Mechanism of eubacterial 6S RNA release and Development of RNA Mango toolkit to study Ribonucleoprotein complexes

by
Shanker Shyam Sundhar Panchapakesan
B.Tech., SASTRA University, 2010

Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

in the
Department of Molecular Biology and Biochemistry Faculty of Science

© Shanker Shyam Sundhar Panchapakesan 2017
SIMON FRASER UNIVERSITY
Spring 2017

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, education, satire, parody, criticism, review and news reporting is likely to be in accordance with the law, particularly if cited appropriately.
Approval

Name: Shanker Shyam Sundhar Panchapakesan
Degree: Doctor of Philosophy
Title: Mechanism of eubacterial 6S RNA release and Development of RNA Mango toolkit to study Ribonucleoprotein complexes

Examiner Committee: Chair: Jenifer Thewalt
                     Professor

Peter Unrau
Senior Supervisor
Professor

Dipankar Sen
Supervisor
Professor

Fiona Brinkman
Supervisor
Professor

Timothy Audas
Internal Examiner
Assistant Professor
Department of Molecular Biology and Biochemistry

Ronald Breaker
External Examiner
Professor
Department of Molecular, Cellular and Developmental Biology
Yale University

Date Defended/Approved: April 13, 2017
Abstract

Ribonucleic acids (RNA) perform diverse biochemical functions in cells ranging from being an intermediate in the flow of genetic information, catalyst performing crucial reactions and as regulators of several processes.

In bacteria, 6S RNA is a global transcription regulator that binds and inhibits housekeeping RNA polymerase holoenzyme (core polymerase + $\sigma^{70}$) under low nutrient conditions and rescues transcription in high nutrient conditions by the synthesis of a short product RNA (pRNA) using itself as a template. I show that a kinetic intermediate containing 6S RNA:core enzyme complex, that is enhanced by the formation of a phylogenetically conserved ‘release’ hairpin arises during 6S RNA release. Using nucleotide feeding experiments to slow down the release and a 6S RNA mutant which precludes the hairpin formation, I found the release process involves intrinsic ‘scrunching’ type mechanism that is modulated by the ability to form a release hairpin during the process of 6S RNA release.

Given the importance of RNA in regulating various cellular processes, a fluorescence tool to track RNA in real time is limiting as RNA lacks intrinsic fluorescence. The Unrau lab has in vitro selected RNA Mango aptamer that binds thiazole orange with nanomolar affinity while increasing its fluorescence up to 1100 fold. To elucidate how this small aptamer exhibits such properties, which make it particularly well suited for studying low-copy cellular RNAs, we, in collaboration with D’Amaré’s lab, determined its co-crystal structure, discovering a three-tiered G-quadruplex connected to a duplex through a GAAA tetraloop-like junction. By combining the compact RNA Mango aptamer with a fluorogenic thiazole orange desthiobiotin (TO1-Dtb) ligand I have created a Mango toolkit that simultaneously enables the purification and characterization of endogenous cellular RNPs in vitro.

Keywords: Bacterial global transcription regulation, 6S RNA, RNA Mango, Fluorescent RNA aptamer, RNA G Quadruplex, RNP Purification
Whatever actions I do with my body, mind and intellect,
I dedicate unto Him.
Acknowledgements

I would like to thank my senior supervisor Dr. Peter Unrau for all the help he has provided since I joined his lab. His support throughout my graduate career helped me to grow as a better scientist. Thanks are due to my supervisory committee members Dr. Dipankar Sen and Dr. Fiona Brinkman for discussions and suggestions throughout my PhD career.

I would like to thank past and present Unrau lab graduate students Sunny Jeng, Razvan Cojocaru, Amir Abdolahzadeh, Mariana Oviedo, Lena Dolgosheina, Neil Dobson, Lyssa Martin and all the Undergraduate and high school students for all the help they rendered in terms of moral support, meaningful discussions and trouble shooting experiments. Thanks are also due to various funding agencies that directly or indirectly supported my research.

I would like to thank Dr. Aaron Hoskins – University of Wisconsin, Dr. Matthew Ferguson and Dr. Eric Hayden – Boise State University for their collaboration on demonstrating the potential of RNA Mango purification tool I have developed. I would like to also thank Dr. Ash Parameswaran and Shruthi Menon, from the School of Engineering, SFU for giving me an opportunity to work with them on the development of electrochemical sensors for detecting bacteria. I would like to thank the MBB office staff, teaching lab staff for all their support. Special thanks to Jenny Lum, teaching lab for lending instruments/chemicals on an urgent basis.

I would like to thank my family, especially my parents Dr. S. Panchapakesan and Dr. B. Shanthi for their love, constant moral support and encouragement. I would also like to thank my fiancé Mahathi Venkataraman and her family for patiently waiting for me to finish my PhD. I would like to thank all my friends in Vancouver, who have supported me throughout my stay and making it memorable.

Last but not least, I would like to thank the Almighty and the Guru lineage for giving me strength and support in terms of their teachings for all the endeavours I undertake.
Table of Contents

Approval.............................................................................................................. ii
Abstract........................................................................................................... iii
Dedication.......................................................................................................... iv
Acknowledgements........................................................................................... v
Table of Contents.............................................................................................. vi
List of Tables...................................................................................................... ix
List of Figures .................................................................................................... x
List of Acronyms ............................................................................................... xiii

Chapter 1. Introduction ...................................................................................... 1
1.1. Regulatory RNAs .......................................................................................... 7
  1.1.1. Riboswitches ............................................................................................. 7
  1.1.2. Small regulatory RNAs ............................................................................. 10
  1.1.3. CRISPR guide RNAs .............................................................................. 12
  1.1.4. Regulatory RNA – Protein complexes ................................................... 13
1.2. Bacterial Transcription ............................................................................... 16
  1.2.1. Bound RNAP Closed promoter complex (RPC) ...................................... 19
  1.2.2. Promoter melting and RNAP Open promoter complex (RPO) .............. 21
  1.2.3. Formation of the elongation complex by abortive transcription .......... 21
1.3. Bacterial Transcription Regulation – Sigma switching ............................. 22
  1.3.1. (p)ppGpp and DskA: ............................................................................. 23
  1.3.2. Rsd: ........................................................................................................ 27
1.4. The 6S RNA: ................................................................................................ 27
  1.4.1. Structural Features of the widely conserved 6S RNA: ......................... 29
  1.4.2. 6S RNA in other bacteria ...................................................................... 30
  1.4.3. Multiple 6S RNAs B. subtilis: ................................................................. 31
  1.4.4. B2 RNA – an RNAP binding RNA in mammalian cells ...................... 32
1.5. RNA splicing................................................................................................ 33
1.6. Fluorescence as a tool for following RNA processes ................................. 38
1.7. RNA complex purification using high affinity fluorescent RNA aptamer tags ................................................. 40
  1.7.1. The importance of high affinity tags in protein complex purification ...... 41
  1.7.2. Existing RNA-Protein Complex purification strategies ........................ 43

Chapter 2. E. coli 6S RNA release from RNA polymerase requires sigma-70 ejection by scrunching, and is orchestrated by a conserved RNA hairpin .................................................................................................................. 47
2.1. Introduction .................................................................................................. 49
2.2. 6S Secondary Structure Rearrangement upon pRNA induced release ........ 51
2.3. 6S:pRNA:E release intermediate found between bound (S2) and final released state (S4) ................................................................. 53
2.4. Removing 6S RNA hairpin does not preclude σ70 ejection ....................... 59
2.5. Discussion .................................................................................................... 61
2.6. Methods: ...................................................................................................... 65
  2.6.1. DNA Constructs: ................................................................................... 65
Chapter 3. Structural characterization and development of biological applications for RNA Mango ................................................................. 69
3.1. INTRODUCTION ........................................................................... 71
  3.1.1. RNA Mango I/TO1-Biotin complex ........................................ 73
  3.1.2. Fluorescent and binding properties of RNA Mango to TO3-Biotin ... 76
  3.1.3. The RNA Mango G-quadruplex is three-tiered ......................... 78
  3.1.4. A tetraloop-like junction connects the quadruplex and the duplex ... 81
  3.1.5. Circular permutation of RNA Mango I .................................. 82
3.2. Using bipartite RNA Mango to develop a fluorescent turn ‘ON’ system for RNA imaging ................................................................. 84
  3.2.1. Rationale ............................................................................... 84
  3.2.2. Bipartite construct with Mango II .......................................... 89
3.3. Integration of RNA Mango into the bacterial 6S transcriptional control RNA .......................................................... 93
3.4. Conclusion: ................................................................................ 94
3.5. MATERIALS AND METHODS .................................................. 96
  3.5.1. RNA synthesis: ..................................................................... 96
  3.5.2. Fluorescence measurement .................................................. 96
  3.5.3. Integration of RNA Mango I into bacterial 6S RNA ............... 96
  3.5.4. In vivo expression of 6S RNA Mango ................................... 97

Chapter 4. A Multipurpose RNA Toolset for Ribonucleoprotein Purification and Characterization ......................................................... 98
4.1. INTRODUCTION .......................................................................... 100
4.2. RNA Mango for native batch RNP purification ................................ 101
4.3. Fluorescence tracking of the Mango purified RNP complexes facilitates further purification ......................................................... 103
4.4. Purification of yeast U1 snRNP using RNA Mango ..................... 105
4.5. Use of RNA Mango in Multi-Wavelength Fluorescence EMSAs of 6S RNP Assembly and Disassembly ...................................... 108
4.7. CONCLUSION ........................................................................... 111
4.8. MATERIAL AND METHODS ..................................................... 112
  4.8.1. Synthesis of Desthiobiotin Derivatives of Thiazole Orange Acetate ... 112
  4.8.2. Determination of the Binding Affinity of RNA Mango for TO1-Dtb and TO3-Dtb ligands .......................................................... 112
  4.8.3. Construction of p6SRDM_T7 plasmid ................................... 113
  4.8.4. Expression of the 6S<sup>RDM</sup> RNA in E. coli and Extract Preparation ... 113
  4.8.5. Preparation of TO1-Dtb derivatized Streptavidin Agarose Affinity Resin ..................................................................................... 113
  4.8.6. Purification of 6S RNA<sup>RDM</sup> RNP Complexes ...................... 114
4.8.7.  $6S^{RDM}$ RNP Complex fractionation by Size Exclusion Chromatography (SEC) ................................................................. 114
4.8.8.  Mass Spectrometry Analysis of $6S^{RDM}$ RNP Complexes and Commercial RNAP.......................................................... 114
4.8.9.  Yeast Strain and Extract Preparation .................................................. 115
4.8.10. SNAP labeling of yeast whole cell extract ........................................ 115
4.8.11. Purification of U1M-Associated snRNPs ........................................... 116
4.8.12. RNA analysis by RT-Primer Extension and Solution Hybridization ...... 116
4.8.13. Partial Purification of eGFP-HE ..................................................... 116
4.8.14. Two Color Gel-imaging Analysis of the Mango and eGFP tagged $6S$:RNAP Complex ......................................................... 117
4.8.15. Two-photon Spectrum and Cross-Correlation Experiments Sample Conditions ........................................................................... 118
4.9.  ACKNOWLEDGEMENTS .................................................................. 119

References .................................................................................................. 120

Appendix A  Mass spectrometry analysis of Biotin eluate .............................. 133

Appendix B Mass spectrometry analysis of SEC purified $6S^{RDM}$:HE ........... 138

Appendix C Mass spectrometry analysis of commercial HE .......................... 139
List of Tables

Table 1.1 Regulatory RNAs and their protein-binding partners ........................................ 14
Table 1.2 Binding affinities of various σ factors to the RNAP core enzyme (E) and number of molecules and fold increase of various σ factors in *E. coli* during exponential phase and stationery phase (Maeda et al., 2000) ................................................................. 23
Table 1.3 Protein components of yeast snRNPs ................................................................. 35
Table 1.4 Existing fluorescent RNA aptamer–fluorophore systems ................................. 40
Table 1.5 Commonly used protein tags and their respective ligands, along with tag size and $K_D$ for binding ........................................................................................................ 42
Table 1.6 RNA aptamers used in the purification of RNP complexes ............................... 45
Table 3.1 RNA aptamer/Fluoromodule Complex Properties ........................................... 72
Table 3.2 RNA Mango variants ......................................................................................... 89
List of Figures

Figure 1.1 Schematic structure of an RNA fragment with sequence
5′…AGCU…3′ .................................................................................................. 2
Figure 1.2 Base and Sugar conformation ...................................................................... 3
Figure 1.3 RNA secondary structure ........................................................................... 5
Figure 1.4 G quadruplex tertiary structure of RNA .......................................................... 6
Figure 1.5 Common ligands sensed by riboswitches ......................................................... 8
Figure 1.6 Organization of riboswitch RNAs ................................................................. 9
Figure 1.7 Gene Arrangement and Regulatory Functions of Base Pairing
Regulatory RNAs ....................................................................................... 10
Figure 1.8 Model for gene silencing by SgrS sRNA ........................................................... 11
Figure 1.9 Mechanism of CRISPR mediated bacterial immunity ........................................ 13
Figure 1.10 6S RNA consensus secondary structure compared with open
promoter DNA ............................................................................................ 15
Figure 1.11 X-Ray crystal structure of E. coli RNA Polymerase with σ^70 subunit
showing the crab-claw structure with the claw formed between β
and β′ subunit (4yg2) .................................................................................. 16
Figure 1.12 Structural transitions during the steps of transcription initiation .................... 18
Figure 1.13 Domain organization, promoter recognition and structural
organization of the σ^70 family ................................................................. 20
Figure 1.14 Summary of the proposed isomerization steps that form the open
promoter complex (RP_0) after recruitment of RNAP to form an
initial complex at the promoter (RP_c) ......................................................... 21
Figure 1.15 (p)ppGpp and regulation of RNAP .............................................................. 25
Figure 1.16 Schematic model showing the two binding sites of ppGpp (Ross et
al., 2013) .................................................................................................. 26
Figure 1.17 Mechanism of 6S RNA action ................................................................. 29
Figure 1.18 Secondary structure and phylogenetic conservation of 6S RNA ..................... 30
Figure 1.19 Two 6S RNA homologues in B. subtilis ........................................................ 31
Figure 1.20 Two-Step mechanism of eukaryotic splicing ............................................... 34
Figure 1.21 Canonical cross-intron assembly and disassembly pathway of the
U2-dependent spliceosome ........................................................................ 36
Figure 1.22 Compositional dynamics of the yeast spliceosome ......................................... 37
Figure 1.23 Schematics of RNA Spinach and RNA Mango ............................................. 40
Figure 2.1 Structural states observed with native gel electrophoresis during 6S
RNA binding and release ........................................................................ 51
Figure 2.2  Presence of hairpin in the released 6S:pRNA complex.......................... 53
Figure 2.3  6S RNA release conditions......................................................................... 54
Figure 2.4  A 6S:pRNA:σ70 intermediate state (S3) revealed by nucleotide feeding experiments.............................................................. 56
Figure 2.5  pRNA is present in the S3 and S4 states of the 6S RNA release complex........................................................................................................... 57
Figure 2.6  Dynamically populating the release intermediate 6S RNA S3 state by titrating AU dinucleotide and GTP concentrations.................................................................................. 58
Figure 2.7  Anti-σ70 antibody supershift assay showing S3 state is devoid of σ70........... 59
Figure 2.8  Abolishing the 6S RNA hairpin increases pRNA length and release time.............................................................. 60
Figure 2.9  Secondary structure changes for the E. coli (left) verses B. subtilis (right) between the S1 (initial unbound) and S4 (final released state).................... 64
Figure 2.10  Model summarizing the major steps of E. coli 6S RNA release (bottom) relative to those of transcriptional initiation (top).................................................. 65
Figure 3.1  RNA Mango I has a high binding affinity to TO1-Biotin. ......................... 74
Figure 3.2  Fluorescent properties of the RNA Mango I:TO1-Biotin complex.............. 75
Figure 3.3  Fluorescent properties of the RNA Mango complex.................................. 76
Figure 3.4  RNA Mango has a high binding affinity to TO3-Biotin.................................. 78
Figure 3.5  Overall structure of RNA Mango I in complex with TO1-Biotin.................. 80
Figure 3.6  The duplex-quadruplex junction of RNA Mango I resembles a GAAA tetraloop................................................................................................................. 81
Figure 3.7  Effect of circular permutation of G-Quadruplexes in RNA Mango .......... 83
Figure 3.8  Imaging scheme using bipartite RNA Mango............................................. 84
Figure 3.9  Schematic representation of different fluorogenic probes for the detection of mRNAs. Donor (green), quencher (gray) and acceptor (yellow) dyes are attached to the probes and interact as indicated. From Tyagi, 2009. .............................................................................................................. 85
Figure 3.10  Design of bipartite Mango system ............................................................. 86
Figure 3.11  Initial screening to find optimal “switch” and the “trigger” constructs for RNA Mango I based bipartite detection system.............................................. 87
Figure 3.12  Bipartite mango probes for b-actin mRNA................................................ 88
Figure 3.13  Bipartite construct with Mango II and inhibitory stem............................. 90
Figure 3.14  Effect of inhibitory stem length on contrast and rate of fluorescence emergence .......................................................................................................................... 91
Figure 3.15  General RNA Mango based bipartite construct. See text for more explanation..................................................................................................................... 92
Figure 3.16  RNA Mango allows monitoring of 6S:holoenzyme complex formation and purification of 6S RNA from total RNA. .......................... 94

Figure 4.1  Native purification of RNP complexes using the fluorescent, high affinity Mango toolset. ................................................................. 102

Figure 4.2  Mango mediated purification of 6S$^{\text{RDM}}$ ......................................................... 103

Figure 4.3  SEC purification of the 6S$^{\text{RDM}}$.HE RNP ........................................................... 104

Figure 4.4  Confirmation of integration of U1$^M$ into yeast genome ........................................ 106

Figure 4.5  Native purification of yeast U1 snRNP complex using Mango ................................ 107

Figure 4.6  Multi-wavelength fluorescence characterization of the 6S$^M$.HE RNP. ...... 109

Figure 4.7  Two-photon fluorescence correlation spectrum of 6SM:TO3 and eGFP-RNAP ................................................................. 110

Figure 4.8  Fluorescence cross correlation spectroscopy (FCCS) analysis of 6S$^M$:TO3-Dtb: eGFP-HE complex .................................................... 111
### List of Acronyms

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>BPS</td>
<td>Branch point site</td>
</tr>
<tr>
<td>CRISPR</td>
<td>Clustered regularly interspaced short palindromic repeats</td>
</tr>
<tr>
<td>DBSB</td>
<td>double-psi β-barrel</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylene diamine tetra acetate</td>
</tr>
<tr>
<td>EMSA</td>
<td>Electrophoretic mobility shift assay</td>
</tr>
<tr>
<td>FCCS</td>
<td>Fluorescence cross correlation spectroscopy</td>
</tr>
<tr>
<td>FCS</td>
<td>Fluorescence correlation spectroscopy</td>
</tr>
<tr>
<td>FMN</td>
<td>Flavin mononucleotide</td>
</tr>
<tr>
<td>FP</td>
<td>Fluorescent protein</td>
</tr>
<tr>
<td>FRET</td>
<td>Fluorescence Resonance Energy Transfer</td>
</tr>
<tr>
<td>GlcN6P</td>
<td>Glucosamine-6-phosphate</td>
</tr>
<tr>
<td>HE</td>
<td>Holoenzyme</td>
</tr>
<tr>
<td>HEPES</td>
<td>(4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid)</td>
</tr>
<tr>
<td>IPTG</td>
<td>Isopropyl β-D-1-thiogalactopyranoside</td>
</tr>
<tr>
<td>LB</td>
<td>Lysogeny broth</td>
</tr>
<tr>
<td>mRNA</td>
<td>Messenger RNA</td>
</tr>
<tr>
<td>NTP</td>
<td>Nucleotide Tri Phosphate</td>
</tr>
<tr>
<td>ORF</td>
<td>Open reading frame</td>
</tr>
<tr>
<td>RBS</td>
<td>Ribosomal binding site</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>RNAP</td>
<td>RNA Polymerase</td>
</tr>
<tr>
<td>RNP</td>
<td>RNA Protein complex</td>
</tr>
<tr>
<td>Rpc</td>
<td>Closed promoter complex</td>
</tr>
<tr>
<td>RPo</td>
<td>Open promoter complex</td>
</tr>
<tr>
<td>RRM</td>
<td>RNA recognition motif</td>
</tr>
<tr>
<td>rRNA</td>
<td>Ribosomal RNA</td>
</tr>
<tr>
<td>SAM</td>
<td>S-adenosylmethionine</td>
</tr>
<tr>
<td>SBP</td>
<td>Streptavidin binding peptide</td>
</tr>
<tr>
<td>SD</td>
<td>Shine Dalgarno</td>
</tr>
<tr>
<td>SDS</td>
<td>Sodium dodecyl Sulphate</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>------------------------------------</td>
</tr>
<tr>
<td>TEC</td>
<td>Transcription elongation complex</td>
</tr>
<tr>
<td>TIC</td>
<td>Transcription initiation complex</td>
</tr>
<tr>
<td>TPP</td>
<td>Thiamine pyrophosphate</td>
</tr>
<tr>
<td>tRNA</td>
<td>Transfer RNA</td>
</tr>
<tr>
<td>TSS</td>
<td>Transcription start site</td>
</tr>
<tr>
<td>UTR</td>
<td>Untranslated region</td>
</tr>
<tr>
<td>ZBP</td>
<td>Zipcode binding protein</td>
</tr>
</tbody>
</table>
Chapter 1. Introduction

In living organisms, deoxyribonucleic acid (DNA) stores genetic information. The information stored in DNA is converted into proteins, which regulate and catalyze many biochemical reactions. The flow of genetic information from DNA to proteins, involves an important intermediate called ribonucleic acid (RNA) (Crick, 1970). RNA is broadly divided into two major classes, coding RNA or messenger RNA (mRNA) that encodes for the amino acid sequence of proteins in codon triplets and non-coding RNA, which does not code for proteins but has other roles including catalysis of several important biochemical reactions such as protein synthesis, mRNA splicing and regulation of cellular processes such as transcription and translation (Cech and Steitz, 2014).

Like protein, RNA has primary, secondary and tertiary structures, which are important for their coding, catalytic and regulatory functions. Nucleotides are the basic repeating units of the RNA polymer. Nucleotides consist of three parts: a nitrogenous base, a ribose sugar and phosphodiester backbone (Figure 1.1). In the repeating units, the ribose sugar is phosphorylated on its C5’ position and substituted by one of the bases at C1’. The primary structure of RNA is the linear sequence of the arrangement of the bases on the phosphate backbone from 5’-3’. There are four types of natural nitrogenous RNA bases. They are adenine (Ade, A), guanine (Gua, G), cytosine (Cyt, C) and uracil (Ura, U). A and G are called purines (R) and C and U are called pyrimidines (Y). N9 of purines and N1 of pyrimidines are connected to C1’ of ribose through a glycosidic linkage, to make a nucleoside.
Figure 1.1  Schematic structure of an RNA fragment with sequence
5’...AGCU...3’
The nitrogenous bases are numbered in blue. The ribose sugar is numbered red. The chain
direction 5’ to 3’ is shown by an arrow.

The base can adopt two major orientations relative to the glycosidic linkage,
called syn and anti. In the syn conformation, the bulk of the bases i.e., the six membered
ring in case of purines and O2 in case of pyrimidines is pointing towards the sugar while
in the anti conformation, it points away from the sugar (Figure 1.2A). The five membered
furanose ring of ribose sugar is generally non-planar and can be puckered in an
envelope form (E) having four atoms in a plane and the fifth atom out of plane or in a
twist (T) form with three atoms on one plane and two adjacent atoms displaced on the
opposite sides (Figure 1.2B). With respect to the C5’ of the ribose, the atom(s) that are
out of plane can assume either an endo or exo puckering. If the out of plane atom is on
the same side as C5’, it is called \( \text{endo} \) and if it is on the opposite side, it is called \( \text{exo} \) (Figure 1.2B). These conformations have important ramifications for RNA structure.

\[ \text{Figure 1.2 Base and Sugar conformation} \]

(A) \textit{anti} and \textit{syn} conformation of 5’ adenosine mono phosphate. (B) Ribose sugar with various sugar puckering as labelled.

The primary sequence is copied directly from the DNA template during the process of transcription. The RNA transcript thus synthesized can undergo various modifications including template independent addition of a cap (like 7me-G) onto the 5’ end in eukaryotes, template independent addition of poly adenosine at the 3’ end and
modified nucleotides such as pseudouridine, dihydrouridine, inosine, hypoxanthine, xanthine depending upon the domain of the organism through RNA editing (Karijolich and Yu, 2015). The linearity of the primary structure is important for seamless transcription of DNA genetic code into RNA and its translation to proteins. Primary sequence is also responsible for specific interaction of RNA to other biomolecules including RNAs and proteins. For instance, in bacteria and archaea, the Shine-Dalgarno (SD) which is present ~8 nucleotides upstream of the start codon binds to 16S ribosomal RNA (rRNA) in the ribosome which is important for the recruitment of ribosomes to mRNA and aligning the active site of the ribosomes to the start codon (Shine and Dalgarno, 1974).

The secondary structure of RNA involves interaction between the bases through hydrogen bonding. There are three edges present in the RNA with the potential to form hydrogen bonding. They are Watson and Crick edge, Hoogstein edge and sugar edge as shown in Figure 1.3A. Interactions through Watson and Crick edge is present in about 60% of the interaction in structured RNAs (Leontis and Westhof, 2001). In general, adenine pairs with uracil forming two hydrogen bonds and cytosine pairs with guanine forming three hydrogen bonds through Watson and Crick base pairing (Figure 1.3B). Formation of intramolecular double stranded stem through Watson and Crick base pairing is the simplest secondary structure present in the RNA. The stem can be present along with a loop in a simple stem loop structure or can have bulges connecting a series of stem loops (Figure 1.3C). Apart from Watson and Crick base pairing scheme, wobble base pairing (such as between G and U) also uses Watson and Crick edges for hydrogen bonding. Hoogstein base pairing involving N7 position of purines as hydrogen bonding acceptor also occurs during unconventional interactions such as triplex and quadruplex formation in which three and four nucleic acids interact respectively. The ability to form secondary structure is important for several regulatory processes such as transcription termination, translation regulation.
Figure 1.3  RNA secondary structure
A. Adenosine monophosphate showing Watson and Crick, Hoogstein and sugar edges that can form hydrogen bonding with other nucleotides to form RNA secondary structure. B. Watson and Crick basepairing pattern between Adenine and Uracil (top) and Guanine and Cytosine (bottom). C. Widely occurring RNA secondary structures. Thick lines represent RNA polymer. Hydrogen bonding is shown with thin lines. Reprinted by permission from Macmillan Publishers Ltd: Nature Reviews Molecular Cell Biology (Tian et al., 2004), copyright (2004).

Tertiary structure is the three-dimensional structure of RNA forming distinct structural motifs. Structural studies involving X-ray crystallography and cryo-electron microscopy is the primary source of information about tertiary structures. Metal ions such as K+, Mg2+ bind and stabilize the tertiary structure of RNAs. One of the important tertiary structures of RNA that has structural and regulatory functions is RNA G quadruplexes (Agarwala et al., 2015). G quadruplexes are formed, when four guanines form a planar...
structure forming hydrogen bonding with both Watson and Crick and Hoogstein edges. Depending upon the directionality of the adjacent RNA strands, G quadruplexes can be parallel or anti-parallel (Figure 1.4). Several G quadruplexes present in a wide variety of 5’ Untranslated regions (UTRs) of mRNA inhibit translation. Apart from translation regulation, G quadruplexes have been shown to regulate mRNA processing, splicing, mRNA localization among others (Song et al., 2016). G quadruplex is also found to be present in synthetic RNA in vitro selected against binding fluorescent dyes such as RNA spinach (Warner et al., 2014) and RNA Mango (Trachman et al.).

Figure 1.4  G quadruplex tertiary structure of RNA
(A) Schematic representation of G quadruplex showing hydrogen bonding interactions in the Watson and Crick edge and Hoogstein edge between four guanine bases. The metal cation shown in the centre as circle would be sandwiched between two layers of G quadruplexes. (B) Different types of strand orientation possible in three layered G quadruplex containing RNA. The orientation is as mentioned below. The red dot represents Guanine residue, the arrows represent directionality from 5’ to 3’ and the dotted lines is used to connect the G residues participating in a particular plane.
1.1. Regulatory RNAs

Having introduced important basic ideas and nomenclature about the structure of RNA, let us focus on one theme of this thesis, regulatory RNAs. Non-coding RNAs can be catalytic or regulatory or have other functions than being an intermediate in the protein synthesis process. Catalytic RNAs include the bacterial 23S ribosomal RNA (rRNA), which catalyzes the protein synthesis process (Nissen et al., 2000), the eukaryotic U6 small nuclear RNA (U6 snRNA), which catalysis mRNA splicing (Valadkhan and Manley, 2001) and are not the focus of this thesis. Regulatory RNAs, on the other hand, regulate several important biochemical reactions. A brief introduction to the regulatory RNAs is given below.

1.1.1. Riboswitches

There are different classes of regulatory RNAs depending upon which molecules the RNAs bind, the process, which it regulates, and how the regulation happens. The first class called riboswitches, are present in the 5′ end of the mRNAs they regulate (Mandal and Breaker, 2004). These riboswitches act as molecular sensors that sense and respond to small molecules, by definition, in the microbial environment by directly binding to the molecule and regulating gene expression. The riboswitches have two domains: an aptamer domain and an expression platform (Mandal and Breaker, 2004). The aptamer domain binds to the small molecule ligand being detected. Riboswitches are classified based upon the ligand it binds. The ligands being detected include coenzyme B12 (Johnson et al., 2012), thiamine pyrophosphate (TPP) (Serganov et al., 2006), S-adenosylmethionine (SAM) (Heppell et al., 2011), purines (Batey, 2012), flavin mononucleotide (FMN) (Pedrolli et al., 2015), L-lysine (Garst et al., 2008), glycine (Mandal et al., 2004) and glucosamine-6-phosphate (GlcN6P) (Ferré-D’Amaré, 2011) (Figure 1.5). The ligand-binding domain of the different variants in the same class of riboswitch is highly conserved, suggesting very early evolutionary origins. The widespread nature of riboswitches that sense TPP, coenzyme B12 and FMN, which are proposed to be the relics of RNA world (Gilbert, 1986; Joyce, 1991), suggests that these
Riboswitches could have a very ancient origin that could go back to the RNA world (Breaker, 2012). Once the ligand binds to the aptamer domain, the expression pattern of the RNA being regulated changes due to structural changes in the expression platform.

![Diagram of common ligands sensed by riboswitches]

**Figure 1.5** Common ligands sensed by riboswitches.

The expression profile of the RNA is normally modulated by changes in the transcription there by regulating the levels of RNA, translation there by regulating the protein synthesis and alternative splicing thereby the process of making two alternate mRNA from the same pre mRNA by differential cleaving. Transcription is controlled by the inhibition or formation of a terminator hairpin, which destabilizes RNA Polymerase (RNAP) and DNA template interaction leading to the dissociation of RNAP while translation is regulated by regulating the availability of the ribosomal binding site (RBS). Splicing, the process in which the pre-mRNAs that are transcribed from DNA are edited
into correct protein coding sequences, on the other hand is regulated by making the splice site – the nucleotides recognized by the spliceosome for cleavage available or not (Figure 1.6). Unlike the ligand-binding domain, the expression platform is not highly conserved.

**Figure 1.6** Organization of riboswitch RNAs.
(A) General architecture of a TPP-sensing riboswitch. The highly conserved aptamer domain binds the ligand, and the expression platform resides downstream of the aptamer but upstream of the adjoining open reading frame. (B) Transcription termination. Some riboswitches induce formation of an intrinsic transcription terminator upon ligand binding. (C) Translation inhibition. Some riboswitches induce formation of a helix that sequesters the RBS, thereby reducing efficiency of translation initiation. (D) Splicing control. The aptamer domain of fungal and plant TPP riboswitches is flanked by consensus splice site sequences, which suggests that TPP binding affects splicing efficiency. Reprinted from (Winkler and Breaker, 2005)
1.1.2. Small regulatory RNAs

The second class of regulatory RNAs called small RNAs (sRNA), regulate gene expression, by directly basepairing to other mRNAs. These sRNAs are either cis-encoded, i.e., transcribed from the opposite strand of the corresponding mRNA thereby having perfect complementarity to the mRNA or can trans-encoded i.e., transcribed from some other regions in the genome and therefore may not be perfectly complementary to the mRNA it regulates. This class of RNA have varied effects such as promoting RNA degradation, site-specific cleavage of the mRNA, transcription inhibition and inhibiting or promoting translation (Figure 1.7).

![Diagram of gene arrangement and regulatory functions of base pairing regulatory RNAs](Figure 1.7)

**Figure 1.7 Gene Arrangement and Regulatory Functions of Base Pairing Regulatory RNAs**

(A) Two possible configurations of cis-encoded antisense sRNAs (red) and their target RNAs (blue) which share extensive complementarity. (Left panel) An sRNA encoded opposite to the 5’ UTR of its target mRNA. Base pairing inhibits ribosome binding and often leads to target mRNA degradation. (Right panels) An sRNA encoded opposite to the sequence separating two genes in an operon (B) Genes encoding trans-encoded antisense sRNAs (red) are located separate from the genes encoding their target RNAs (blue) and only have limited complementarity. Trans-encoded sRNA can block transcription (left panel) and/or induce RNA degradation (middle panel). Trans-encoded sRNA can act positively by preventing the formation of an inhibitory structure, which sequesters the ribosome binding site (RBS) (right panel). Reprinted from Cell, 136/4, Waters and Storz, Regulatory RNAs in bacteria, 615-628, Copyright (2009), with permission from Elsevier.
While the cis-encoded sRNA can interact with the target mRNA without the need for any additional proteins to facilitate binding, trans-encoded sRNAs, due to their inability to form perfect base pairing, requires additional proteins like Hfq for efficient target binding and regulation (Valentin-Hansen et al., 2004). For example, glucose uptake during glucose-6-phosphate accumulation is regulated by trans-encoded sRNA (Aiba, 2007) (Figure 1.8). Sugar transport related sRNA (SgrS) negatively regulates ptsG mRNA which encodes glucose transporter protein PtsG. Upon high glucose-6-phosphate accumulation inside the cell, the expression of SgrS RNA is upregulated by SgrR, which binds DNA and promotes transcription. SgrS RNA, binds to the RNA chaperone Hfq which is in complex with RNase E. Hfq facilitates the basepairing of SgrS to ptsG mRNA, upon which RNase E degrades ptsG mRNA down regulating its expression (Aiba, 2007).

**Figure 1.8  Model for gene silencing by SgrS sRNA.**
Hfq associates with the C-terminal scaffold region of RNase E. SgrS is induced in response to accumulation of glucose-6-phosphate. A specific transcription factor, SgrR, is required for the induction. SgrS forms a ribonucleoprotein complex with the Hfq–RNase E to act on the ribosome-binding site of target mRNAs through imperfect base-pairing. The base-pairing and recruitment of Hfq–RNase E on the target mRNAs leads to translation inhibition and RNase E-dependent rapid degradation of the target mRNAs. Reprinted from Current Opinion in Microbiology, 10/2, Aiba Mechanism of RNA silencing by Hfq-binding small RNAs, 134-139, Copyright (2007), with permission from Elsevier.
1.1.3. CRISPR guide RNAs

Clustered regularly interspaced short palindromic repeats (CRISPR) RNA, providing adaptive immunity to bacteria against bacteriophages (Barrangou et al., 2007) is another major regulatory RNA that has profound applications in gene editing (Jiang et al., 2013). When a bacteriophage infects bacteria, sometimes, a piece of bacteriophage DNA gets inserted into the CRISPR locus of the host (Barrangou et al., 2007) through CRISPR associated 1 (Cas1) – Cas2 complex in the process called immunization. Thereby the infection leaves a genetic footprint in the host genome, which is then used in the immunity stage (Figure 1.9A). During immunity stage, a guide RNA called small CRISPR RNA (crRNA), which is complementary to the bacteriophage DNA are synthesized from the CRISPR locus guides Cas DNA nuclease complex to the viral DNA. The viral DNA is degraded with the help of Cas nuclease thereby providing immunity to the bacteria. Depending upon what Cas proteins are involved and how the guide RNAs are synthesized, there are three types of CRISPR system as outlined in Figure 1.9B (Marraffini, 2015). In type I systems, the pre-crRNA that is transcribed from the CRISPR locus is identified and bound by a complex known as cascade which cleaves it into crRNA. The cascade – crRNA complex then scans the target DNA for protospacer adjacent motif (PAM) – a specific sequence in the DNA that can be recognized by the cascade proteins. Upon binding to PAM Cas3 nuclease is recruited to the complex, which then cleaves DNA in upstream. In the type II system, Cas9 protein identifies and binds pre-crRNA with the help of trans encoded crRNA (tracrRNA). RNAse III then cleaves precrRNA to crRNA. The tracrRNA – crRNA – Cas9 complex then binds PAM sequence in the target DNA, upon which target DNA is cleaved upstream of PAM. In type III systems, the pre-crRNA repeats is identified and processed by cas6 protein, which then loads the RNA into Cas10 complex. The Cas10-crRNA complex binds to the nascent RNA transcript that is being synthesized by RNA Polymerase. After binding the non-template strand is cleaved by Cas10 complex. Of the three systems, the type II system is being exploited successfully for genome editing purposes in a variety of organisms.
### Figure 1.9 Mechanism of CRISPR mediated bacterial immunity

General mechanism of CRISPR system showing immunization and immunity steps. (B) Types of CRISPR systems. Black arrows indicate cleavage sites to convert pre-crRNA to crRNA. Red arrows indicate the cleavage of DNA. PAM motif is shown in green. Reprinted by permission from Macmillan Publishers Ltd: Nature (Marraffini), copyright (2015).

#### 1.1.4. Regulatory RNA – Protein complexes

Another class of non-coding RNA regulate cellular processes in association with proteins. These RNA-Protein complexes regulate a variety of cellular processes including genome maintenance, transcription, RNA editing. Though these regulatory RNAs function through several mechanisms, one thing common in all is that they have a variety of binding partners (Table 1.1).
### Table 1.1 Regulatory RNAs and their binding partners

<table>
<thead>
<tr>
<th>RNA</th>
<th>Known binding partners</th>
<th>Known functions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Acting at genomic / DNA level</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Telomerase RNA component (TERC)</td>
<td>Telomerases</td>
<td>Scaffold for proteins, maintains chromosomal ends</td>
</tr>
<tr>
<td>X-inactive specific transcript (XIST RNA)</td>
<td>Total of 81 proteins, Breast cancer 1 (BRCA1)</td>
<td>X chromosome inactivation, Maintains gene dosage</td>
</tr>
<tr>
<td>CRISPR RNA (crRNA)</td>
<td>Cas nuclease</td>
<td>Bacterial immunity</td>
</tr>
<tr>
<td>HOTAIR RNA</td>
<td>Polycomb repressive complex 2, histone demethylase LSD1</td>
<td>Epigenetic regulation. Implied on cancer metastasis.</td>
</tr>
<tr>
<td>HoxA transcript at the distal tip (HOTTIP) RNA</td>
<td>WD repeat containing protein 5 (WDR5), Histone-lysine N-methyl transferase 2A</td>
<td>H3K9 methylation</td>
</tr>
<tr>
<td>Extra coding CEBPA (ecCEBPA) RNA</td>
<td>DNA Methyl transferase I (DNMT1)</td>
<td>Prevents CpG methylation in DNA</td>
</tr>
<tr>
<td><strong>Acting at transcription level</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6S RNA</td>
<td>Bacterial RNA Polymerase</td>
<td>Global transcription regulation during stationery phase</td>
</tr>
<tr>
<td>7SK RNA</td>
<td>Several proteins including P-TEFb</td>
<td>Inactivates</td>
</tr>
<tr>
<td>B2 RNA</td>
<td>Eukaryotic RNA polymerase II</td>
<td>Regulates RNA polymerase</td>
</tr>
<tr>
<td>T-box</td>
<td>tRNA</td>
<td>Regulation of tRNA synthesis</td>
</tr>
<tr>
<td>Drosophila Rox RNAs</td>
<td>Male specific Lethal (MSL) complex</td>
<td>Inactivation of Positive transcription elongation factor b (P-TEFb)</td>
</tr>
<tr>
<td><strong>RNA editing</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RNase P RNA</td>
<td>Ribonuclease P protein</td>
<td>Pre-tRNA processing</td>
</tr>
<tr>
<td>Uridine rich small nuclear RNAs (snRNAs) (U1, U2, U4, U5, U6)</td>
<td>Various components of spliceosome</td>
<td>RNA splicing</td>
</tr>
<tr>
<td><strong>Protein synthesis associated steps</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ribosomal RNAs (rRNAs)</td>
<td>Ribosomal proteins</td>
<td>Protein synthesis</td>
</tr>
<tr>
<td>IRES</td>
<td>Initiation factors (IF), ribosomes</td>
<td>Translation initiation</td>
</tr>
<tr>
<td>CsrB RNA</td>
<td>Carbon storage regulator protein A</td>
<td>Carbon metabolism regulation in bacteria</td>
</tr>
<tr>
<td><strong>Miscellaneous</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7SL RNA</td>
<td>Various signal recognition proteins</td>
<td>Protein localization</td>
</tr>
<tr>
<td>Gas5 ncRNA</td>
<td>Glucocorticoid receptor</td>
<td>Regulation of apoptosis, tumour suppression</td>
</tr>
<tr>
<td>Y RNA</td>
<td>Ro60 Protein</td>
<td>Initiation of DNA replication</td>
</tr>
<tr>
<td>Vault (Vt) RNA</td>
<td>Vault complex</td>
<td>Drug resistance, post transcription gene silencing</td>
</tr>
</tbody>
</table>
In general the regulatory RNPs can be divided into two major categories. While some of the RNA-protein complexes have intrinsic activities like the Uridine rich small nuclear RNAs (snRNAs) in the eukaryotic spliceosome, in other cases the RNA competes with the protein-binding partner by mimicking some features of the binding partner to bind and antagonize the protein. For example in *E. coli*, the secondary structure of the transcription regulating 6S RNA mimics DNA in an intermediate complex during transcription (Barrick et al., 2005) inside RNA Polymerase hence competes to bind to *E. coli* RNA Polymerase holoenzyme leading to reversible inhibition of the holoenzyme thereby regulating transcription (Wassarman and Saecker, 2006) (Figure 1.10).

![Diagram of DNA template in the open promoter complex with RNA polymerase as described elsewhere (Murakami et al., 2002a). -10 and -35 refers to the regions in DNA identified by the RNA Polymerase holoenzyme. Bottom - Secondary structure model for *E. coli* 6S RNA.]

**Figure 1.10** 6S RNA consensus secondary structure compared with open promoter DNA.

Early in my graduate career, I worked on 6S RNA regulation of *E. coli* RNA polymerase. In order to understand the importance and the mechanism of 6S RNA mediated regulation, understanding the major steps of bacterial transcription is important as it has parallels with the process of transcription initiation. Hence, a brief overview of bacterial transcription is given below.
1.2. Bacterial Transcription

In bacteria a multi domain DNA dependent RNA Polymerase (RNAP) is responsible for transcription. The *E. coli* core enzyme comprises of a total of 5 subunits (Figure 1.11) with two identical $\alpha$ (36.5 kDa) subunits along with $\beta$ (150.6 kDa) and $\beta'$ (155.2 kDa) and a $\omega$ (10 kDa). The core enzyme has a crab-claw structure with the two arms of the claws being $\beta$ and $\beta'$ (Zhang et al., 1999). The $\alpha$ subunits play key roles in enzyme assembly, contacting transcription factors thus regulating the RNAP and contacting upstream promoter DNA. The catalytic centre of the enzyme consists of a $\text{Mg}^{2+}$ binding pocket with absolutely conserved $\text{–NADFDGD–}$ motif (Zhang et al., 1999). The space in between the two claws form a 27 Å wide basic internal central channel and helps to hold on to the template DNA. The $\omega$ subunit wraps around the C terminal domain of the $\beta'$ subunit, suggesting that it acts as a chaperone in the assembly of core RNAP (Zhang et al., 1999). Consistent with the hypothesis that the $\omega$ plays a role in the assembly of RNAP and not the catalysis of NTP addition, 60kDa chaperonin has been shown to bind and assist assembly of the RNAP core in the absence of $\omega$ subunit (Mukherjee et al., 1999). Overall, the structure of the multi subunit RNAP is conserved across all the three domains of life as judged by X-ray crystal structures.

![X-Ray crystal structure of *E. coli* RNA Polymerase with $\sigma^{70}$ subunit showing the crab-claw structure with the claw formed between $\beta$ and $\beta'$ subunit (4yg2)](image)

The core enzyme is the minimal structure required for the catalysis of DNA dependent RNA synthesis, but it is not sufficient to initiate transcription from DNA. For
transcriptional initiation to occur, bacterial RNAPs require another protein factor called the sigma (σ) factor (Burgess et al., 1969). *E. coli* K-12 has about 4,288 genes coded in ~4.6 million base pairs of its genome (Blattner et al., 1997). However, the total number of RNAP core enzyme is estimated to be ~2000 in a growing cell. With the help of σ factors, RNAP identifies which gene to transcribe at the particular time and where to initiate transcription. The σ factor identifies and binds to specific DNA called promoter elements and serves to initiate transcription downstream. *E. coli* has 7 sigma factors of which σ70 is responsible for the majority of transcription during exponential growth and hence it is called the house keeping sigma factor. The core enzyme binds σ70 with 0.26 nM affinity (Maeda et al., 2000). This results in holoenzyme (HE) complex, which then diffuses along the double stranded DNA until it locates and binds its cognate DNA promoter thereby initiating transcription (Guthold et al., 1999). By regulating each σ factor, the genes that are present downstream of the corresponding promoter can in turn be regulated. In bacteria, a total of five structurally stable steps have been identified involving the DNA and the enzyme during transcription, each playing an important role in the process (Figure 1.12). These steps, which take the polymerase from a static complex bound to a promoter to a fully functional elongation complex, and eventual termination are described below.
Figure 1.12  Structural transitions during the steps of transcription initiation
Shown are cross-sectional views of the RNAP holoenzyme (β flap, blue; α, orange; rest of RNAP, gray; catalytic Mg$^{2+}$, yellow sphere), promoter DNA (template strand, dark green; non-template strand, light green; −10 and −35 elements, yellow) and the RNA transcript (red) at the RP$_c$, intermediate (I), RP$_o$ and abortive initiation, end of abortive initiation, promoter clearance and TEC stages of transcription initiation. The view is looking down on top of the β subunit, but with most of β removed, revealing the inside of the RNAP active site channel. Reprinted from Current Opinion in Structural Biology, 13/1, Murakami and Darst, Bacterial RNA Polymerase: the wholo story, 31-39, Copyright (2003), with permission from Elsevier.
1.2.1. **Bound RNAP Closed promoter complex (RPc)**

The first step in the process of transcription is the recognition of DNA promoter by the RNAP holoenzyme. In *E. coli*, the promoter region consists of two parts, a ‘−10’ and a ‘−35’ region of sequence approximately located ~10 and ~35 nucleotides upstream of the transcription start site (TSS). In general bacteria have two distinct families of σ factors called σ\(^70\) family and σ\(^54\) family, based upon the homology with *E. coli* σ factors (Paget, 2015). σ\(^70\) family members consist of up to 4 conserved domains that are separated by flexible linkers (Figure 1.13). They are σR1.1 (σ1 domain), σR1.2, 2.1-2.4 (σ2 domain), σR3.0-3.2 (σ3 domain) and σR4.1-4.2 (σ4 domain). σ2 contacts the -10 promoter region, σ3 domain contacts the extended -10 promoter region and σ4 contacts the -35 promoter region. The σ\(^70\) family is further subdivided into 4 groups based upon the presence or absence of the above said domains as shown in the Figure 1.13. Since σ\(^70\) is responsible for the majority of transcription, most experiments including structure determination by X Ray crystallography have been performed using the σ\(^70\) HE complex. Crystal structures and single molecule studies along with biochemical data with *Thermus* and *E. coli* RNAP complexes have helped in understanding the structural dynamics of DNA template, RNA product and the enzyme in the above complexes during transcription. RNAP σ\(^70\) binds to -10 and -35 regions of DNA promoter defined by the consensus ‘TTGACA’ and ‘TATAAT’ respectively (Harley and Reynolds, 1987) forming the closed promoter complex. Upon σ\(^70\) binding, due to the structural rearrangement, the width of the internal channel narrows from 27 Å to 10 Å (Murakami et al., 2002a) tightly holding the DNA.
Figure 1.13  Domain organization, promoter recognition and structural organization of the $\sigma^{70}$ family

(A) The domain organization of $\sigma$ factors from Groups 1, 3 and 4 are illustrated above $\sigma^{70}$ (Group 1) consensus E. coli promoter DNA. Structural domains are colored: $\sigma_1$, white; $\sigma_2$, green/orange; $\sigma_3$, blue; $\sigma_4$, red. Within each domain, conserved $\sigma$ regions are indicated for Group 1 $\sigma$s. Non-template (NT) strand DNA is colored magenta and template (T) strand cyan, with key consensus promoter elements contacted by $\sigma$ indicated in yellow: “−35”, −35 element; “ext −10”, extended −10 element; “−10”, −10 element; “disc”, discriminator. Transcription initiates at +1. Note that $\sigma_2$ is colored green and orange to distinguish $\sigma$ regions 2.1–2.4 and 1.2. The nonconserved region (NCR; pink) located between 1.2 and 2.1 (pink) is variable in size and structure among Group 1 $\sigma$ factors. (B) Organization of E. coli $\sigma^{70}$ in an RNA polymerase transcription initiation complex. The model was based on the crystal structure of an E. coli transcription initiation complex (PDB: 4YLN). $\sigma^{70}$ domains (surface representation) and promoter DNA (spheres) are colored as in (A), as indicated in the panel. For clarity the $\beta$, 2$\alpha$ and $\omega$ subunits of RNA polymerase are omitted. The model indicates the location of the $\sigma$ finger and its close proximity to nascent RNA (4 nt, yellow) and template strand DNA. From Paget, 2015.
1.2.2. Promoter melting and RNAP Open promoter complex (RPo)

The conversion of RPc to RPo involves a series of processes collectively called isomerization and involves two kinetically significant intermediates I₁ and I₂ together with a third predicted intermediate I₃ (Figure 1.14). A consensus adenine residue at the -11 position is flipped outwards forming the first intermediate I₁ (Heyduk et al., 2006). Immediately DNA from -11 to +2 is unwound. The σ2 region in the σ factor binds to the single stranded non-template region from -11 to -7 (Figure 1.14). With the help of σ factor, template strand is brought into the active site of holoenzyme (Bae et al., 2015) forming an open promoter complex (Figure 1.12 and Figure 1.14). The RPo is capable of binding to NTPs complementary to the template strand and initiate polymerization. The σ₃.₂ loop of σ₇₀ physically occupies the path of the elongating RNA and would pose a steric hindrance for the full RNA extension (Murakami et al., 2002b; Vassylyev et al., 2002).

![Diagram](image)

**Figure 1.14** Summary of the proposed isomerization steps that form the open promoter complex (RPo) after recruitment of RNAP to form an initial complex at the promoter (RPc).

Formation of the closed complex RPc triggers a series of subsequent large-scale conformational changes. The RNAP molecular machine places start-site duplex DNA in the active-site cleft in I₁, opens it to form I₂, and stabilizes the open form by assembling a clamp in I₃ and RPo. Once promoter DNA is open, NTPs can bind, and transcription initiates. I₂ and I₃ are open complexes. α-CTDs are show in cyan, β and β': gray; σ1.1 purple σ1.2 in orange, σ2.1, 2.2, ad 2.4 in yellow and σ3.1 in red. From Ruff et al., 2015.

1.2.3. Formation of the elongation complex by abortive transcription

When an RNA product of ~8-15 nt has been synthesized, in the vast majority of cases, the transcription cycle is aborted because of steric hindrance, leading to the accumulation short RNA abortive initiation products, *in vitro* (Hsu, 2009) as well as *in*
vivo (Goldman et al., 2009). Several models have been proposed for the abortive initiation such as RNAP translocation, RNAP inch worming with a flexible element and DNA scrunching. Single molecule Fluorescence Resonance Energy Transfer (FRET) experiments however have established that the DNA template gets scrunched inside the central canal (Kapanidis et al., 2006) accumulating strain energy upto ~58-75 kcal/mol (Revyakin et al., 2006).

After several rounds of abortive initiation, the nascent RNA is able to push on the $\sigma^{3.2}$ loop, making its way to the exit channel, causes a structural change in the $\sigma^{70}$ interaction with the RNAP (Mukhopadhyay et al., 2001) which sometimes even leads to its ejection. When the full length RNA is synthesized, the transcription terminates either through an intrinsic sequence dependent termination or assisted by a protein factor Rho in Rho dependent termination.

1.3. Bacterial Transcription Regulation – Sigma switching

E. coli has 7 $\sigma$ factors of which $\sigma^{70}$ is responsible for transcription of a majority of house keeping genes. The other sigma factors are, $\sigma^{38}$ responsible for transcription during stationery phase, $\sigma^{54}$ playing a role depending upon nitrogen availability, $\sigma^{32}$ responsible for transcribing heat shock proteins, $\sigma^{28}$ for flagellar gene synthesis and during chemotaxis, $\sigma^{24}$ for the synthesis of heat shock and exocytoplasmic genes and $\sigma^{fci}$ for the synthesis of ferric citrate transport genes (Maeda et al., 2000). By regulating sigma factors the transcription of gene encoded under a specific family of promoters can be tightly controlled. The binding affinities of the $\sigma$ factors to the core enzyme and their relative abundance during exponential phase has been determined and are shown in Table 1.2.
Table 1.2 Binding affinities of various σ factors to the RNAP core enzyme (E) and number of molecules and fold increase of various σ factors in E. coli during exponential phase and stationery phase (Maeda et al., 2000)

<table>
<thead>
<tr>
<th>σ subunit</th>
<th>$K_0$ (nM)</th>
<th>No. of molecules per cell during active growth (x10^3)</th>
<th>Fold increase during starvation or stationery phase</th>
</tr>
</thead>
<tbody>
<tr>
<td>E</td>
<td>-</td>
<td>2.5</td>
<td>1</td>
</tr>
<tr>
<td>$\sigma^{70}$</td>
<td>0.26</td>
<td>7.2</td>
<td>1</td>
</tr>
<tr>
<td>$\sigma^{54}$</td>
<td>0.3</td>
<td>1.1</td>
<td>1</td>
</tr>
<tr>
<td>$\sigma^{38}$</td>
<td>4.26</td>
<td>ND</td>
<td>~1000</td>
</tr>
<tr>
<td>$\sigma^{32}$</td>
<td>1.24</td>
<td>0.12</td>
<td>Not known</td>
</tr>
<tr>
<td>$\sigma^{24}$</td>
<td>2.43</td>
<td>5.5</td>
<td>1</td>
</tr>
</tbody>
</table>

The probability that a specific σ factor associates with the core enzyme is dependent upon the concentrations of the core enzyme, along with the binding affinity with which they interact. As seen from Table 1.2, the total number of core enzyme units available per cell is limiting and is relatively constant throughout the growth states of E. coli. Even though the relative concentration of $\sigma^{38}$ increases by 1000 fold, the levels of $\sigma^{70}$ do not decrease. Given the relatively higher affinity for $\sigma^{70}$ than $\sigma^{38}$ this would suggest that there must be some other mechanism that bacteria adopt to switch the majority of transcription from $\sigma^{70}$ to $\sigma^{38}$ dependent promoters during stationery phase in addition to the increase in $\sigma^{38}$ levels. This process is called ‘transcription switching’ and is achieved through four major molecules in E. coli: (p)ppGpp, DskA, Rsd and the 6S RNA.

1.3.1. (p)ppGpp and DskA:

Guanosine tetraphosphate and guanosine pentaphosphate, collectively known as (p)ppGpp are well known alarmones involved in the stringent response of bacteria during nutrient deprivation (Wagner, 2002). These small molecules are present at basal levels during exponential phase, where amino acids are not limiting (Sarubbi et al., 1988). During stationery phase, when there is nutrient deprivation, ppGpp and pppGpp are rapidly synthesised by RelA or spoT from GDP and GTP respectively (Hauryliuk et al., 2015). (Figure 1.15A). (p)ppGpp decreases life times of all open promoter complexes in general. However, since some promoters such as rnb P1 and P2, PpyrB1, rnrDP1 and
P2, rnb P2, PhisR, rpoD P1 and P2 PleuV form relatively short lived open promoter complexes, it is these promoters that are affected by (p)ppGpp. Therefore, the inhibitory effects of (p)ppGpp is dependent upon the intrinsic kinetic parameters of the promoters (Barker et al., 2001). These short-lived open promoters are responsible for the transcription of rRNA and ribosomal protein genes, which are required for cell proliferation. In *E. coli* (p)ppGpp directly binds to RNAP in the interface of β′ and ω subunits (Zuo et al., 2013) (Figure 1.15B). The nucleobase interacts with the double-psi β-barrel (DBSB) of the β′ subunit and the N-terminus of the ω subunit. While the β′ subunit is responsible for the nucleobase recognition, the ω subunit interacts with the phosphate groups making the ω subunit required for (p)ppGpp mediated regulation. Since the binding cavity is ~30 Å away from the active site of the RNAP, this suggests an allosteric mode of transcription regulation.
Figure 1.15  (p)ppGpp and regulation of RNAP.
(A) The synthesis of guanosine tetraphosphate and guanosine pentaphosphate (collectively referred to as (p)ppGpp) by RelA–SpoT homologue (RSH) family enzymes is shown. The γ-phosphate moiety of GTP and pppGpp are boxed in blue; these moieties are not present in GDP and ppGpp. (B) An overview of ppGpp bound to the E. coli RNAP holoenzyme. The RNAP holoenzyme is shown in a tube and arrow cartoon representation with each subunit colored differently: α I – blue; α II – light blue; β – orange; β’ – green; ω – yellow; σ70 – red. A metal ion at the active site is shown as a magenta sphere, and ppGpp is shown in red as space-filling model. (C) Details of the residues of RNAP that are contacting ppGpp. RNAP subunits are shown as cartoon representation with each subunit colored differently as in panel B. ppGpp and the side chains of the residues that contact ppGpp are shown as ball-and-stick models. Reprinted from Molecular Cell, 50/3, Zuo et al., The Mechanism of E. coli RNA Polymerase regulation by ppGpp is suggested by the structure of their complex, 430-436, Copyright (2013), with permission from Elsevier.
DskA and ppGpp have been shown to synergistically regulate transcription from rRNA promoters (Paul et al., 2004). However, the identification of ppGpp binding site in the RNAP away from DskA binding site, made the synergistic regulation of DskA and ppGpp hard to explain (Chatterji et al., 1998; Toulokhonov et al., 2001). Recently, an RNAP crystal structure showing another binding site for ppGpp at the interface of RNAP and DskA binding site (called site 2, Figure 1.16) was published which help to explain the synergistic transcription regulation (Ross et al., 2013). It is hypothesized that at low ppGpp concentrations, during early stationary phase, ppGpp binds to site 1, and at high ppGpp concentrations, during late stationary phase of bacterial growth, ppGpp binds to Site 2 along with DskA, regulating transcription (Ross et al., 2013).

Figure 1.16  Schematic model showing the two binding sites of ppGpp. Reprinted from Molecular Cell, 62/6, Ross et al. ppGpp Binding to a Site at the RNAP-DksA Interface Accounts for Its Dramatic Effects on Transcription Initiation during the Stringent Response, 811-823, Copyright (2016), with permission from Elsevier.
1.3.2. **Rsd:**

Rsd is an anti-sigma factor, which binds and inhibits free $\sigma^{70}$ in *E. coli* (Jishage and Ishihama, 1998). By inhibiting $\sigma^{70}$, it promotes transcription from alternative sigma factors mainly $\sigma^{38}$ (Mitchell et al., 2007). Overexpression of Rsd protein resulted in up-regulation of genes that are controlled by $\sigma^{38}$ promoters. The amount of Rsd protein reaches its maximum during stationery phase and could be co-purified with $\sigma^{70}$ (Jishage and Ishihama, 1998). Rsd was shown to bind regions 2, 3 and 4 of $\sigma^{70}$ (Jishage et al., 2001). However, deletion of Rsd does not have a detectable effect on the growth rate or gene expression pattern during stationery phase *in vivo*, suggesting other mechanisms by which transcription switching from $\sigma^{70}$ to $\sigma^{38}$ is effected.

1.4. **The 6S RNA:**

6S RNA, one of the first non-coding regulatory RNAs to be sequenced (Brownlee, 1971), widely present in all branches of eubacteria (Barrick et al., 2005). During its discovery, it was known that the 6S RNA participated in a protein complex but the binding partner was unknown until Wassarman showed that it binds and inhibits $E\sigma^{70}$ (Wassarman and Storz, 2000). 6S RNA competes with DNA promoters to bind *E. coli* RNAP with $\sigma^{70}$ ($E\sigma^{70}$). 6S RNA was found to sequester the RNAP HE, which is the dominant sigma factor during the exponential phase of *E. coli* growth. 6S RNA accumulates through out the growth with its concentration peaking at the stationery phase, when the growth medium runs out of nutrients (Wassarman and Storz, 2000). When the cells reach stationary phase, the transcription from the housekeeping sigma factor is shut off and in *E. coli* transcription occurs mainly by $\sigma^{38}$ sigma factor. Thus 6S RNA was hypothesized to help in switching transcription from the house keeping sigma factor to alternate sigma factors such as $\sigma^{38}$ by the physical sequestration of the house keeping HE. By sequestering the housekeeping RNAP HE, 6S RNA helps in efficient usage of cellular resources during nutrient starved conditions. The inhibition of RNAP HE by 6S RNA is relieved when the nutrient conditions improve by the synthesis of a
small product RNA called pRNA, templated by the 6S RNA, by the RNAP HE (Wassarman and Saecker, 2006). Once pRNA mediated release of the RNAP inhibition occurs, the RNAP is now free to associate with the house keeping sigma factor to transcribe genes responsible for proliferation. The fact that, RNA Polymerase, which has evolved to bind DNA to make RNA, binds to 6S RNA and makes an RNA product in the process of global transcription regulation makes it a very interesting non-coding RNA to study. It also demonstrates the transient and highly dynamic nature of the RNA: protein association during the regulatory process.

Since the 6S RNA has been predicted to be present in a wide variety of bacteria including pathogenic bacteria and since this competition with DNA promoter is very unique to bacteria, this process could be potentially targeted in future for the synthesis of antimicrobial agents.
Figure 1.17 Mechanism of 6S RNA action

Levels of free core enzyme (E, large circle) and free sigma factors (σ^{70} and σ^{S} shown) determine the formation rate of Eσ^{70} complex. This complex is depleted by the formation of either a bound 6S RNA:Eσ^{70} complex (top path) or DNA:Eσ^{70} complex (bottom path). This competitive binding provides a 6S RNA-dependent mechanism to distinguish between strong and weak DNA promoters that depend on 6S RNA intrinsic binding affinity and release rate as well as its cellular concentration. From Shephard et al., 2010.

1.4.1. Structural Features of the widely conserved 6S RNA

Mfold, an RNA folding algorithm, predicted an extended double stranded secondary structure with a central bubble for the *E. coli* 6S RNA. Using sequence based BLAST searches and more sophisticated covariance model searches 6S RNA homologues from all the families of eubacteria were found (Barrick et al., 2005). There are about 3611 sequences from 2726 species comprising a wide variety of bacteria has been annotated in the Rfam database. Upon sequence alignment, they all share a similar secondary structure of an extended double stranded RNA with a central bulge mimicking DNA in an open promoter complex (Figure 1.18). The regions of the 6S RNA
are identified throughout this thesis as, upstream helix, central bubble and downstream helix (Figure 1.18)

![diagram of 6S RNA structure](image1.png)

**Figure 1.18  Secondary structure and phylogenetic conservation of 6S RNA**
The 6S RNA biological consensus showing only residues in a minimal functional construct of 6S RNA. This structure was derived from the 6S RNA sequence found in three orders of the gammaproteobacteria namely, the enterobacteriaceae, alteromonadales, and vibronales. Two phylogenetically conserved upstream regions are labeled ‘-35’ and ‘-10’, respectively, for their correspondence to the equivalent regions in a bound DNA promoter complex Ref. Barrick et al. In all panels, primary sequence conservation and base-pair conservation is summarized by the color bar. Nucleotide numbering corresponds to full length *E. coli* 6S RNA. From Shephard et al., 2010.

In general, 6S RNA has three islands of phylogenetically conserved residues (Figure 1.18). They are the ‘-35’ region (in the upstream helix), ‘-10’ region (in the top strand of central bubble) because of their geometrical similarity to the -35 and -10 DNA promoter region. A region in the downstream helix is also conserved along with the transcription start site (TSS). *In vitro* selection identified ‘-35’ region to be important for binding Ec$^{70}$. The importance of the phylogenetic conservation of the region in the downstream helix was not initially fully understood. I discovered that the top strand of the downstream helix is complementary to the ‘-10’ region and can make a hairpin when the pRNA is made. I also found that formation of this release hairpin is important for the rapid release of the RNAP HE *in vitro*. These finding are elaborated in Chapter 2.

1.4.2. 6S RNA in other bacteria

Based upon sequence similarity and secondary structure predictions, 6S RNA homologue was predicted to be in all classes of bacteria. In several cases, using RNA
seq and in vitro binding to RNAP HE, the putative 6S RNA has been biochemically validated such as in Coxiella burnetti (Warrier et al., 2014), Aquificales (Lechner et al., 2014), Rickettsia prowazekii (Schroeder et al., 2016), Legionella pneumophila (Faucher et al., 2010), Aquifex aeolicus (Köhler et al., 2015) In L. pneumophila, it has been found that 6S RNA is important for intracellular multiplication although the sigma factor of the HE to which 6S RNA binds could not be determined (Faucher et al., 2010). Ms1 RNA in Mycobacterium smegmatitis, also found to be present in all Mycobacterium spp. was found through in silico search as a 6S RNA homologue. Interestingly, this RNA was found to interact with RNAP core enzyme without any sigma factor instead of HE like other 6S RNAs (Hnilicová et al., 2014).

1.4.3. Multiple 6S RNAs B. subtilis:

In some bacteria such as B. subtilis, two homologues of the 6S RNA exist (Barrick et al., 2005; Trotochaud and Wassarman, 2005). They have similar secondary structure. In B. subtilis, the two homologues are named 6S-1 and 6S-2 RNAs (Figure 1.19). In vitro, both of the homologues were found to bind house keeping RNAP HE (Eσ^A in B. subtilis) and initiate pRNA synthesis (Beckmann et al., 2011). Interestingly, in vivo, the 6S RNA homologues get expressed during different growth conditions. 6S-1 RNA gets accumulated throughout stationery phase while the 6S-2 RNA remains fairly constant through out different conditions tested. 6S-1 RNA is similar to E. coli 6S RNA in expression pattern and the deletion strains are delayed in their ability to restart growth.
upon nutrient upshift. In addition, 6S-1 RNA was found to be important for timing the sporulation process in *B. subtilis* (Cavanagh and Wassarman, 2013). Cells lacking 6S-1 RNA sporulate earlier and was found to be less efficient in utilizing nutrients. 6S-2 RNA was long thought not to release from Eσ^A through pRNA synthesis (Beckmann et al., 2011). However, it was later shown that *in vitro* pRNA mediated release occurs and through RNA seq and primer extension it was also shown that *in vivo*, pRNA from 6S-2 RNA can be found (Hoch et al., 2016). However, the evolutionary and physiological advantage of having two 6S RNA variants is not completely understood.

### 1.4.4. B2 RNA – an RNAP binding RNA in mammalian cells

Unlike prokaryotes, eukaryotes have three different RNA Polymerases called as RNA polymerase I (Pol I), Pol II and Pol III. Each polymerase is responsible for transcription of a distinct class of RNA, such as Pol I is responsible for the synthesis of all rRNA except 5S rRNA. Pol II is responsible for transcription of all mRNA, some small nuclear RNA (snRNA), small interfering RNA (siRNA) and micro RNA (miRNA) while Pol III is responsible for the synthesis of all tRNA, 5S rRNA, and other small RNAs. Unlike bacterial RNAP, which has five core subunits, eukaryotic RNAP are relatively bigger that Pol I, II and III respectively have 14, 12 and 17 subunits. Eukaryotic RNAPs require transcription factors (TF) for promoter recognition and transcription initiation and this process is tightly regulated through various mechanisms (Cramer, 2008).

In mammalian cells, a non-coding RNA called B2 RNA was found to bind RNAP II and inhibit Pol II transcription (Espinoza et al., 2004). B2 RNA, itself a product of Pol III, binds to an RNA docking site of core Pol II and inhibit early steps of transcription. In contrast to bacterial 6S RNA, which competes with DNA promoters to bind RNA Polymerase, B2 RNA doesn’t prevent promoter binding. B2 RNA was found not to interfere with the binding of transcription factors but prevents the establishment of several important contacts between the RNA promoter and the Polymerase. Physiologically, B2 RNA is hypothesized to inhibit Pol II transcription in response to heat shock (Yakovchuk et al., 2009).
That this phenomenon of RNA directly binding to RNA polymerase so as to regulate transcription, suggests a strong evolutionary advantage to this type of transcription regulation.

The bacterial 6S RNA is an example of a relatively simple RNA protein complex, wherein the RNA regulator participates in one complex. On the other hand, there are important RNA regulators that participate in multiple complexes sequentially as a part of regulatory process such as RNA splicing.

1.5. RNA splicing

RNA splicing is a process in which pre-mRNA gets matured into mRNA by removing ‘introns’. In eukaryotes, a highly dynamic, multi-megadalton spliceosome complex normally catalyzes this process. In most eukaryotes, U2-dependent and U12-dependent spliceosomes coexist (Sharp and Burge, 1997) of which the former is responsible for a majority of the splicing reactions and hence will be briefly discussed.

The splicing site is defined by 3’ splice site (3’SS), 5’ splice site (5’SS) and a branch site (BS). Pre mRNA splicing has two catalytic steps (Figure 1.20). In the first step the 2’ OH of the adenosine in the BS carries out a nucleophilic attack on the 5’SS leading to formation of a lariat. In the second step, the 3’SS is attacked by the 3’OH of the 5’ exon leading to the ligation of the 5’ and 3’ exons thereby forming mRNA and the release of the intron (Wahl et al., 2009). Since the intermediates and the products of the pre mRNA splicing are similar to those of group II self-splicing introns, it was hypothesized that the splicing of the pre mRNA is also RNA based (Wahl et al., 2009).
RNA splicing entails two transesterification steps. In the first, the 2’ hydroxyl group of an intronic adenosine residue (branch-point site, BPS) carries out a nucleophilic attack on the phosphate group between 5’ exon and the intron, generating a 2’-5’ phosphodiester bond and, consequently, a lariat intermediate. In the second step, the free 3’ OH of the 5’ exon attacks the phosphate group between the intron and the 3’ exon, splicing the two exons together and releasing the intron lariat. Reprinted from Trends in Biochemical Sciences, 1/13, Papasaikas and Valcárcel, The Spliceosome: The ultimate RNA chaperone and sculptor, 33-45, Copyright (2016), with permission from Elsevier.

Figure 1.20 Two-Step mechanism of eukaryotic splicing

The U2-dependent spliceosome is assembled from 4 major (small nuclear RNA Protein complexes) snRNPs. They are the U1, U2, U5, U4:U6 snRNPs. The RNA involved are U1, U2, U4, U5 and U6 snRNAs. Table 1.3 shows the snRNP complexes involved in the U2-dependent splicing.
Table 1.3  Protein components of yeast snRNPs

<table>
<thead>
<tr>
<th>SnRNA</th>
<th>Sm proteins</th>
</tr>
</thead>
<tbody>
<tr>
<td>U1</td>
<td>B, D3, G, E, F, D2, D1, Snp1, Mud1, Yhc, Prp39, Prp40, Prp42, Snu71, Nam8, Snu56, Urn1</td>
</tr>
<tr>
<td>U2</td>
<td>B, D3, G, E, F, D2, D1, Lea1, Msi1, Prp9, Prp11, Prp21, Rds3, Snu17, Hsh155, Cus1, Rse1, Hsh49, Ysf3, U2AF35, Mud2 and Msl5</td>
</tr>
<tr>
<td>U4–U6</td>
<td>U4: B, D3, G, E, F, D2, D1, U6: Lsm2–8; Prp3, Prp31, Prp4 and Snu13</td>
</tr>
<tr>
<td>U5</td>
<td>B, D3, G, E, F, D2, D1, Prp8, Prp6, Prp28, Br2, Snu114, U5–40K, Dib1, Snu23, Prp38, Prp2, Spp2, Yju2 and Cbc2</td>
</tr>
<tr>
<td>U4–U6+U5</td>
<td>U4 and U5: B, D3, G, E, F, D2 and D1 (i.e. two sets); U6: Lsm2–8; Prp3, Prp31, Prp4, Snu13/15.5K, Prp8, Prp6, Prp28, Br2, Snu114, U5–40K, snRNP27, Dib1, Snu23, Prp38, Prp2, Spp2, Yju2, Snu66 and Sad1</td>
</tr>
</tbody>
</table>

The splicing reaction is a dynamic process where the snRNPs sequentially associate / dissociate with the pre-mRNA to efficiently perform the two catalytic reactions (Figure 1.21). Firstly, U1 snRNP associate with the 5′SS of the pre-mRNA forming complex A. U2 snRNP then associates with complex E at the 3′ SS to from pre-spliceosomal complex A. U4/U6.U5 tri snRNP complex forms independently wherein U4 and U6 snRNA interacts through base-pairing. U5 snRNP then associates through proteins contacts. This tri snRNP associates with complex A to form a pre-catalytic spliceosome complex B. U1 and U4 snRNPs dissociate from complex B to form activated Complex B (B*act). Catalytically activated complex B* then forms which performs the actual catalysis of the first step forming complex C. The complex C then performs the second catalytic step leading to the release of the spliced mRNA and the snRNPs. Each of these steps also requires a number of other non-snRNP proteins as depicted in the figure. As can be noticed from Figure 1.21 and Figure 1.22, the pre-mRNA transiently associates with many different RNP complexes in each step in this elegantly orchestrated process (Will and Lührmann, 2011).
Figure 1.21  Canonical cross-intron assembly and disassembly pathway of the U2-dependent spliceosome

For simplicity, the ordered interactions of the snRNPs (indicated by circles), but not those of non-snRNP proteins, are shown. The various spliceosomal complexes are named according to the metazoan nomenclature. Exon and intron sequences are indicated by boxes and lines, respectively. The stages at which the evolutionarily conserved DExH/D-box RNA ATPases/helicases Prp5, Sub2/UAP56, Prp28, Brr2, Prp2, Prp16, Prp22 and Prp43, or the GTPase Snu114, act to facilitate conformational changes are indicated. From Will and Lührmann, 2011.
Figure 1.22  Compositional dynamics of the yeast spliceosome
Proteins identified by mass spectrometry in *S. cerevisiae* B, B$^{act}$, and C spliceosomal complexes are shown. Proteins are grouped according to their function or association with an snRNP, protein complex or spliceosomal complex. The relative abundance of the indicated proteins is indicated by light (substoichiometric) or dark (stoichiometric) lettering. From Will and Lührmann, 2011.
Thus I have outlined with specific examples one from prokaryotes and another from eukaryotes, how RNA performs various functions by associating with other proteins. With the advent of high throughput RNA sequencing techniques, several potential candidates for functional RNAs have been proposed. While some of the regulatory RNAs such as ribosomal RNAs, spliceosomal RNAs have been widely characterized in terms of the binding partners, physiological functions, a vast majority of potential functional RNA candidates are yet to be characterized. As a major part of my graduate research, I have developed tools for fluorescent characterization and purification of RNP complexes that will help in the rapid characterization of RNP complexes.

1.6. Fluorescence as a tool for following RNA processes.

Fluorescence is a process in which an excited molecule reaches its ground state by emitting a photon of higher wavelength after absorbing a lower wavelength photon. Since the initial discovery of Green Fluorescent Protein (GFP) (Shimomura et al., 1962), it has been used as a fluorescent tag for visualizing a wide variety of proteins. GFP has also been genetically engineered to alter its excitation and emission properties enabling orthogonal fluorescent tagging of several proteins (Shaner et al., 2005). Due to its widespread applications, in 2008, the Nobel Prize in chemistry was awarded for the “initial discovery of GFP and a series of important developments which have led to its use as a tagging tool in bioscience” jointly to Osamu Shimomura, Martin Chalfie and Roger Y. Tsien. Thus the discovery of Fluorescent Proteins (FP) has revolutionized the field of protein biology for decades.

RNA, on the other hand lacks strong intrinsic fluorescence making it difficult to use for intrinsic fluorescence tagging (Kwok et al., 2013; Mendez and Szalai, 2009). Small fluorescent molecules have been chemically attached to the RNAs of interest and used for imaging purposes, however, this makes imaging of endogenous RNA difficult. One of the strategies to fluorescently tag endogenous RNA is to use RNA aptamers that can be co-transcribed with the RNA of interest, which when binds to a fluorogenic molecule, would increase its fluorescence several fold. While the tagged RNA can be expressed through a plasmid or introduced into the genome of the organism through
recombination, the cell permeable, non-toxic small molecule fluorogenic ligand can be added externally making it suitable to study endogenous RNA processes. Two important properties of RNA aptamer/fluorophore complexes govern the strength of this type of RNA tracking system: These are the fluorescent enhancement of the fluorophore and the binding affinity of the complex. A strong fluorescent enhancement allows the fluorescent signal to overcome non-specific noise from unbound fluorophore, which has been deliberately selected to have very low intrinsic fluorescence in its unbound form. Likewise an acceptably low $K_D$ is necessary to insure that the RNA dye binding is saturated at very low concentrations of both RNA and dye. Together fluorescent enhancement and binding affinity define a fluorescence efficiency ($E = F_E / K_D$) for the complex that should be maximized. A common tool in RNA tracking uses a nanomolar affinity interaction (~3 nM) between the MS2 RNA hairpin and MS2 coat protein attached to a fluorescent protein so as to recruit a fluorescent signal to the MS2 tagged RNA (Bertrand et al., 1998). However, the unquenched fluorescence of unbound proteins generates an undesirable background signal in cells limiting the strength of this technique as an RNA tracking system. As such, RNA aptamers to bind compounds that generate fluorescence upon binding (fluoromodules) have been selected to optimize both fluorescent enhancement upon binding as well as $K_D$ of the aptamer/fluoromodule interaction (Table 1.4). An early aptamer/fluorophore interaction pair was the malachite green aptamer, which binds and enhances fluorescence upon a triphenylmethane dye: malachite green, or its analogs (Babendure et al., 2003). The aptamer enhances fluorescence of the dye by 2,340 fold and possesses a binding affinity of ~117 nM. However, in vivo application of this system is challenging due to the toxicity of the dye (Srivastava et al., 2004).

RNA Spinach, which binds to a GFP-like fluorophore DMHBI as well its analogs, was the first applicable aptamer/fluorophore pair method in RNA tracking due to the dyes non-toxic properties and strong fluorescent enhancement of the fluorophore upon binding of approximately 2,000-fold (Figure 1.23A). However, RNA Spinach and DMHBI presents a relatively weak binding affinity of ~464 nM (Paige et al., 2011), which has a marked impact on its fluorescence efficiency. In contrast, the recently selected RNA Mango (Figure 1.23B), binds to a biotinylated thiazole orange dye (TO1-Biotin), and provides a solution to this problem with a significantly stronger binding constant of 3.2
nM while maintaining fluorescent enhancement of about 1,100 fold (Dolgosheina et al., 2014). The details of RNA Mango including crystal structure are discussed in Chapter 3.

### Table 1.4  
Existing fluorescent RNA aptamer–fluorophore systems

<table>
<thead>
<tr>
<th>RNA</th>
<th>Motifa</th>
<th>Motif size (nt)</th>
<th>$K_D$</th>
<th>$F_E$</th>
<th>Relative E</th>
</tr>
</thead>
<tbody>
<tr>
<td>Malachite green</td>
<td>Bulge</td>
<td>38</td>
<td>117 nM</td>
<td>2340</td>
<td>62</td>
</tr>
<tr>
<td>MS2</td>
<td>Stem loop</td>
<td>18</td>
<td>~3 nM</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>RNA spinach</td>
<td>G quadruplex</td>
<td>51</td>
<td>300-540 nM</td>
<td>2000</td>
<td>11 to 20</td>
</tr>
<tr>
<td>RNA Mango</td>
<td>G Quadruplex</td>
<td>19</td>
<td>3.2 nM</td>
<td>1100</td>
<td>970</td>
</tr>
</tbody>
</table>

a The minimal size of the RNA that binds to the ligand with the specified $K_D$ and fluorescence enhancement ($F_E$). $F_E$ and the fluorescence efficiency $E = F_E/K_D$ are as defined in the text and are normalized to that of the MS-2 system.

---

**Figure 1.23**  
Schematics of RNA Spinach and RNA Mango  
The dye is depicted as red. Adapted from Panchapakesan et al., 2015.

1.7. **RNA complex purification using high affinity fluorescent RNA aptamer tags**

Interactions between RNAs and other cellular factors are at the center of nearly all biochemical processes (Cech and Steitz, 2014b), yet RNA-dependent interactions are notoriously difficult to study (McHugh et al., 2014). Many RNA-protein (RNP) interactions can be transient and are likely to be dependent on the specific subcellular spatial context.
of the RNP interaction. When it is also considered that many regulatory RNP complexes are found at low concentration, the difficulty of RNA complex purification can be appreciated. The solution to these intrinsically difficult problems requires the development of new, higher sensitivity methodologies, and appears likely to be aided by the recent development of high affinity fluorescent dye binding aptamers that can significantly extend the limits of RNP detection while simultaneously serving as an RNA purification tag (Panchapakesan et al., 2015).

In comparison, the study of protein complexes was significantly advanced by the development of affinity tags that can be simply incorporated into a protein of interest at either its amino or carboxyl terminus. Such tags are essential for the native or denaturing purification of proteins and have been instrumental in the exponential increase in determining protein structures since the first crystallization of myoglobin in 1958 (Kendrew et al., 1958). Equivalent methodologies are largely lacking for the purification of RNA and while there are some notable exceptions, RNA complexes are still widely considered less tractable for purification than protein complexes.

1.7.1. The importance of high affinity tags in protein complex purification

Purification of specific protein complexes has been greatly simplified by the development of peptide tag sequences that can specifically bind to a ligand target (Table 1.5). Such tags, which vary dramatically in molecular weight, have made possible a broad range of experiments ranging from the purification of overexpressed protein to the identification of low abundant protein complexes (Young et al., 2012; Zhao et al., 2013). These tags, when attached to a protein of interest, allow purification if the ligand is attached to a solid support. Column or bead immobilized proteins can be stringently washed and then released in native conditions provided that the tag-ligand binding interaction can be competitively displaced (typically by the addition of free ligand), chelated (i.e. addition of EDTA to displace His-tagged proteins) or proteolytically cleaved from the support (such as trypsin, TEV Protease). Several tags like Tandem Affinity Purification (TAP Tag (Puig et al., 2001)) combine ligand binding based purification with subsequent enzymatic cleavage from the solid support so as to yield homogenous protein complexes devoid of tags. In contrast to proteins purified on a solid support, tags
for which an antibody or other high affinity protein ligand exist can be immunoprecipitated even when they exist in low quantities (Björck and Kronvall, 1984). While a straightforward approach, which results in considerable enrichment in the targeted protein, precipitated proteins are typically difficult to recover in a native form. This complicates the analysis of protein complexes, particularly if the tagged protein forms complexes with more than one cellular species.

Table 1.5  Commonly used protein tags and their respective ligands, along with tag size and $K_D$ for binding

<table>
<thead>
<tr>
<th>Protein tag</th>
<th>Ligand</th>
<th>Tag size (kDa)</th>
<th>$K_D$</th>
<th>Existing mode of native elution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hexa-histidine</td>
<td>Ni-NTA</td>
<td>0.84</td>
<td>~1µM</td>
<td>Imidazole</td>
</tr>
<tr>
<td>S-tag</td>
<td>S-Protein</td>
<td>1.75</td>
<td>110 nM</td>
<td></td>
</tr>
<tr>
<td>Streptavidin binding peptide</td>
<td>Streptavidin</td>
<td>4.3</td>
<td>2.5-4.9 nM</td>
<td>Biotin</td>
</tr>
<tr>
<td>GST</td>
<td>Anti-GST antibody</td>
<td>26</td>
<td>1 nM</td>
<td>GST</td>
</tr>
<tr>
<td></td>
<td>Glutathione</td>
<td>26</td>
<td>~100 µM</td>
<td></td>
</tr>
<tr>
<td>GFP</td>
<td>Anti-GFP antibody</td>
<td>27</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Maltose-binding protein (MBP)</td>
<td>Maltose, maltotriose</td>
<td>42</td>
<td>Wild type: 1.2 µM</td>
<td>Maltose</td>
</tr>
</tbody>
</table>

The tag-ligand binding affinity has a strong influence on these purification strategies, with commonly used systems having affinities that range over three orders of magnitude (Table 1.5). Several widely used tags such as the His tag or the S-tag have relatively weak binding affinities. As a consequence (presumably due to their rapid off-rates), these tags are typically used to purify overexpressed proteins in large quantities for crystallographic and in vivo studies. If however the goal is to study low abundance protein complexes, then high affinity ligands are essential. This presents a double-edged sword: antibody based systems, for example, have excellent binding affinities, which typically range from $10^{-9}$ M to $10^{-11}$ M making them capable of stringent purification. However, elution of a protein complex bound to an antibody normally requires conditions such as SDS and / or boiling, which denatures the complex under study making it difficult or impossible to recover a native protein complex. While antibody binding can
sometimes be displaced by competitive displacement such strategies are not universally applicable to all tag-antibody systems.

1.7.2. Existing RNA-Protein Complex purification strategies.

Demonstrating the existence of an RNP complex by tagging either the protein or the RNA components and purifying the resulting complex in native conditions has such obvious benefits that a number of strategies have been developed to tag RNAs in analogy to the methodologies developed for protein purification:

I. Nucleic Acid Hybridization based approaches. One simple strategy if either the 3' or 5' end of the RNA of interest is unpaired, is to use the complementary sequence immobilized on beads to pull down the RNA (Kroiss et al., 2009; Pereira; Prongidi Fix et al., 2013). Such systems are capable of achieving very high affinities and specificities and if a cleavable linker between the complimentary oligonucleotide and the solid support is introduced it can then be used to elute native complexes (Pereira; Prongidi Fix et al., 2013). If an unpaired 5' or 3' end is not present naturally, then it needs to be introduced co-transcriptionally and such modifications may have an unpredictable effect on RNA stability. One of which is susceptibility to poly adenylation of free 3' termini and subsequent recruitment to the RNA degradosome (Blum et al., 1999). For this reason, the introduction of artificial unpaired RNA sequence to RNA of interest may not always be desirable.

II. RNA aptamer tags to protein targets. Another RNA based strategy is to use an RNA aptamer that binds to a ligand. The RNA aptamer can be attached co-transcriptionally to the RNA of interest and ideally should be small, forming a simple structural motif and should in addition not disrupt the native folding of the RNA of interest. If studying low-abundance RNA, then the motif should bind to the ligand with high affinity matching that typically seen with antigen-antibody interactions required for immunoprecipitation. Several natural or artificially selected RNA aptamers are currently in use that have such high binding affinities. Naturally occurring aptamers bind a range of protein targets including: the viral MS2 protein (Slobodin and Gerst, 2010) (Batey and Kieft, 2007), thermostable Thermotoga maritima bacterial M – domain (Kieft and Batey, 2004), lambda N protein (Tomasso et al., 2011) or imidazole activatable Csy4 protein of
yeast CRISPR complex RNA (Lee et al., 2013). The artificially selected S1m RNA binds to streptavidin and has also been used (Leppek and Stoecklin, 2013). These protein-binding partners are normally immobilized onto a solid support as a part of the purification strategy. In some cases these proteins are overexpressed as a chimeric protein that, in turn can bind to another ligand commercially available in immobilized form like streptavidin binding Peptide (SBP) (Slobodin and Gerst, 2010) - Streptavidin or GST – GSH (Tomasso et al., 2011) systems. Even though protein binding RNA aptamers have strong binding affinity, a number of issues are faced by this general approach. These include release of the bound RNA complex from the immobilized protein ligand, together with an inability to simply visualize the ligand bound RNA complex during the purification process.
Table 1.6  RNA aptamers used in the purification of RNP complexes

<table>
<thead>
<tr>
<th>RNA</th>
<th>RNA tag size (nt)(^a)</th>
<th>Ligand</th>
<th>Ligand size (^b)</th>
<th>K(_D)</th>
<th>Mode of ligand immobilization</th>
<th>Existing mode of non-denaturating elution</th>
</tr>
</thead>
<tbody>
<tr>
<td>MS2 stem loop</td>
<td>13</td>
<td>MS2 coat protein</td>
<td>10.7 kDa</td>
<td>3 nM</td>
<td>Fusion protein with SBP; streptavidin column</td>
<td>Eluting with biotin</td>
</tr>
<tr>
<td>(\lambda) BoxB</td>
<td>15</td>
<td>(\lambda) N</td>
<td>12.3 kDa</td>
<td>&gt;10 pM</td>
<td>Fusion protein with GST on GSH sepharose beads</td>
<td>GlcN6P-induced self-cleavage of ARiBo tag</td>
</tr>
<tr>
<td>Csy4 hairpin</td>
<td>20</td>
<td>Imidazole-inducible H29A Csy4</td>
<td>20.4 kDa</td>
<td>50 pM</td>
<td>Biotinylated Csy4 on avidin–agarose beads</td>
<td>RNA cleavage by activation of Csy4 with imidazole</td>
</tr>
<tr>
<td>SRP RNA</td>
<td>46</td>
<td>TmaM protein</td>
<td>~15 kDa</td>
<td>N/A</td>
<td>TmaM protein in Affigel–10 matrix through chemical coupling</td>
<td>Imidazole-induced self-cleavage of C75U hepatitis (\delta) virus ribozyme</td>
</tr>
<tr>
<td>D8 Sephadex binding motif</td>
<td>33</td>
<td>Sephadex</td>
<td>N/A</td>
<td></td>
<td>Sephadex matrix</td>
<td>Competition with dextran</td>
</tr>
<tr>
<td>S1</td>
<td>45</td>
<td>Streptavidin</td>
<td>52.8 kDa</td>
<td>~70 nM</td>
<td>Streptavidin matrix</td>
<td>Competition with biotin</td>
</tr>
<tr>
<td>S1m</td>
<td>60</td>
<td>Streptavidin</td>
<td>52.8 kDa</td>
<td>29 nM</td>
<td>Streptavidin matrix</td>
<td>Competition with biotin</td>
</tr>
<tr>
<td>Streptotag</td>
<td>46</td>
<td>Streptomycin</td>
<td>581 Da</td>
<td>1 (\mu)M</td>
<td>Streptomycin-coupled matrix</td>
<td>Competition with streptomycin</td>
</tr>
</tbody>
</table>

\(^a\)The minimal size of the RNA that binds to the ligand. Additional RNA sequences may be required, depending on other required functions. \(^b\)Ligand sizes are theoretically calculated from primary sequences in Uniprot.

The high affinity fluorescent RNA Mango offers to be a valuable RNA purification tag since only RNA Mango binds with nM affinities and has been demonstrated to allow single molecule visualization (Dolgosheina et al., 2014). Since RNA Mango binds tightly to a range of thiazole orange dye derivatives that can include a biotin or equivalent purification handles, this system offers the potential to purify RNA complexes using streptavidin magnetic beads or native gel shifts (Dolgosheina et al., 2014). By providing a method to selectively elute the dye ligand from a solid support, the RNA Mango system would allow native RNA complexes to be separated from each other, for
example, on a size exclusion column. The complexes could then be tracked using their intrinsically bright fluorescent properties, as the RNA complex would remain tagged with its small fluorescent dye ligand. In Chapter 4, I have shown the development of RNA Mango mediated native purification tool for purification and characterization of ribonucleoprotein complexes.
Chapter 2.  

*E. coli* 6S RNA release from RNA polymerase requires sigma-70 ejection by scrunching, and is orchestrated by a conserved RNA hairpin

This chapter is based on the manuscript: Shanker Shyam S. Panchapakesan and Peter J. Unrau 2012 “*E. coli* 6S RNA release from RNA polymerase requires sigma-70 ejection by scrunching, and is orchestrated by a conserved RNA hairpin”. RNA 18(12): 2251-2259

I planned and performed experiments and contributed to preparation of the manuscript. PJU planned experiments and contributed to writing the manuscript.
Abstract

The 6S RNA plays a key role in the global regulation of eubacterial transcription. This RNA in *E. coli* suppresses housekeeping transcription by binding to RNA polymerase holoenzyme (core polymerase + \(\sigma^{70}\)) under low nutrient conditions and rescues \(\sigma^{70}\)-dependent transcription in high nutrient conditions by the synthesis of a short product RNA (pRNA) using itself as a template. Previous studies have observed core polymerase bound to 6S RNA and lacking \(\sigma^{70}\). Here we show that this state is a kinetic intermediate that arises during the process of 6S RNA release and that is enhanced by the formation of a ‘release’ hairpin that is conserved across the \(\gamma\)-proteobacteria. Deliberately slowing the intrinsic 6S RNA release rate by nucleotide feeding experiments reveals that \(\sigma^{70}\) ejection occurs abruptly once a pRNA length of 9-nt is reached. After ejecting \(\sigma^{70}\), an additional 4-nt of pRNA synthesis is required before the 6S:pRNA complex is released from core polymerase. Intriguingly, changing the *E. coli* 6S RNA sequence to preclude formation of the release hairpin dramatically slows the speed of 6S RNA release but surprisingly does not alter the abruptness of \(\sigma^{70}\) ejection. Rather, the pRNA size required to trigger \(\sigma^{70}\) release increases from 9-nt to 14-nt. The precision in which the pRNA length triggers \(\sigma^{70}\) release either with or without a hairpin implicates an intrinsic ‘scrunching’ type mechanism that is modulated by the ability to form a release hairpin during the process of 6S RNA release. We speculate this release hairpin serves a dual purpose in the \(\gamma\)-proteobacteria as its formation appears highly likely to strip ‘-10’ 6S RNA region interactions away from \(\sigma^{70}\) just as the forming hairpin helps to eject \(\sigma^{70}\). This scrunching type mechanism appears likely to be used by other eubacterial 6S RNAs that lack this release hairpin but nevertheless appear highly likely to undergo significant secondary structure changes during release.
2.1. Introduction

The 6S RNA is a small non-coding RNA that plays a key role in regulating eubacterial transcription at the global level. One of the first non-coding RNAs to be sequenced (Brownlee, 1971), it accumulates to a high level in stationary phase and regulates transcription by competitively binding to E \( \sigma^{70} \) in the place of \( \sigma^{70} \) dependent DNA promoters (Wassarman and Storz, 2000). The resulting stable RNP complex precludes DNA dependent transcription in \( E. coli \) and inhibits all but the strongest \( \sigma^{70} \) dependent promoters \textit{in vivo} (Trotochaud and Wassarman, 2006). When starving cells from stationary phase encounter rich nutrient conditions, a short pRNA templated by the 6S RNA itself is produced using the higher NTP levels made available by the beneficial environment that allows outgrowth (Wassarman and Storz, 2000). This pRNA synthesis in turn triggers the ejection of the 6S:pRNA from RNA polymerase (Wassarman and Saecker, 2006) allowing normal transcription of \( \sigma^{70} \) housekeeping genes to resume. Thus the 6S RNA transcript, which is itself initiated by both \( \sigma^{70} \) and \( \sigma^{38} \) (stationary phase) promoters (Kim and Lee, 2004), provides an elegant negative feedback mechanism to directly regulate the global cellular transcriptional state via complex interactions with RNA polymerase.

The 6S RNA resembles an open form DNA promoter (Figure 1.10) and binds to E \( \sigma^{70} \) via interactions with specific regions of its sequence that correspond to parts of a \( \sigma^{70} \) DNA promoter (Barrick et al., 2005). A phylogenetically conserved ‘-35’ region (Barrick et al., 2005) (Figure 1.18) that is located upstream of the large open bubble structure has been demonstrated by \textit{in vitro} selection as well as truncation experiments to be essential for 6S RNA binding to \( E. coli \) E \( \sigma^{70} \) (Shephard et al., 2010). This site interacts with the C-terminus of \( \sigma^{70} \) as deleting the 4.2 subdomain, which normally recognizes the -35 region of a \( \sigma^{70} \) DNA promoter, precludes 6S RNA binding (Klocko and Wassarman, 2009). Since this region of the \( \sigma^{70} \) interacts with both the 6S RNA and the -35 DNA promoter elements, a mechanism for the 6S RNA to competitively bind to E \( \sigma^{70} \) is naturally suggested (Shephard et al., 2010; Wassarman and Storz, 2000). A poorly conserved bulged helix in the 6S RNA spaces this ‘-35’ binding element away from a large open bubble that contains a ‘-10’ region that is highly conserved in the \( \gamma \)-proteobacteria (Barrick et al., 2005; Shephard et al., 2010). This region, would by
geometric analogy to the -10 DNA promoter recognition element (Barrick et al., 2005), make interactions with σ\textsuperscript{70} regions 2.3, 2.4 and potentially region 3 (Murakami et al., 2002b). Consistent with this picture mutating the ‘-10’ region can either increase or decrease 6S RNA binding. The conserved downstream region has only a marginal influence on the initial binding of the 6S RNA to E σ\textsuperscript{70}. Notably in the γ-proteobacteria the top strand of this region is the exact reverse complement of the ‘-10’ region. Therefore, when pRNA invades into the downstream helix by base pairing to the bottom strand, the complementary top strand, which would now be unpaired, can potentially base pair with the conserved ‘-10’ region forming a 8–9 bp hairpin loop (Figure 1.18). Together, these data suggest a potential role for the formation of a hairpin during 6S RNA release from E σ\textsuperscript{70} in the γ-proteobacteria.

The pRNA-dependent formation of this potential hairpin structure is only predicted for bacteria that either are in or are closely related to the γ-proteobacteria. Helicobacter pylori, a member of the ε-proteobacteria, has such a potential hairpin, while the 6S RNAs from members of the α- and δ-proteobacteria do not make appropriate downstream pairing interactions with their ‘-10’ sequence (Barrick et al., 2005). Interestingly, phylogenetic evidence suggests that smaller preformed hairpins in the ‘-10’ region are quite common in other eubacteria. The 6S-1 RNA in B. subtilis for example appears to have such a preformed hairpin in the ‘-10’ region and when hybridized to a synthetic pRNA undergoes significant secondary structure rearrangements between the downstream duplex and the template strand of the open RNA bubble (Beckmann et al., 2012). These rearrangements leave the preformed hairpin structure found in the ‘-10’ region invariant before and after pRNA hybridization in contrast to the potential hairpin formation mechanism predicted in the γ-proteobacteria. Such variations suggest either the existence of several distinct 6S RNA release mechanisms across the eubacteria or that a common, highly conserved release mechanism exists that is itself regulated by several types of 6S RNA secondary structure rearrangement.

In support of the second hypothesis, we show that a hairpin structure formed between the ‘-10’ region of the γ-proteobacteria 6S RNA and the top strand of the downstream duplex in the course of pRNA synthesis, helps to coordinate first the release of σ\textsuperscript{70}, and then in a second pRNA dependent step the ejection of the 6S:pRNA
complex from core polymerase. Remarkably, we find that removing the hairpin delays both steps in this process but does not change the fundamental order of the 6S RNA release process nor the sharp nature of these transitions as pRNA extension proceeds. The abruptness of these transitions indicates that a universal mechanism, similar to that of scrunching during the transition from transcriptional initiation to elongation, exists for 6S RNA release in all eubacteria.

2.2. 6S Secondary Structure Rearrangement upon pRNA induced release

We find evidence for a total of four distinct structural states during pRNA-induced 6S RNA release. The unbound 6S RNA (S1 state) and the final released 6S:pRNA complex (the S4 state) are devoid of protein and provide simple evidence that a structural rearrangement occurs during the process of 6S RNA release from *E. coli* RNA polymerase (Figure 2.1).

![Figure 2.1](image.png)

**Figure 2.1** Structural states observed with native gel electrophoresis during 6S RNA binding and release
S1 is unbound 6S RNA, S2 is 6S RNA bound to RNAP holoenzyme, S3 is 6S RNA bound to RNAP core enzyme and S4 is the released 6S RNA:pRNA complex. In this work, the states S3 and S4 are characterized.

Having shown that the '-10' region can alter both binding and release rate (Shephard et al., 2010) we noticed, together with others (Beckmann et al., 2012; Wurm et al., 2010), that the top strand of the conserved downstream helix is complementary to the conserved '-10' region (Figure 2.2A). Although this hairpin forming potential has been noticed previously, there has been no evidence to date that this hairpin actually forms during 6S RNA release and hence the importance of this release hairpin is unknown. We hypothesized that the 6S RNA, after binding holoenzyme to form the
bound S2 state, could potentially form a hairpin as a consequence of pRNA synthesis that would preclude binding interactions between the `-10' region and σ^70. G136, which makes a UV crosslink to the holo polymerase in the S2 state (Gildehaus et al., 2007) (Figure 2.2A), is predicted to be unpaired in the unbound central bubble of the 6S RNA and should become strongly protected against T1 RNase digestion upon formation of a 6S RNA hairpin as T1 RNase cuts only unpaired G residues. Indeed, T1 RNase digestion of the free 6S RNA (S1) under native conditions revealed G136 to be sensitive to T1 RNase (Figure 2.2B), consistent with previous secondary structure models of the 6S RNA (Barrick et al., 2005; Shephard et al., 2010). After binding and then releasing the 6S RNA from the holoenzyme via pRNA synthesis so as to form the S4 state, G136 became significantly resistant to T1 RNase while the cleavage pattern of G97, G88, G82, G80 and G79, which were not expected to undergo any secondary structure rearrangement, remained unaffected. G143, adjacent to the initial base pair of the downstream duplex in the free S1 state, became more sensitive to T1 RNase in the S4 state, as would be expected if found in the proposed hairpin tri-loop after formation of the 6S RNA hairpin (Figure 2.2A). Having established this important change in secondary structure between the S1 and S4 states, we next explored potential intermediate states of the release process that could depend on pRNA length.
Figure 2.2 Presence of hairpin in the released 6S:pRNA complex.

(A) Highly conserved residues in the unbound 6S RNA (S1 state) from \( \gamma \)-proteobacteria are shown in purple. pRNA is shown in blue. The hairpin forms in the released 6S RNA:pRNA complex (S4 state) between the phylogenetically conserved “~10” region and the top strand of the downstream helix. Green circles indicate G residues that change their protection pattern upon release, while black circles indicate single-stranded G residues that are unaffected by the release process. Residues previously shown to form UV crosslinks with holoenzyme in the bound Eo\(^{70} \) (S2 state) are marked with asterisks (Gildehaus et al. 2007). (B) Native T1 RNase digestion of 5' \( ^{32}P \) end-labeled 6S RNA before binding (S1) and after pRNA induced release (S4). Analyzed using 10% denaturing PAGE. T1 indicates denaturing RNase T1 digest of the unbound 6S RNA.

2.3. 6S:pRNA:E release intermediate found between bound (S2) and final released state (S4)

The bound 6S:Eo\(^{70} \) (S2) complex is capable of being very rapidly released both \textit{in vitro} and \textit{in vivo} (Wurm et al., 2010). Using aggressive release conditions that simulate the early stages of outgrowth from stationary phase (3.75 mM MgCl\(_2\), 148 mM of each NTP, and stoichiometric amounts of heparin) the 6S:pRNA complex can be released from RNA polymerase with a half time of \(~50\) seconds and produces a 6S:pRNA complex containing a pRNA that is 13-nt long (Figure 2.3). Titrating magnesium has a dramatic effect on release rate but leaves the length of the pRNA
invariant. The addition of heparin during binding and 6S RNA release simulates the nonspecific interactions of endogenous RNA during the release process, and results in the in vitro production of pRNAs that are very comparable in size to that observed in vivo. In the absence of heparin, pRNA sizes increase (Figure 2.3C) but appear to have limited in vivo significance; as pRNA of length ~10 to 14-nt are known to be produced within 15 to 30 s of rapid outgrowth in E. coli (Wurm et al., 2010) and are very similar in size to the pRNAs observed in H. pylori (12-nt long) (Sharma et al., 2010) and B. subtilis (8 to 12-nt long) (Beckmann et al., 2011). Longer RNAs complementary to a pRNA probe have been detected by Northern blot from E. coli, but start to accumulate only after 1-2 min of outgrowth (Wurm et al., 2010). Such RNAs could reflect the rapid in vivo degradation of the 6S RNA itself or alternatively longer pRNAs could have a role secondary to the initial and very rapid burst of short pRNAs synthesis observed at the start of outgrowth. Both in vivo and in vitro, the high intrinsic speed of the 6S RNA release process has made it difficult to explore potential intermediate states.

**Figure 2.3  6S RNA release conditions**

Magnesium concentration dramatically modulates release rate but not pRNA size. (A). Release kinetics of 25 nM 6S RNA bound to 200 nM Holoenzyme released with 225 µM each NTP in the presence of the indicated amounts of MgCl₂. Percentage bound to Holoenzyme plotted against time. The curves are fitted to eqn. 2.4. (B). pRNA sizes in the S4 released state as a function of magnesium concentration, Lanes 2 to 8: 0, 1, 1.5, 2, 2.5, 3, and 3.5 mM MgCl₂. Lane 1: alkaline hydrolysis ladder. (C) Addition of Heparin simulates in vivo pRNA size observed during 6S RNA release. 6S RNA was bound to the holoenzyme in the presence or absence of heparin as indicated and the release was induced by the addition of 147 µM each NTPs and 4 mM MgCl₂ with a trace amount of γ-[³²P]-ATP to ensure pRNA labeling. The products were loaded onto a 23% Denaturing PAGE gel. H: alkaline hydrolysis ladder for the lane to the immediate right.
In order to slow down the 6S RNA release process pRNA synthesis conditions were modified. Using AU dinucleotide (25 mM) to initiate RNA synthesis (Wassarman and Saecker, 2006), we found that removing ATP results in a new intermediate state between the S2 and S4 states. Analysis by native gel revealed the conversion of the S2 state into a new intermediate state (S3), that initially grows and then declines as the S2 state is depleted and the S4 state is produced (Figure 2.4). A simple two step kinetic model (see Equation 2.1 -Equation 2.6), where the S2 state proceeds to state S3 with a first order rate $k_{23}$ and then from state S3 to state S4 with a rate $k_{34}$, fit acceptably to the S3 state time course data (Figure 2.4B) and gave $k_{23} = 0.36 \pm 0.04 \text{ min}^{-1}$ and $k_{34} = 0.94 \pm 0.11 \text{ min}^{-1}$ under the slow pRNA synthesis conditions used for this assay. An independent fit to the S2 kinetics gave for this experiment $k_{23} = 0.48 \pm 0.07 \text{ min}^{-1}$. While the $k_{23}$ values were not in perfect agreement between the S2 and S3 fits, it is notable that the S2 state data did not fit perfectly to a single exponential, suggesting that the S2 state is not completely homogeneous, and that it contains a subpopulation (we estimate 10-15% for our enzyme) that releases slower and/or incompletely relative to the majority species. If this subpopulation does not utilize the S3 state during release the increased $k_{23}$ value determined for the S2 fit relative to the S3 fit could be explained.
Figure 2.4 A 6S:pRNA:Ec70 intermediate state (S3) revealed by nucleotide feeding experiments.

(A) Native gel analysis of 32P body-labeled 6S RNA (6S*) showing the emergence and decline of the S3 state when pRNA synthesis is initiated with AU dinucleotide (GTP concentration 25.9 µM, ATP withheld). Time points are 0.25, 0.5, 1, 2, 4, 8, 16, and 32 min. F indicates free 6S* (S1); B, bound Ec70 complex (S2). (B) Fitted data for the S2 and S3 state kinetics shown in panel A (for fitting details, see Materials and Methods) using Equations 2.4 and 2.5, respectively.

If all steps of this 6S RNA release process are pRNA-dependent then the length of the pRNA in the S2, S3 and S4 states should be in strict and increasing progression. Tracking pRNA-induced release by labeling AU with γ-32P]-ATP and polynucleotide kinase revealed no evidence for pRNA in the S2 state whereas bound 6S RNA is clearly evident (Figure 2.5A). The S3 state contained both 6S RNA and a shorter 9-nt long pRNA (Figure 2.5A and B). This was initially unexpected as pRNA synthesis initiated with AU in the absence of ATP requires ATP for the 9th pRNA addition against template residue U36 and should therefore only contain a pRNA 8-nt long. If however E. coli RNA polymerase can add GTP to form a wobble pair across from U36 then the observed pattern of extension makes complete sense and implies that the S3 state and its 9-nt long pRNA are indeed intermediate to the S2 and S4 states.
Figure 2.5  

pRNA is present in the S3 and S4 states of the 6S RNA release complex.

(A, lanes 1–4) Time course using $^{32}$P radiolabeled AU (*pAU) dinucleotide, time points 2, 5, 10, and 30 min. (A, lanes 6–9) $^{32}$P body-labeled 6S RNA (6S*) showing bound (S2) complex followed by the emergence of the S3 and S4 states after addition of 25 µM AU, 147 µM of each CTP, GTP, and UTP. B indicates bound Eø70 complex (lane 5). (B) Size characterization of *pAU-labeled pRNA extracted from the S3 and S4 states. Recovered RNA was analyzed using 23% denaturing PAGE. S4 H indicates alkaline hydrolysis ladder of the S4 state pRNA.

Further support for this model was found by titrating GTP. Our two state model predicts that GTP concentration should have a pronounced effect on the $k_{34}$ rate constant as multiple additions of GTP are required to convert the 9-nt long pRNA in the S3 state to the 13-nt long pRNA found in S4 state. As predicted increasing GTP concentration decreased the amount of S3 state observed (Figure 2.6). In no conditions tested did we observe a change in pRNA size within either the S3 or the S4 states; they always remained 9-nt and 13-nt, respectively.
Figure 2.6  Dynamically populating the release intermediate 6S RNA S3 state by titrating AU dinucleotide and GTP concentrations.  
(A) 25 nM of body-labeled 6S RNA was bound to an excess (200 nM) holoenzyme and was released with the indicated amounts of AU. CTP, GTP and UTP were held at 145 µM and MgCl$_2$ was at 3.75 mM. F: free 6S RNA, B: bound E$_{70}$ complex. Time points were 2, 5, 10, and 30 min and loaded in a 5% native PAGE gel with 5% glycerol. (B) Bound body labeled 6S RNA was released with 25 µM AU, 147 µM each CTP, UTP. GTP concentrations were as indicated. Time points were taken at 2, 5, 10 and 30 min and processed as in panel A.

The S3 band migrates with a mobility previously attributed to 6S RNA bound to core polymerase (Wassarman and Saecker, 2006; Wurm et al., 2010). To confirm that the S3 state is devoid of $\sigma^{70}$, a gel supershift assay with mouse monoclonal anti $\sigma^{70}$ antibody (Neoclonel) was performed. Addition of anti $\sigma^{70}$ antibody to 6S RNA resulted in a supershift of only the S2 band, and not the S1, S3 or S4 bands (Figure 2.7). This would be expected if only the S2 state contains $\sigma^{70}$. Consistent with this observation, adding antibody prior to the initiation of pRNA-dependent release using the AU-ATP nucleotide feeding approach lowered the overall 6S RNA release rate while simultaneously lowering the amount of S3 state observed. This would be predicted if the anti-sigma antibodies bind to $\sigma^{70}$ in the S2 state, so as to lower the pRNA dependent
rate constant \( k_{23} \) from state S2 to S3, but not the rate constant \( k_{34} \) out of S3 and into S4 (where we presume that owing to the absence of \( \sigma^{70} \), the antibody would no longer have an inhibitory effect on the S3 to S4 state kinetics).

**Figure 2.7** Anti-\( \sigma^{70} \) antibody supershift assay showing S3 state is devoid of \( \sigma^{70} \). (A) 25 nM radiolabeled 6S RNA bound to 200 nM holoenzyme was incubated with mouse anti-\( \sigma^{70} \) antibody and was released with 147 \( \mu \)M of each NTP and 3.75 mM MgCl\(_2\). F indicates free 6S; B and B', bound E\( \sigma^{70} \)complex; B', E\( \sigma^{70} \) Ab; and SS, super shifted band. Time points of 2, 4, 8, 16, and 32 min were loaded in a 5% native gel with 5% glycerol. (B) The 25 nM body-labeled 6S RNA bound to 200 nM holoenzyme was incubated with or without mouse anti-\( \sigma^{70} \) antibody as indicated and was released with 147 \( \mu \)M of each CTP, GTP, and UTP; 1.5 mM MgCl\(_2\); and 25 \( \mu \)M AU. Time points of 1, 2, 4, 8, 16, and 32 min were loaded as in panel A.

**2.4. Removing 6S RNA hairpin does not preclude \( \sigma^{70} \) ejection**

In order to further understand the role of the hairpin formation in the release process, a 6S RNA variant, 6S_S, was made by swapping the top and bottom strand of the downstream helix so that the hairpin could not be formed (Figure 2.8A). This construct leaves the pRNA initiation site unchanged as well as the stability of the downstream helix largely invariant. On comparing the release kinetics of the 6S and 6S_S RNAs, we saw that the 6S_S RNA construct releases at least 10 times slower than the wild type 6S RNA under aggressive release conditions. Whereas the 6S RNA is
fully liberated in less than 2 min, the 6S_S is still forming the S4 state after 10 min of incubation (Figure 2.8B). Intriguingly, the prominent emergence of the S3 intermediate state band implied that the 6S_S construct could still trigger the rapid release of σ70 while requiring a longer time to convert the S3 state so formed, to the S4 state.

Figure 2.8 Abolishing the 6S RNA hairpin increases pRNA length and release time.
(A) The 6S RNA swapped construct (6S_S) conserves pairing interactions in the downstream region but lacks the ability to form a hairpin. The top strand in the 6S RNA (green box) was swapped with the bottom strand (blue box) to generate the 6S_S construct. (B) Radiolabeled 6S_S (6S_S*) release assay showing a dramatic accumulation of the S3 state. (Lanes 1–6) 6S_S RNA. (Lanes 7–12) 6S RNA. Time points are 2, 5, 10, and 30 min; F indicates free RNA; B, bound RNA. (C) pRNA lengths in the S3 and S4 states of 6S_S relative to 6S RNA S4 state. All pRNAs were initiated with γ−[32P]-ATP. H indicates alkaline hydrolysis ladder. Samples were analyzed using 23% denaturing PAGE.

Not only did the 6S_S RNA eject σ70 prior to release, but it also produced a S3 state containing pRNA dominated by a specific size of pRNA. Whereas the 6S RNA
produced a pRNA of length 9-nt in the S3 state using modified NTP feeding conditions, the 6S_S when released with high magnesium (3.75 mM) and all four NTPs produced a pRNA distribution that was nearly completely dominated by a 14-nt long pRNA (Figure 2.8C). Just as for the 6S RNA no evidence for pRNA in the S2 state was observed. The pRNAs found in the released 6S_S RNA S4 complex, in contrast to the S3 state, were no longer homogeneous. At least 5 distinct sizes where observed with the shortest being 18-nt long and the longest being ~28-nt long (Figure 2.8C). These RNAs were all found to have related sequence as the hydrolysis ladder for the entire set of bands produced a homogenous set of bands. The two longest pRNA bands are dominant in the S4 state (22X the signal seen for the 14-nt long pRNA band) but are also seen weakly in a complex having a mobility very similar to that of the S3 state (the two longer bands have 0.64X the signal seen relative to 14-nt band in the excised S3 state pRNA). This data together implies that these longer pRNAs can form rapidly after σ70 ejection, but do not immediately trigger 6S_S:pRNA release from holoenzyme. As expected in the aggressive release conditions used for the 6S_S RNA release, the wild type 6S RNA did not manifest an obvious S3 band, but did robustly convert the S2 state to S4 state (presumably with an unobserved S3 transient intermediate), with the S4 state containing a pRNA of 13-nt size (Figure 2.5B) consistent with our previous observations using altered nucleotide release conditions.

2.5. Discussion

During pRNA synthesis in the γ-proteobacteria, a secondary structure rearrangement occurs involving the phylogenetically conserved regions of the 6S RNA and that results in the formation of a hairpin in the top strand of the 6S RNA. This hairpin forms once pRNA invasion destabilizes the downstream duplex sufficiently to allow the top strand to pair with the ‘-10’-sequence region. We have demonstrated that when pRNA synthesis is slowed, a pRNA precisely 9-nt long is sufficient to trigger the abrupt formation of the S3 state, which lacks σ70. This length of pRNA appears to be important for two reasons: First, a 9-nt pRNA is precisely that required to make all top strand residues of the conserved downstream helix (Figure 2.2) available to pair with the conserved ‘-10’ region. Second, once a pRNA 9-nt has been synthesized the polymerase active site is now located precisely at the start of a downstream bubble
(defined by UUACA – residues 154-158 on the top strand and CAGU – residues 28-31 on the bottom strand). This bubble could in principle facilitate the abrupt movement of the now unpaired top strand residues of the downstream helix with respect to the polymerase active site and allow these residues to pair with the ‘-10’ region so as to form the release hairpin.

Hairpin formation strips the 6S RNA ‘-10’ sequence away from protein contacts in $\sigma^{70}$ so as to destabilize $\sigma^{70}$ binding to the 6S:pRNA:E complex. This interpretation is consistent with the phenotypes of a number of ‘-10’ sequence mutants that manifest release defects (Shephard et al., 2010). A C132A mutant introduces a net increase in S2 state binding and exhibits a significant release defect after addition of NTPs. This mutation destabilizes the 6S RNA hairpin by one terminal bp and since it binds initially more tightly to $E\sigma^{70}$ than wild type 6S RNA, its phenotype is as expected if the formation of the weakened hairpin has difficulty competing for a stronger ‘-10’:protein interaction. Conversely a U134A mutant decreases S2 binding but also releases more rapidly than wild type 6S RNA. This would be expected if the bulged hairpin resulting from this mutation can still compete effectively for a weakened RNA:protein interaction. Together these two mutants support the idea that one function of hairpin formation is to destabilize interactions between the ‘-10’ region and $E\sigma^{70}$ that are initially present in the S2 bound state. Since the ‘-10’ region appears likely to make contacts specifically with $\sigma^{70}$ (Barrick et al., 2005; Murakami et al., 2002b; Shephard et al., 2010) we conclude that one important role of hairpin formation is to weaken 6S RNA interactions with $\sigma^{70}$ and hence facilitate entry into the S3 state.

The 6S release process appears very similar to that of ‘scrunching’ during transcriptional initiation. After $\sigma^{70}$ ejection to form the S3 state, the 6S:pRNA complex is still firmly attached to core polymerase and survives entry into a native gel implying that downstream interactions, that initially are quite weak in the S2 state (Shephard et al., 2010), are dramatically enhanced as a result of pRNA synthesis. The $\beta'$ jaw domain of the core polymerase is well known to play a critical role during transcriptional initiation and holds downstream dsDNA tightly as DNA is packed into the enzyme complex as a result of the ratcheting effect of NTP incorporation (Ederth et al., 2002). The jaw domain together with “hood” regions of the polymerase defined by the $\beta$ and $\beta'$ subunits that
define the top of a channel containing both the template and nontemplate stands appears likely to be fully envelope both the downstream and bubble regions of the 6S RNA. Since the 6S RNA is initially bound via its highly conserved −35 domain to the C terminus of σ^{70}, pRNA synthesis prior to formation of the S3 state could therefore be expected to the build-up strain inside the RNP complex that together weaken interactions between σ^{70} and the bound 6S:pRNA:core polymerase. This strain, as just discussed, is made more potent by the abrupt loss of −10 interactions that result from formation of the release hairpin. More critically, the formation of a hairpin in the −10 region would simultaneously “scrunch” RNA into a region of the holoenzyme complex that has previously been implicated in the accumulation of bulged ssDNA during the process of transcriptional initiation (Figure 2.9) (Revyakin et al., 2006). This interpretation is consistent with our findings that removing the hairpin does not prevent σ^{70} release or make the process less crisp, rather pRNA length increases by ∼5 nt so as create a 14-bp homoduplex at the moment of σ^{70} ejection. Such an observation fits naturally with a scrunching-type model, where additional ssRNA from both the top and bottom strands of the open 6S RNA bubble must accumulate inside the holoenzyme complex during pRNA synthesis in order to accumulate strain equivalent to that afforded by the release hairpin.

From an evolutionary perspective, release hairpin formation on the top strand of the 6S RNA open bubble of the sort found in E. coli 6S RNA occurs predominantly in the γ-proteobacteria (Shephard et al., 2010). In contrast, many eubacteria such as B. subtilis appear to utilize a smaller preformed hairpin in the top strand ‘−10’ region (Figure 2.9). This preformed structure is correlated with an alternative secondary structure pairing rearrangement between the bottom strand of the 6S RNA open bubble and downstream residues that become available during pRNA synthesis (Barrick et al., 2005; Beckmann et al., 2012; Cavanagh et al., 2012). These two distinct structural changes are both entirely consistent with the scrunching model just suggested for 6S RNA release in the γ-proteobacteria and suggest that this mechanism may be common to all eubacteria (Figure 2.9). Consistent with this, final pRNA sizes upon formation of the S4 state appear to be largely invariant (12 to 14-nt) across the eubacteria observed in vivo to date (Cavanagh et al., 2012; Sharma et al., 2010; Wurm et al., 2010). Since the γ-proteobacteria form a top strand release hairpin that simultaneously decreases ‘−
10’ binding to $\sigma^{70}$, it would appear that these bacteria may be able to trigger 6S RNA release more abruptly than other eubacteria. This remains to be demonstrated, but it is a curious fact in this respect, that the $\gamma$-proteobacteria are highly enriched in pathogens (i.e. the *Vibrionaceae*, *Enterobacteriaceae* and the *Pseudomonadaceae*). An abrupt transition from stationary phase into rapid growth triggered by a 6S RNA release hairpin may be of considerable benefit to these organisms, which have diverse metabolic activities.

Figure 2.9 Secondary structure changes for the *E*. (left) versus *B*. *subtilis* (right) between the S1 (initial unbound) and S4 (final released state).

More generally, the ejection of $\sigma^{70}$ prior to 6S:pRNA release summarized in Figure 2.10, offers the potential to study transcriptional initiation and its RNA dependent regulation in a new light. If the S3 state can only be populated in certain cellular conditions, then two independent classes of transcriptional regulators normally thought to act on DNA:RNA polymerase complexes could act on the 6S RNA regulatory system in a variety of ways. The first class would interact with the bound S2 complex and modulate transcriptional initiation. The second class of regulators could act on the 6S:pRNA:E (S3) complex, which we presume would more closely resemble an elongation complex. Gre, NusA, ppGpp, which are known to modulate elongation (Roberts et al., 2008), could have pronounced effects on the S3 state but would only be expected to be observed when the transition from S3 state to S4 state is rate limiting. Most excitingly, pRNA dependent 6S RNA release with its many detailed parallels to transcription initiation offers the opportunity to extract common mechanistic features shared between transcriptional initiation and RNA dependent transcriptional regulation.
These fundamental mechanisms are key to fully understand bacterial gene expression and regulation, and promise to unify understanding of this important subject.

**Figure 2.10** Model summarizing the major steps of *E. coli* 6S RNA release (**bottom**) relative to those of transcriptional initiation (**top**). Conserved sequence of 6S RNA in γ-proteobacteria is shown in purple; pRNA is in blue; and template strand is light brown and non-template is dark brown. Initiation and elongation factors have the potential to interact with and regulate 6S RNA release as indicated.

### 2.6. Methods:

#### 2.6.1. DNA Constructs:

The 6S RNA gene was originally isolated from *E. coli* genomic DNA (12). The 6S RNA swapped construct (6S_S) was made by PCR (10 mM TRIS pH 8.3, 50 mM KCl, 1.5 mM
MgCl$_2$, 0.1% Gelatin, 200 mM each dNTP, 2.5 U/100 ml, 0.5 mM primers), using P1: 5' GAA TCT CCG AGA TGC CGC CGC AGG CTG TAA GGG GAC TCG GCT TGG TTC AAG GT and P2: 5' ttc taa tac gac tca cta tAg GAT TTC TCT GAG ATG TTC GCA AGC GGG CCA GTG GGA ACT TGG GAT ATT TCA TAC CAC AAG A. The lowercase residues indicate the T7 RNA Polymerase promoter for \textit{in vitro} transcription of the RNA, residues in italics are the regions of the inverted downstream helix.

2.6.2. 6S RNA Binding:

25 nM $^{32}$P body-labeled PAGE purified, T7 RNA Polymerase \textit{in vitro} transcribed 6S RNA was heated to 80 °C for 2 min and cooling to 50 °C for 5 min prior to binding to 200 nM \textit{E. coli} RNA Polymerase Holoenzyme (Epicentre) in 15 mM HEPES at pH 7.5, 90 mM KCl, 0.75 mM DTT, 75 mg/ml Heparin at 37 °C for 30 min.

2.6.3. 6S Release:

Aggressive release conditions - pRNA synthesis was initiated by the addition of 148 uM of each NTP and 3.75 mM MgCl$_2$ and incubation at 37 °C for 30 min. Modified release conditions - pRNA synthesis was initiated by the addition of 25 mM AU, 148 mM of each CTP, GTP and UTP and 3.5 mM or 1.5 mM MgCl$_2$ as indicated and incubated at 37 °C for 30 min. Reactions were quenched by the addition of 2X native gel loading dye containing 0.025% Bromophenol Blue and 0.025 % Xylene Cyanol, 50% Glycerol, 40 mM HEPES pH 7.5, 120 mM KCl and 8 mM EDTA and loaded onto a 5% Polyacrylamide gel with 5% Glycerol ran with 1X TBE at 4 °C.

2.6.4. T1 RNase Digestion:

5' labeled 6S RNA was bound to the RNAP H.E. and pRNA synthesis was initiated by adding NTPs and MgCl$_2$. For digestion in native conditions, T1 RNase was serially diluted in 20 mM HEPES pH 7.5, 120 mM KCl and 50% glycerol and used to digest end labeled 6S RNA so as to find optimal digestion conditions. Initial and released 6S RNA:pRNA complexes were added to the appropriate dilution of T1 RNase and incubated at 50 °C for 10 min. In the case of denaturating T1 RNase digestion, the
enzyme was diluted in 20 mM Sodium Citrate with 6 M Urea. The reactions were quenched by adding a denaturating loading dye with 0.025% Bromophenol Blue, 0.025% Xylene Cyanol, 5 mM EDTA and 90% Formamide and loaded onto a 10% sequencing PAGE with 6 M Urea.

2.6.5. Curve fitting:

A model summarized in Figure 2.10 and defined by two first order rate constants $k_{23}$ and $k_{34}$ was found to fit the S2 and S3 data well (Equation 2.4 and Equation 2.5). Phosphorimager gel images were analyzed using ImageQuant V5.2 and then analyzed using Prism V software.

2.6.6. Equations:

The differential equations for 6S RNA release are:

Equation 2.1
\[
\frac{dS_2}{dt} = -k_{23}S_2
\]

Equation 2.2
\[
\frac{dS_3}{dt} = k_{23}S_2 - k_{34}S_3
\]

Equation 2.3
\[
\frac{dS_4}{dt} = -k_{34}S_3
\]

The solutions for 6S release are:

Equation 2.4
\[
S_2 = S_2^i e^{-k_{23}t}
\]

Equation 2.5
\[ S_3 = \frac{k_{23}}{(k_{34} - k_{23})} S_2 \left( e^{-k_{23}t} - e^{-k_{34}t} \right) \]

Equation 2.6

\[ S_4 = -\frac{k_{23}k_{34}}{(k_{34} - k_{23})} S_2 \left( \frac{e^{-k_{34}t}}{k_{34}} - \frac{e^{-k_{23}t}}{k_{23}} - \frac{1}{k_{34}} + \frac{1}{k_{23}} \right) \]
Chapter 3. Structural characterization and development of biological applications for RNA Mango

This chapter is based on the following manuscripts:


2. Robert J. Trachman III, Natalia A. Demeshkina, Matthew W.L. Lau, Shanker Shyam S. Panchapakesan, Sunny C.Y. Jeng, Peter J. Unrau & Adrian R. Ferré-D'Amaré “Structural basis for high-affinity thiazole orange binding and fluorescence activation by a small in vitro selected RNA” Nature Chemical Biology, manuscript in revision wherein I was responsible for circular permutation of RNA Mango.

3. Unpublished work performed by Amir Abdolahzadeh under my supervision during his Independent Undergraduate Study Semester in the Unrau lab wherein I was responsible for design and general guidance.
Abstract

Since RNA lacks strong intrinsic fluorescence, it has proven challenging to track RNA molecules in real time. To address this problem and to allow the purification of fluorescently tagged RNA complexes, the Unrau lab has isolated RNA Mango I using *in vitro* selection. This aptamer binds a series of thiazole orange (fluorophore) derivatives with nanomolar affinity, while increasing fluorophore fluorescence by up to 1,100-fold. X-Ray crystal structure of the RNA Mango along with its binding ligand has been solved in collaboration with Ferre-D’Amare lab.

I have contributed to the structure-guided analysis of RNA Mango-I. With inspirations from the crystal structure, I also have designed and demonstrated *in vitro*, a fluorescence ‘turn on’ system that can be used for fluorescent imaging of endogenous RNAs without co-transcriptional / synthetic tagging. By inserting RNA Mango into a stem loop of the bacterial 6S RNA and biotinylating the fluorophore, I demonstrate that the aptamer can also be used to simultaneously fluorescently label and purify biologically important RNAs.
3.1. INTRODUCTION

The detection of low abundance RNA in vitro or in vivo has proven to be difficult and much effort has been devoted to developing nanostructured fluorescent probes for imaging RNA in vivo (Armitage, 2011; Boutorine et al., 2013; Song et al., 2014; Tyagi and Alsmadi, 2004; Urbinati and Long, 2011). Since strongly fluorescent RNAs do not exist (Kwok et al., 2013; Mendez and Szalai, 2009), fluorescent reporter molecules must be delivered to RNA, normally through non-covalent binding to a compatible tag region (e.g. sequence or tertiary structure) incorporated into the RNA target. As such, both the binding affinity ($K_D$) and the fluorescent efficiency (defined here as $E = F_E/K_D$, where $F_E$ is the fluorescent enhancement observed between bound and unbound fluorophore) are important parameters to optimize, when tracking RNA in low concentrations.

In an ideal RNA tracking system, $K_D$ should be minimized, as only then will low concentrations of RNA be fully complexed with fluorophore, while $E$ should be maximized, so that the signal-to-noise ratio of the system is as high as possible. Systems, such as the GFP-MS2 (Bertrand et al., 1998) and the recent Spinach aptamer (Paige et al., 2011; Song et al., 2014; Strack et al., 2013), optimize one or the other of these two constraints, but fail to simultaneously optimize both (Table 3.1). The MS2 system fuses a fluorescent protein reporter to a peptide sequence recognized by an RNA motif with nanomolar affinity and has proven successful in RNA live cell imaging (Forrest and Gavis, 2003; Golding and Cox, 2004; Haim et al., 2007; Querido and Chartrand, 2008; Rook et al., 2000; Shav-Tal et al., 2004; Weil et al., 2006; Zimyanin et al., 2008). Yet, it suffers since the intrinsic fluorescence of the unbound protein reporter is not enhanced upon binding to the target RNA (i.e. $F_E = n$, where $n$ is the number of encoded MS2 RNA sequence elements), making it potentially difficult to discern the signal coming from the free fluorophore itself or the fluorophore bound to RNA. Conversely, despite the very high fluorescent enhancement of the Spinach aptamer ($F_E = 2,000$), the system has a poor $K_D$ (Paige et al., 2011). This places the fluorescent efficiency of the Spinach aptamer some six times lower than that of the toxic malachite green aptamer (Table 3.1), which has both a higher $F_E$ and lower $K_D$ (Babendure et al., 2003; Fessard et al., 1999; Srivastava et al., 2004).
For single molecule imaging and for the purification of RNA complexes, low $K_D$, or more specifically a slow off-rate from the RNA fluoromodule, takes on an additional significance: A single bound complex should persist for the duration of imaging and/or complex purification.

Table 3.1  **RNA aptamer/Fluoromodule Complex Properties**

<table>
<thead>
<tr>
<th>Fluoromodule complex</th>
<th>Rel.E</th>
<th>$F_E$</th>
<th>$K_D$ (nM)</th>
<th>Ex (nm)</th>
<th>Em (nm)</th>
<th>$\varepsilon$ (M$^{-1}$ cm$^{-1}$)</th>
<th>$\phi$</th>
</tr>
</thead>
<tbody>
<tr>
<td>GFP-MS2</td>
<td>1*</td>
<td>1</td>
<td>~3</td>
<td>395</td>
<td>508</td>
<td>27,600</td>
<td>0.79</td>
</tr>
<tr>
<td>Malachite Green</td>
<td>62</td>
<td>2,400</td>
<td>117</td>
<td>630</td>
<td>650</td>
<td>150,000</td>
<td>0.19</td>
</tr>
<tr>
<td>(Babendure et al., 2003)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RNA Spinach</td>
<td>11-20</td>
<td>2,000</td>
<td>300-540</td>
<td>469</td>
<td>501</td>
<td>24,300</td>
<td>0.72</td>
</tr>
<tr>
<td>(Paige et al., 2011)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RNA Mango (TO1-Biotin)</td>
<td>970</td>
<td>1,100</td>
<td>3.2</td>
<td>510</td>
<td>535</td>
<td>77,500</td>
<td>0.14</td>
</tr>
</tbody>
</table>

*Fluorescent efficiencies, defined as $E = F_E/K_D$, were normalized to the GFP-MS2 construct assuming only one MS2 RNA binding element (n=1). $F_E$ is the fluorescent enhancement of bound complex relative to unbound fluorophore, $K_D$ the dissociation constant for the complex, $\varepsilon$ extinction coefficient, $\phi$ - fluorescence quantum yield.

RNA Mango I is an RNA aptamer with nanomolar binding affinity to its fluorogenic ligand, TO1-Biotin comparable to that found in the MS2 system (Johansson et al., 1998) and with fluorescent enhancement similar to that found in the Spinach system (Paige et al., 2011) (Table 3.1). In terms of fluorescent efficiency, RNA Mango I is nearly three orders of magnitude superior to the MS2 system and nearly two orders of magnitude superior to the Spinach system. As we demonstrate, the high binding affinity of RNA Mango I to its fluorophore makes possible single molecule RNA visualization and offers the potential to fluorescently monitor RNA complexes while simultaneously using the fluorophore as a purification tag.
3.1.1. RNA Mango I/TO1-Biotin complex

RNA Mango I, isolated by Lena Dolgosheina, a previous graduate student in the Unrau lab, after 12 rounds of \textit{in vitro} selection, binds rapidly to TO1-Biotin with nanomolar affinity (Figure 3.1). Fluorescent response fit well to a simple one-to-one aptamer-fluorophore binding model and yielded a $K_D$ of $3.2 \pm 0.7$ nM. As characterized by Sunny Jeng, a graduate student in the Unrau lab, the pattern of four regularly spaced G pairs is strongly suggestive of a G-quadruplex (G4) based ligand binding core. Potassium, which is known to stabilize G-quadruplexes much more effectively than other monovalent ions, was found to be essential for fluorescence. CD spectra of the potassium fluorophore-bound complex are suggestive of a parallel-stranded G-quadruplex (Vorlíčková et al., 2012), since a negative rotation in the 220-240 nm region appeared only after the addition of fluorophore in the presence of KCl. This negative rotation was accompanied by a positive rotation in the 260 nm region. Consistent with the CD data, native T1 RNase probing experiments indicated that the core G residues only become strongly protected when potassium and TO1-Biotin are present simultaneously. Taken together these data indicate that a compact, potassium-dependent, parallel-stranded G-quadruplex structure forms upon fluorophore binding.
RNA Mango I has a high binding affinity to TO1-Biotin.

(A) Structure of TO1-Biotin. (B) RNA Mango I aptamer binds to TO1-Biotin over a broad concentration range. Fluorophore concentrations: 5 nM - right triangles, 10 nM - circles, 20 - nM squares, 40 - nM diamonds, 100 nM - triangles, 1000 nM - inverted triangles. The initial unselected pool when incubated with 10 nM fluorophore is shown by open circles. Error bars are standard deviations about the mean of three measurements.

RNA Mango I complexed with TO1-Biotin generated a fluorescent enhancement of 1,100-fold relative to unbound fluorophore with strong excitation peaks observed at both 260 nm and 510 nm (Figure 3.2). Emission is centered at 535 nm. The extinction coefficient for the bound complex at 510 nm was determined to be 77,500 M⁻¹cm⁻¹. The quantum yield for fluorescence was 0.14 as judged relative to a fluorescein standard in 0.1 mM NaOH. This value is slightly higher than previously observed for TO1 intercalation into double-stranded DNA, where quantum yields of 0.11 have been reported (Carreon et al., 2004; Nygren et al., 1998). The overall brightness of RNA Mango I was found to be ~11,000 M⁻¹cm⁻¹, implying that three repeats of RNA Mango
would generate fluorescence with intensity comparable to enhanced green fluorescent protein (brightness of 34,000 M$^{-1}$ cm$^{-1}$) (Shaner et al., 2005).

**Figure 3.2** Fluorescent properties of the RNA Mango I:TO1-Biotin complex.
Excitation and emission scans were performed with TO1-Biotin (40 nM) bound to an excess of RNA Mango I. Excitation curves are in green and emission curves are in red and blue with the corresponding peak wavelength indicated in the top.
3.1.2. Fluorescent and binding properties of RNA Mango to TO3-Biotin

![Fluorescent properties of the RNA Mango complex.
(A) TO3 fluorophore, R: Biotin-PEG₃-Amine. (B) Excitation and emission scans were performed with TO3–Biotin (40 nM) bound to an excess of RNA Mango I. Excitation curves are in green and emission curves are in red and blue with the corresponding peak wavelength indicated in the top. (C) Direct visualization of the RNA Mango I:TO1-Biotin or TO3-Biotin complexes using a hand held short wavelength UV lamp and 2 µM of either fluorophore.]

If TO1 derivatives bind to RNA Mango with nanomolar affinity, is it possible that our aptamer can induce fluorescence with other thiazole orange derivatives? Such fluorophores are of particular interest, if they allow fluorescent imaging further into the infrared region, where biologically-derived fluorescence background is greatly attenuated.
(Escobedo et al., 2010). To explore this possibility, Razvan Cojocaru, a graduate student in the Unrau lab, synthesized TO3-Biotin (Figure 3.3A), which is significantly red-shifted with respect to the TO1 fluorophore. RNA Mango I:TO3-Biotin complex (Figure 3.3B) has an excitation at 637 nm that is fully compatible with standard solid state 640 nm red lasers. Emission was found to be maximal at 658 nm, 10 nm beyond that of the far-red fluorescent protein mPlum (Shaner et al., 2005). Just as for TO1-Biotin, excitation at 260 nm together with a high wavelength shoulder (295 nm) resembling the absorbance spectrum of guanine was observed suggesting that, upon binding, TO3 fluorophore is also electronically coupled to the guanine face of the G-quadruplex found in RNA Mango I. The extinction coefficient of TO3-Acetate was found to be 9,300 M$^{-1}$cm$^{-1}$, considerably lower than that observed for TO1-Biotin. Nevertheless, a short wavelength UV handheld lamp readily generated fluorescence from both 2 μM RNA Mango I:TO1-Biotin complex and 2 μM RNA Mango I:TO3-Biotin complex (Figure 3.3C).
Figure 3.4  RNA Mango has a high binding affinity to TO3-Biotin.
Log-log plot of RNA titrations with varying concentrations of TO3-Biotin (shown in insert). Data points are connected by dotted lines. Solid lines of the corresponding color represent fits using Eq. 1 for one to one binding model. Fluorescent response at plateau has been normalized to one to highlight the divergence from a one to one binding model at increasing fluorophore concentrations.

When TO3-Biotin concentrations were lower than 50 nM a one-to-one binding model fit well to data and suggested a $K_D$ of 6-8 nM. At higher fluorophore concentrations, an enhancement in fluorescence at low RNA concentrations was observed that steadily diverged from the expected one-to-one binding model as fluorophore concentrations were increased (Figure 3.4). The source of this fluorophore-dependent effect has not been determined, but may be related to the propensity of the TO3 fluorophore to dimerize at micromolar concentrations.

3.1.3. The RNA Mango G-quadruplex is three-tiered

X-Ray crystal structure of RNA Mango I:TO1-Biotin complex has been recently solved in collaboration with D’Amare lab (Trachman et al.) and was found to have several interesting features (Figure 3.5). The RNA Mango G-quadruplex is comprised of
three G-quartet tiers (T1, T2, T3; Figure 3.5B). T1 and T2 are connected in parallel, and all guanine residues in these tiers are in anti conformation. In T1, the four nucleotides have alternating 2'-endo and 3'-endo puckers, while in T2, all four guanines adopt the 2'-endo pucker. T3, the third quartet, is unusual in that three of its four guanines (G16, G21, G26) are antiparallel to the adjacent guanines in T2 (these three residues all have anti- glycosidic bond angles and adopt the 3'-endo pucker), while the remaining guanine (G10) is parallel relative to T2 (and adopts a syn- glycosidic angle and a 3'-endo pucker) (Figure 3.5C).

While the nucleobases of T1, T2 and the three antiparallel guanines of T3 are nearly coplanar within their corresponding tiers, the base of G10 is buckled, its long axis subtending an angle of ~30° relative to the mean plane defined by the other three nucleotides of T3. This unusual orientation of G10 allows its exocyclic amine to hydrogen bond to its own pro-RP non-bridging phosphate oxygen (Figure 3.5D), and improves its packing against the biotin of the chromophore (Figure 3.5A). RNA Mango folds into its mixed parallel and antiparallel G-quadruplex using six loops to connect the twelve guanine residues of its three G-quartets. Nucleotides in four of these loops make noteworthy interactions in addition to providing connectivity. A22, which is part of a "propeller" loop connecting T3 with T1, augments T1 into a pentad by making two hydrogen bonds between its Hoogsteen face and the sugar edge of G18. U15, A20 and A25, each of which forms one of three loops that allow the RNA chain to reverse direction above T3, participate in forming the TO1-Biotin binding site (Figure 3.5A,B).
Figure 3.5  Overall structure of RNA Mango I in complex with TO1-Biotin.

(A) Cartoon representation of the RNA Mango I:TO1-complex. Curved arrows indicate direction of chain, 5' to 3'. Orange mesh depicts a simulated-annealing omit |Fo|-|Fc| map (TO1-biotin was omitted from the calculation using the final refined atomic coordinates) contoured at 3.0 σ. Purple and red spheres represent K+ ions and water molecules, respectively. (B) Side view of the RNA Mango I:TO1-Biotin complex. (C) Connectivity and stereochemistry of the RNA Mango G-quadruplex. Except for G10, which adopts the syn conformation, all nucleotides are anti- (dark nucleobase outlines). Circles denote the pucker of successive backbone riboses (open and black circles, 3'-endo and 2'-endo, respectively). The four guanine stacks are denoted by white lowercase Roman numerals. (D) Detail of a side-view of the quadruplex showing a ribose-zipper-like interaction between G8, G19 and G24 (in tiers T1, T2, and T3, respectively). Buckling of G10 allows its nucleobase to hydrogen bond to its backbone phosphate.
3.1.4. A tetraloop-like junction connects the quadruplex and the duplex

There are no tertiary interactions between the A-form duplex and the quadruplex moieties of RNA Mango I, and they connect through a junction that resembles a GAAA tetraloop (Figure 3.5B, Figure 3.6A). Canonically, such tetraloops feature a sugar-edge-to-Hoogsteen closing base pair between first and fourth residues, a chain direction reversal between the guanine and the first adenine, and stacking of the three adenines on the 3’ side of the tetraloop. In the case of RNA Mango I, G5 and A27 make the closing base pair, and the chain reverses between G5 and A6. A6 then stacks on A7 on the 3’ side of the loop. However, rather than stacking underneath A7, the next residue of RNA Mango I (G8) points ~120° away and forms part of the G-quadruplex. The fourth position in the tetraloop is instead occupied by A27, which lies immediately 3’ to the G-quadruplex (Figure 3.6B) rather than being adjacent in sequence to the second adenine of the GAAA motif. Thus, the G-quadruplex interrupts the GAAA tetraloop-like element of RNA Mango I between the second and third adenines. As in a canonical tetraloop, the 2’-OH of the guanine hydrogen bonds with the N7 of the second adenine (A7), and the Watson-Crick face of this guanine hydrogen bonds to the phosphate that follows the second adenine.

Figure 3.6 The duplex-quadruplex junction of RNA Mango I resembles a GAAA tetraloop.

(A) Cartoon representation of the junction with one flanking Watson-Crick base pair from the duplex (gray), and adjacent residues from the G-quadruplex (G8 and G26). (B) Hydrogen bonding pattern within the junction. (C) Junction of RNA Mango superimposed on a canonical GAAA tetraloop (gray; PDB 4FNJ).

However, in the junction, this is the phosphate of A27, rather than that of the next residue in the RNA chain (G8). Despite the interruption in the backbone, the GAAA motif of RNA Mango I superimposes closely (rmsd = 0.29 Å for all non-hydrogen atoms,
excluding the phosphate of A27) on a conventional tetraloop (Figure 3.6C). Unlike many GAAA tetraloops, which are involved in A-minor and stacking tertiary interactions, the junction of RNA Mango makes none except crystal contacts.

3.1.5. Circular permutation of RNA Mango I

To further explore the importance of the GAAA junction, I circularly permuted its site of attachment to the G-quadruplex from its wild-type locus (G8 and G26 in stacks i and iv, respectively; stack numbering defined in Figure 3.5C) to each of the other three faces of the G-quadruplex (Figure 3.7). Remarkably, all three circular permutations decreased binding affinity by less than 2-fold and had only marginal effects on fluorescence enhancement. Further, removing the junction sequence and stem from these circular permutations resulted in a further 2 to 8 fold decrease in binding affinity. Since these circular permutations systematically disrupted each of the three propeller loops connecting the adjacent guanine stacks of the G-quadruplex, neither the site of the GAAA junction insertion nor the detailed sequence of the propeller loops responsible for linking adjacent stacks of the G-quadruplex are essential for binding. Circular permutation and alteration of the loops did significantly impact fluorescence enhancement, reminiscent of the altered intrinsic fluorescence observed for loop connectivity in model G-quadruplexes \(^{30,32}\).
Figure 3.7  Effect of circular permutation of G-Quadruplexes in RNA Mango
(A) through (D), effect on $K_d$ and relative fluorescence enhancement ($F_E^\ddagger$) of circular permutation (CP) of the connectivity of RNA Mango by either moving the attachment point of the tetraloop-like junction in a full-length (FL) 35-nt construct, or by deleting the duplex and junction (Δ). Guanines of the three tiers are represented as squares, colored as in Figure 3.5. The four guanine stacks are numbered as in Figure 3.5C. The two adenosines that cover the two heterocycles of the TO moiety of the fluorophore are represented as yellow rectangles.
3.2. Using bipartite RNA Mango to develop a fluorescent turn ‘ON’ system for RNA imaging

3.2.1. Rationale

Since its publication, RNA Mango has attracted attention of international RNA research community leading to tagging of several RNAs with RNA Mango. However, in its current state the RNA of interest needs to be co-transcriptionally tagged and introduced into the cells either through chromosomal recombination of the DNA template or through transfection. Expressing the tagged RNA of interest through plasmid or through transfection will alter the gene expression levels thereby it might not accurately reflect the biologically relevant states in the wild type cells. Therefore, we wanted to develop a system, where using RNA Mango, endogenous RNAs can be fluorescently tracked (Figure 3.8)

In order to have a specific RNA based signal, such system should be fluorescently ‘off’ on the absence of the RNA of interest and should get turned ‘on’ in its presence. Since this system will require two components. An RNA mango module as well as the endogenous RNA for a fluorescence turn ‘on’, this system is referred as bipartite Mango. Potential applications of the bipartite Mango system not only include imaging of the endogenous RNA of interest in vivo, but also include in vitro detection of the RNA of interest that has immense diagnostic application.

![Image](image_url)

**Figure 3.8 Imaging scheme using bipartite RNA Mango.** Bipartite RNA Mango construct targeting an RNA of interest and TO1-B can be transfected directly into Wild type cells and can be fluorescently tracked.
Currently for RNA based visualization of the unmodified endogenous RNA of interest, several modification of RNA Fluorescence In situ Hybridization (FISH) is being used (Tyagi, 2009) (Figure 3.9). In all those methods, a fluorogenic nucleic acid probe that is complementary to a region in the RNA of interest is introduced into the cells. The fluorogenic probe has a fluorescent dye and a quencher molecule in a confirmation that keeps both of them spatially together to quench the fluorescence in the unbound state. Once the probe binds the RNA of interest, the fluorescent dye and the quencher are spatially separated to turn the fluorescence ‘on’ specifically upon binding to the RNA of interest. This methodology requires chemical synthesis of nucleic acid probes with both the fluorescent dye as well as the quencher, which raises the cost of the probes considerably. We wondered if such a probe could be constructed with RNA Mango, as it doesn’t require covalent attachment of the dye to the RNA Mango.

![Schematic representation of different fluorogenic probes for the detection of mRNAs. Donor (green), quencher (gray) and acceptor (yellow) dyes are attached to the probes and interact as indicated. From Tyagi, 2009.](image)

We noted in the X-Ray crystal structure of the RNA Mango I that the ‘GAAA’ tetraloop motif with ‘GAA’ on the 5’ end of the G quadruplex core and the remaining ‘A’ on its 3’ end separates the two phosphates in the end of the core by 8 Å. We
hypothesized that this 8 Å separation is crucial for fluorescence enhancement with RNA Mango I, as it is places the G quadruplex at optimal configuration for ligand binding. We tried to recreate the optimal separation of the two phosphates through complementary base pairing of the Mango tag (called the ‘switch’) to a linear RNA (called the ‘trigger’), which would mimic the RNA of interest (Figure 3.10).

![Diagram of Mango system](image)

**Figure 3.10 Design of bipartite Mango system**
Mango core is in red, flanked by complementary regions, also called as ‘bait’ (in green) to the RNA of interest (blue). Experiments are done to find the junction which gives optimal 8 Å separation for maximum fluorescence turn ‘on’ upon binding to the RNA of interest.

An ideal turn ‘on’ switch will have the following characteristics,

- It has minimal fluorescence in the absence of a ‘trigger’ RNA
- It has maximum fluorescence upon binding the ‘trigger’ RNA
- It could be used to tag a wide variety of RNA of interest
Figure 3.11  Initial screening to find optimal “switch” and the “trigger” constructs for RNA Mango I based bipartite detection system

Red line is the core Mango I G-quadruplex “GGG AC GGUG C GGAG A GGAG”, in bold are the nucleotide likers that connect the G-quadruplex and trigger binding region tested.

<table>
<thead>
<tr>
<th>Schematic</th>
<th>RFU</th>
<th>Schematic</th>
<th>RFU</th>
<th>Schematic</th>
<th>RFU</th>
<th>Schematic</th>
<th>RFU</th>
</tr>
</thead>
</table>

The Figure 3.11 shows various Mango switches and triggers tested. We found that the Mango switch with two A residues each flanking on the either side of the G – quadruplex core gave maximal response to the binding of the trigger RNA, the reason for which is currently unknown but could be explained by with a crystal structure of the bipartite constuct. We also found that those triggers, which are continuously complementary to the bait region, without any uncomplimentary base in the middle to be optimal for fluorescence turn on.

As a proof of principle, we did our experiments using β-actin mRNA from mouse embryo fibroblasts as a model. β-actin mRNA, because of its Zipcode region binds to
Zipcode Binding Protein (ZBP) (Kislauskis et al., 1994) which then binds to cytoskeleton proteins and get transported to the leading edges of lamellipodia in growing fibroblasts. (Figure 3.12A). We have developed RNA Mango constructs that can target β-actin mRNA in those regions which has previously shown not to interfere with the localization process (Briley et al., 2015) (Figure 3.12B). We have developed another RNA Mango construct that target the Zipcode binding region, which when targeted causes delocalization of β-actin mRNA (Lynette Schedlich et al., 1997) (Figure 3.12B).

**Figure 3.12 Bipartite mango probes for b-actin mRNA**

(A) Model for β-actin mRNA transport. mRNA synthesized in the nucleus binds to Zipcode binding protein (ZBP1) which then associates with other proteins to bind to actin. The RNP then translocates to the leading edge of lamellipodia. (B). Fluorescence response of Mango I constructs that target b-actin mRNA. The Bipartite switch (50 nM) was added (at 10 min) to (100 nM) TO1-B. After 60 mins, the complementary *in vitro* transcribed ‘trigger’ RNA (100 nM) was added. (C) shows the sequence of the probe and the target sequences used.

In order to quantitatively compare different bipartite constructs, we define a term called contrast as the ratio of fluorescent increase upon binding of the bipartite construct with dye to the target. Higher contrast means low background thus high signal to noise ratio. For instance, in the above Mango I based bipartite construct targeting β-actin from 1087-1121, the equilibrium RFU after the addition of TO1-Biotin to the Mango construct is 3 (point a in Figure 3.12B) while after adding the target it raised up to 321 (point b in
Figure 3.12B), giving raise to contrast of ~110. In the same manner the contrast for the Mango construct targeting the zipcode region (2955-2990) of β-actin, is ~80.

### 3.2.2. Bipartite construct with Mango II

In collaboration with Michael Ryckelynck’s lab in France, using fluorescence-activated droplet sorting (FADS) (Baret et al., 2009), several improved variants of Mango were selected. These were characterized by Sunny Jeng, a graduate student in the Unrau lab. Three interesting candidates were selected and were found to have superior binding and fluorescence enhancement. We predicted that the superior dye binding and fluorescence enhancement properties of the improved variants are due to their ability to form stronger G-Quadruplex cores. Of the three improved variants, named Mango II to Mango IV, Mango II was predicted to have similar overall structure to Mango I. Since the Mango II had improved binding affinity and fluorescence when compared Mango I (Table 3.2), we sought to develop bipartite Mango constructs with Mango II. Surprisingly, we found that fluorescence of the Mango II construct by itself were significantly fluorescent, decreasing the signal to noise ratio (Figure 3.13, MII).

<table>
<thead>
<tr>
<th>RNA</th>
<th>Sequence</th>
<th>KD</th>
<th>Rel. $F_{\text{max}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mango I</td>
<td>GGCACGTAC GAAAGGACGGTGGAGGAGGAGA GTACGTGC</td>
<td>2.6</td>
<td>1</td>
</tr>
<tr>
<td>Mango II</td>
<td>GGCACGTAC GAAAGGACGGGAGGAGGAGA GTACGTGC</td>
<td>0.35</td>
<td>1.5</td>
</tr>
</tbody>
</table>

Table 3.2 RNA Mango variants

Table showing sequence of RNA Mango variants with Gs predicted to be participating in the G-Quadruplex in red along with binding affinity (KD) in nM and Fluorescence enhancement relative to Mango I (Rel. $F_{\text{max}}$).
Figure 3.13 Bipartite construct with Mango II and inhibitory stem
(A and B) Schematic representation of Mango II construct without inhibitory stem (MII) and with inhibitory stem (MII +4i). (C) Fluorescence response of the bipartite constructs. The Bipartite switch (50 nM) was added (at 10 min) to (100 nM) TO1-B. At 70 mins, the complementary in vitro transcribed ‘trigger’ RNA (100 nM) was added. Right - Table showing the contrast of bipartite constructs with and without inhibitory stem.

We hypothesized that Mango II core partially folds without the trigger, leading to a stronger fluorescence enhancement. Hence we decided to add an inhibitory stem that would destabilize the G quadruplex core in the absence of mRNA trigger. The inhibitory stem was designed in such a way that one of the stacks would be complementary to the inhibitory stem in the absence of the trigger. Upon the addition of the trigger, due to the toehold on the 3’ end of the bait region in the bipartite switch, the interactions between
the G stack and the inhibitory stem will be destabilized. We indeed found that the addition of inhibitory stem not only decreased fluorescence without the trigger, but also resulted in increased total fluorescence upon the addition of the trigger (Figure 3.13).

The length of the inhibitory stem is critical for two reasons. If the length is too short, it would not adequately destabilize the G-quadruplex, thus decreasing the contrast. If it were too long, then the inhibited structure would become very stable thus taking more time for the proper folding of G-quadruplex upon trigger binding. The net effect would be either increase in the response time or permanently destabilizing the G-quadruplex core, both of which are undesirable. Hence, we sought to determine optimal inhibitory arm length to use with Mango II type constructs.

![Figure 3.14 Effect of inhibitory stem length on contrast and rate of fluorescence emergence](image)

(A) The number associated with the MII annotation refers to the number of interfering nucleotides. The Bipartite switch (50 nM) was added (at 20 min) to (100 nM) TO1-B. After 80 mins, the complementary in vitro transcribed ‘trigger’ RNA (100 nM) was added. (B) Sequences of the constructs used in the panel (A). In bold are potentially base pairing nucleotides that destabilizes the G-quadruplex in the absence of trigger.

Increasing the inhibitory arm length by increments of one nucleotide revealed that an inhibitory arm containing five nucleotides gave high contrast with minimal
decrease in the rate of fluorescence increase (Figure 3.14). Thus a general RNA Mango based bipartite construct has three parts (Figure 3.15).

Figure 3.15 General RNA Mango based bipartite construct. See text for more explanation.

I. RNA mango core (Red) – Aptamer part that binds to TO1-B and becomes fluorescent. Can be either Mango I or Mango II core.

II. An inhibitory arm (Orange) – region complementary to the RNA Mango so that it binds and destabilizes RNA Mango core in the absence of a ‘trigger’ RNA, decreasing fluorescence in the absence of the RNA of interest.

III. Bait (IIIa, IIIb and IIIc Green) – A region, which is complementary to the RNA of interest, sequence depending upon RNA of interest.

Initial testing of the bipartite system in Drosophila embryos is promising. More work is currently underway in the Unrau lab for testing these constructs in vivo in tissue culture systems.
3.3. Integration of RNA Mango into the bacterial 6S transcriptional control RNA

Many biological RNAs contain stem loops that are not essential for function. For example, the 6S transcriptional control RNA, which binds and releases from bacterial RNA polymerase (H.E.) via an NTP-dependent mechanism (Panchapakesan and Unrau, 2012; Wassarman and Saecker, 2006), has a nonconserved stem loop structure (Barrick et al., 2005) (Figure 3.16A). We replaced this loop sequence with RNA Mango (Figure 3.16B) and found that, in vitro, the construct allowed the RNA Mango-dependent monitoring of 6S RNA binding and release. The properties of the RNA Mango-tagged 6S was indistinguishable from untagged 6S RNA monitored as judged by radiolabelling of both constructs (Figure 3.16C). The robustness of this binding interaction was further reinforced by a native gel shift experiment wherein the 6S RNA Mango, when bound to TO1-Biotin can be bound to streptavidin (Figure 3.16D) as well as capturing the RNA Mango tagged 6S RNA construct onto streptavidin beads using TO1-Biotin as a handle. After three washes with WB, 87% of the construct remained bound to beads, while unmodified 6S RNA binding could not be detected as judged by liquid scintillation counting. As a final demonstration we overexpressed the RNA Mango tagged 6S RNA in *E. coli* and cleanly purified the expressed construct away from total RNA extract (Figure 3.16E). Together these experiments strongly suggest that RNA Mango can be used to simultaneously track and purify RNA complexes via its high affinity binding to TO1-Biotin.
Figure 3.16  RNA Mango allows monitoring of 6S:holoenzyme complex formation and purification of 6S RNA from total RNA.

(A) The 6S RNA has been associated with three states corresponding to unbound 6S RNA (F), 6S RNA bound to E. coli holoenzyme (B) and 6S RNA:pRNA released from holoenzyme by addition of NTPs (R). (B) RNA Mango (red) replaced the 5-nt stem loop found in the 6S RNA. The "-35" binding and "-10" regions correspond to highly conserved sequence within the 6S RNA and are shown in purple. (C) Radiolabelled 6S RNA and 6S RNA Mango were prepared in the F, B and R states and found to behave identically in a native gel assay (phosphorimager scan, bottom panel). RNA Mango fluorescence was captured by a Typhoon imager (488 nm excitation laser, PMT at 600 V, 526 SP filter, top panel). (D). 25 nM 6S RNA or the 6S RNA Mango was incubated with combinations of 250 nM TO1-Biotin and or an excess of streptavidin and run into a 5% native gel. The image in sub panel i. was recorded using a Syngene U:Genius3 UV transilluminator, while sub panel ii. was recorded by a Typhoon Trio+ Variable Mode Imager using a FUJI phosphorimager plate. (D) Overexpressed 6S RNA Mango could be purified from bacterial total RNA, using streptavidin beads loaded with TO1-Biotin.

3.4. Conclusion:

RNA Mango offers significantly higher fluorescent efficiency than either RNA Spinach or the MS2 systems (Table 3.1). The small size and simple closing stem structure of RNA Mango indicates that it will be straightforward to incorporate this aptamer into arbitrary biological RNAs without disrupting biological function, as we have demonstrated here using the 6S RNA as an example. The stability of RNA Mango within
C. elegans gonads, the established permeability of TO1 fluorophores into cells (Weston and Parish, 1990) and bacteria together with the very low intrinsic fluorescence of unbound TO1 fluorophores suggests the potential for RNA Mango to track the spatial and temporal expression of low copy number RNAs within living organisms. Perhaps equally important, the robust binding interaction between RNA Mango and its fluorophore provides a means to not only visualize RNA complexes, but to also purify such constructs using a single small RNA aptamer tag. In the next chapter using a modified version of the fluorophore, I have developed a native purification methodology for the purification of RNP complexes using RNA Mango.
3.5. MATERIALS AND METHODS

3.5.1. RNA synthesis:

Circular permutation constructs without duplex and junction (Figure 3.7, CP Δ) were chemically synthesized from IDT. All other RNA constructs were in vitro transcribed with T7 RNA polymerase using standard protocol.

3.5.2. Fluorescence measurement

All fluorescence measurement were done in 140 mM KCl, 10 mM Sodium Phosphate Buffer, 1 mM MgCl₂, 0.05% TWEEN-20. Fluorescence titration for measuring $K_D$ in Figure 3.7 was done with a Varian Cary eclipse Spectrofluorometer with 10 nM TO1-Biotin and 2-fold dilution of the corresponding RNA from 512 nM. Fluorescence data from three independent titrations were plotted and fit to a one to one binding model and binding affinity was obtained. Fluorescence kinetics were obtained in Varian Cary eclipse Spectrofluorometer with indicated amounts of the bipartite switch, dye and the trigger as indicated in the figure legend. Fluorescence data in Figure 3.11 was obtained from Spectramax M5.

3.5.3. Integration of RNA Mango I into bacterial 6S RNA

Loop nucleotides 94–98 (sequence 5'-UCCGU) were removed from the *E. coli* 6S RNA and replaced with the core of RNA Mango: 5'-UAC GAA GG GAC GG UGC GG AGA GG AGA GUA together with three additional base pairs of closing stem (in italics). Binding and release assays were performed as described elsewhere using 25 nM RNA, but with the addition of 250 nM TO1-Biotin. 6S RNA construct release was performed for 15 min, after the addition of NTPs and 4 mM MgCl₂ and stopped on ice after the addition of an equivolume of loading dye (50% glycerol, 30 mM HEPES pH 7.5, and 180 mM KCl). Samples were analyzed by 5% native PAGE, supplementing the gel with 0.5% glycerol. The gel was made and run using 15 mM HEPES pH 7.5, and 90 mM KCl at 4 °C.
3.5.4. In vivo expression of 6S RNA Mango

6S RNA mango (RNA Mango core shown in purple, 6S RNA in red) is flanked by SgrAI restriction site (green) and a 6S RNA σ\textsuperscript{70} promoter (brown) on the 5’ end and intrinsic terminator sequence (orange) and Clal site (blue) is inserted into pEcoli-Cterm-6xHN plasmid between SgrAI and Clal sites as above.

```
CGC CGG TGA AAT TTT GAA TGA CAC TTT TCG GTT TAC TGT GGT AGA GTA
ACC GTG AAG ACA AAA ATT TCT CTG AGA TGT TCG CAA GCG GGC CAG TCC
CCT GAG CCG ATA TTT CAT ACC ACA AGA ATG TGG CGC TCC GCG GTT GGT
GAG CAT GCT CGG TAC GAA GGG ACG GTG CGG AGA GGA GAG TAC CGA
GAA GCC TTA AAA CTG CGA CGA CAC ATT CAC CTT GAA CCA AGG GGT CAA
GGG TTA CAG CCT GCG GCG GCA TCT CGG AGA TTC CAA AAC AAA AGG
CTC AGT CGG AAG ACT GGG CCT TTT GTT TTA TCG AT
```

\textit{E. coli} cells were transformed with the resulting pEcoli-6S RNA Mango plasmid. A colony was picked for overnight inoculation in LB media containing 100 μg/ml ampicillin. It was diluted 50 fold in LB/Amp and RNA was extracted by RNAsnap\textsuperscript{TM} method (10) when OD600 reached 2.7. The resulting RNA was precipitated with ethanol for better quantification. A 100 pmol portion of TO1-Biotin was added to 25 μL of Dynabeads M-270 streptavidin that was prepared as for in vitro selection. Total RNA (5 μg) was supplemented with 100 μg/mL heparin, 15 mM HEPES pH 7.5, and 90 mM KCl and added to the dye-bound beads. After two washes in 15 mM HEPES pH 7.5 and 90 mM KCl, the bound 6S RNA Mango was recovered in 100 μL of 98% formamide and 10 mM EDTA after a 10 min incubation on a rotator, and 1/10th of the feed, flow through, washes, and eluted RNA were analyzed via 8% denaturing PAGE. The gel was stained with 1X SYBRGreen (Invitrogen) and imaged using an UV trans-illuminator.
Chapter 4. A Multipurpose RNA Toolset for Ribonucleoprotein Purification and Characterization

This chapter is based on the following manuscript:

Shanker Shyam S. Panchapakesan, Matthew Ferguson, Eric Hayden, Xin Chen, Aaron A. Hoskins, and Peter J. Unrau “A Multipurpose RNA Toolset for Ribonucleoprotein Purification and Characterization” (Manuscript in preparation) wherein I conceived and executed the concept native RNP purification with desthiobiotinylated TO1 and RNA Mango and demonstrated RNP purification of both bacterial and yeast snRNP complexes. In addition, I have also performed the two-color, in gel fluorescent analysis of dual labelled RNP complexes.
Abstract

The characterization of RNA-protein complexes (RNPs) is a difficult and increasingly important problem in modern biology. By combining the compact RNA Mango aptamer with a fluorogenic thiazole orange desthiobiotin (TO1-Dtb) ligand, we have created an RNA tagging system that simplifies the purification and subsequent fluorescent biophysical characterization of endogenous RNPs from native cellular extracts. The Mango:TO1-Dtb tagged RNP complex can be immobilized and washed on streptavidin solid support and recovered by the addition of free biotin. Since the eluted complex(es) are fluorescent they can be easily tracked during further purification steps. Using RNA Mango, we purified *E. coli* 6S RNA:RNA polymerase holoenzyme (HE) RNPs as well as a small nuclear RNP component of the *S. cerevisiae* spliceosome. In addition to purification, the unique fluorescence properties of the RNA Mango/ligand complex facilitates the biophysical characterization of RNP complexes by both bulk fluorescence and single-molecule fluorescence cross-correlation spectroscopy. Together, we demonstrate how the Mango toolkit simultaneously enables the purification and characterization of endogenous cellular RNPs *in vitro*.
4.1. INTRODUCTION

RNP complexes play essential roles in gene expression and regulation in all domains of life (Cech and Steitz, 2014a; Wang and Chang, 2011), yet many RNPs remain poorly characterized due to the challenges inherent in their purification and subsequent biochemical characterization. While protein-based tags can be used for both purification (Lichty et al., 2005) and fluorescent labeling of RNPs (Cranfill et al., 2016), few options exist that simultaneously facilitate affinity-based native purification and/or fluorescent labeling of the RNA components of an RNP complex (Panchapakesan et al., 2015). Small RNA aptamer tags that are easy to insert into an RNA of interest and that bind to derivatizable fluorophore tags offer a potential solution to this problem. To be effective, such an approach requires high affinity aptamers able to bind their fluorophore ligands so that that the fluorescent aptamer/fluorophore complex can also serve as purification handle. A key problem in this respect has been the difficulty in isolating aptamer fluorophore pairs that have sufficiently high binding affinity and brightness to serve as a dual use RNA tag.

We have used Mango (Dolgosheina et al., 2014) to purify RNPs complexes from both prokaryotic and eukaryotic sources using a modified RNA Mango fluorophore ligand. E. coli 6S RNA binds HE and inhibits σ70 dependent transcription (Wassarman and Storz, 2000). Release of this regulatory RNA from the HE is triggered by the synthesis of a short pRNA templated by the 6S RNA itself (Panchapakesan and Unrau, 2012; Wassarman and Saecker, 2006). Using a release defective (Oviedo Ovando et al., 2014) Mango tagged 6S RNA (6SRDM), we purified an intact 6S RNP containing all of the HE subunits in two steps. Mango insertions can maintain biological function of endogenous cellular RNPs (Dolgosheina et al., 2014), and we show that a single Mango tag insertion into the S. cerevisiae (yeast) U1 small nuclear RNA (snRNA) maintains yeast viability and enables purification of snRNPs. In addition to purification, Mango also enables the fluorescent characterization of Mango-containing RNP complexes by both bulk and single-molecule methods.
4.2. RNA Mango for native batch RNP purification

The Mango aptamer binds with nanomolar affinity to thiazole orange (TO) derivatives such as TO1-Biotin and can increase the fluorescence of such fluorophores by ~1,000 fold (Dolgosheina et al., 2014). It therefore appeared to be an ideal aptamer/fluorophore system for native RNP purification and fluorescent characterization. Our interest in this system was increased by the recent finding that the Mango aptamer consists of a compact 19-nt fluorophore binding G-quadruplex domain (‘...GAA GG GAC GG UGC GG AGA GG AGA...’) that is physically connected to an arbitrary stem by a novel GAAA loop-like motif in bold (Trachman et al.). Since many biological RNAs use simple GNRA tetraloops (Uhlenbeck, 1990) or other simple motifs to allow the formation of an RNA hairpin, we reasoned that the Mango tag might be biologically well tolerated if it was used to replace pre-existing stemloops whose sequences are not biologically essential (Figure 4.1).

In order to enable native purification of RNPs, we synthesized TO1-Desthiobiotin (TO1-Dtb, Figure 4.1B) as desthiobiotin can be displaced from streptavidin by the addition of free biotin (Hirsch et al., 2002). TO1-Dtb binds to the RNA Mango tag with nanomolar affinity, similar to what has been reported for other TO1 derivatives (Dolgosheina et al., 2014) (Figure 4.1C) and has fluorescence excitation (Ex) / fluorescence emission (Em) centered at 510 nm and 535 nm, respectively. Using this ligand, we developed a protocol for the native purification of RNP complexes using the Mango tag (Figure 4.1D). Native extract containing the Mango-tagged RNA is first bound to a TO1-Dtb saturated streptavidin solid support. Binding can be performed either at RT or at 4 °C depending upon the stability of the RNA: protein complex. The solid support is then washed to remove non-specific components of the extract with the stringency of washing being dependent upon the stability of the complex under study. After washing, the complexes are eluted by the addition of biotin.
Figure 4.1  Native purification of RNP complexes using the fluorescent, high affinity Mango toolset.

(A) RNA Mango aptamer core quadruplex and the GAAA tetraloop-like motif sequence (in brown). N and N’ are complementary base pairs of arbitrary sequence forming a RNA duplex (Trachman et al.) (B) Structures of Dtb-derivatized TO1 (n=1) or TO3 (n=3) fluorophores. (C) Determination of the equilibrium binding constants ($K_D$) for TO1-Dtb and TO3-Dtb. Error bars at each data point represent standard deviation from three independent measurements. See methods section for details of the $K_D$ determination. (D) General schematic for purification of native RNP complexes using Mango.

To purify endogenous 6S RNPs formed in vivo, we tagged a release-defective mutant of the 6S RNA (Oviedo Ovando et al., 2014) with Mango ($6S^{RDM}$) and overexpressed $6S^{RDM}$ in E. coli cells. Cell lysate from the $6S^{RDM}$-expressing bacteria was incubated with TO1-Dtb derivatized streptavidin agarose beads and washed, and the desthiobiotin-bound material was then eluted by addition of biotin (Figure 4.2A). LC-
MS/MS analysis revealed that this single-step purification yielded the four core RNA polymerase proteins as the highest ranked polypeptides, consistent with enrichment of intact HE (Appendix A).

![Figure 4.2](image_url)

**Figure 4.2  Mango mediated purification of 6S$^{RDM}$**

Purification of 6S$^{RDM}$ RNA from *E. coli* extracts using TO1-Dtb derivatized streptavidin agarose beads. Samples from the native extract (N.E.), column flow through (F.T.), wash (W) and biotin eluate (E) from non-denaturing purification of 6S$^{RDM}$ were phenol:chloroform extracted and loaded alongside *in vitro* synthesized 6S$^{RDM}$ RNA (I) before analysis by 6% denaturing PAGE (19:1 acrylamide:bis-acrylamide). Nucleic acids were visualized by staining the gel after electrophoresis with SYBR Green.

**4.3. Fluorescence tracking of the Mango purified RNP complexes facilitates further purification**

As the biotin eluted complex(es) should be both fluorescent and contain intact RNPs, we reasoned that size exclusion chromatography with fluorescence detection could be used to further purify the 6S RNA: HE complex away from unbound 6S RNA. Three TO1-Dtb dependent fluorescent peaks were observed (Figure 4.3). SEC analysis identified two of these peaks as free TO1-Dtb and 6S$^{RDM}$:TO1-Dtb (data not shown), while the earliest-eluting peak was consistent with expected size of the 6S$^{RDM}$:HE RNP. LC-MS/MS analysis confirmed this hypothesis with all five components of the HE being present with only two additional proteins being identified (Appendix B). Thus, Mango provides a versatile tool for rapid visualization, purification and analysis of RNPs.
The two additional proteins found after Mango purification of the 6S RNP were the 60 kDa chaperonin protein (GroEL) and thioredoxin, and our evidence suggests that these proteins make stable interactions with HE. Previously, it has been shown that in the absence of RNAP ω subunit, the 60 kDa chaperonin facilitates assembly of the core enzyme (Mukherjee et al., 1999). Hence, it could be possible that in a fraction of RNAP HE bound to 6S\textsuperscript{RDM}, 60 kDa chaperonin might substitute for the ω subunit. Consistent with this, 60 KDa chaperonin was also present in a commercial holoenzyme preparation (Appendix C). Thioredoxin is a small \textit{E. coli} protein that has previously been shown to bind to the thioredoxin binding domain of T7 DNA Polymerase (T7 DNAP), exposing some of its positively charged basic residues to the negatively charged backbone of DNA, thereby stabilizing the protein:primer template complex (Ghosh et al., 2008). Thioredoxin has also been shown to increase the processivity of T7 DNAP and insertion of the thioredoxin binding domain from T7 DNAP into \textit{E. coli} DNAP (Bedford et al., 1997). Since thioredoxin was found in both the 6S\textsuperscript{RDM}:HE RNP and in commercial HE, the combined evidence suggests that these proteins...
two proteins are able to make stable association with at least fractions of both HE and the 6S:HE RNP found within an *E. coli* cell lysate.

### 4.4. Purification of yeast U1 snRNP using RNA Mango

To study the utility of RNA Mango based purification of eukaryotic RNPs, we inserted the Mango motif into the stem loop VII of the 568 nt *Saccharomyces cerevisiae* (yeast) U1 sRNA (*U1^M*). The U1 snRNA is the RNA component of the U1 snRNP, initially recognizes 5' splice sites of introns during spliceosome assembly (Seraphin and Rosbash, 1989), and is essential for eukaryotic cell growth. This *U1^M* gene, which replaced the 4-nt loop of U1 stem loop VII with the core Mango sequence (Figure 4.4A), also replaced the wild type (WT) U1 snRNA allele in a U1 deletion (*snr19D*) strain and was able to maintain yeast viability. The presence of the *U1^M* gene and absence of the WT allele was confirmed by PCR and DNA sequencing (Figure 4.4B and C). The survival of *U1^M* yeast demonstrates that Mango can be incorporated into a non-coding RNA essential for eukaryotic cell proliferation while maintaining the RNA’s biological function.
Figure 4.4  Confirmation of integration of U1\textsuperscript{M} into yeast genome
(A) Schematic secondary structure of U1\textsuperscript{M}. Secondary structure of U1\textsuperscript{M} snRNA with Mango (M, red) inserted into stem loop VII. (B) Nucleotide base calls from DNA sequencing results of a PCR product resulting from amplification of the U1 (WT) or U1\textsuperscript{M} genes. (C) Chromatograph from the U1\textsuperscript{M} strain showing the Mango sequence insertion and absence of the U1 WT allele.
Since U1 snRNA associates with 17 proteins to form the yeast U1 snRNP (Gottschalk et al., 1998), we used the Mango/TO1-Dtb toolset to purify both the snRNA and snRNP proteins using native conditions. The U1\textsuperscript{M} snRNA was highly enriched following Mango/TO1-Dtb purification as judged by \textit{in vitro} solution hybridization (Figure 4.5A). A dual-labeled yeast strain containing both U1\textsuperscript{M} and a fast SNAP-tagged U1 snRNP protein (Snp1-SNAP\textsubscript{f}) was also prepared in order to easily detect co-purification of snRNA and protein components of the U1. The U1\textsuperscript{M} snRNA was enriched only in Mango-tagged strains while another non-coding snRNA (U2) was not, indicating that enrichment was specific for the Mango tag (Figure 4.5B). U1\textsuperscript{M} and Snp1-SNAP\textsubscript{f} could only be detected together from extracts prepared from strains containing both U1\textsuperscript{M} and Snp1-SNAP\textsubscript{f} (Figure 4.5C). Thus, not only is U1\textsuperscript{M} viable in yeast but the Mango aptamer appears to be stable, folded, and functional for RNP purification from eukaryotic cell lysates.

**Figure 4.5 Native purification of yeast U1 snRNP complex using Mango.**

(A) Solution hybridization analysis of snRNAs present in native extract (N.E.) from U1\textsuperscript{M} tagged yeast and following Mango batch purification using TO1-Dtb (Biotin Eluate). The probed snRNAs are labeled on the right and correspond to the number labels above each lane. U1\textsuperscript{M} migrates as a doublet following RNA deproteinization and denaturation, but as a single band upon Mango based purification. (B) Primer extension analysis to detect U1\textsuperscript{M} or U2 snRNAs following TO1-Dtb bead based purification. (C) SDS-PAGE analysis of Snp1-SNAP\textsubscript{f} following Mango purification.
4.5. Use of RNA Mango in Multi-Wavelength Fluorescence EMSAs of 6S RNP Assembly and Disassembly

Next, we tested the functionality of the Mango system for multi-wavelength fluorescence imaging. We reasoned that since the Mango:TO3-Dtb complex (Ex/Em 637/658 nm, Figure 4.1B) and enhanced GFP (eGFP, Ex/Em: 488/509 nm) are spectrally distinguishable, this pair could be used to dual-label RNP complexes. We combined the TO3-Dtb ligand with a Mango labeled 6S RNA (Dolgosheina et al., 2014) (6S\textsuperscript{M}) and prepared partially purified HE tagged on the β' subunit with eGFP (eGFP-HE) (Bratton et al., 2011). We mixed eGFP-HE with 6S\textsuperscript{M} and TO3-Dtb and analyzed the assembly and pRNA dependent release of 6S\textsuperscript{M} from the HE complex by an EMSA (Wassarman and Saecker, 2006). As little as 10 pmol of the 6S\textsuperscript{M}:HE complex could be clearly imaged by dual-color fluorescence imaging. Upon the addition of partially purified eGFP-HE, 6S\textsuperscript{M} formed a single dual labeled RNP complex (Figure 4.6). As expected, the addition of NTPs and MgCl\textsubscript{2} triggered release of 6S\textsuperscript{M} RNA from eGFP-HE and produced a faster mobility fluorescent 6S\textsuperscript{M}:pRNA band (Figure 4.6). Mango:TO3-Dtb in combination with eGFP-labeled proteins therefore enables biochemical characterization of RNP complex formation without radioactive labeling of RNA transcripts or the purchase of more expensive, chemically derivatized fluorescent RNA oligonucleotides.
Figure 4.6  Multi-wavelength fluorescence characterization of the 6S$^{M}$:HE RNP. EMSA assay of the 6S$^{M}$:eGFP-HE complex visualized using eGFP-HE (green, top) and 6S$^{M}$ :TO3-Dtb (red, middle) fluorescence. The bottom panel is the composite of the green and red images. Product RNA (pRNA) release was induced by the addition of NTPs and MgCl$_2$.


Finally, we set out to test the effectiveness of RNA Mango for single-molecule analysis of dual-labeled RNPs. Fluorescence cross-correlation spectroscopy (FCCS) is a single-molecule approach that can be used to study biomolecular interactions in real-time (Bacia and Schwille, 2007). Utilizing a two-photon excitation source has further advantages, serving to minimize photo bleaching and photo damage (Zipfel et al., 2003). We carried out a two-photon FCCS analysis of RNPs containing in vitro synthesized 6S$^{M}$ bound to the TO3-Dtb ligand and reconstituted with partially purified eGFP-HE. Analysis of the two-photon fluorescence excitation spectrum for 6S$^{M}$:TO3-Dtb identified a peak near ~840 nm (Figure 4.7A). At this two-photon wavelength, we were able to determine the diffusion coefficient of 6S$^{M}$:TO3-Dtb ($D_{6SM} = 140 \pm 8 \, \text{µm}^2/\text{s}$) and eGFP-HE ($D_{HE} = 58 \pm 2.1 \, \text{µm}^2/\text{s}$) in isolation (Figure 4.7B).
Figure 4.7  Two-photon fluorescence correlation spectrum of 6SM:TO3 and eGFP-RNAP

(A) Two-photon action spectrum spectrum for 6S\textsuperscript{M}:TO3-Dtb. Fluorescence intensity of a solution containing 1 µM each 6S\textsuperscript{M} and TO3-Dtb is shown over a range of two-photon excitation wavelengths. Data (red) were acquired at 100 Hz. The black line represents averaging of the data using a moving block average function with a 0.5 s block size. (B) Fluorescence correlation spectrum of 6S\textsuperscript{M}:TO3-Dtb and eGFP-HE. Fluorescence auto and cross correlation functions and fits of separate solutions of 6S\textsuperscript{M}:TO3-Dtb at 33.6 ± 0.7 nM (left), and eGFP-HE (from partially purified extract, right) at 34.4 ± 0.4 nM. Red (auto, 6S\textsuperscript{M}:TO3-Dtb) and Yellow (cross) correlations functions are shown on the left axis with Green (auto, eGFP-HE) are shown on the right axis for better scaling. Residuals from these fits are shown in the middle panel. The bottom pair of panels show detected counts per second as a function of time.
Next, we added increasing amounts of eGFP-HE (green channel) solution to the 6S<sup>M</sup>:TO3 (red channel) solution. As expected, we observed cross-correlation between the fluorescent signals in the red and green channels upon the addition of holoenzyme (Kim et al., 2005) (Figure 4.8), indicating the detection of dual-labeled complexes. Further, the abundance of the dual-labeled complexes, likely representing the 6S RNP, increased as more eGFP-HE was added to the 6S<sup>M</sup>:TO3 sample (Figure 4.8, see methods for further details). These results demonstrate the compatibility of RNA Mango for enabling two-photon FCCS single molecule characterization of RNP formation.

**Figure 4.8** Fluorescence cross correlation spectroscopy (FCCS) analysis of 6S<sup>M</sup>:TO3-Dtb: eGFP-HE complex

(A) Fluorescence auto and cross correlation spectroscopy of 6S<sup>M</sup>:TO3-Dtb (red) in the presence of 25.4 ± 0.3 nM free eGFP-HE (green), 1% 6S<sup>M</sup>:eGFP-HE complex (yellow - cross correlation amplitude). (B) 6S<sup>M</sup>:TO3-Dtb (red) in the presence of 127.5 ± 0.9 nM (right) free eGFP-HE (green), 20% 6S<sup>M</sup>:eGFP-HE complex (yellow - cross correlation amplitude). Residuals are shown below each graph.

4.7. CONCLUSION

In summary, we have developed a multi-functional Mango tag for purification and biochemical analysis of cellular RNA and RNPs. The applications of the Mango toolset demonstrated here are made possible due to the unique combination of high affinity between TO derivatives and RNA Mango, the large fluorescence enhancement in TO that occurs upon binding to the RNA, and the small size of the aptamer. It is likely that many biological RNAs in addition to the U1 snRNA and the 6S RNA contain modifiable
stem-loop structures or will permit insertion of new, Mango-containing stem-loops. Therefore, we expect this Mango toolset to be useful for diverse studies of a broad range of prokaryotic and eukaryotic RNP complexes. As comprehensive libraries of eGFP tagged proteins already exist for a range of organisms (Buszczak et al., 2007; Huh et al., 2003; Kitagawa et al., 2006), the Mango system should allow the rapid and systematic study of RNPs by multi-wavelength fluorescence in the future.

4.8. MATERIAL AND METHODS

4.8.1. Synthesis of Desthiobiotin Derivatives of Thiazole Orange Acetate

TO1 PEG$_4$ desthiobiotin (TO1-Dtb) and TO3 PEG$_4$ desthiobiotin (TO3-Dtb) were synthesized using TO1-acetate and TO3-acetate, respectively, together with EZ link amine PEG$_4$ desthiobiotin (Thermo Fisher) as precursors following an established protocol (Dolgosheina et al., 2014). Product identity was confirmed by electrospray ionization – mass spectrometry in the positive mode. Expected mass (in Da) for TO1-Dtb: C$_{40}$H$_{55}$N$_6$O$_7$S$^+$: 763.3847, obtained mass: 763.3855 and expected mass for TO3-Dtb: C$_{42}$H$_{57}$N$_6$O$_7$S$^+$: 789.4003, obtained mass: 789.3999.

4.8.2. Determination of the Binding Affinity of RNA Mango for TO1-Dtb and TO3-Dtb ligands

The indicated fluorophore was held constant at 10 nM in 10 mM sodium phosphate pH 7.2, 140 mM KCl, 1 mM MgCl$_2$ and 0.05% (v/v) Tween-20 and titrated by addition of RNA Mango. Fluorescence (in Relative Fluorescence Units - RFU) was measured using a quartz cuvette with 1 mm path length (Starna cells) in a Varian Cary Eclipse Spectrofluorometer. Binding affinities were obtained as described (Dolgosheina et al., 2014) using the following equation

$$F([\text{apt}]) = \frac{F(\text{K_D}[\text{apt}]+[\text{fluorophore}])}{\sqrt{([\text{apt}]-[\text{fluorophore}])^2 + \text{K_D}[\text{K_D}+2[\text{apt}]+2[\text{fluorophore}]])}}$$
where [apt] and [fluorophore] are the RNA Mango and the initial dye concentrations, respectively, and $F'$ is the maximum fluorescence per mole of RNA:Dye complex. Data from three independent titrations were plotted and fitted using Prism 5.0b software.

4.8.3. **Construction of p6SRDM T7 plasmid**

The 6SRDM plasmid was created as previously described (Dolgosheina et al., 2014) but with point mutations corresponding to the R9-33 construct (Oviedo Ovando et al., 2014). The 6SRDM is flanked by a T7 promoter and lac operator at its 5' end and an intrinsic terminator at its 3' end and was cloned into the pEcoli-Cterm-6xHN (Clontech) plasmid between the SgrAI and ClaI sites.

4.8.4. **Expression of the 6SRDM RNA in E. coli and Extract Preparation**

The 6SRDM RNA was expressed in *E. coli* BL21 (DE3) cells transformed with the p6SRDM T7 plasmid. *E. coli* cells transformed with p6SRDM T7 were grown overnight at 37 °C in a shaking incubator in Luria-Bertani (LB) media containing 100 µg/ml ampicillin. A portion of the overnight culture (10 ml) was then used to inoculate 1 l of LB media containing 100 µg/ml ampicillin in a 4 l Erlenmeyer flask. The cells were then grown at 37 °C in a shaking incubator at 250 rpm until the OD$_{600}$ reached ~1, at which point IPTG was added to a final concentration of 50 µM to induce 6SRDM RNA transcription. After 40 min of induction, cells were harvested by centrifugation at 3800xg for 10 min at 4 °C and resuspended in 20 ml native extract buffer (20 mM Tris pH 8, 150 mM KCl, 1 mM MgCl$_2$ and 1 mM DTT). The resuspended cells were then lysed by two passages through a French press pre-chilled to 4 °C and at 1000 psi. The lysate was centrifuged at 12,000xg for 10 min at 4 °C to remove cell debris to yield native extract (N.E.). The extract was then frozen in liquid N$_2$ and stored at -80 °C.

4.8.5. **Preparation of TO1-Dtb derivatized Streptavidin Agarose Affinity Resin**

In all experiments, streptavidin agarose beads (Invitrogen) were used for batch purification in either 1.7 ml or 5 ml tubes, as required, and solutions containing the beads
were mixed using a rotator. The beads (400 µl) were washed twice with 0.1 M NaOH and 0.05 M NaCl (Buffer A) as per the manufacturer’s protocol and three times with 15 mM HEPES pH 7.5, 90 mM KCl (Buffer B). TO1-Dtb was added (22.5 nmole) in 400 µl Buffer B and incubated with the washed beads for 15 min at RT. The beads were then washed once with 400 µl Buffer B, to remove any unbound dye.

4.8.6. Purification of 6S RNA<sup>RDM</sup> RNP Complexes

For the purification of 6S<sup>RDM</sup> associated RNPs, 3.8 ml of E. coli native extract, was made up to 5 ml final volume by the addition of 1 ml of 5X Buffer B and supplemented with final concentrations of Heparin at 75 µg/ml and DTT at 0.75 mM. This solution was mixed with the TO1-Dtb derivatized agarose beads for 15 min while incubating at RT. The beads were washed twice with WB (15 mM HEPES pH 7.5, 90 mM KCl, 75 µg/ml Heparin and 0.75 mM DTT) at RT for 2 min each prior to elution with 400 µl of WB containing 20 mM free biotin (Sigma-Aldrich) at 37 °C for 20 min (called Biotin Eluate).

4.8.7. 6S<sup>RDM</sup> RNP Complex fractionation by Size Exclusion Chromatography (SEC)

The biotin eluate (380 µl) was loaded onto a manually packed Superdex 200 Prep Grade (GE healthcare) column (30 x 1 cm) and eluted with Buffer B using a flow rate of 0.1 ml/min. Fractions (500 µl) were collected and TO1-Dtb fluorescence was measured using a Spectramax M5 fluorescent plate reader (Ex/Em 495/535 nm, PMT at Medium) reading from the bottom of a 96-well plate (Greiner Bio-one).

4.8.8. Mass Spectrometry Analysis of 6S<sup>RDM</sup> RNP Complexes and Commercial RNAP

Samples were analyzed at the University of British Columbia Proteomics Core Facility. LC-MS/MS analysis used a quadrupole – time of flight mass spectrometer (Impact II; Bruker Daltonics) coupled to an Easy nano LC 1000 HPLC precolumn (ThermoFisher Scientific). Analysis of Mass Spectrometry Data was performed using
MaxQuant 1.5.1.0. The peptide search was performed against a database comprised of the protein sequences from *E. coli* K12. Commercial holoenzyme was from Epicentre.

### 4.8.9. Yeast Strain and Extract Preparation

The yeast U1\(^M\) strain was prepared from the BJ2168 derivative yAAH0441 [Mat a, prc1-407, prb1-1122, pep4-3, leu2, trp1, ura3-52, his3::loxP gal2 snr19::loxP (pMKU1-7 URA CEN)], a kind gift of Dr. Magda Konarska. The U1\(^M\) gene along with 500 bp of upstream and 723 bp of downstream genomic DNA sequence was constructed by Genewiz, sub cloned into the BamHI and NotI sites of the pRS413 plasmid (HIS3 CEN6), and sequenced. The U1\(^M\) plasmid was then transformed into yAAH0441. Transformants were selected on dropout media, and single colonies were streaked onto medium containing 5-fluoroorotic acid (5-FOA) to select for loss of the WT U1 URA3-marked plasmid. Presence of the U1-Mango gene and absence of the WT allele in the resulting strain (yAAH1204) was confirmed by using PCR to amplify the U1 gene and DNA sequencing (Supplementary Figure S5). A double-tagged strain containing both U1\(^M\) and a SNAP-tagged Snp1 protein (yAAH1362) was created by homologous recombination to insert the SNAP\(_1\) gene and a downstream hygromycin selection marker as previously described (Hoskins et al., 2011). Transformants were confirmed by PCR and by labeling of Snp1-SNAP\(_1\) with fluorescent SNAP ligands. Yeast whole cell extracts were prepared from strains yAAH1204 and yAAH1362 as previously described using the liquid N\(_2\) method and a ball mill (Hoskins et al., 2011).

### 4.8.10. SNAP labeling of yeast whole cell extract

Yeast whole cell extracts (500 µl) were first fluorescently tagged with the SNAP-Surface® 549 tag (NEB) by incubating the extracts with 1 U/µl Murine RNase inhibitor (NEB), Protease Inhibitors (cOmplete mini at the recommended concentration, EDTA Free, Sigma), 5 mM DTT together with 5 µM SNAP-Surface® 549 tag in Buffer B at RT for 30 min.
4.8.11. Purification of U1\textsuperscript{M}-Associated snRNPs

Yeast whole cell extracts (either with or without SNAP label) were added to 200 µl of TO1-Dtb saturated agarose beads and incubated at 4 °C for 1 hour in a rotator. The beads were washed twice each with 5 ml Buffer B for 15 min at 4 °C. The beads were washed once more with buffer B at 30 °C for 5 min. The snRNP was then eluted by the addition of 200 µl Buffer B supplemented with 20 mM biotin at 37 °C for 30 min. For visualizing the fluorescent labeled SNAP\textsubscript{r} tag, 100 µl of the biotin eluate was concentrated in a speed vac to 10 µl and loaded directly onto a 10% SDS PAGE gel. The gel was imaged using a Typhoon scanner using 532 nm excitation and 580 nm emission filter settings and with PMT kept at 1000 V. The remaining 100 µl of the biotin eluate was phenol extracted once, chloroform extracted twice and ethanol precipitated using 2 µg glycogen. The pellet was resuspended in 5.1 µl of water and analyzed by RT-primer extension assay.

4.8.12. RNA analysis by RT-Primer Extension and Solution Hybridization

Either 1 µl of the precipitated RNA or 1 µg of total RNA isolated from WT yeast extract was added to the corresponding 5' \textsuperscript{32}P-labeled primers and the RT extension reaction was then carried out using the Maxima Reverse Transcriptase (Thermo Fisher Scientific) at 45 °C for 1 h. Reactions were then loaded onto a 8% (19:1 acrylamide:bis-acrylamide) denaturing polyacrylamide gel and run at 500 V at RT. For solution hybridization, 5' \textsuperscript{32}P-labeled primers (Supplementary Table 4) were added directly to the precipitated RNAs and incubated for 15 min at RT prior to loading onto a 5% (37.5:1 acrylamide:bis-acrylamide) native polyacrylamide gel and electrophoresis at 4 °C in 1X TBE at 350 V.

4.8.13. Partial Purification of eGFP-HE

\textit{E. coli} eGFP-HE was partially purified from \textit{E. coli} strain RL1314 which was the kind gift from Dr. Robert Landick, UW Madison (Bratton et al., 2011). The partial purification was adapted from the protocol of Burgess and Jendrisak (Burgess and Jendrisak, 1975). Briefly, 1 l of RL1314 \textit{E. coli} cells at OD ~ 1 were harvested and
resuspended in 8 ml of 50 mM Tris pH 8, 5% (v/v) glycerol, 2 mM EDTA, 0.1 mM DTT and 233 mM NaCl along with protease inhibitors (cOmplete mini at the recommended concentration, EDTA Free, Sigma) and lysed through a French press (pre-chilled to 4 °C) at 1000 psi. The crude lysate was diluted with 8 ml of TGED Buffer (10 mM Tris pH 8, 5% v/v glycerol, 0.1 mM EDTA, and 0.1 mM DTT) supplemented with 0.2 M NaCl and was centrifuged at 12,000xg for 45 min to remove cell debris. To this cleared lysate, 10% v/v Polymin P at pH 8 was added to 0.175 ml/100 ml of the cleared lysate and stirred for 5 min, before pelleting at 4300xg for 15 min. The resulting pellet was washed once for 10 min with 10 ml TGED + 0.5 M NaCl and pelleted at 4300xg for 15 min. The washed pellet was once again resuspended in 10 ml TGED + 1 M NaCl and stirred for 10 min and the insoluble proteins were removed by spinning at 4300xg for 15 min. The resulting material was precipitated with 50% saturated ammonium sulfate (35 g/100 ml) and resuspended in 500 µl dialysis buffer (TGED + 0.1 M NaCl but with 50% w/v glycerol) and was dialyzed overnight against 1 l of dialysis buffer. All the above steps were done either at 4 °C or on ice.

4.8.14. Two Color Gel-imaging Analysis of the Mango and eGFP tagged 6S:RNAP Complex

*In vitro* synthesized 6S\(^M\) RNA (1 µM) was added to 2 µM TO3-Dtb and one-tenth of the reaction volume of the partially purified eGFP-HE in 15 mM HEPES pH 7.5, 90 mM KCl, 0.75 mM DTT and 75 µg/ml heparin. This solution was then incubated at 37 °C for 30 min to form bound complex. To induce pRNA synthesis dependent 6S\(^M\) release, 500 µM NTPs and 5 mM MgCl\(_2\) were then added and incubation continued at 37 °C. Samples were loaded into a 5% (37.5:1 acrylamide: bis-acrylamide) native polyacrylamide gel using 15 mM HEPES pH 7.5 and 90 mM KCl as both the gel and running buffer. Electrophoresis was carried out for 90 min at 4 °C at 100 V. The gel was then imaged using a Typhoon Trio+ Variable Mode Imager (GE LifeSciences) using a 488 nm laser and 526 BP filter for eGFP-HE visualization and a 644 nm laser and 670 BP filter for imaging TO3-Dtb:6S\(^M\) complex. PMT settings were kept at 1000 V and 600 V for eGFP and Mango/TO3-Dtb, respectively. Image data were analyzed and a composite image was made using ImageQuant TV v8.1.0.0 (GE LifeSciences).
4.8.15. Two-photon Spectrum and Cross-Correlation Experiments

Sample Conditions

The two-photon spectrum of 6S\textsuperscript{M}:TO3-Dtb was obtained by measuring fluorescence intensity in photons/s over a range of two-photon excitation wavelengths. The solution contained 1 µM \textit{in vitro} transcribed 6S\textsuperscript{M}, 1 µM TO3-Dtb, 15 mM HEPES pH 7.5, 90 mM KCl, 0.75 mM DTT and 75 µg/ml heparin. The wavelength was first set to 800 nm, and then tuned to higher wavelengths at a constant rate up to 900 nm. During this tuning, data were acquired in the red channel at a rate of 100 Hz.

Fluorescence cross correlation spectroscopy was performed by adding increasing amounts of RNAP-eGFP extract to a fixed amount of 6S\textsuperscript{M} and TO3-Dtb. The starting mixture (128 µL) contained 1 µM \textit{in vitro} transcribed 6S\textsuperscript{M}, 1 µM TO3-Dtb and in 15 mM HEPES pH 7.5, 90 mM KCl, 0.75 mM DTT and 75 µg/ml heparin. To this was added 12.5 µl of partially purified \textit{E. coli} eGFP-HE (see above), which was pre-diluted 5 fold in 15 mM HEPES pH 7.5, 90 mM KCl, 0.75 mM DTT and 75 µg/ml heparin. The sample was mixed and cross correlation spectroscopy was performed. This was repeated for five sequential 12.5 µl additions, and two final 62.5 µl additions of the pre-diluted eGFP-HE. The molecular brightness of the 6S\textsuperscript{M}:TO3 alone was determined by correlation spectroscopy on the mixture prior to adding any eGFP-HE. The molecular brightness of eGFP-HE alone was determined by adding 12.5 µl to 128 µl water and performing fluorescence correlation spectroscopy. A mixture containing 6S\textsuperscript{M} in 15 mM HEPES pH 7.5, 90 mM KCl, 0.75 mM DTT and 75 µg/ml heparin, without TO3-Dtb, was measured and showed no fluorescence above background.

Fluorescence cross correlation spectroscopy was performed on an ISS Alba microscope controlled using ISS Vista Vision 4.0 Software. Two-photon laser excitation was achieved using a Titanium Sapphire laser (Spectra Physics, Tsunami) mode locked at a median wavelength of 840 nm and a laser power of 45 mW unless otherwise specified. Light was collected using two Perkin Elmer SPCM-ARQ Avalanche Photodiodes. Laser light was filtered using a 780 short pass filter (Semrock) and split using a 561 long pass dichroic filter (Semrock) before directing the fluorescence emission into the detectors. The excitation lasers was focused onto the sample using a Nikon 60x water immersion objective with a 1.2 Numerical Aperture. Data were acquired
at 100 MHz for 100 s using a 16-bit ISS FCS PCI card to transfer data between the photodiodes and computer. Each measurement was repeated three times, correlated and fit independently using Craig Markwardt’s MPFIT library in IDL 8.3 (Exellis) using custom written software. Details on data analysis can be found in Supplementary Methods.

4.9. ACKNOWLEDGEMENTS

We would like to thank Dr. Sen, Dr. Audas and the members of the Unrau Lab (Sunny Jeng, Razvan Cojocaru and Amir Abdolahzadeh) for the critical reading of the manuscript. We would also like to thank Robert Landick for providing us with eGFP tagged HE containing *E. coli* strain.
References


Pereira Nondenaturing Purification of Co-Transcriptionally Folded RNA Avoids Common Folding Heterogeneity.


## Appendix A

### Mass spectrometry analysis of Biotin eluate

<table>
<thead>
<tr>
<th>Fasta headers</th>
<th>Number of proteins</th>
<th>Peptides</th>
<th>Unqiue peptides</th>
<th>MOL. weight [kDa]</th>
<th>Sequence length</th>
<th>Sequence coverage [%]</th>
<th>Intensity</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;sp</td>
<td>P0A8V2</td>
<td>RP0B_ECOLI DNA-directed RNA polymerase subunit beta OS=Escherichia coli (strain K12) GN=rp0B PE=1 SV=1</td>
<td>1</td>
<td>77</td>
<td>77</td>
<td>150.63</td>
<td>1342</td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P0A877</td>
<td>RP0C_ECOLI DNA-directed RNA polymerase subunit beta OS=Escherichia coli (strain K12) GN=rp0C PE=1 SV=1</td>
<td>1</td>
<td>83</td>
<td>83</td>
<td>155.16</td>
<td>1407</td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P0D579</td>
<td>RP0D_ECOLI RNA polymerase sigma factor RpoD OS=Escherichia coli (strain K12) GN=rpoD PE=1 SV=2</td>
<td>1</td>
<td>23</td>
<td>22</td>
<td>70.263</td>
<td>613</td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P0A724</td>
<td>RP0A_ECOLI DNA-directed RNA polymerase subunit alpha OS=Escherichia coli (strain K12) GN=rpoA PE=1 SV=1</td>
<td>1</td>
<td>16</td>
<td>16</td>
<td>36.51</td>
<td>329</td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P0A656</td>
<td>CH06_ECOLI 60 KDa chaperonin OS=Escherichia coli (strain K12) GN=groL PE=1 SV=2</td>
<td>1</td>
<td>19</td>
<td>19</td>
<td>57.328</td>
<td>548</td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P0AEU7</td>
<td>SKP_ECOLI Chaperone protein Skp OS=Escherichia coli (strain K12) GN=skp PE=1 SV=1</td>
<td>1</td>
<td>7</td>
<td>7</td>
<td>17.688</td>
<td>161</td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P0CE48</td>
<td>EFTU2_ECOLI Elongation factor Tu 2 OS=Escherichia coli (strain K12) GN=tufB PE=1 SV=1</td>
<td>2</td>
<td>18</td>
<td>18</td>
<td>43.313</td>
<td>394</td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P14182</td>
<td>ACCC_ECOLI Biotin carboxylase OS=Escherichia coli (strain K12) GN=acc PE=1 SV=1</td>
<td>1</td>
<td>15</td>
<td>15</td>
<td>49.32</td>
<td>449</td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P00579</td>
<td>RS15_ECOLI 30S ribosomal protein S15 OS=Escherichia coli (strain K12) GN=rpsO PE=1 SV=2</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>10.269</td>
<td>89</td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P0A6Y8</td>
<td>DNAK_ECOLI Chaperone protein DnaK OS=Escherichia coli (strain K12) GN=dnaK PE=1 SV=2</td>
<td>1</td>
<td>6</td>
<td>6</td>
<td>96.031</td>
<td>876</td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P0A6A8</td>
<td>RS4_ECOLI 30S ribosomal protein S4 OS=Escherichia coli (strain K12) GN=rpsD PE=1 SV=1</td>
<td>1</td>
<td>18</td>
<td>18</td>
<td>61.157</td>
<td>557</td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P0A8B2</td>
<td>G3P1_ECOLI Glyceraldehyde-3-phosphate dehydrogenase A OS=Escherichia coli (strain K12) GN=gapA PE=1 SV=1</td>
<td>1</td>
<td>12</td>
<td>11</td>
<td>35.532</td>
<td>331</td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P0A8M8</td>
<td>EF1_G_ECOLI Elongation factor G OS=Escherichia coli (strain K12) GN=ef1A PE=1 SV=2</td>
<td>1</td>
<td>15</td>
<td>15</td>
<td>77.58</td>
<td>704</td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P0A7V8</td>
<td>RS4_ECOLI 30S ribosomal protein S4 OS=Escherichia coli (strain K12) GN=rpsD PE=1 SV=2</td>
<td>1</td>
<td>9</td>
<td>9</td>
<td>29.86</td>
<td>273</td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P0A733</td>
<td>RLA_ECOLI 30S ribosomal protein L4 OS=Escherichia coli (strain K12) GN=rla PE=1 SV=1</td>
<td>1</td>
<td>7</td>
<td>7</td>
<td>22.086</td>
<td>201</td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P0A7T0</td>
<td>RL1_ECOLI 30S ribosomal protein L1 OS=Escherichia coli (strain K12) GN=rla PE=1 SV=2</td>
<td>1</td>
<td>6</td>
<td>6</td>
<td>24.729</td>
<td>234</td>
</tr>
<tr>
<td>Accession</td>
<td>Description</td>
<td>E. coli (strain K12)</td>
<td>ORF Length (aa)</td>
<td>Molecular Weight (kDa)</td>
<td>Peptide Mass (Da)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>-------------</td>
<td>---------------------</td>
<td>-----------------</td>
<td>------------------------</td>
<td>------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>sp</td>
<td>P60438</td>
<td>RL3_ECOLI 50S ribosomal protein L3 O5=Escherichia coli (strain K12) GN=rplC PE=1 SV=1</td>
<td>1 5 5 22.243</td>
<td>209</td>
<td>24.4</td>
<td>8326900</td>
<td></td>
</tr>
<tr>
<td>sp</td>
<td>P0A7V0</td>
<td>RS2_ECOLI 30S ribosomal protein S2 O5=Escherichia coli (strain K12) GN=rpsB PE=1 SV=2</td>
<td>1 5 5 26.743</td>
<td>241</td>
<td>21.6</td>
<td>8014400</td>
<td></td>
</tr>
<tr>
<td>sp</td>
<td>P0A7V3</td>
<td>RS3_ECOLI 30S ribosomal protein S3 O5=Escherichia coli (strain K12) GN=rpsC PE=1 SV=2</td>
<td>1 10 10 25.983</td>
<td>231</td>
<td>51.9</td>
<td>7565200</td>
<td></td>
</tr>
<tr>
<td>sp</td>
<td>P0A7W1</td>
<td>RS5_ECOLI 30S ribosomal protein S5 O5=Escherichia coli (strain K12) GN=rpsE PE=1 SV=2</td>
<td>1 6 6 17.603</td>
<td>167</td>
<td>44.9</td>
<td>7001900</td>
<td></td>
</tr>
<tr>
<td>sp</td>
<td>P0A6P1</td>
<td>EFTS_ECOLI Elongation factor Ts O5=Escherichia coli (strain K12) GN=tsx PE=1 SV=2</td>
<td>1 8 8 26.743</td>
<td>241</td>
<td>21.6</td>
<td>8014400</td>
<td></td>
</tr>
<tr>
<td>sp</td>
<td>P0A7X3</td>
<td>RS9_ECOLI 30S ribosomal protein S9 O5=Escherichia coli (strain K12) GN=rpsI PE=1 SV=2</td>
<td>1 3 3 14.856</td>
<td>130</td>
<td>23.8</td>
<td>6789900</td>
<td></td>
</tr>
<tr>
<td>sp</td>
<td>P0A7J3</td>
<td>RL10_ECOLI 50S ribosomal protein L10 O5=Escherichia coli (strain K12) GN=rplJ PE=1 SV=2</td>
<td>1 6 6 17.711</td>
<td>165</td>
<td>47.3</td>
<td>6762400</td>
<td></td>
</tr>
<tr>
<td>sp</td>
<td>P0A7R1</td>
<td>RL9_ECOLI 50S ribosomal protein L9 O5=Escherichia coli (strain K12) GN=rplI PE=1 SV=1</td>
<td>1 12 12 48.192</td>
<td>432</td>
<td>31</td>
<td>6912800</td>
<td></td>
</tr>
<tr>
<td>sp</td>
<td>P0A7K6</td>
<td>RL19_ECOLI 50S ribosomal protein L19 O5=Escherichia coli (strain K12) GN=rplM PE=1 SV=2</td>
<td>1 7 7 15.769</td>
<td>149</td>
<td>48.3</td>
<td>6676800</td>
<td></td>
</tr>
<tr>
<td>sp</td>
<td>P0A7R5</td>
<td>RS10_ECOLI 30S ribosomal protein S10 O5=Escherichia coli (strain K12) GN=rpsJ PE=1 SV=1</td>
<td>1 4 4 11.735</td>
<td>103</td>
<td>40.8</td>
<td>4934500</td>
<td></td>
</tr>
<tr>
<td>sp</td>
<td>P02359</td>
<td>RS7_ECOLI 30S ribosomal protein S7 O5=Escherichia coli (strain K12) GN=rpsG PE=1 SV=3</td>
<td>1 6 6 20.019</td>
<td>179</td>
<td>43.6</td>
<td>5449400</td>
<td></td>
</tr>
<tr>
<td>sp</td>
<td>P33012</td>
<td>SBMC_ECOLI DNA gyrase inhibitor O5=Escherichia coli (strain K12) GN=sbmC PE=1 SV=2</td>
<td>1 3 3 11.806</td>
<td>105</td>
<td>33</td>
<td>5302800</td>
<td></td>
</tr>
<tr>
<td>sp</td>
<td>P0A7X6</td>
<td>RL19_ECOLI 50S ribosomal protein L19 O5=Escherichia coli (strain K12) GN=rplP PE=1 SV=2</td>
<td>1 4 4 13.133</td>
<td>115</td>
<td>33</td>
<td>5098300</td>
<td></td>
</tr>
<tr>
<td>sp</td>
<td>P0A810</td>
<td>RL13_ECOLI 50S ribosomal protein L13 O5=Escherichia coli (strain K12) GN=rpmPE=1 SV=1</td>
<td>1 4 4 16.018</td>
<td>142</td>
<td>26.8</td>
<td>4828900</td>
<td></td>
</tr>
<tr>
<td>sp</td>
<td>P02413</td>
<td>RL15_ECOLI 50S ribosomal protein L15 O5=Escherichia coli (strain K12) GN=rplO PE=1 SV=1</td>
<td>1 4 4 14.98</td>
<td>144</td>
<td>24.3</td>
<td>5049300</td>
<td></td>
</tr>
<tr>
<td>sp</td>
<td>P0A461</td>
<td>EFTS_ECOLI Elongation factor Ts O5=Escherichia coli (strain K12) GN=tsx PE=1 SV=2</td>
<td>1 8 8 26.743</td>
<td>241</td>
<td>21.6</td>
<td>8014400</td>
<td></td>
</tr>
<tr>
<td>sp</td>
<td>P02413</td>
<td>RL15_ECOLI 50S ribosomal protein L15 O5=Escherichia coli (strain K12) GN=rplO PE=1 SV=1</td>
<td>1 4 4 14.98</td>
<td>144</td>
<td>24.3</td>
<td>5049300</td>
<td></td>
</tr>
<tr>
<td>sp</td>
<td>P0A850</td>
<td>TIG_ECOLI Trigger factor O5=Escherichia coli (strain K12) GN=tig PE=1 SV=1</td>
<td>1 12 12 48.192</td>
<td>432</td>
<td>31</td>
<td>6912800</td>
<td></td>
</tr>
<tr>
<td>sp</td>
<td>P0A7J3</td>
<td>RL10_ECOLI 50S ribosomal protein L10 O5=Escherichia coli (strain K12) GN=rplJ PE=1 SV=2</td>
<td>1 6 6 17.711</td>
<td>165</td>
<td>47.3</td>
<td>6762400</td>
<td></td>
</tr>
<tr>
<td>sp</td>
<td>P0A7R1</td>
<td>RL9_ECOLI 50S ribosomal protein L9 O5=Escherichia coli (strain K12) GN=rplI PE=1 SV=1</td>
<td>1 7 7 15.769</td>
<td>149</td>
<td>48.3</td>
<td>6676800</td>
<td></td>
</tr>
<tr>
<td>sp</td>
<td>P0A7X3</td>
<td>RS9_ECOLI 30S ribosomal protein S9 O5=Escherichia coli (strain K12) GN=rpsG PE=1 SV=3</td>
<td>1 6 6 20.019</td>
<td>179</td>
<td>43.6</td>
<td>5449400</td>
<td></td>
</tr>
<tr>
<td>sp</td>
<td>P0A7K6</td>
<td>RL19_ECOLI 50S ribosomal protein L19 O5=Escherichia coli (strain K12) GN=rplM PE=1 SV=2</td>
<td>1 4 4 13.133</td>
<td>115</td>
<td>33</td>
<td>5098300</td>
<td></td>
</tr>
<tr>
<td>sp</td>
<td>P0A7R5</td>
<td>RS10_ECOLI 30S ribosomal protein S10 O5=Escherichia coli (strain K12) GN=rpsJ PE=1 SV=1</td>
<td>1 4 4 11.735</td>
<td>103</td>
<td>40.8</td>
<td>4934500</td>
<td></td>
</tr>
<tr>
<td>sp</td>
<td>P0A7X6</td>
<td>RL19_ECOLI 50S ribosomal protein L19 O5=Escherichia coli (strain K12) GN=rplP PE=1 SV=2</td>
<td>1 4 4 13.133</td>
<td>115</td>
<td>33</td>
<td>5098300</td>
<td></td>
</tr>
<tr>
<td>sp</td>
<td>P0A7R1</td>
<td>RL9_ECOLI 50S ribosomal protein L9 O5=Escherichia coli (strain K12) GN=rplI PE=1 SV=1</td>
<td>1 7 7 15.769</td>
<td>149</td>
<td>48.3</td>
<td>6676800</td>
<td></td>
</tr>
<tr>
<td>sp</td>
<td>P0A7X3</td>
<td>RS9_ECOLI 30S ribosomal protein S9 O5=Escherichia coli (strain K12) GN=rpsG PE=1 SV=3</td>
<td>1 6 6 20.019</td>
<td>179</td>
<td>43.6</td>
<td>5449400</td>
<td></td>
</tr>
<tr>
<td>sp</td>
<td>P0A7K6</td>
<td>RL19_ECOLI 50S ribosomal protein L19 O5=Escherichia coli (strain K12) GN=rplM PE=1 SV=1</td>
<td>1 4 4 16.018</td>
<td>142</td>
<td>26.8</td>
<td>4828900</td>
<td></td>
</tr>
<tr>
<td>sp</td>
<td>P0A7X6</td>
<td>RL19_ECOLI 50S ribosomal protein L19 O5=Escherichia coli (strain K12) GN=rplM PE=1 SV=1</td>
<td>1 4 4 16.018</td>
<td>142</td>
<td>26.8</td>
<td>4828900</td>
<td></td>
</tr>
<tr>
<td>Accession</td>
<td>Description</td>
<td>Organism</td>
<td>Length</td>
<td>Mw</td>
<td>Id</td>
<td>Pe</td>
<td>SV</td>
</tr>
<tr>
<td>-----------</td>
<td>-------------</td>
<td>----------</td>
<td>--------</td>
<td>-----</td>
<td>----</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>sp</td>
<td>P0A77</td>
<td>RL11_ECOLI 50S ribosomal protein L11 OS=Escherichia coli (strain K12) GN=rplK PE=1 SV=2</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>14.875</td>
<td>142</td>
</tr>
<tr>
<td>sp</td>
<td>P0A836</td>
<td>SUCC_ECOLI Succinyl-CoA ligase [ADP-forming] subunit beta OS=Escherichia coli (strain K12) GN=sucC PE=1 SV=2</td>
<td>1</td>
<td>10</td>
<td>10</td>
<td>41.392</td>
<td>388</td>
</tr>
<tr>
<td>sp</td>
<td>P0A759</td>
<td>R513_ECOLI 30S ribosomal protein S13 OS=Escherichia coli (strain K12) GN=rpsM PE=1 SV=2</td>
<td>1</td>
<td>6</td>
<td>6</td>
<td>13.099</td>
<td>118</td>
</tr>
<tr>
<td>sp</td>
<td>P60624</td>
<td>RL24_ECOLI 50S ribosomal protein L24 OS=Escherichia coli (strain K12) GN=rplX PE=1 SV=2</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>11.316</td>
<td>104</td>
</tr>
<tr>
<td>sp</td>
<td>P0A7S9</td>
<td>RS13_ECOLI 30S ribosomal protein S13 OS=Escherichia coli (strain K12) GN=rpsM PE=1 SV=2</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>14.126</td>
<td>130</td>
</tr>
<tr>
<td>sp</td>
<td>P0A6Z3</td>
<td>HTPG_ECOLI Chaperone protein HtpG OS=Escherichia coli (strain K12) GN=htpG PE=1 SV=1</td>
<td>1</td>
<td>7</td>
<td>7</td>
<td>32.337</td>
<td>312</td>
</tr>
<tr>
<td>sp</td>
<td>P0A77</td>
<td>RL11_ECOLI 50S ribosomal protein L11 OS=Escherichia coli (strain K12) GN=rplK PE=1 SV=2</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>14.875</td>
<td>142</td>
</tr>
<tr>
<td>sp</td>
<td>P61889</td>
<td>MDH_ECOLI Malate dehydrogenase OS=Escherichia coli (strain K12) GN=mdh PE=1 SV=2</td>
<td>1</td>
<td>7</td>
<td>7</td>
<td>32.337</td>
<td>312</td>
</tr>
<tr>
<td>sp</td>
<td>P0A759</td>
<td>R513_ECOLI 30S ribosomal protein S13 OS=Escherichia coli (strain K12) GN=rpsM PE=1 SV=2</td>
<td>1</td>
<td>6</td>
<td>6</td>
<td>13.099</td>
<td>118</td>
</tr>
<tr>
<td>sp</td>
<td>P0A7S9</td>
<td>RS13_ECOLI 30S ribosomal protein S13 OS=Escherichia coli (strain K12) GN=rpsM PE=1 SV=2</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>14.126</td>
<td>130</td>
</tr>
<tr>
<td>sp</td>
<td>P0A7R9</td>
<td>RS11_ECOLI 30S ribosomal protein S11 OS=Escherichia coli (strain K12) GN=rpsK PE=1 SV=2</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>13.845</td>
<td>129</td>
</tr>
<tr>
<td>sp</td>
<td>P0A77</td>
<td>RL11_ECOLI 50S ribosomal protein L11 OS=Escherichia coli (strain K12) GN=rplX PE=1 SV=2</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>20.301</td>
<td>179</td>
</tr>
<tr>
<td>sp</td>
<td>P0A6Z3</td>
<td>HTPG_ECOLI Chaperone protein HtpG OS=Escherichia coli (strain K12) GN=htpG PE=1 SV=1</td>
<td>1</td>
<td>7</td>
<td>7</td>
<td>32.337</td>
<td>312</td>
</tr>
<tr>
<td>sp</td>
<td>P0A77</td>
<td>RL11_ECOLI 50S ribosomal protein L11 OS=Escherichia coli (strain K12) GN=rplX PE=1 SV=2</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>20.301</td>
<td>179</td>
</tr>
<tr>
<td>sp</td>
<td>P0A77</td>
<td>RL11_ECOLI 50S ribosomal protein L11 OS=Escherichia coli (strain K12) GN=rplX PE=1 SV=2</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>20.301</td>
<td>179</td>
</tr>
<tr>
<td>sp</td>
<td>P0A77</td>
<td>RL11_ECOLI 50S ribosomal protein L11 OS=Escherichia coli (strain K12) GN=rplX PE=1 SV=2</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>20.301</td>
<td>179</td>
</tr>
<tr>
<td>sp</td>
<td>P0A77</td>
<td>RL11_ECOLI 50S ribosomal protein L11 OS=Escherichia coli (strain K12) GN=rplX PE=1 SV=2</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>20.301</td>
<td>179</td>
</tr>
<tr>
<td>sp</td>
<td>P0A77</td>
<td>RL11_ECOLI 50S ribosomal protein L11 OS=Escherichia coli (strain K12) GN=rplX PE=1 SV=2</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>20.301</td>
<td>179</td>
</tr>
<tr>
<td>sp</td>
<td>P0A77</td>
<td>RL11_ECOLI 50S ribosomal protein L11 OS=Escherichia coli (strain K12) GN=rplX PE=1 SV=2</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>20.301</td>
<td>179</td>
</tr>
<tr>
<td>sp</td>
<td>P0A77</td>
<td>RL11_ECOLI 50S ribosomal protein L11 OS=Escherichia coli (strain K12) GN=rplX PE=1 SV=2</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>20.301</td>
<td>179</td>
</tr>
<tr>
<td>sp</td>
<td>P0A77</td>
<td>RL11_ECOLI 50S ribosomal protein L11 OS=Escherichia coli (strain K12) GN=rplX PE=1 SV=2</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>20.301</td>
<td>179</td>
</tr>
<tr>
<td>sp</td>
<td>P0A77</td>
<td>RL11_ECOLI 50S ribosomal protein L11 OS=Escherichia coli (strain K12) GN=rplX PE=1 SV=2</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>20.301</td>
<td>179</td>
</tr>
<tr>
<td>sp</td>
<td>P0A77</td>
<td>RL11_ECOLI 50S ribosomal protein L11 OS=Escherichia coli (strain K12) GN=rplX PE=1 SV=2</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>20.301</td>
<td>179</td>
</tr>
<tr>
<td>sp</td>
<td>P0A77</td>
<td>RL11_ECOLI 50S ribosomal protein L11 OS=Escherichia coli (strain K12) GN=rplX PE=1 SV=2</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>20.301</td>
<td>179</td>
</tr>
<tr>
<td>sp</td>
<td>P0A77</td>
<td>RL11_ECOLI 50S ribosomal protein L11 OS=Escherichia coli (strain K12) GN=rplX PE=1 SV=2</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>20.301</td>
<td>179</td>
</tr>
<tr>
<td>sp</td>
<td>P0A77</td>
<td>RL11_ECOLI 50S ribosomal protein L11 OS=Escherichia coli (strain K12) GN=rplX PE=1 SV=2</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>20.301</td>
<td>179</td>
</tr>
<tr>
<td>sp</td>
<td>P0A77</td>
<td>RL11_ECOLI 50S ribosomal protein L11 OS=Escherichia coli (strain K12) GN=rplX PE=1 SV=2</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>20.301</td>
<td>179</td>
</tr>
<tr>
<td>sp</td>
<td>P0A77</td>
<td>RL11_ECOLI 50S ribosomal protein L11 OS=Escherichia coli (strain K12) GN=rplX PE=1 SV=2</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>20.301</td>
<td>179</td>
</tr>
<tr>
<td>Accession</td>
<td>Description</td>
<td>Source</td>
<td>Genbank ID</td>
<td>Peptide PepSLV</td>
<td>E-value</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>-------------</td>
<td>--------</td>
<td>------------</td>
<td>----------------</td>
<td>---------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P0A9P0</td>
<td>DLDH_ECOLI Dihydrolipoyl dehydrogenase OS=Escherichia coli (strain K12) GN=lpdA PE=1 SLV=2</td>
<td>1 10</td>
<td>10 50.688</td>
<td>474 25.7</td>
<td>2960700</td>
<td></td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P0ADY3</td>
<td>RL14_ECOLI S05 ribosomal protein L14 OS=Escherichia coli (strain K12) GN=rpn PE=1 SLV=1</td>
<td>1 3</td>
<td>3 13.541</td>
<td>123 27.6</td>
<td>2721400</td>
<td></td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P0A70S</td>
<td>IF2_ECOLI Translation initiation factor IF-2 OS=Escherichia coli (strain K12) GN=infb PE=1 SLV=1</td>
<td>1 8</td>
<td>8 97.349</td>
<td>890 9.8</td>
<td>2622200</td>
<td></td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P62707</td>
<td>GPMA_ECOLI 2,3W bisphosphoglycerate-dependent phosphoglycerate mutase OS=Escherichia coli (strain K12) GN=gpmA PE=1 SLV=2</td>
<td>1 10</td>
<td>10 25.7</td>
<td>296 15.2</td>
<td>2721400</td>
<td></td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P0A717</td>
<td>KPRS_ECOLI Ribose-phosphate pyrophosphokinase OS=Escherichia coli (strain K12) GN=prs PE=1 SLV=1</td>
<td>1 3</td>
<td>3 41.118</td>
<td>387 24</td>
<td>2511100</td>
<td></td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P0A705</td>
<td>IF2_ECOLI Translation initiation factor IF-2 OS=Escherichia coli (strain K12) GN=infb PE=1 SLV=1</td>
<td>1 8</td>
<td>8 97.349</td>
<td>890 9.8</td>
<td>2622200</td>
<td></td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P62707</td>
<td>GPMA_ECOLI 2,3W bisphosphoglycerate-dependent phosphoglycerate mutase OS=Escherichia coli (strain K12) GN=gpmA PE=1 SLV=2</td>
<td>1 10</td>
<td>10 25.7</td>
<td>296 15.2</td>
<td>2721400</td>
<td></td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P0A717</td>
<td>KPRS_ECOLI Ribose-phosphate pyrophosphokinase OS=Escherichia coli (strain K12) GN=prs PE=1 SLV=1</td>
<td>1 3</td>
<td>3 41.118</td>
<td>387 24</td>
<td>2511100</td>
<td></td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P0A705</td>
<td>IF2_ECOLI Translation initiation factor IF-2 OS=Escherichia coli (strain K12) GN=infb PE=1 SLV=1</td>
<td>1 8</td>
<td>8 97.349</td>
<td>890 9.8</td>
<td>2622200</td>
<td></td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P62707</td>
<td>GPMA_ECOLI 2,3W bisphosphoglycerate-dependent phosphoglycerate mutase OS=Escherichia coli (strain K12) GN=gpmA PE=1 SLV=2</td>
<td>1 10</td>
<td>10 25.7</td>
<td>296 15.2</td>
<td>2721400</td>
<td></td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P0A717</td>
<td>KPRS_ECOLI Ribose-phosphate pyrophosphokinase OS=Escherichia coli (strain K12) GN=prs PE=1 SLV=1</td>
<td>1 3</td>
<td>3 41.118</td>
<td>387 24</td>
<td>2511100</td>
<td></td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P0A705</td>
<td>IF2_ECOLI Translation initiation factor IF-2 OS=Escherichia coli (strain K12) GN=infb PE=1 SLV=1</td>
<td>1 8</td>
<td>8 97.349</td>
<td>890 9.8</td>
<td>2622200</td>
<td></td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P62707</td>
<td>GPMA_ECOLI 2,3W bisphosphoglycerate-dependent phosphoglycerate mutase OS=Escherichia coli (strain K12) GN=gpmA PE=1 SLV=2</td>
<td>1 10</td>
<td>10 25.7</td>
<td>296 15.2</td>
<td>2721400</td>
<td></td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P0A717</td>
<td>KPRS_ECOLI Ribose-phosphate pyrophosphokinase OS=Escherichia coli (strain K12) GN=prs PE=1 SLV=1</td>
<td>1 3</td>
<td>3 41.118</td>
<td>387 24</td>
<td>2511100</td>
<td></td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P0A705</td>
<td>IF2_ECOLI Translation initiation factor IF-2 OS=Escherichia coli (strain K12) GN=infb PE=1 SLV=1</td>
<td>1 8</td>
<td>8 97.349</td>
<td>890 9.8</td>
<td>2622200</td>
<td></td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P62707</td>
<td>GPMA_ECOLI 2,3W bisphosphoglycerate-dependent phosphoglycerate mutase OS=Escherichia coli (strain K12) GN=gpmA PE=1 SLV=2</td>
<td>1 10</td>
<td>10 25.7</td>
<td>296 15.2</td>
<td>2721400</td>
<td></td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P0A717</td>
<td>KPRS_ECOLI Ribose-phosphate pyrophosphokinase OS=Escherichia coli (strain K12) GN=prs PE=1 SLV=1</td>
<td>1 3</td>
<td>3 41.118</td>
<td>387 24</td>
<td>2511100</td>
<td></td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P0A705</td>
<td>IF2_ECOLI Translation initiation factor IF-2 OS=Escherichia coli (strain K12) GN=infb PE=1 SLV=1</td>
<td>1 8</td>
<td>8 97.349</td>
<td>890 9.8</td>
<td>2622200</td>
<td></td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P62707</td>
<td>GPMA_ECOLI 2,3W bisphosphoglycerate-dependent phosphoglycerate mutase OS=Escherichia coli (strain K12) GN=gpmA PE=1 SLV=2</td>
<td>1 10</td>
<td>10 25.7</td>
<td>296 15.2</td>
<td>2721400</td>
<td></td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P0A717</td>
<td>KPRS_ECOLI Ribose-phosphate pyrophosphokinase OS=Escherichia coli (strain K12) GN=prs PE=1 SLV=1</td>
<td>1 3</td>
<td>3 41.118</td>
<td>387 24</td>
<td>2511100</td>
<td></td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P0A705</td>
<td>IF2_ECOLI Translation initiation factor IF-2 OS=Escherichia coli (strain K12) GN=infb PE=1 SLV=1</td>
<td>1 8</td>
<td>8 97.349</td>
<td>890 9.8</td>
<td>2622200</td>
<td></td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P62707</td>
<td>GPMA_ECOLI 2,3W bisphosphoglycerate-dependent phosphoglycerate mutase OS=Escherichia coli (strain K12) GN=gpmA PE=1 SLV=2</td>
<td>1 10</td>
<td>10 25.7</td>
<td>296 15.2</td>
<td>2721400</td>
<td></td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P0A717</td>
<td>KPRS_ECOLI Ribose-phosphate pyrophosphokinase OS=Escherichia coli (strain K12) GN=prs PE=1 SLV=1</td>
<td>1 3</td>
<td>3 41.118</td>
<td>387 24</td>
<td>2511100</td>
<td></td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P0A705</td>
<td>IF2_ECOLI Translation initiation factor IF-2 OS=Escherichia coli (strain K12) GN=infb PE=1 SLV=1</td>
<td>1 8</td>
<td>8 97.349</td>
<td>890 9.8</td>
<td>2622200</td>
<td></td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P62707</td>
<td>GPMA_ECOLI 2,3W bisphosphoglycerate-dependent phosphoglycerate mutase OS=Escherichia coli (strain K12) GN=gpmA PE=1 SLV=2</td>
<td>1 10</td>
<td>10 25.7</td>
<td>296 15.2</td>
<td>2721400</td>
<td></td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P0A717</td>
<td>KPRS_ECOLI Ribose-phosphate pyrophosphokinase OS=Escherichia coli (strain K12) GN=prs PE=1 SLV=1</td>
<td>1 3</td>
<td>3 41.118</td>
<td>387 24</td>
<td>2511100</td>
<td></td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P0A705</td>
<td>IF2_ECOLI Translation initiation factor IF-2 OS=Escherichia coli (strain K12) GN=infb PE=1 SLV=1</td>
<td>1 8</td>
<td>8 97.349</td>
<td>890 9.8</td>
<td>2622200</td>
<td></td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P62707</td>
<td>GPMA_ECOLI 2,3W bisphosphoglycerate-dependent phosphoglycerate mutase OS=Escherichia coli (strain K12) GN=gpmA PE=1 SLV=2</td>
<td>1 10</td>
<td>10 25.7</td>
<td>296 15.2</td>
<td>2721400</td>
<td></td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P0A717</td>
<td>KPRS_ECOLI Ribose-phosphate pyrophosphokinase OS=Escherichia coli (strain K12) GN=prs PE=1 SLV=1</td>
<td>1 3</td>
<td>3 41.118</td>
<td>387 24</td>
<td>2511100</td>
<td></td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P0A705</td>
<td>IF2_ECOLI Translation initiation factor IF-2 OS=Escherichia coli (strain K12) GN=infb PE=1 SLV=1</td>
<td>1 8</td>
<td>8 97.349</td>
<td>890 9.8</td>
<td>2622200</td>
<td></td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P62707</td>
<td>GPMA_ECOLI 2,3W bisphosphoglycerate-dependent phosphoglycerate mutase OS=Escherichia coli (strain K12) GN=gpmA PE=1 SLV=2</td>
<td>1 10</td>
<td>10 25.7</td>
<td>296 15.2</td>
<td>2721400</td>
<td></td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P0A717</td>
<td>KPRS_ECOLI Ribose-phosphate pyrophosphokinase OS=Escherichia coli (strain K12) GN=prs PE=1 SLV=1</td>
<td>1 3</td>
<td>3 41.118</td>
<td>387 24</td>
<td>2511100</td>
<td></td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P0A705</td>
<td>IF2_ECOLI Translation initiation factor IF-2 OS=Escherichia coli (strain K12) GN=infb PE=1 SLV=1</td>
<td>1 8</td>
<td>8 97.349</td>
<td>890 9.8</td>
<td>2622200</td>
<td></td>
</tr>
<tr>
<td>Accession</td>
<td>Description</td>
<td>Organism</td>
<td>Identity</td>
<td>Length</td>
<td>Estart</td>
<td>End</td>
<td>Score</td>
</tr>
<tr>
<td>-----------</td>
<td>-------------------------------------------------------</td>
<td>----------------------</td>
<td>----------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>-------</td>
</tr>
<tr>
<td>sp</td>
<td>P0ABH7</td>
<td>Citrate synthase OS=Escherichia coli (strain K12) GN=gltA</td>
<td>1 6 6</td>
<td>48.014</td>
<td>427</td>
<td>15.9</td>
<td>1877500</td>
</tr>
<tr>
<td>sp</td>
<td>P45577</td>
<td>RNA chaperone OS=Escherichia coli (strain K12) GN=proQ</td>
<td>1 4 4</td>
<td>25.892</td>
<td>232</td>
<td>21.1</td>
<td>1678500</td>
</tr>
<tr>
<td>sp</td>
<td>P0A9M8</td>
<td>Phosphate acetyltransferase OS=Escherichia coli (strain K12) GN=pta</td>
<td>1 7 7</td>
<td>77.171</td>
<td>714</td>
<td>11.3</td>
<td>1577400</td>
</tr>
<tr>
<td>sp</td>
<td>P04370</td>
<td>Malodextrin phosphorylase OS=Escherichia coli (strain K12) GN=malP</td>
<td>1 6 6</td>
<td>90.521</td>
<td>797</td>
<td>8.9</td>
<td>1439000</td>
</tr>
<tr>
<td>sp</td>
<td>P0A9B0</td>
<td>ATP synthase subunit alpha OS=Escherichia coli (strain K12) GN=atpA</td>
<td>1 8 8</td>
<td>55.221</td>
<td>513</td>
<td>16.4</td>
<td>1392300</td>
</tr>
<tr>
<td>sp</td>
<td>P23585</td>
<td>Ribosomal protein S6 OS=Escherichia coli (strain K12) GN=rpsF</td>
<td>1 3 3</td>
<td>15.703</td>
<td>135</td>
<td>22.2</td>
<td>1319800</td>
</tr>
<tr>
<td>sp</td>
<td>P0AGD3</td>
<td>Superoxide dismutase OS=Escherichia coli (strain K12) GN=sodB</td>
<td>1 5 5</td>
<td>21.265</td>
<td>193</td>
<td>32.6</td>
<td>1254300</td>
</tr>
<tr>
<td>sp</td>
<td>P0ABB0</td>
<td>Acetyl-CoA carboxylase OS=Escherichia coli (strain K12) GN=accA</td>
<td>1 4 4</td>
<td>17.835</td>
<td>168</td>
<td>42.9</td>
<td>1220400</td>
</tr>
<tr>
<td>sp</td>
<td>P68919</td>
<td>Transketolase OS=Escherichia coli (strain K12) GN=tktA</td>
<td>1 2 2</td>
<td>10.693</td>
<td>94</td>
<td>13.8</td>
<td>1179900</td>
</tr>
<tr>
<td>sp</td>
<td>P0ACA3</td>
<td>Stringent starvation protein A OS=Escherichia coli (strain K12) GN=sspA</td>
<td>1 3 3</td>
<td>10.387</td>
<td>97</td>
<td>42.3</td>
<td>1178100</td>
</tr>
<tr>
<td>sp</td>
<td>P03580</td>
<td>6-phosphogluconate dehydrogenase OS=Escherichia coli (strain K12) GN=gnd</td>
<td>1 6 6</td>
<td>51.481</td>
<td>468</td>
<td>17.3</td>
<td>1146100</td>
</tr>
<tr>
<td>sp</td>
<td>P0AF6</td>
<td>Transcription termination/antitermination protein NusA OS=Escherichia coli (strain K12) GN=nusA</td>
<td>1 3 3</td>
<td>15.703</td>
<td>135</td>
<td>22.2</td>
<td>1319800</td>
</tr>
<tr>
<td>sp</td>
<td>P09737</td>
<td>Formate acetyltransferase OS=Escherichia coli (strain K12) GN=fbtA</td>
<td>1 3 3</td>
<td>15.703</td>
<td>135</td>
<td>22.2</td>
<td>1319800</td>
</tr>
<tr>
<td>sp</td>
<td>P0A685</td>
<td>Formate-dependent protease OS=Escherichia coli (strain K12) GN=hslU</td>
<td>1 3 3</td>
<td>24.305</td>
<td>212</td>
<td>19.3</td>
<td>1080500</td>
</tr>
<tr>
<td>sp</td>
<td>P0A687</td>
<td>IscS OS=Escherichia coli (strain K12) GN=iscS</td>
<td>1 3 3</td>
<td>45.089</td>
<td>404</td>
<td>7.7</td>
<td>1063600</td>
</tr>
<tr>
<td>sp</td>
<td>P0ACB3</td>
<td>Aspartate ammonia-lyase OS=Escherichia coli (strain K12) GN=aspA</td>
<td>1 4 4</td>
<td>52.356</td>
<td>478</td>
<td>8.6</td>
<td>991770</td>
</tr>
</tbody>
</table>
### Appendix B
Mass spectrometry analysis of SEC purified 6S<sup>RDm</sup>:HE

<table>
<thead>
<tr>
<th>Fasta headers</th>
<th>Number of proteins</th>
<th>Peptides</th>
<th>Unique peptides</th>
<th>Mol. weight [kDa]</th>
<th>Sequence length</th>
<th>Sequence coverage [%]</th>
<th>Intensity</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;sp</td>
<td>P0ABV2</td>
<td>RPOB_ECOLI DNA-directed RNA polymerase subunit beta OS=Escherichia coli (strain K12) GN=rpoB PE=1 SV=1</td>
<td>1</td>
<td>51</td>
<td>51</td>
<td>150.63</td>
<td>1342</td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P0ABT7</td>
<td>RPOC_ECOLI DNA-directed RNA polymerase subunit beta OS=Escherichia coli (strain K12) GN=rpoC PE=1 SV=1</td>
<td>1</td>
<td>51</td>
<td>51</td>
<td>155.16</td>
<td>1407</td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P0A724</td>
<td>RPOA_ECOLI DNA-directed RNA polymerase subunit alpha OS=Escherichia coli (strain K12) GN=rpoA PE=1 SV=1</td>
<td>1</td>
<td>13</td>
<td>13</td>
<td>36.511</td>
<td>329</td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P00579</td>
<td>RPOD_ECOLI RNA polymerase sigma factor RpoD OS=Escherichia coli (strain K12) GN=rpoD PE=1 SV=2</td>
<td>1</td>
<td>13</td>
<td>13</td>
<td>70.263</td>
<td>613</td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P0A6F5</td>
<td>CH60_ECOLI 60 kDa chaperonin OS=Escherichia coli (strain K12) GN=groL PE=1 SV=2</td>
<td>1</td>
<td>13</td>
<td>13</td>
<td>57.328</td>
<td>548</td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P0A25</td>
<td>THIO_ECOLI Thioredoxin-1 OS=Escherichia coli (strain K12) GN=trxA PE=1 SV=2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>11.806</td>
<td>109</td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P0A800</td>
<td>RPOZ_ECOLI DNA-directed RNA polymerase subunit omega OS=Escherichia coli (strain K12) GN=rpoZ PE=1 SV=1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>10.236</td>
<td>91</td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P0A7K2</td>
<td>RL7_ECOLI 50S ribosomal protein L7/L12 OS=Escherichia coli (strain K12) GN=rplP PE=1 SV=2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>12.295</td>
<td>121</td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P21599</td>
<td>KPY2_ECOLI Pyruvate kinase II OS=Escherichia coli (strain K12) GN=pykA PE=1 SV=3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>51.357</td>
<td>480</td>
</tr>
</tbody>
</table>
## Appendix C

Mass spectrometry analysis of commercial HE

<table>
<thead>
<tr>
<th>Fasta headers</th>
<th>Number of proteins</th>
<th>Peptides</th>
<th>Unique peptides</th>
<th>Mol. weight [kDa]</th>
<th>Sequence length</th>
<th>Sequence coverage [%]</th>
<th>Intensity</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;sp</td>
<td>P0A8V2</td>
<td>RPOB_ECOLI</td>
<td>DNA-directed RNA polymerase subunit beta OS=Escherichia coli (strain K12) GN=rpoB PE=1 SV=1</td>
<td>1</td>
<td>53</td>
<td>53</td>
<td>150.63</td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P0A8T7</td>
<td>RPOC_ECOLI</td>
<td>DNA-directed RNA polymerase subunit alpha OS=Escherichia coli (strain K12) GN=rpoA PE=1 SV=1</td>
<td>1</td>
<td>55</td>
<td>55</td>
<td>155.16</td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P0A6F5</td>
<td>CH60_ECOLI</td>
<td>60kDa chaperonin OS=Escherichia coli (strain K12) GN=groL PE=1 SV=2</td>
<td>1</td>
<td>18</td>
<td>18</td>
<td>109.77</td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P0A800</td>
<td>RPOZ_ECOLI</td>
<td>DNA-directed RNA polymerase subunit omega OS=Escherichia coli (strain K12) GN=rpoZ PE=1 SV=1</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>10.236</td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P0A6J5</td>
<td>DADA_ECOLI</td>
<td>D-amino acid dehydrogenase OS=Escherichia coli (strain K12) GN=dadA PE=1 SV=1</td>
<td>1</td>
<td>12</td>
<td>12</td>
<td>54.87</td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P0AES4</td>
<td>GYRA_ECOLI</td>
<td>DNA gyrase subunit A OS=Escherichia coli (strain K12) GN=gyrA PE=1 SV=1</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>96.962</td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P08394</td>
<td>RECB_ECOLI RecB/CDC enzyme subunit RecB OS=Escherichia coli (strain K12) GN=rebc PE=1 SV=1</td>
<td>1</td>
<td>6</td>
<td>6</td>
<td>133.96</td>
<td>1180</td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P13445</td>
<td>RPOE_ECOLI</td>
<td>RNA polymerase sigma factor RpoE OS=Escherichia coli (strain K12) GN=rpoE PE=1 SV=1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>32.468</td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P22531</td>
<td>MUKB_ECOLI</td>
<td>Chromosome partition protein MukB OS=Escherichia coli (strain K12) GN=mukB PE=1 SV=2</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>170.23</td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P42632</td>
<td>TDCD_ECOLI</td>
<td>PIF-like enzyme TcdE OS=Escherichia coli (strain K12) GN=tcdE PE=1 SV=2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>85.935</td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P0A6F6</td>
<td>NUSG_ECOLI</td>
<td>Transcription termination/antitermination protein NusG OS=Escherichia coli (strain K12) GN=nusG PE=1 SV=2</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>20.531</td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P0AFG0</td>
<td>NUSG_ECOLI</td>
<td>Transcription termination/antitermination protein NusG OS=Escherichia coli (strain K12) GN=nusG PE=1 SV=2</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>20.531</td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P0AFG6</td>
<td>NUSG_ECOLI</td>
<td>Transcription termination/antitermination protein NusG OS=Escherichia coli (strain K12) GN=nusG PE=1 SV=2</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>20.531</td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P0AFG6</td>
<td>NUSG_ECOLI</td>
<td>Transcription termination/antitermination protein NusG OS=Escherichia coli (strain K12) GN=nusG PE=1 SV=2</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>20.531</td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P0496</td>
<td>OOD2_ECOLI</td>
<td>Dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex OS=Escherichia coli (strain K12) GN=sudB PE=1 SV=3</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>44.011</td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P0AFG6</td>
<td>OOD2_ECOLI</td>
<td>Dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex OS=Escherichia coli (strain K12) GN=sudB PE=1 SV=3</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>44.011</td>
</tr>
<tr>
<td>&gt;sp</td>
<td>P0AFG6</td>
<td>OOD2_ECOLI</td>
<td>Dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex OS=Escherichia coli (strain K12) GN=sudB PE=1 SV=3</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>44.011</td>
</tr>
</tbody>
</table>