td>Chapter 2: Materials and Methods</td>
<td>69</td>
</tr>
<tr>
<td>2.6.1</td>
<td>69</td>
</tr>
<tr>
<td>2.6.2</td>
<td>70</td>
</tr>
<tr>
<td>Chapter 2: Acknowledgements</td>
<td>73</td>
</tr>
<tr>
<td>Chapter 2: References</td>
<td>74</td>
</tr>
<tr>
<td>Chapter 2: Supporting Information</td>
<td>78</td>
</tr>
<tr>
<td>Chapter 3: Sequence-Dependent Structural Dynamics of Primate A-to-I Editing Substrates</td>
<td>81</td>
</tr>
<tr>
<td>3.1</td>
<td>81</td>
</tr>
<tr>
<td>3.1.1</td>
<td>82</td>
</tr>
<tr>
<td>3.1.2</td>
<td>85</td>
</tr>
<tr>
<td>3.1.3</td>
<td>87</td>
</tr>
<tr>
<td>3.2</td>
<td>89</td>
</tr>
<tr>
<td>3.2.1</td>
<td>89</td>
</tr>
<tr>
<td>3.2.2</td>
<td>90</td>
</tr>
<tr>
<td>3.2.3</td>
<td>90</td>
</tr>
<tr>
<td>3.2.4</td>
<td>92</td>
</tr>
<tr>
<td>3.2.4.1</td>
<td>95</td>
</tr>
<tr>
<td>3.3</td>
<td>100</td>
</tr>
<tr>
<td>3.3.1</td>
<td>100</td>
</tr>
<tr>
<td>3.3.3</td>
<td>102</td>
</tr>
<tr>
<td>3.3.4</td>
<td>103</td>
</tr>
<tr>
<td>3.3.4.1</td>
<td>103</td>
</tr>
<tr>
<td>3.3.5</td>
<td>106</td>
</tr>
<tr>
<td>3.4</td>
<td>108</td>
</tr>
<tr>
<td>3.4.1</td>
<td>108</td>
</tr>
<tr>
<td>3.4.1.1</td>
<td>108</td>
</tr>
<tr>
<td>3.4.1.2</td>
<td>109</td>
</tr>
<tr>
<td>3.4.2</td>
<td>110</td>
</tr>
<tr>
<td>3.4.2.1</td>
<td>110</td>
</tr>
<tr>
<td>3.4.2.2</td>
<td>110</td>
</tr>
<tr>
<td>3.4.3</td>
<td>111</td>
</tr>
<tr>
<td>3.4.4</td>
<td>112</td>
</tr>
<tr>
<td>3.4.4.1</td>
<td>112</td>
</tr>
<tr>
<td>3.4.5</td>
<td>113</td>
</tr>
<tr>
<td>3.4.6</td>
<td>114</td>
</tr>
<tr>
<td>3.5</td>
<td>115</td>
</tr>
<tr>
<td>Appendix: Chapter 3 Thermodynamic Parameters</td>
<td>118</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>4.1 Functional Intermediates in Non-Coding RNAs</td>
<td>127</td>
</tr>
<tr>
<td>4.2 Purine Riboswitches: Closing Remarks and Future Experiments</td>
<td>128</td>
</tr>
<tr>
<td>4.2.1 Prospective Studies of Putative Riboswitches</td>
<td>129</td>
</tr>
<tr>
<td>4.3 A-to-I editing: Closing Remarks and Future Experiments</td>
<td>131</td>
</tr>
<tr>
<td>4.3.1 Prospective 2-Aminopurine-Based Studies of A-to-I Editing Constructs</td>
<td>131</td>
</tr>
<tr>
<td>4.4 References</td>
<td>132</td>
</tr>
</tbody>
</table>
List of Tables

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 $G_A$RNA ..........................................................51

Table 2-2. Parameters from a representative experiment monitoring the fluorescence lifetime of 500 nM Ap as a function of $G_A$RNA 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 $\alpha_u$, $\alpha_o$, and $\alpha_i$ are the amplitudes, and $\tau_u$, $\tau_o$, and $\tau_i$ 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). ........................................................................................................79

Table 2-3. Parameters from a representative experiment monitoring the fluorescence lifetime of 500 nM Ap as a function of $Mg^{2+}$ concentration (~ 3 $\mu$M $G_A$RNA, 20°C). Fluorescence decay fit to: $I(t) = \alpha_u \tau_u + \alpha_o \tau_o + \alpha_i \tau_i$ where $\alpha_u$, $\alpha_o$, and $\alpha_i$ are the amplitudes, and $\tau_u$, $\tau_o$, and $\tau_i$ 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). ......................................................................................................80

Table 3-1: Thermodynamic parameters for primate constructs from multi-state model proposed in scheme 3-1 (shown again below).................................101
List of Schemes

**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. .................................................................47

**Scheme 3-1:** Proposed folding/unfolding pathway for ChAp5 construct ........99

**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.................................................................105
List of Figures

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

Figure 1-2. Modified form of a Jablonski diagram. S₀, S₁, and S₂ represent ground, first singlet, and second singlet states, respectively. Numbers 0, 1, and 2, adjacent to S₀ represent vibrational levels; they exist in other states, but are excluded for clarity. The dashed line from S₂ to S₁ represents an internal conversion. hνₐ, hν₇ and hνₚ represent energy of absorbance, fluorescence, and phosphorescence emission, respectively.....................................................3

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. Copyright © 2007 American Chemical Society. All rights reserved.....................................5

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.........................................................8

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).......................................................................................................................9

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)..................................................................................11

Figure 1-7. Cartoon of ARNA (pink) and G₃RNA (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 G₃RNA (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..............................................................18
Figure 1-8. RNA titration and temperature variation curves of 2Ap fluorescent lifetime fractions (C₁ = open conformer, C₂ = intermediate, C₃ = closed) bound to *Bacillus subtilis* ARNA aptamer domain. C₃ 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 Tₘ 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.

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.

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 are noted.

Figure 2-2: Fluorescence decay profile of free Ap, and Ap in the presence of either ARNA or GₐRNA (500 nM Ap, 4 µM RNA, 20°C, 10 mM Mg²⁺) 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 a linear scale.

Figure 2-3: Ligand binding to wild-type and mutant purine riboswitches. Top panels show the variation in the fraction of Ap-bound conformers, C₀, Cᵢ, and Cₙ, of ARNA (a) and GₐRNA (b) as a function of RNA concentration (500 nM Ap, 10 mM Mg²⁺, 20°C). Bottom panels show the variation in total population bound for wild type, loop- and P1 stem-mutant ARNA (c) and GₐRNA (d) as a function of RNA concentration (500 nM Ap, 10 mM Mg²⁺, 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 Kᵢ values of the individual conformers, and the ensemble average.
Figure 2-4. Mg$^{2+}$-dependence of ligand-mediated folding of the three conformers of ARNA (a) and $G_A$RNA (b). Shown are the variation in the fraction of Ap-bound conformers, $C_0$, $C_i$, and $C_c$, together with the total population bound ($\Sigma C_n$) as a function of Mg$^{2+}$ concentration (500 nM Ap, RNA:Ap $\sim$ 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 $\sim$ 20% fluorescence quenching observed upon addition of $G_A$RNA, but not ARNA to Ap in the absence of Mg$^{2+}$……………54

Figure 2-5. Dependence of the fraction of liganded-RNA folded, $f_{\text{folded}}$ (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)……………………………………………………………………59

Figure 2-6. Schematized energy landscapes for ligand-bound conformers of the purine riboswitches. In the proposed model, liganded-RNA exchanges between three states ($C_0$, $C_i$, 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 $C_0 > C_i > C_c$). The relative conformer energies measured for ARNA (pink) and $G_A$RNA (blue) are estimated based on the equilibrium constants, $K_{-1}$ and $K_2$ measured for the $C_i$-$C_0$ and $C_i$-$C_c$ equilibria, respectively, from $fC_n$ values near the maximum $f_{\text{folded}}$' Sequence-dependent partitioning of ligand-bound $C_i$ leads to greater unfolding to $C_0$ in the more dynamic ARNA complex, and drives the more preorganized $G_A$RNA towards the fully folded $C_c$ in the bound state……………………………………………………………………………….63

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 nRNA with stars representing edit sites. ………………………………………83

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
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.

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

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.

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, \( \tau_n \), and fractions of each lifetime, \( f_n \), displayed on the right.

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.

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.
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 Angstrom = 0.1 nanometre
A-T adenine- thymine base pair
A-to-I adenosine-to-inosine
ADAR adenosine deaminase acting on double-stranded RNA
ad ARNA adenine riboswitch
Alu element approximately 300 nt primate-specific region recognized by Alu enzyme

ARNA adenine riboswitch
ATP adenosine triphosphate
B. subtilis Bacillus 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 αl5 of the COB pre-mRNA

C\(_d\)/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

$f_{\text{fold}}$ fraction folded

$f_i$ fraction intermediate

$f_m, f_j$ fractions of intermediates m and j, respectively

FRET Förster resonance energy transfer

$f_{\text{SQ}}$ fraction static quenching

$f_{\text{TQ}}$ 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

| **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 coefficient |
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
Rho (ρ): prokaryotic transcription factor
RNA: ribonucleic acid
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNAP</td>
<td>RNA polymerase</td>
</tr>
<tr>
<td>rRNA</td>
<td>ribosomal RNA</td>
</tr>
<tr>
<td>s⁻¹</td>
<td>per second</td>
</tr>
<tr>
<td>SAM</td>
<td>S-adenosylmethionine</td>
</tr>
<tr>
<td>SRL</td>
<td>sarcin-ricin loop</td>
</tr>
<tr>
<td>SSF</td>
<td>steady-state fluorescence</td>
</tr>
<tr>
<td>ssRNA</td>
<td>single-stranded RNA</td>
</tr>
<tr>
<td>T</td>
<td>thymine</td>
</tr>
<tr>
<td>TAR</td>
<td>transactivating response element</td>
</tr>
<tr>
<td>TAT</td>
<td>transactivator protein</td>
</tr>
<tr>
<td>TCSPC</td>
<td>time correlated single photon counting</td>
</tr>
<tr>
<td>TE</td>
<td>tris ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>TEMED</td>
<td>Tetramethylethylenediamine</td>
</tr>
<tr>
<td>TM</td>
<td>melting temperature (half population in single-strand state)</td>
</tr>
<tr>
<td>TPP</td>
<td>thiamine pyrophosphate</td>
</tr>
<tr>
<td>TRF</td>
<td>time-resolved fluorescence</td>
</tr>
<tr>
<td>TRIS-HCl</td>
<td>tris(hydroxymethyl)aminomethane hydrochloride</td>
</tr>
<tr>
<td>tRNA</td>
<td>transfer RNA</td>
</tr>
<tr>
<td>U</td>
<td>uracil</td>
</tr>
<tr>
<td>UTP</td>
<td>uridine triphosphate</td>
</tr>
<tr>
<td>UV</td>
<td>ultraviolet</td>
</tr>
<tr>
<td>V. vulnificus</td>
<td>Vibrio vulnificus</td>
</tr>
</tbody>
</table>
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

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, \( S_0 \), 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 \( S_1 \) 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. 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. 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. $h\nu_A$, $h\nu_F$ and $h\nu_P$ 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.\textsuperscript{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.\textsuperscript{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).\textsuperscript{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.\textsuperscript{4,14,15}

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

![Figure 1-3](image-url)

**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. Copyright © 2007 American Chemical Society. All rights reserved.

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,\textsuperscript{16} nucleic acid conformational transitions,\textsuperscript{17} and nuclease behaviours.\textsuperscript{18}

In order to monitor the fluorescence of 2Ap, investigators utilize two broad techniques: steady-state fluorescence (SSF) and time-resolved fluorescence (TRF).\textsuperscript{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.\textsuperscript{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.\textsuperscript{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\(^2\)\(^2\), mismatch repair\(^2\)\(^3\), and protein kinase activity\(^2\)\(^4\). 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 \(\beta\)-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\(^2\)\(^4\). 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\(^2\)\(^5\). 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,\textsuperscript{26} with the latter leading to a
G-C mutagenesis (Figure 1-4).\textsuperscript{27}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure1.png}
\caption{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.}
\end{figure}

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. \textit{In vivo}, hairpin stems can be hundreds of bps, with intervening bulges, and capping loops averaging fifteen
nucleotides (Figure 1-5).\textsuperscript{28} Although they serve imperative functions at the DNA level, such as telomere organization,\textsuperscript{29} and antibody generation in higher eukaryotes,\textsuperscript{30} hairpins are most abundant in non-coding (nc) RNA.

\textbf{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.\textsuperscript{31,32} Eukaryotic hairpins are involved in processes such as RNA interference,\textsuperscript{33,34} and spliceosome assembly.\textsuperscript{35} In prokaryotes hairpins may function independently to regulate Rho-independent transcription,\textsuperscript{36} and temperature sensitive genes (e.g. RNA thermometers).\textsuperscript{37} Furthermore, in viruses
including human immunodeficiency virus-1 (HIV-1), West Nile and influenza, hairpins serve as protein binding sites,\textsuperscript{38,39} and are key elements in orchestrating viral replication.\textsuperscript{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 \textit{in vitro} and \textit{in vivo}.

\textbf{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.\textsuperscript{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,\textsuperscript{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),\textsuperscript{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 \textit{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\(^\text{1,13,46}\), which are about or less than the timescale by which RNA folds (ps-min).\(^\text{47-49}\) Investigations of small RNA hairpins using 2Ap as a probe have provided fascinating results, demonstrating the amazing capability of ncRNAs.

![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).](image)

**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\textsuperscript{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,\textsuperscript{52} and interrogating non-native conformers provides alternative avenues for drug design.\textsuperscript{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 \textit{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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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. 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. 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.
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. 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, and a C74U conversion of GRNA in *B. subtilis* with 2Ap or adenine as the binding partner does not alter gene regulation *in vivo*. Here RT-qPCR assays demonstrated that the levels of elongated transcripts in GRNA and G₄RNA 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]

![Cartoon of ARNA (pink) and G$_R$RNA (blue) aptamer domain folding with and without ligand binding. When no cognate ligand is present, the \textit{pbuE} gene downstream of ARNA is turned off, while \textit{xpt-pbuX} is turned on by G$_R$RNA (top), by formation and disruption of terminator stems, respectively. When ligand is present, both riboswitches are ‘funelled’ into their folding pathways, whereby \textit{pbuE} is turned on and \textit{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. \textit{Biopolymers} 2009, 91, 953. Copyright © 2009, John Wiley & Sons, Inc. All rights reserved.][Figure 1-7]
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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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\textsubscript{A}RNA 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 heterogeneous 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.\textsuperscript{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\textsubscript{1}), intermediate (C\textsubscript{2}) and closed (C\textsubscript{3}) conformers, respectively. Fractions of each conformer were subsequently monitored upon varying temperature, concentrations of Mg\textsuperscript{2+}, and RNA. As anticipated, the highest melting temperature corresponded to C\textsubscript{C} due to its presumably compact structure consisting of an array of hydrogen bonds and Van Der Waal's interactions. C\textsubscript{C} also displayed the highest Mg\textsuperscript{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).\textsuperscript{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.\textsuperscript{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.

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure1-8.png}
\caption{RNA titration and temperature variation curves of 2Ap fluorescent lifetime 
fractions (C_1 = open conformer, C_2 = intermediate, C_3 = closed) bound to \textit{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 \textit{Journal of the American Chemical Society} \textbf{2007},\textit{129}, 11308. 
Copyright © 2007 American Chemical Society. All rights reserved}
\end{figure}

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 riboswitches.

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.\textsuperscript{84,88-90} Interestingly, the Rechavi group performed a genome vs. transcriptome analysis of several known \textit{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.\textsuperscript{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,\textsuperscript{92} and this diversity may be compounded by post-translational modifications \textit{in vivo}.\textsuperscript{93} Another valid

\begin{figure}
\centering
\includegraphics[width=\textwidth]{adarnucleotides.png}
\caption{ADAR-mediated adenosine-to-inosine conversion. Note: This event occurs within a dsRNA duplex, but 3' terminal nucleotides are shown here for clarity.}
\end{figure}
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. 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. 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. These
observations are consistent with a mechanism whereby the modified base is flipped out from the duplex, 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. 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 accommodates 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


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 \textit{xpt-pbuX} gene.

Using 2Ap as a cognate ligand mimic for ARNA and GRNA with a C74U mutations (G\textsubscript{A-RNA}), which does not affect riboswitch function \textit{in vivo}, we have probed the conformational dynamics of the ligand-bound purine riboswitch aptamer domains from \textit{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\textsuperscript{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\textsubscript{A-RNA}. 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\textsuperscript{2+} titrations on ARNA and G\textsubscript{A}RNA 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\textsubscript{o}, C\textsubscript{i}, C\textsubscript{c}) of the liganded G- and A-sensing riboswitches from Bacillus subtilis. The ligand binding affinities, and sensitivity to Mg\textsuperscript{2+}, together with results from mutagenesis, suggest that C\textsubscript{o} and C\textsubscript{i} are partially unfolded species compromised in key loop-loop interactions present in the fully folded C\textsubscript{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\textsuperscript{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,
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,\(^2\) and being uncovered in eukaryotes,\(^6\) 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.\(^10\)\(^\text{-}\)\(^12\) 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 riboswitch 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. 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.

Considerable sequence variation amongst A and GRNA exists outside the binding pocket (Fig. 2-1). These variations have little effect on secondary and ensemble-averaged tertiary structure, or the overall thermodynamics of binding. Yet, they impact function, since mutation of the specificity nucleotide N74 alone is insufficient to interchange G/A regulation of gene expression in vivo. 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 ions, the purine riboswitches are heterogeneous mixtures of conformers in solution. 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.\textsuperscript{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 \textit{Bacillus subtilis}, GRNA, and ARNA, respectively, and the tertiary structure of the aptamer domain of GRNA, and the adenine riboswitch from \textit{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*. Ap is a natural probe of A-responsive riboswitches as its interactions with these RNA molecules mimic those of A structurally, thermodynamically, and kinetically. From the fluorescence lifetime distribution we resolved three distinct ligand-bound species of ARNA, assigned to three RNA conformers, C₁, C₂, and C₃. The ligand binding affinity, and sensitivity of folding to Mg²⁺, was substantially higher for C₃, than for C₁ or C₂. 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. Consequently, we proposed that C₁, C₂, and C₃ 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, C₃ = C₃) and partially unfolded (intermediate and open, C₁ = C₂)
and \( C_0 = C_1 \) conformers through mutagenesis. We test our prediction by comparative investigation of the Ap-bound conformers of ARNA and C74U GRNA; the latter (abbreviated as \( G_{\text{A}} \)RNA) 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 \( G_{\text{A}} \)RNA. 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 \((t = 11.5 \text{ ns, } \chi^2 \sim 1.2)\), the time-resolved fluorescence of Ap in the presence of saturating concentrations of \( G_{\text{A}} \)RNA, 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,

\[ I(t) = f_{C_u}e(-t/\tau_u) + f_{C_o}e(-t/\tau_o) + f_{C_i}e(-t/\tau_i) + f_{C_c}e(-t/\tau_c) \]  

where \( f_{C_c} \) is the statically quenched population with a lifetime, \( t_c < 300 \text{ ps} \). The lifetimes (\( t_u, t_o, \) and \( t_i \)) and relative populations (\( f_{C_u}, f_{C_o} \) and \( f_{C_i} \)) of the other species are obtained from the time-resolved decay, and their populations are adjusted to account for \( f_{C_c} \) (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. 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, \( t_u = 11.5 \text{ ns} \), 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, 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 (\( f_{C_n} \)) 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 \( C_2 \)), 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.

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 \((0.5 \pm 0.2)\) and \((2.0 \pm 0.2)\) ns, respectively. The

**Figure 2-2:** Fluorescence decay profile of free Ap, and Ap in the presence of either ARNA or G\(_A\)RNA (500 nM Ap, 4 \( \mu \)M 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 \((0.5 \pm 0.2)\) and \((2.0 \pm 0.2)\) ns, respectively. The
results described below aim to test these assignments and compare the conformer distribution in ARNA and GₐRNA.

Under the conditions represented in Figure 2-2 (10 mM Mg²⁺, 20°C, excess RNA, Ap:RNA ≤ 1:8), the population of partially quenched species constitutes ~ 25%: $f_{C_o} = 0.14 \pm 0.02$, $f_{C_i} = 0.13 \pm 0.02$, $f_{C_c} = 0.62 \pm 0.02$, and $f_u = 0.10 \pm 0.02$, the same within experimental error for ARNA and GₐRNA. However, the population of each bound conformer varies with the ratio of Ap:RNA and Mg²⁺: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 Mg²⁺ 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.](image)

conformer (each conformer) to Mg$^{2+}$, $K_{Mg}$. Both experiments yield values for $K_{-1}$

$$K_{-1} = \frac{[Co-L-(Mg^{2+})_x]}{[Ci-L-(Mg^{2+})_y]} = f_{Co}/f_{Ci}$$

and $K_2$

$$K_2 = \frac{[Cc-L-(Mg^{2+})_y]}{[Ci-L-(Mg^{2+})_y]} = f_{Cc}/f_{Ci}$$

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 $Co$, $Ci$, and $Cc$. 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.\textsuperscript{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\textsuperscript{2+} is required for ligand binding (\textit{vide}
\textit{infra}), values of $K_L$ presented are those in the presence of Mg\textsuperscript{2+}.

\subsection*{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\textsubscript{A}RNA, 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 $\sim$ 4-fold larger than those
derived from fitting to a sigmoidal model.\textsuperscript{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 $G_A$ RNA (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 $C_C$. 2) The ensemble $K_L$, evaluated from time-resolved Ap fluorescence, is ~ 2-fold smaller for $G_A$ RNA than for ARNA, predominantly due to tighter binding of $G_A$ RNA $C_c$ (Table 2-1). As a result, the variation in relative binding affinity of the conformers is larger for $G_A$ RNA (~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 $K_L = (300 \pm 20)$ nM for either ARNA or $G_A$ RNA (Fig. 2-4), comparable to previous reports. 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, $C_0$, $C_i$, and $C_c$, of ARNA (a) and GARN (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 GARN (d) as a function of RNA concentration (500 nM Ap, 10 mM Mg$^{2+}$, 20°C). Comparison with wild-type GARN 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.
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 GA RNA

<table>
<thead>
<tr>
<th>Arna</th>
<th>Sequence</th>
<th>Ap</th>
<th>$K_L$ (µM)$^b$</th>
<th>$K_{rel}$</th>
<th>$K_{Mg}$ (µM)$^d$</th>
<th>$K_{Mg,rel}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>wt</td>
<td>En.</td>
<td>G37C</td>
<td>3.2 ± 0.5</td>
<td>16</td>
<td>3.9 ± 0.5</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G13C</td>
<td>Not detected</td>
<td>&gt;100</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>G78C</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>wt</td>
<td>En.</td>
<td>Mg$^{2+}$</td>
<td>300 ± 20</td>
<td>1</td>
<td>280 ± 20</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>200 ± 20</td>
<td>1</td>
<td>150 ± 20</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1500 ± 300</td>
<td>7</td>
<td>1100 ± 200</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2700 ± 400</td>
<td>13</td>
<td>2000 ± 400</td>
<td>13</td>
</tr>
</tbody>
</table>

$^a$ Averages and standard errors of three to six experiments in Phosphate buffer 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).

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, and is among the most deleterious loop mutations to ligand binding affinity, and presumably loop-loop interactions, in ARNA. 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. In GA RNA, the corresponding P1 stem C-C mismatch is provided by the G78C mutant. Each mutation substantially reduces (~20 to > 100 fold) the ligand binding affinity of ARNA and GA RNA (Fig. 2-3c,d). We report only the ensemble averages.
because very little $C_l$ 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 $G_A$RNA, within experimental error, although the loss in affinity relative to wild type is greater for $G_A$RNA (Table 2-1). In contrast, while G78C-$G_A$RNA 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* (*pbe*) ARNA and A to *B. subtilis* (*xpt-***p**uX*) GRNA has been detected by NMR,$^{30}$ and binding of Ap to GRNA is only $\sim$ 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 $C_o$, $C_l$, and $C_c$ with Mg$^{2+}$ concentration (at a constant Ap:RNA ratio) is shown for wild type ARNA and $G_A$RNA in Figure 2-4. In the absence of Mg$^{2+}$, addition of saturating concentrations of ARNA (RNA:Ap $\sim$ 8) leaves the Ap fluorescence largely unchanged; the fraction bound is effectively zero. In contrast, addition of $G_A$RNA
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. These data may be fit to a Hill model (Materials and Methods) to extract the midpoint of each folding transition, K$_{Mg}$. 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$^{2+}$. The latter includes the affinity-weighted average for binding at specific sites, detected by crystallography and NMR as well as non-specific association by the diffuse ion atmosphere 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. For both riboswitches, K$_{Mg}$ values (Table 2-1) vary as: C$_c$ << C$_i$ < C$_o$, each slightly smaller for G$_A$RNA 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.
Figure 2-4. Mg\(^{2+}\)-dependence of ligand-mediated folding of the three conformers of ARNA (a) and G\(_A\)RNA (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 (\(\Sigma C_n\)) as a function of Mg\(^{2+}\) concentration (500 nM Ap, RNA:Ap ~ 8, 20\(^\circ\)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\(_A\)RNA, 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.\textsuperscript{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.\textsuperscript{26,30,34} Although presumably representing intermediates on route to the fully folded structure, the kinetic\textsuperscript{34} and thermodynamic\textsuperscript{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\textsubscript{A}RNA observed by time-resolved fluorescence correspond to two such partially unfolded intermediates, C\textsubscript{o} and C\textsubscript{i}, that are in reversible conformational exchange with the fully folded structure, C\textsubscript{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\textsubscript{A}RNA 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,\textsuperscript{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 ($f_{C_n}$) from the fluorescence lifetime distribution as described above. This provides the fraction folded within the bound state $f_{\text{folded}}$, (eq. 2-2),

$$f_{\text{folded}} = \frac{fC_c}{(fC_c + fC_i + fC_o)}$$

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 $G_A$RNA, it is possible to measure the $f_{\text{folded}}$, $K_1$ and $K_2$ over a large binding range ($f_B \sim 0.2 - 0.8$); these vary as expected for ligand-mediated folding of $C_o$ and $C_i$ to $C_c$. The $f_{\text{folded}}$ 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 $f_{\text{folded}} (> 0.6)$ is reached at the lowest measurable $f_B$. We expect the $f_{\text{folded}}$ 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 $f_{\text{folded}}$ 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 $G_A$ RNA, respectively, for the same range of Ap:RNA ratios (higher $f_B$ for $G_A$ RNA, due to its tighter ensemble $K_D$). These findings match well with recent investigations of ligand-induced folding of add-ARNA from *V. vulnificus*. While folding could be partially driven by Mg$^{2+}$, subsequent binding of A facilitates additional conformational changes, detected in the remote L2-L3 region.
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 $f_{\text{folded}}$ (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^\circ 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^\circ 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.\textsuperscript{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.\textsuperscript{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\textsuperscript{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.\textsuperscript{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,\textsuperscript{34} and structure-based simulations\textsuperscript{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,\textsuperscript{26} providing a platform to initiate ligand binding, while P1 is likely (partially) unfolded.\textsuperscript{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 \textasciitilde 4 kcal/mol above the fully folded structure.\textsuperscript{34} Due to the \textasciitilde 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,\textsuperscript{12} while studies by Lafontaine and coworkers find a stem length of at least 2 base pairs is required for detectable binding to ARNA.\textsuperscript{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\textsubscript{A} RNA 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\textsuperscript{26} or a mix of open and closed conformers in rapid exchange on the electrophoresis time scale.
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

**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.
Similar conclusions may be drawn from time-resolved NMR and stopped-flow fluorescence 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 Cc, tight ligand binding is maintained also through strong L2-L3 interactions, while these interactions are weakened in Ci and Co. Analogous to the interplay between pseudo-knot and P1 stem formation observed for the SAM riboswitch, 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” 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 Ci and Co. 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 Cc. The L2-L3 interactions existing in Cc appear to be stronger in G_ARNA than ARNA, and their loss in Ci and Co, or the G37C mutant,
is accompanied by a larger reduction in binding affinity for $G_A$-RNA ($K_{rel}$, 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$^{2+}$, $G_A$-RNA 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 $G_A$-RNA ($f_{folded, max} \sim 0.95$ at $\sim 10 \mu M$ Mg$^{2+}$) than for ARNA ($f_{folded, max} \sim 0.85$ at $500 \mu M$ Mg$^{2+}$). The weaker requirement on the divalent cation for ligand-mediated folding may be a consequence of intrinsic preorganization in $G_A$-RNA, 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 $G_A$-RNA, 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 $G_A$ RNA are also amplified by mutation. For instance, for the same fraction of RNA bound, the liganded state of loop mutant, G37C-$G_A$ RNA 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 $G_A$ RNA, 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 $G_A$ RNA $C_o$ 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 $\sim 6-9$ kcal/mol,\textsuperscript{22,34} folding and unfolding of the ligand-bound riboswitches involves relatively small free energy changes. Values of $\sim 1-2$ kcal/mol are calculated
from $K_2$ and $K_1$ (Scheme 2-1) and corroborated by $k_{\text{open}}/k_{\text{close}}$ and force spectroscopy for single ARNA molecules.\textsuperscript{34,42} Consequently, interconversion of ligand-bound C\textsubscript{o}, C\textsubscript{i} and C\textsubscript{c} may exhibit heightened sensitivity\textsuperscript{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\textsubscript{o}, while the shorter loops in G\textsubscript{A}, RNA (7 bases) may be the source of less fluctuating L2-L3 interactions and the higher population of C\textsubscript{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\textsuperscript{2+} to stabilize L2-L3 interactions.\textsuperscript{29} The sequence of both L3 and the P1 stem of the adenine-sensing riboswitch from \textit{V. vulnificus} (add-ARNA) more closely resembles the \textit{xpt-pbuX} \textit{B. subtilis} GRNA than the \textit{pbuE} ARNA investigated here. One might predict that folding/unfolding of add-ARNA\textsuperscript{27} may be more similar to \textit{xpt-pbuX} GRNA. Unlike \textit{pbuE} ARNA (Fig. 2-1), exchange between expression platform conformations of \textit{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.\textsuperscript{54}

For some, riboswitches, including the \textit{xpt-pbuX} GRNA and the TPP riboswitch,\textsuperscript{55} there appears to exist an intrinsic level of preorganization set at the sequence level that resists further folding of unbound RNA by Mg\textsuperscript{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\textsuperscript{56,57} Mg\textsuperscript{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\textsuperscript{48} where folding appears to be mediated by coupled binding of both
Mg\textsuperscript{2+} and the specific ligand. As for glmS, the conformationally-dynamic free
RNA requires Mg\textsuperscript{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\textsuperscript{26} and, at least
for ARNA, are likely stronger than secondary interactions within the P1 stem.\textsuperscript{34}
The former may serve to organize the pocket for ligand binding, as \textit{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 MgCl₂, 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 Mg²⁺ 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°) 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 ≥ 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 Mg²⁺, or addition of Mg²⁺ 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 $f_{C_C} = 1-(I'/I'_o)$, 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 Mg²⁺, 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 $\Sigma f_{C_n} = 1-(I/I_o)$, 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 Mg²⁺, respectively.

As the full lifetime decay of Ap bound to the purine riboswitches is not completely resolved in our experiments, $f_{C_c}$ is evaluated from the extent of static
quenching, while \( f_{Ci} \) and \( f_{Co} \) are evaluated from their amplitude in the fluorescence lifetime decay, \( \alpha_n \) in eq. 2-3 (\( \Sigma \alpha_n = 1 \)).

\[
I(t) = \alpha_u e(-t / \tau_u) + \alpha_o e(-t / \tau_o) + \alpha_i e(-t / \tau_i)
\]

These relative amplitudes, \( a_o \) and \( a_i \), were adjusted to account for \( f_{Cc} \) according to equation 2-4:

\[
f_{C_{o,i}} = a_{o,i} \times (1 - f_{Cc})
\]

The relative amplitude of unbound Ap, \( \alpha_u \), is similarly adjusted to yield the unbound fraction, \( f_u \), and \( \Sigma f_{C_i} + f_u = 1 \). This analysis accurately (errors \( \leq 5\% \)) measures \( f_{Cc} \), but less accurately quantifies \( f_{Ci} \) and \( f_{Co} \), 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.

\[
\Sigma (f_{Ci} + f_{Co}) = fTQ - fSQ
\]

At values of \( \Sigma (f_{Ci} + f_{Co}) < 5 \% \) we do not reliably detect the presence of bound species in our lifetime decay. This precludes independent evaluation of \( f_{Ci} \) and
\( f_{C_0} \) (and thus their \( K_L \) values, and the equilibrium constants, \( K_{-1} \) and \( K_2 \)) for the mutant riboswitches.

The variation in \( f_{C_n} \), or total fluorescence quenching, \( \Sigma f_{C_n} \), 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:

\[
f_{C_n} = \frac{(f_{C_n})_{\text{max}}}{2[f_{Ap}]} \left\{ [f_{Ap}]+[\text{RNA}]+K_L - \left( ([f_{Ap}]+[\text{RNA}]+K_L)^2 - 4[f_{Ap}][\text{RNA}] \right)^{1/2} \right\}
\]

\[
f_{C_n} = (f_{C_n})_{\text{max}*} \frac{[\text{Mg}^{2+}]^n}{[K_{\text{Mg}}]+[\text{Mg}^{2+}]^n}
\]

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 \( K_{\text{Mg}} \), 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.

73
2.8 References


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
Table 2-2. Parameters from a representative experiment monitoring the fluorescence lifetime of 500 nM Ap as a function of G_RNA 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 $\alpha_u$, $\alpha_o$, and $\alpha_i$ are the amplitudes, and $\tau_u$, $\tau_o$, and $\tau_i$ 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).

<table>
<thead>
<tr>
<th>$[\text{Mg}^{2+}]$ (µM)</th>
<th>Total counts</th>
<th>Peak counts</th>
<th>$\tau_u$</th>
<th>$\alpha_u$</th>
<th>$\tau_o$</th>
<th>$\alpha_o$</th>
<th>$\tau_i$</th>
<th>$\alpha_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>4.8734e+06</td>
<td>45935</td>
<td>11.8</td>
<td>0.9253</td>
<td>6.13</td>
<td>0.0427</td>
<td>1.39</td>
<td>0.032</td>
</tr>
<tr>
<td>0.0159</td>
<td>4.6445e+06</td>
<td>43972</td>
<td>11.78</td>
<td>0.9241</td>
<td>5.42</td>
<td>0.0479</td>
<td>1.07</td>
<td>0.028</td>
</tr>
<tr>
<td>0.0395</td>
<td>4.5062e+06</td>
<td>42879</td>
<td>11.76</td>
<td>0.9371</td>
<td>4.38</td>
<td>0.0331</td>
<td>1.075</td>
<td>0.0298</td>
</tr>
<tr>
<td>0.0782</td>
<td>4.1598e+06</td>
<td>40200</td>
<td>11.86</td>
<td>0.9166</td>
<td>4.00</td>
<td>0.0498</td>
<td>1.098</td>
<td>0.0336</td>
</tr>
<tr>
<td>0.1562</td>
<td>3.6016e+06</td>
<td>36284</td>
<td>11.88</td>
<td>0.9043</td>
<td>5.00</td>
<td>0.0653</td>
<td>0.962</td>
<td>0.0404</td>
</tr>
<tr>
<td>0.2338</td>
<td>3.1455e+06</td>
<td>32941</td>
<td>11.78</td>
<td>0.9126</td>
<td>3.13</td>
<td>0.0499</td>
<td>0.734</td>
<td>0.0375</td>
</tr>
<tr>
<td>0.3111</td>
<td>2.7337e+06</td>
<td>30137</td>
<td>11.77</td>
<td>0.8978</td>
<td>2.92</td>
<td>0.058</td>
<td>0.676</td>
<td>0.0441</td>
</tr>
<tr>
<td>0.3881</td>
<td>2.3450e+06</td>
<td>27761</td>
<td>11.76</td>
<td>0.8763</td>
<td>2.72</td>
<td>0.0728</td>
<td>0.566</td>
<td>0.051</td>
</tr>
<tr>
<td>0.4649</td>
<td>2.0650e+06</td>
<td>27010</td>
<td>11.71</td>
<td>0.847</td>
<td>2.23</td>
<td>0.0945</td>
<td>0.545</td>
<td>0.059</td>
</tr>
<tr>
<td>0.6174</td>
<td>1.6028e+06</td>
<td>27010</td>
<td>11.67</td>
<td>0.7566</td>
<td>2.01</td>
<td>0.1601</td>
<td>0.543</td>
<td>0.0833</td>
</tr>
<tr>
<td>0.844</td>
<td>1.2166e+06</td>
<td>26145</td>
<td>11.61</td>
<td>0.6504</td>
<td>2.07</td>
<td>0.2162</td>
<td>0.564</td>
<td>0.1334</td>
</tr>
<tr>
<td>1.216</td>
<td>8.8569e+05</td>
<td>24303</td>
<td>11.48</td>
<td>0.499</td>
<td>2.22</td>
<td>0.2933</td>
<td>0.609</td>
<td>0.2076</td>
</tr>
<tr>
<td>1.5945</td>
<td>7.7288e+05</td>
<td>23828</td>
<td>11.23</td>
<td>0.44</td>
<td>2.05</td>
<td>0.342</td>
<td>0.535</td>
<td>0.218</td>
</tr>
<tr>
<td>2.3493</td>
<td>6.6840e+05</td>
<td>22601</td>
<td>11.08</td>
<td>0.372</td>
<td>2.07</td>
<td>0.378</td>
<td>0.524</td>
<td>0.25</td>
</tr>
<tr>
<td>3.85</td>
<td>5.9569e+05</td>
<td>22416</td>
<td>10.9</td>
<td>0.2957</td>
<td>2.24</td>
<td>0.3942</td>
<td>0.564</td>
<td>0.3101</td>
</tr>
<tr>
<td>[Mg$^{2+}$] ($\mu$M)</td>
<td>Total counts</td>
<td>Peak counts</td>
<td>$\tau_U$</td>
<td>$\alpha_U$</td>
<td>$\tau_O$</td>
<td>$\alpha_O$</td>
<td>$\tau_i$</td>
<td>$\alpha_i$</td>
</tr>
<tr>
<td>---------------------</td>
<td>--------------</td>
<td>-------------</td>
<td>----------</td>
<td>-----------</td>
<td>----------</td>
<td>-----------</td>
<td>----------</td>
<td>-----------</td>
</tr>
<tr>
<td>0</td>
<td>4.9421e+06</td>
<td>21844</td>
<td>11.6</td>
<td>0.97</td>
<td>5.5</td>
<td>0.015</td>
<td>0.55</td>
<td>0.015</td>
</tr>
<tr>
<td>0.001</td>
<td>3.9197e+06</td>
<td>17734</td>
<td>11.67</td>
<td>0.955</td>
<td>3.8</td>
<td>0.033</td>
<td>0.38</td>
<td>0.012</td>
</tr>
<tr>
<td>0.00498</td>
<td>3.8991e+06</td>
<td>17624</td>
<td>11.69</td>
<td>0.954</td>
<td>3.9</td>
<td>0.034</td>
<td>0.38</td>
<td>0.0131</td>
</tr>
<tr>
<td>0.0109</td>
<td>3.7904e+06</td>
<td>17203</td>
<td>11.69</td>
<td>0.954</td>
<td>3.9</td>
<td>0.034</td>
<td>0.48</td>
<td>0.013</td>
</tr>
<tr>
<td>0.018</td>
<td>3.6680e+06</td>
<td>16844</td>
<td>11.7</td>
<td>0.955</td>
<td>3.2</td>
<td>0.0318</td>
<td>0.3</td>
<td>0.013</td>
</tr>
<tr>
<td>0.027</td>
<td>3.5258e+06</td>
<td>16116</td>
<td>11.68</td>
<td>0.954</td>
<td>3.5</td>
<td>0.032</td>
<td>0.4</td>
<td>0.0134</td>
</tr>
<tr>
<td>0.066</td>
<td>3.2255e+06</td>
<td>15140</td>
<td>11.65</td>
<td>0.953</td>
<td>2.9</td>
<td>0.032</td>
<td>0.324</td>
<td>0.015</td>
</tr>
<tr>
<td>0.104</td>
<td>2.8865e+06</td>
<td>13886</td>
<td>11.65</td>
<td>0.9497</td>
<td>2.3</td>
<td>0.033</td>
<td>0.25</td>
<td>0.0175</td>
</tr>
<tr>
<td>0.201</td>
<td>2.3081e+06</td>
<td>11774</td>
<td>11.61</td>
<td>0.9376</td>
<td>2.11</td>
<td>0.037</td>
<td>0.24</td>
<td>0.0254</td>
</tr>
<tr>
<td>0.58</td>
<td>2.2188e+06</td>
<td>11350</td>
<td>11.68</td>
<td>0.9285</td>
<td>3</td>
<td>0.042</td>
<td>0.31</td>
<td>0.0292</td>
</tr>
<tr>
<td>0.96</td>
<td>9.8615e-05</td>
<td>8198</td>
<td>11.62</td>
<td>0.8292</td>
<td>2.2</td>
<td>0.079</td>
<td>0.287</td>
<td>0.092</td>
</tr>
<tr>
<td>1.53</td>
<td>6.6121e-05</td>
<td>7179</td>
<td>11.53</td>
<td>0.753</td>
<td>1.8</td>
<td>0.153</td>
<td>0.281</td>
<td>0.133</td>
</tr>
<tr>
<td>2.1</td>
<td>5.3991e-05</td>
<td>6932</td>
<td>11.51</td>
<td>0.6889</td>
<td>1.75</td>
<td>0.1509</td>
<td>0.25</td>
<td>0.1602</td>
</tr>
<tr>
<td>3.097</td>
<td>4.2329e-05</td>
<td>6944</td>
<td>11.4</td>
<td>0.5825</td>
<td>1.9</td>
<td>0.1853</td>
<td>0.294</td>
<td>0.2322</td>
</tr>
<tr>
<td>5.88</td>
<td>3.8451e-05</td>
<td>6727</td>
<td>11.5</td>
<td>0.5377</td>
<td>2</td>
<td>0.2035</td>
<td>0.31</td>
<td>0.269</td>
</tr>
<tr>
<td>7.72</td>
<td>3.6656e-05</td>
<td>6757</td>
<td>11.3</td>
<td>0.5046</td>
<td>1.96</td>
<td>0.221</td>
<td>0.292</td>
<td>0.274</td>
</tr>
<tr>
<td>10.51</td>
<td>3.6919e-05</td>
<td>6925</td>
<td>11.25</td>
<td>0.4948</td>
<td>1.93</td>
<td>0.2354</td>
<td>0.276</td>
<td>0.2698</td>
</tr>
</tbody>
</table>

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$_A$RNA, 20 °C). Fluorescence decay fit to: $I(t) = \alpha_U\tau_U + \alpha_O\tau_O + \alpha_i\tau_i$ where $\alpha_U$, $\alpha_O$, and $\alpha_i$ are the amplitudes, and $\tau_U$, $\tau_O$, and $\tau_i$ 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,\textsuperscript{1,2} while proteomes experience post-translational modifications such as methylations, glycosylations, phosphorylations,\textsuperscript{3} and even truncations.\textsuperscript{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.\textsuperscript{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.\textsuperscript{7} Furthermore, aberrant A-to-I editing events are linked to psychiatric disorders,\textsuperscript{8} neurodegenerative disease,\textsuperscript{9} and cancers.\textsuperscript{10} A-to-I edits across all organisms are most prominent in ncRNAs,\textsuperscript{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.\textsuperscript{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,\textsuperscript{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.

\textit{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.\textsuperscript{6,17-19} The ncRNA element that almost singlehandedly facilitates A-to-I editing is the \textit{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). *Alu* elements 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.](image-url)
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,\textsuperscript{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?

\textbf{Figure 3-2:} Qualitative comparison of A-to-I editing levels in selected primate substrates. The \textit{Alu} element we have directed our studies towards corresponds to \textit{Nup50}, boxed and shown in light blue. Other genes of interest are also shown (below). Reproduced with permission from \textit{Proceedings of the National Academy of Sciences of the United States of America} \textbf{2010}, \textbf{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), a 50 kDa nucleoporin gene, which regulates the flow of molecules and ions across the nucleus. Nup50 Alu A-to-I conversion in human brain tissues was approximately five times higher compared to chimp/rhesus.

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, 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.\textsuperscript{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).

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

**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.
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).
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 $0.5(\tau_1)$, $3.0(\tau_2)$, and $8.5(\tau_3)$ 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 $\tau_1$, $\tau_2$, and $\tau_3$,
alongside their corresponding fractions (data not shown), represent a dynamic
equilibrium between folded, intermediate, and base-exposed conformers,
respectively.
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 \( \tau_1 \) and \( \tau_2 \) 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

**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, \( \tau_n \), and fractions of each lifetime, \( f_n \), displayed on the right.

\[
\begin{align*}
\tau_{\text{Ap}5} & = 9.6 \text{ ns} \\
\tau_1 & = 0.4 - 0.6 \text{ ns} \\
\tau_2 & = 2.4 - 3.8 \text{ ns} \\
\tau_3 & = 8.0 - 9.1 \text{ ns} \\
\hat{f}_1 & = 0.74 - 0.78 \\
\hat{f}_2 & = 0.17 - 0.21 \\
\hat{f}_3 & = 0.03 - 0.05
\end{align*}
\]
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 ($\tau_3$), 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 ($\sum f_n = 1$) 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 \(\pi\)-stack with one another causing a decrease in fluorescence intensity.\textsuperscript{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 ($K_n$, $\Delta H_n$ and $\Delta S_n$), see Table 3-1) according to scheme 3-1.
A) Human Ap5

B) Chimp Ap5

Scheme

F \leftarrow K_1 \rightarrow M \leftarrow K_2 \rightarrow U
F \leftarrow K_3 \rightarrow J \leftarrow K_4 \rightarrow U
M \leftarrow K_j \rightarrow J
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 $K_3$.

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.
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 \(T_M\)s 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 \(T_M\)s. 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.

<table>
<thead>
<tr>
<th>Construct</th>
<th>Pathway/Thermodynamic Parameters (ΔHᵢ kJ/mol, ΔSᵢ J/mol K, Kᵢ at 37 °C)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ChAp5</strong></td>
<td>F→M</td>
</tr>
<tr>
<td>K₁ = 2.16x10⁴ ± 9 x10⁶</td>
<td>K₂ = 0.320 ± 1.5 x10²</td>
</tr>
<tr>
<td>ΔH₁ = 303.85 ± 0.06</td>
<td>ΔH₂ = 57.34 ± 0.09</td>
</tr>
<tr>
<td>ΔS₁ = 909.5 ± 0.3</td>
<td>ΔS₂ = 175.4 ± 0.3</td>
</tr>
</tbody>
</table>

| **HuAp5** | F→M | M→U | F→J | J→U | M→J |
| K₁ = 8.65 x10⁻¹¹ ± 7.2 x10⁻¹² | K₂ = 0.560 ± 2.53 x10⁻² | K₃ = 0.290 ± 8.08 x10⁻³ | K₄ = 1.64 x10⁻¹⁰ ± 1.62 x10⁻¹¹ | K_J = 3.40 x10⁹ ± 2.98 x10⁹ |
| ΔH₁ = 596.63 ± 0.17 | ΔH₂ = 36.46 ± 0.09 | ΔH₃ = -5.02 ± 0.05 | ΔH₄ = 638.1 ± 0.2 | ΔH_J = -601.7 ± 0.2 |
| ΔS₁ = 1731.1 ± 0.4 | ΔS₂ = 112.7 ± 0.3 | ΔS₃ = -26.4 ± 0.2 | ΔS₄ = 1870.2 ± 0.5 | ΔS_J = -1757.5 ± 0.5 |

| **RhAp5** | F→M | M→U | F→J | J→U | M→J |
| K₁ = 2.91 x10⁻³ ± 1.69 x10⁻⁴ | K₂ = 0.888± 3.45 x10⁻² | K₃ = 0.31751 ± 9.58 x10⁻³ | K₄ = 0.00812 ± 6.2 x10⁻⁴ | K_J = 109 ± 7.2 |
| ΔH₁ = 291.04 ± 0.11 | ΔH₂ = 46.18 ± 0.06 | ΔH₃ = -18.74 ± 0.06 | ΔH₄ = 356.0 ± 0.1 | ΔH_J = -309.8 ± 0.1 |
| ΔS₁ = 889.9 ± 0.3 | ΔS₂ = 147.9 ± 0.3 | ΔS₃ = -70.0 ± 0.2 | ΔS₄ = 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).

![Diagram](image-url)
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.\textsuperscript{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, $K_1 (F \rightarrow M)$ and $K_3 (F \rightarrow J)$ are less than 1, indicating that 2Ap is in an overall quenched/folded state. Interestingly, the relatively high $K_J (M \rightarrow J)$ 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 $f_l$ 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 $\tau_1$ value does not correspond to the lowest temperature (data not shown). Since equation 3-7 assumes that $\tau_1$ is shortest at the lowest temperature, our data sometimes translate to negative $f_l$ 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, $K_1$ ($F \rightarrow M$) is extremely low, whereas $K_3$ ($F \rightarrow J$) is similar compared to rhesus and chimp constructs (Table 3-1). The most notable difference is in the value of $K_J$ ($M \rightarrow J$), 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 $37 \ ^\circ C$, the $K_1$ and $K_3$ values generally favour a folded state. Interestingly, the $F \rightarrow J$ 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 \(F \rightarrow J\) transition. Since ADAR is relatively conserved in sequence across primates, it should interact with the human construct with the highest affinity, facilitating the \(F \rightarrow J\) transition.

![Diagram](image)

**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 \(M \rightarrow J\) 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 \textit{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 \textit{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\textsubscript{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\textsubscript{Ms} within experimental error compared to our Ap5 constructs, we may be able to draw a conclusion similar to Schwalbe and colleagues,\textsuperscript{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\textsubscript{M} that would roughly mirror Ap5 and Ap13 T\textsubscript{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
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:

\[
S(T) = \frac{\Delta S_{\text{max}} + T^n}{T_M^n + T^n}
\]

where S(T) represents the absorbance as a function of temperature, S_{\text{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 ($X$),

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

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 ($K_{\text{unfold}}$) by the following equation:

$$K_{\text{unfold}} = [U]/[F] = f_u/(1 - f_u) \quad 3-3$$

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

$$\ln K_{\text{unfold}} = \frac{-\Delta H}{R} * \frac{1}{T} + \frac{\Delta S}{R} \quad 3-4$$
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°). 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 (χ² ≤ 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).

\[ I(t) = f e\left( -\frac{t}{\tau} \right) \]  \hspace{1cm} (3-5)

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

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

\[ I(t) = f_1 e^{-t/\tau_1} + f_2 e^{-t/\tau_2} + f_3 e^{-t/\tau_3} \]  \hspace{1cm} (3-6)

with \( f_1, f_2 \) and \( f_3 \) (\( \Sigma f_n =1 \)) representing the fractions of short, intermediate, and long lifetimes (\( \tau_1, \tau_2, \tau_3 \), 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.
where \( f_u(\tau_1) \) is the fraction unfolded as a function of \( \tau_1 \) determined by the difference between \( \tau_1 \) at a given temperature, \( \tau_1(T) \) and \( \tau_1 \) at the lowest (presumably most folded temperature) \( \tau_{1\text{folded}} \), divided by the maximum difference between \( \tau_1 \) across the given temperature range, \( \Delta \tau_{\text{max}} \).

\( f_u(\tau_1) \) provides us with the proportion of the folded conformer, as a function of \( \tau_1 \) 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.

\[
f_u = f_u(\tau_1) \times f_1 \tag{eq 3-8}
\]

Since \( \Sigma f_i = 1 \), 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.

\[
f_f = 1 - f_u - f_i \tag{eq 3-9}
\]

### 3.5 References


### Appendix: Chapter 3 Thermodynamic Parameters

<table>
<thead>
<tr>
<th>Substance</th>
<th>Temperature (°C)</th>
<th>Heat Capacity (J/K)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Substance 1</td>
<td>250</td>
<td>450</td>
</tr>
<tr>
<td>Substance 2</td>
<td>300</td>
<td>500</td>
</tr>
<tr>
<td>Substance 3</td>
<td>350</td>
<td>550</td>
</tr>
</tbody>
</table>

### Equations

1. \( Q = \int_{T_1}^{T_2} C_p \, dT \)
2. \( T = \frac{1}{\sum \frac{1}{\Delta C_p}} \)
3. \( \Delta H = \int_{T_1}^{T_2} C_p \, dT \)

### Diagram

- [Diagram of thermodynamic processes]

- [Legend for thermodynamic processes]
| 00   | 01   | 02   | 03   | 04   | 05   | 06   | 07   | 08   | 09   | 10   | 11   | 12   | 13   | 14   | 15   | 16   | 17   | 18   | 19   | 20   | 21   | 22   | 23   | 24   | 25   | 26   | 27   | 28   | 29   | 30   | 31   | 32   | 33   | 34   | 35   | 36   | 37   | 38   | 39   | 40   | 41   | 42   | 43   | 44   | 45   | 46   | 47   | 48   | 49   | 50   | 51   | 52   | 53   | 54   | 55   | 56   | 57   | 58   | 59   | 60   | 61   | 62   | 63   | 64   | 65   | 66   | 67   | 68   | 69   | 70   | 71   | 72   | 73   | 74   | 75   | 76   | 77   | 78   | 79   | 80   | 81   | 82   | 83   | 84   | 85   | 86   | 87   | 88   | 89   | 90   | 91   | 92   | 93   | 94   | 95   | 96   | 97   | 98   | 99   | 100  | 101  |
|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|---
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.\textsuperscript{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,\textsuperscript{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.\textsuperscript{4}  Furthermore, the aI5\gamma RNA exists in a collapsed state at physiological Mg\textsuperscript{2+} levels, but can be driven toward the native state in the presence of proteins.\textsuperscript{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 \textit{in vivo}.

\subsection*{4.2.1 Prospective Studies of Putative Riboswitches}

Bioinformatics-based identification has demonstrated the utility and abundance of riboswitches across prokaryotes and yeast\textsuperscript{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\textsuperscript{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 \textit{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.\textsuperscript{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.\textsuperscript{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. 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


