LIGAND RECOGNITION AND CONFORMATIONAL DYNAMICS OF SAM-BINDING RIBOSWITCHES

by Changrui Lu

LIGAND RECOGNITION AND CONFORMATIONAL DYNAMICS OF SAM-BINDING RIBOSWITCHES

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Changrui Lu

August 2010

STUDY OF REGULATION MECHANISMS OF SAM-BINDING RIBOSWITCHES USING X-RAY CRYSTALLOGRAPHY AND CHEMICAL PROBING

Changrui Lu, Ph.D.

Cornell University 2010

Riboswitches are recently discovered regulatory RNAs that directly relay environmental cues to the genetic regulatory machinery. S box (SAM-I) and the S_{MK} box (SAM-III) RNA are both *S*- Adenosyl -*L*-methionine (SAM) responsive riboswitches that regulate bacterial gene expression at the levels of transcription attenuation or translation inhibition respectively. The S box RNA is the most wide-spread transcriptional riboswitch in Gram-positive bacteria, while the S_{MK} RNA is a translational riboswitch in members of the Lactobacillales. Both RNAs bind specifically to SAM and feedback regulate the SAM biosynthetic pathway. However, these riboswitches share no sequence, structural or mechanistic homology. Our lab employs both X-ray crystallography and chemical probing to analyze their ligand recognition and mechanisms of gene repression.

Our X-ray structures show that both the S_{MK} and S box riboswitches fold into compact structures once bound to SAM, and that both RNAs adopt unique yet elegantly simple ways to specifically recognize SAM in favor of its close homolog, S-Adenosyl-homocysteine (SAH). Further mutagenesis and chemical probing results show that both the S_{MK} and S box RNA adopt extremely dynamic conformations in the absence of SAM, and certain structural elements outside the SAM-binding domain have profound effects on the conformation dynamics. The thermal melting experiments also showed that both riboswitches undergo unique unfolding processes when subject to increasing temperatures. Our findings suggest that these SAM

riboswitches have independently evolved conformational equilibriums that accurately respond to changes in their respective environments.

BIOGAPHICAL SKETCH

Changrui (Ray) Lu, the only child to Yongchang Lu and Yun Dong, was born in Shanghai, the largest city and economic center in China on June 6th, 1983. Except for the fifteen months Ray lived in Moscow, Russia visiting with his parents, Ray spent most of his childhood in the city until he was 18. He attended Shanghai Foreign Language School where he excelled in natural sciences and foreign languages, before admitted to Colgate University in New York, USA with full scholarship.

During his undergraduate studies in Colgate, Ray was encouraged to pursue a research career in biological sciences. He spent six months on researching human adenosine receptors with Dr. Kenneth Jacobson in National Institute of Health as a Hughes scholar in 2004. He also completed two independent research projects with Dr. Nancy Pruitt on freeze tolerance in *Eurosta solidaginis* larvae, and Dr. Jeffery Buboltz on creating artificial bio-membrane, at Colgate University. After graduating *magna cum laude* towards his Bachelor's degree in Molecular Biology, Ray started his graduate research in biochemistry and molecular biology at Cornell University towards his Ph.D.

After rotated with four different labs in biochemistry, biophysics and physics, Ray joined Dr. Ailong Ke's lab to study structures and mechanisms of RNA and RNA binding proteins, such as archaeal exosome complex, telomerase, riboswitches and pRNA complex. His training and research with Dr. Ke has fully utilized Ray's problem solving ability and interest in multiple disciplines including chemistry, physics, biology and computer science. In Dr. Ke's lab, Ray resorted to many different techniques to reveal detailed mechanisms behind gene regulation by RNAs, also known as riboswitches.

Dedicated to my parents who have supported me unconditionally through this first journey of many.

ACKNOWLEDGEMENTS

I would like to express my extreme gratitude to my thesis advisor, Dr. Ailong Ke and his wife Fang Ding, for the unbelievable intellectual stimulation and support throughout my graduate career. I would also like to thank my committee members, Dr. Eric Alani, Dr. Steven Ealick and Dr. Bik Tye for their expert support, advice and conversations. Several colleagues have provided assistance and encouragement over the years that are indispensible to my life in Cornell, including the awesome members of the Ealick lab, Feigenson lab, Ke lab, Roberts lab, Sondermann lab and Tumbar lab. Last but not least, I would like to thank Anirban Chowdhury, a very intelligent undergraduate researcher in my lab, for performing many experiments among providing other assistance. This work would not be possible without the generous support from each and every one of them.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
INTRODUCTION

BACKGROUND

A wide variety of mechanisms regulate genes that respond to environmental changes. Historically, most analysis of gene regulation has focused primarily on regulatory proteins that bind to cellular metabolites (e.g., SAM, cAMP etc.) and DNA, which in turn affect the behavior of RNA polymerase (RNAP) at the promoter region of targeted genes. These protein-DNA complexes may act to increase or decrease transcription of the promoters, and the binding efficacy of the protein regulator is often controlled by an effector molecule that activates or represses DNA binding (e.g., *lac* repressor binding is inhibited by lactose, catabolite gene activator protein is activated by cAMP, *trp* repressor is activated by tryptophan). Multiple regulatory proteins with each of its respective ligands may affect a single promoter in order to integrate multiple signals.

Historically, protein regulators are well studied and considered as primary means of gene regulation. They are specific and effective in complex regulatory systems, but such mechanism always requires that the cell encode and express these proteins on a regular basis, so that the cellular regulator concentration is maintained. This mechanism is ideal for complex systems in higher organisms, but often considered superfluous for many prokaryotes. Since protein regulators usually operate at transcription, regulation only at the level of transcription level is not always sufficient, as transcription occurs only under appropriate conditions and may not provoke a rapid response in case of need. Obviously, regulation at translation level relays a much faster response because the mRNA is already made and is ready to translate into proteins. Similar to DNA-binding proteins, mRNA-binding proteins,

such as spliceosomes, editosomes and poly-A binding proteins, are also required to be expressed constitutively. In higher organisms where the complicity of the proteome and intersections of complex metabolic pathways require intricate and precise gene regulation, multiple regulators will have to collaborate on a single location in order to correctly modulate a response. However, such elaborate mechanism is considered too wasteful for simple organisms such as bacteria and viruses, which values accuracy and efficacy over complicity. Bacterial cells require a faster yet more cost effective mechanism that rapidly responds to environmental cues.

The identification of self-cleaving ribozymes two decades ago has led to the RNA World hypothesis, an early point in evolution when RNA stored genetic information, catalyzed reactions and possibly self-replicated. For such an RNA World to be completely functional, RNA must have also predated protein as principle regulator of metabolism. The ability of RNA to regulate gene expression have eluded discovery until recent discovery of "riboswitches", *cis*-acting regulatory RNAs located in the 5′-untranslated region (5′-UTR) that can form mutually exclusive alternate conformations, wherein one of the conformations leads to greater expression of the associated genes while the other configuration represses the expression (Figure 1.1). Riboswitches are environment-sensing RNAs, which detects changes in metabolic or environmental conditions in its "aptamer domain" and relay a structural response in its "expression platform" (Garst and Batey, 2009, Serganov, 2009, Dambach and Winkler, 2009, Henkin, 2008, Montange and Batey, 2008, Edwards et al., 2007, Fuchs et al., 2006, Tucker and Breaker, 2005, Grundy and Henkin, 2006, Vitreschak et al., 2004). Being simple and efficient, riboswitches are wide-spread in bacterial species. Recent studies of published literature showed that at least 4% of the *Bacillus subtilis* genome is regulated through *cis*-acting regulatory RNAs (Winkler, 2005a), and similar

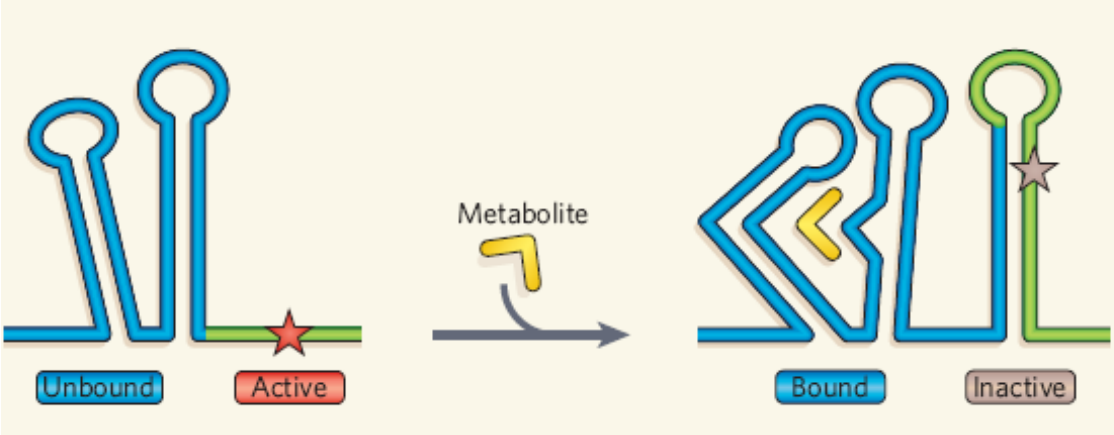


Figure 1.1. Gene expression *cis*-regulated by riboswitches.

The RNA of a riboswitch contains two functional domains: a metabolite-sensing aptamer domain (blue) and a expression platform (green). In response to a particular metabolite (yellow), the aptamer domain switches conformation and induces structural change at the expression platform, shutting off protein production. (Adapted from Reichow and Varani, 2006)

proportion in *Escherichia coli* is regulated by *trans*-acting small RNAs (Storz et al., 2005, Gottesman, 2004). Riboswitches exist in diverse bacterial species (Vitreschak et al., 2004) and associate with coenzymes that have ancient origins (White, 1976). Combined with the RNA world hypothesis, riboswitch could predate all protein regulators and could be modern relics from primordial organisms. Given their importance in metabolisms of both modern and ancient organisms, riboswitches have recently become subjects of intense genetic, biochemical, biophysical and structural studies.

RIBOSWITCH DISCOVERY

Historically, both genetic and bioinformatic tools have contributed towards the discovery of riboswitches. Prior to the invention of term "riboswitch", the Henkin group identified a conserved non-coding 5′-UTR upstream of the S-box gene family which is involved in the biosynthesis of methionine and cysteine in some Gram-positive bacteria (Grundy and Henkin, 1998). These motifs consist of a 5' portion of complicated 2nd structures and a 3' element resembling an intrinsic transcriptional terminator. The S box system responds to methionine starvation without any involvement of promoter sequence and protein regulators, yet it is dependent upon the conserved 5′ leader sequence. Mutational analysis combined with genome-wide searches showed that the S box leader mRNA changes its conformation in response to metabolite concentrations.

In 2002, the first comprehensive proof of multiple classes of riboswitches was published by the Breaker group, demonstrating that mRNAs can bind metabolites directly in the absence of proteins (Nahvi et al., 2002). In-line probing analyses

demonstrated that the *Escherichia coli btuB* mRNA changes its conformation upon specific binding to coenzyme B_{12} (AdoCbl). This novel RNA genetic control element was termed “riboswitch”.

Although recent advances in genomics and bioinformatics have greatly increased the efficiency in sifting through billions of bases in genome sequences for riboswitches, the basic strategy and algorithm behind these automated comparative genomics searches remained the same. The most modern approach clusters intergenic regions (IGRs) by BLAST (Altschul et al., 1997) and predicts secondary structures by using CMfinder (Yao et al., 2006). Then the resulting identified homolog structures are refined and realigned. The highest scoring alignments are then analyzed manually to identify the most promising candidates for further genetic and biochemical analyses (Weinberg et al., Weinberg et al., 2007, Eddy and Durbin, 1994). Many of these candidate RNAs are later experimentally proven to regulate cell metabolism, such as glmS (Wilkinson and Been, 2005, McCarthy et al., 2005, Jansen et al., 2006, Klein and Ferre-D'Amare, 2006), glycine (Mandal and Breaker, 2004a) and PreQ1-I (Roth et al., 2007, Klein et al., 2009) riboswitches.

FUNCTIONS AND MECHANISM OF RIBOSWITCHES

It is believed that riboswitch is extremely prevalent in bacteria (Winkler, 2005a), with a few identification is plants (Bocobza et al., 2007), fungi and Archaea (Sudarsan et al., 2003a). A number of important pathogens including *Bacillus anthracis, Staphylococcus, Enterococcus, Streptococcus, Listeria, Clostridium,* and *Mycobacterium* extensively utilizes riboswitch to regulate key metabolic pathways (Winkler and Breaker, 2005, Grundy and Henkin, 2006) (Figure 1.2).

Riboswitch RNAs have evolved to regulate a plethora of gene families within the same pathways, such as the metabolism and transport of vitamins (Winkler et al., 2002a, Wickiser et al., 2005, Serganov et al., 2006, Winkler et al., 2002b, Sudarsan et al., 2005, Thore et al., 2006), nucleotides (Christiansen et al., 1997, Mandal and Breaker, 2004a, Mandal et al., 2003, Roth et al., 2007, Klein et al., 2009, Kim et al., 2007, Meyer et al., 2008), amino acids (Mandal et al., 2004, Rodionov et al., 2003, Sudarsan et al., 2003b, Grundy et al., 2003), cofactors (Winkler et al., 2003, Wickiser et al., 2005, McDaniel et al., 2003, Epshtein et al., 2003, Corbino et al., 2005, Fuchs et al., 2006) and metal ions (Groisman et al., 2006, Coppins et al., 2007). It is also possible that riboswitches bind to additional ligands that have not yet been discovered (Barrick et al., 2004). Considering the effectiveness, extent and spread of riboswitch-regulated pathways in prokaryotes, the small number of metabolite-binding RNAs found in eukaryotic systems only suggests that such mechanisms may be more widespread than currently appreciated (Sudarsan et al., 2003a, Thore et al., 2006, Cheah et al., 2007, Winkler, 2005a).

Most riboswitch RNA is comprised of two separate elements, a ligand-binding aptamer domain and a gene expression platform, with the exception of some single-domain thermosensor RNAs (Grundy and Henkin, 2004, Morita et al., 1999). The aptamer domain is responsible for specific recognition of the correct ligand and repels similar compounds. Binding of the ligand promotes a structural shift in this domain that affects the structure of downstream expression platform, usually resulting in sequestration of a segment of RNA that would otherwise bind to another part of the RNA. In systems where binding of the effector inhibits gene expression, the ligand is often the end product of the biosynthetic pathway, and the regulated operon consists of many, if not all, proteins involved. This feedback regulation is a common theme in

Figure 1.2. Selected human bacterial pathogens that carry riboswitches as of 2007.

The number of representatives of each riboswitch class that is found in each species is given, followed in parentheses by the total number of genes regulated by given riboswitches. The numbers in red indicate that at least one of the genes regulated by that riboswitch is predicted to be essential for survival or virulence. (Adapted from Blount and Breaker, 2006)

Riboswitch target metabolite:	TPP	FMN	AdoCbl	Purine	SAM1	SAM2	SAM3	Lysine	GlcN6P	Glycine	$PreQ_1$
Acinetobacter baumannii	1(3)	1(1)	1(1)	–	–	–	–	–	–	1*(1)	–
Bacillus anthracis	6(19)	2(5)	1(1)	6(9)	17(36)	–	–	4(4)	1(2)	1(1)	2(5)
Brucella melitensis	2(11)	1(1)	2(5)	–	–	1(1)	–	–	–	1*(3)	–
Enterococcus faecalis	2(5)	–	2(4)	1(2)	–	–	1(1)	1(1)	1(1)	–	2(3)
Escherichia coli	3(11)	1(1)	1(2)	–	–	–	–	1(1)	–	–	–
Francisella tularensis	1(1)	1(5)	–	–	–	1(1)	–	–	–	–	1(1)
Haemophilus influenza	3(11)	1(1)	–	–	–	–	–	1(1)	–	1*(1)	1(1)
Helicobacter pylori	1(2)	–	–	–	–	–	–	–	–	–	–
Listeria monocytogenes	2(5)	1(1)	2(20)	2(3)	7(14)	–	–	1(1)	1(1)	1(3)	1(1)
Mycobacterium tuberculosis	2(6)	–	2(4)	–	–	–	–	–	–	2*(3)	–
Pseudomonas aeruginosa	1(1)	1(2)	5(24)	–	–	–	–	–	–	–	–
Salmonella enterica	3(11)	1(1)	2(22)	–	–	–	–	–	–	–	–
Staphylococcus aureus	2(7)	2(5)	–	1(2)	4(6)	–	–	2(2)	1(1)	1(3)	2(4)
Streptococcus pneumoniae	4(11)	1(4)	–	1(2)	–	–	1(1)	–	–	1(1)	–
Vibrio cholerae	2(7)	1(1)	1(1)	–	–	–	–	3(3)	–	1(1)	–
Yersinia pestis	3(8)	1(1)	1(2)	–	–	–	–	–	–	–	–

nearly all riboswitch controlled mechanisms. One exception is that some riboswitch mRNA inhibits gene expression in the absence of ligand (e.g., the *gcv* element) (Mandal and Breaker, 2004b), promoting gene expression once the ligand is associated with the riboswitch.

Similar to protein regulators, riboswitch mediates gene expression at both translational and transcriptional levels. However, while many proteins regulate the transcription initiation, none of the riboswitches discovered to date *cis*-regulate the promoter region of the gene, probably because it is much more complicated for mRNA to regulate transcription initiation by directly interacting with DNA. So far, the mechanisms riboswitches adopt to regulate gene expression can be roughly classified in to following six categories.

Transcription termination

This class consists of some of the most prevalent riboswitches found to date, including but not limited to the S box (SAM-1), SAM-IV, T box riboswitches. In the absence of the effector, the ligand-binding domain is unoccupied, and the RNA is in a conformation that allows expression of the downstream coding sequence, through formation of an antiterminator (AT) that prevents formation of the terminator stem loop and therefore allows transcription to continue into coding sequences. In the presence of the riboswitch's cognitive ligand, the riboswitch binds to the ligand, resulting in a structural change in the aptamer domain which is relayed downstream to allow the terminator to form (Figure 1.3A). This configuration results in premature termination of transcription.

Translation initiation

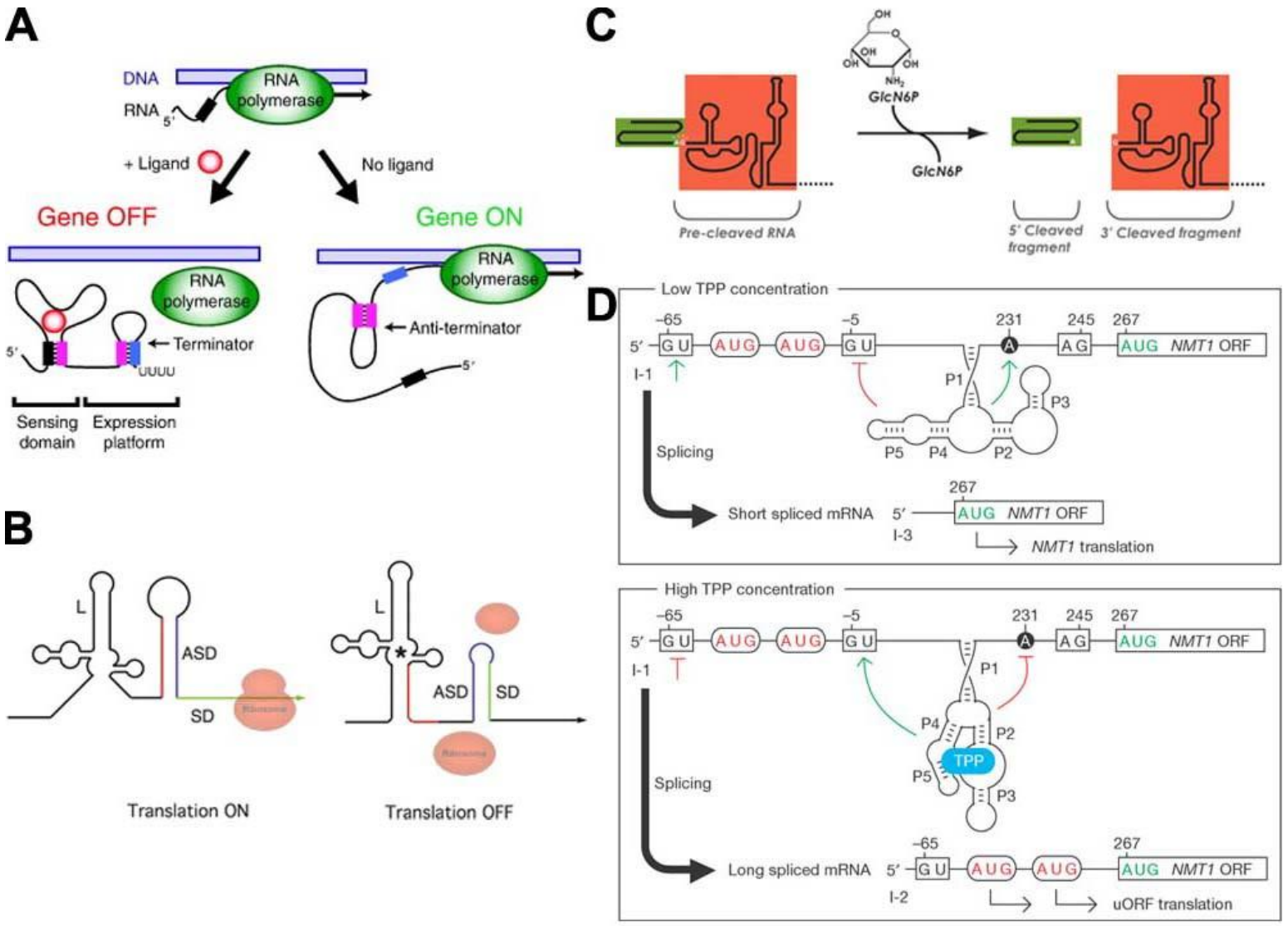
The translation riboswitch, such as the S_{MK} box (SAM-III) (Lu et al., 2008, Fuchs et al., 2006, Fuchs et al., 2007) and SAM-II (Corbino et al., 2005) riboswitches are generally similar to the transcriptional switches in principle, except that instead of blocking transcription, they sequester ribosome binding site and prevent translation initiation. In the absence of ligands, the RNA captures of the anti-Shine-Delgarno (ASD) into a structure that liberates the Shine-Delgarno (SD) sequence and allows translation to initiate. After the ligand concentration has reached desired levels, the ligand-binding aptamer domain is occupied, sequestration of the SD sequence in an ASD–SD helix that prevents translation initiation (Figure 1.3B).

Self-cleavage

In some cases, the riboswitch can be a ribozyme that cleaves itself before the open reading frame (ORF) in the presence of sufficient concentrations of its metabolite, exemplified by the Glucosamine-6-phosphate activated ribozyme (GlmS). The GlmS RNA is at the 5′-UTR of the mRNA and regulates biosynthesis of glucosamine-6-phosphate (GlcN6P) by controlling the production of the GlmS enzyme, which catalyzes the production of GlcN6P. When levels of GlcN6P are high, the RNA folds into a special conformation that catalyzes its own cleavage using GlcN6P as a cofactor, which leads to the degradation of both fragments of the mRNA that contains the coding sequence and the ribozyme, respectively (Figure 1.3C). When GlcN6P levels are low, the ribozyme does not cleave, and GlmS enzymes are produced (Barrick et al., 2004, Blount et al., 2006, Cochrane et al., 2007, Cochrane and Strobel, 2008, Klein et al., 2007, Klein and Ferre-D'Amare, 2006, Winkler et al., 2004).

Figure 1.3. Four examples of riboswitch mechanism.

(A) When the ligand is present, the transcriptional riboswitch adopts a stable metabolite-bound conformation to induce a transcription terminator. If the ligand is absent, mRNA forms an anti-terminator conformation which allows the transcription to proceed through the entire gene. (B) In the case of translational riboswitch, the presence of ligand allows sequestration of the SD sequence in an ASD–SD helix that prevents translation initiation. (C) The GlcN6P-induced self-cleavage reaction of glmS riboswitch. The green-shaded region denotes the portion of the sequence that is released upon self-cleavage by the ribozyme. The red-shaded region denotes the region corresponding to the GlcN6P-sensing ribozyme (D) Mechanism for alternative splicing by Thiamine pyrophosphate (TPP) riboswitch. Green lines indicate activation and red lines indicate inhibition during different occupancy states of the aptamer. (Adapted from Serganov and Patel, 2009, Henkin, 2008, Collins et al., 2007, Cheah et al., 2007) Both the transcription and translation riboswitches are used to control similar pathways across different species, such as the SAM biosynthetic pathway. In contrast to transcriptional switches, the translation riboswitch promotes a faster and more direct response (Lu et al., 2008).


A
DNA
RNA
RNA polymerase
+ Ligand
No ligand
Gene OFF
Gene ON
RNA polymerase
Terminator
UUUU
Anti-terminator
Sensing domain
Expression platform
B
L
ASD
SD
Translation ON
Translation OFF
C
GlcN6P
Pre-cleaved RNA
5′ Cleaved fragment
3′ Cleaved fragment
D
Low TPP concentration
High TPP concentration
Splicing
Short spliced mRNA
Long spliced mRNA
NMT1 ORF
NMT1 translation
uORF translation
TPP

Alternative splicing

The thiamine pyrophosphate (TPP) riboswitch, also known as the THI element and Thi-box riboswitch, regulates thiamine-related genes through a variety of splicing-related mechanisms in fungi (Cheah et al., 2007) and plants (Bocobza et al., 2007). The TPP riboswitches in the filamentous fungus *Neurospora crassa* control gene expression by mRNA splicing in a TPP-dependent base-pairing change (Cheah et al., 2007). The TPP-mediated splicing product has a frame-shifted ORF that does not produce active gene products and therefore quickly degraded (Figure 1.3D). In plants, such as *Arabidopsis Thaliana,* TPP-bound thiamine riboswitch generates an unstable transcript through differential processing of the 3' end of pre-mRNA (Bocobza et al., 2007).

Anti-sense mRNA competition

Recent analysis of a sulfur metabolic operon from bacteria *Clostridium acetobutylicum* revealed a complex multi-riboswitch mediated controlling mechanism (Andre et al., 2008). The *ubiGmccBA* operon of *Clostridium acetobutylicum* is flanked by a sense 5'-T box riboswitch and 3'-antisense S box riboswitch. In the presence of methionine, glutathione or cystine, the abundance of the *ubiG* transcript was inversely correlated to that of antisense transcripts as shown by Northern blot experiments and quantitative RT–PCR analyses in *B. subtilis* and *C. acetobutylicum*. In the presence of methionine, premature transcription termination at the anti-sense S-box riboswitch leads to a low level of antisense RNAs, and therefore an increase of sense mRNA. Therefore, the level of anti-sense mRNA transcripts is inversely correlated to the sense *ubiG* mRNA encoding the enzymes (Andre et al., 2008). However, the molecular mechanism of this sense-anti-sense competition is still under investigation.

In trans mRNA interference

Just several months before, for the first time, the Johansson group reported that two *S*-adenosylmethionine (SAM) riboswitches, SreA and SreB, also function in-*trans* in *Listeria monocytogenes* (Loh et al., 2009). Apart from regulating the expression of their own downstream genes, SreA and SreB control expression of the virulence regulator PrfA, not located in the immediate proximity of SreA and SreB, unexpected by previous studies. In the absence of SAM, SreA binds to, possibly in direct contact with, the distal side of the prfA untranslated RNA, thereby causing decreased expression of PrfA (Loh et al., 2009). This mechanism links low nutrient levels to higher bacterial virulence in *L. monocytogenes*.

STRUCTURE AND DYNAMICS

The flexibility and versatility of RNA has enabled RNA to form diverse and creative folds to accommodate a variety of ligands and structural functions. Recent advancements in RNA crystallography and chemical probing methods have provided many atomic resolution insights into how RNA folds and binds to effectors, as well as dynamics information related to the "switching" actions, as demonstrated, but not limited to, the following examples.

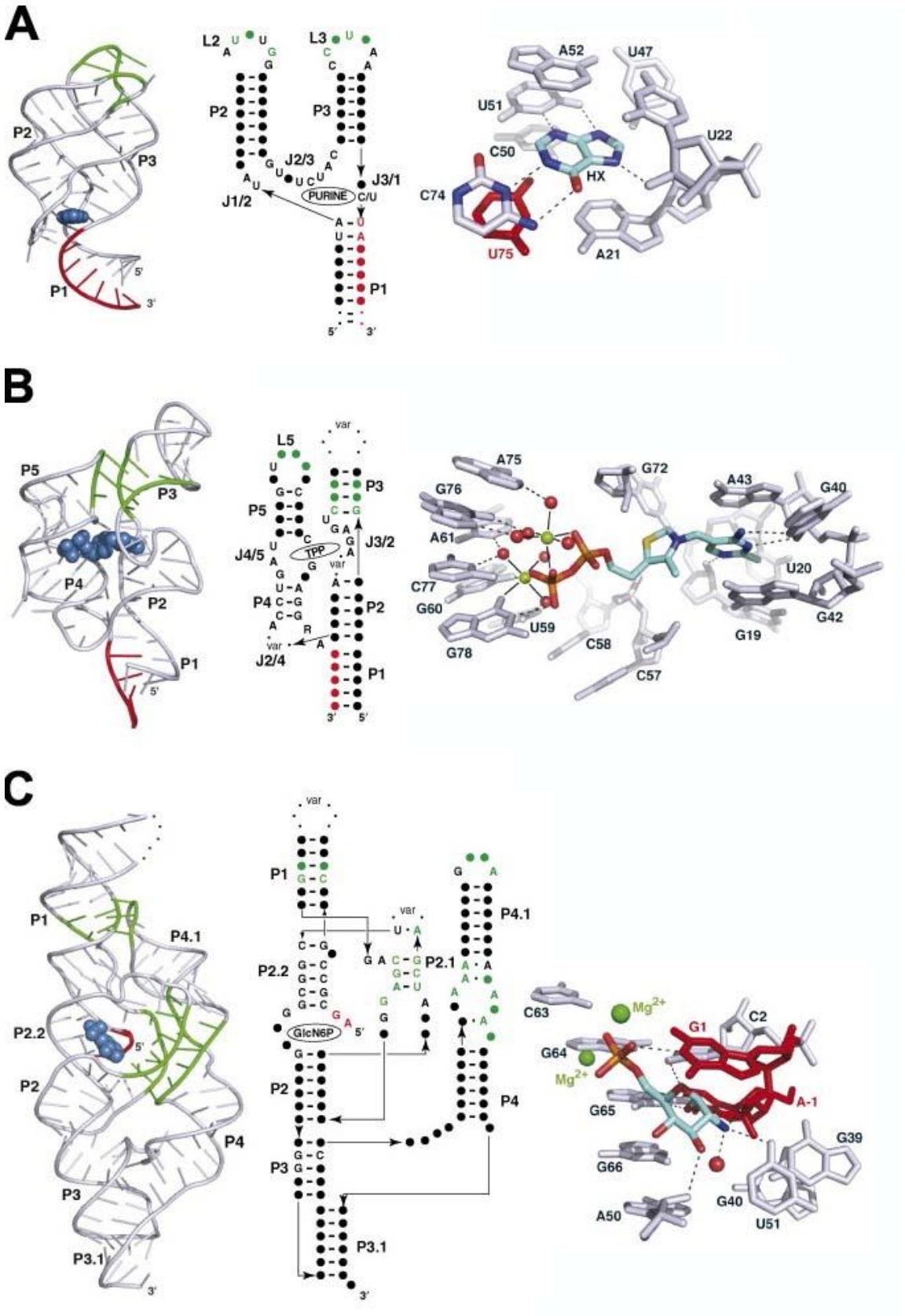
The first atomic resolution riboswitch structure solved with ligand is the guanine riboswitch (Batey et al., 2004), immediately followed by the adenine variant (Serganov et al., 2004). The RNA folds into three helices or paired regions, P1, P2 and P3, which are arranged as an 'h', with P1 on top of P3, and P2 pack side-by-side with P3. The terminal loops of P2 and P3 (the bottom ends of the "h") forms pseudoknot

interactions to stabilize the P2-P3 packing. The purine-binding pocket is located at the three-way junction of P1, P2 and P3 and is solvent-inaccessible (Figure 1.4A, left, middle) (Batey et al., 2004, Serganov et al., 2004). The 3′ strand of the riboswitch, which forms P1 together with the 5′ end forms part of an alternative structure within the expression platform when purine concentration is low and the riboswitch aptamer domain is at the unbound form. The purine riboswitch recognizes its ligand almost exclusively through hydrogen-bonding interactions that satisfy nearly all possible acceptors and donors of the purine, through Watson–Crick pairing (Figure 1.4A, right). By recognizing the Watson–Crick (as opposed to Hoogsteen) face of the ligand, the cell can easily utilize same RNA fold can recognize either adenine or guanine (Batey et al., 2004, Serganov et al., 2004).

The overall folds of the metabolite-binding aptamer domain of the TPP riboswitch from the eukaryote *Arabidopsis thaliana* (Thore et al., 2006) and the bacterium *Escherichia coli* (Edwards and Ferre-D'Amare, 2006, Serganov et al., 2006) have essentially the same architecture as the purine riboswitch (Figure 1.4B, left, middle). TPP-bound riboswitch adopts a similar "h" architecture with two parallel sets of coaxially stacked helices joined by a three-way junction. An A-minor interaction between P3 and L5 is functionally analogous to the loop–loop pseudoknot interaction of the purine riboswitch, stabilizes the bottom end of this "h"-fold. Instead of burying the ligand at the three-way junction, TPP bridges the two vertical helical stacks in the "h" (Thore et al., 2006). Therefore, in this case the ligand indirectly stabilizes with the expression platform (3′ of P1 as in the purine riboswitch) while making no direct contact with it. Ligand TPP is recognized through interactions from two helical stacks: the "pyrimidine sensor helix" (the P1-P2-P3 stack) and, the "pyrophosphate sensor helix" (the P4-P5 stack) (Figure 1.4B, right). The pyrimidine moiety is recognized

Figure 1.4. Crystal structures of riboswitch–effector complexes, schematic depiction of conserved primary and secondary structures and ligand interactions.

(A) Purine riboswitch (PDB code 1U8D), (B) TPP riboswitch (PDB code 2GDI), (C) *glmS* ribozyme-riboswitch (PDB code 2H0Z). The expression platform is colored red, and important tertiary interactions are green. Bound metabolites are shown in blue; PK and KT denote pseudoknots and K-turns, respectively; 'var' indicates variable RNA length and composition. Filled spheres indicate W-C base pairing. (Adapted from Edwards et al., 2007)


A
P2
P3
P1
5'
3'
L2
L3
J2/3
J3/1
J1/2
PURINE
C/U
A52
U47
U51
U22
C50
HX
C74
U75
A21
B
P5
P3
P4
P2
P1
L5
var
J4/5
TPP
J3/2
J2/4
A75
G72
G76
A43
G40
A61
U20
C77
G60
U59
G42
G78
C58
G19
C57
C
P1
P4.1
P2.2
P2
P4
P3
P3.1
P2.1
GlcN6P
C63
Mg2+
G1
C2
G64
A-1
G65
G39
G66
A50
G40
U51

through Watson–Crick face hydrogen bonding, while the positively charged pyrophosphate group is coordinated by two solvated metal ions (Figure 1.4B, right) (Thore et al., 2006).

The structure of the *glmS* riboswitch consists of three parallel helical stacks (P1/P2.2/P2/P3/P3.1, P2.1 and P4/P4.1) tied together by two pseudoknots (Figure 1.4C, left, middle) (Klein and Ferre-D'Amare, 2006, Cochrane et al., 2007). The riboswitch also adopts its active structure in the absence of its metabolite ligand (Hampel and Tinsley, 2006, Klein and Ferre-D'Amare, 2006). Crystal structures of the *glmS* riboswitch show that glucose-6-phosphate (Glc6P, an isosteric competitive inhibitor of GlcN6P) and GlcN6P bind equivalently in the *glmS* riboswitch active site (Cochrane et al., 2007, Klein et al., 2007). The sugar hydroxyl groups hydrogen bond to G1, C2, A50 and G65 in P2. The amine of GlcN6P hydrogen bonds to a water molecule, U51 and the 5′ oxygen of G1. Finally, the phosphate of the metabolite interacts with the RNA through two solvated divalent metal ions similar to the TPP riboswitch (Figure 1.4C, right) (Klein and Ferre-D'Amare, 2009, Cochrane et al., 2007, Klein et al., 2007). The ligand makes a direct contact with the leaving group in the ribozyme/riboswitch.

S-adenosylmethionine(SAM)-binding riboswitches are possibly the most diverse and wide-spread riboswitch discovered to date, both in terms of structures and regulation mechanisms. They all bind SAM to regulate SAM biosynthesis and transport. Across different organisms, four distinct SAM riboswitches are known: S box (later termed SAM-I), SAM-II, the S_{MK} box (later termed SAM-III) and SAM-IV riboswitch. These four varieties of SAM riboswitch have no obvious similarities in terms of genetics, sequence or structure, except that SAM-IV appears to have a similar

ligand-binding core to that of SAM-I riboswitches, but in the context of a different tertiary fold.

Two S box family riboswitches crystal structure has been reported ((Montange et al., 2009, Montange and Batey, 2006) and Chapter 5 of this thesis). Both structures are architecturally similar and have the same ligand recognition mechanism. The RNA folds into two sets of helical stacks spatially arranged by tertiary interactions including a K-turn and a pseudoknot at the four-way junction (Figure 1.5A, left, middle). The tertiary structure is further stabilized by extensive ribose zipper H-bonds between helices and metal coordination (Figure 1.5A, left). The SAM binding core of the *YitJ* S box, located at the interface of P1/P4 and P2/P3 coaxial stacks, is conserved across species. SAM adopts a "U" shaped conformation inside the binding pocket (Figure 1.5A, right). The Watson-Crick and Hoogsteen faces of the adenosine base of SAM are collectively recognized by four hydrogen bonds in a universally conserved, sheared A46 ŀU78 pair. The methionine tail of SAM makes hydrogen bond contacts with G79-C45•G11 triple, and the positively charged sulfonium group in SAM is specified by two ~ 4Å favorable electrostatic interactions to the O4 carbonyl oxygens of U7 and U112. The sulfonium ion also makes a very strong ~ 3.0 Å intramolecular electrostatic interaction with the O4′ of its ribose (see Chapter 2, 5 for details). Further chemical probing data strongly suggests that the aptamer domain of the S box RNA is largely preformed, and SAM enters the binding pocket through J1/2 and J3/4 (see Chapter 5).

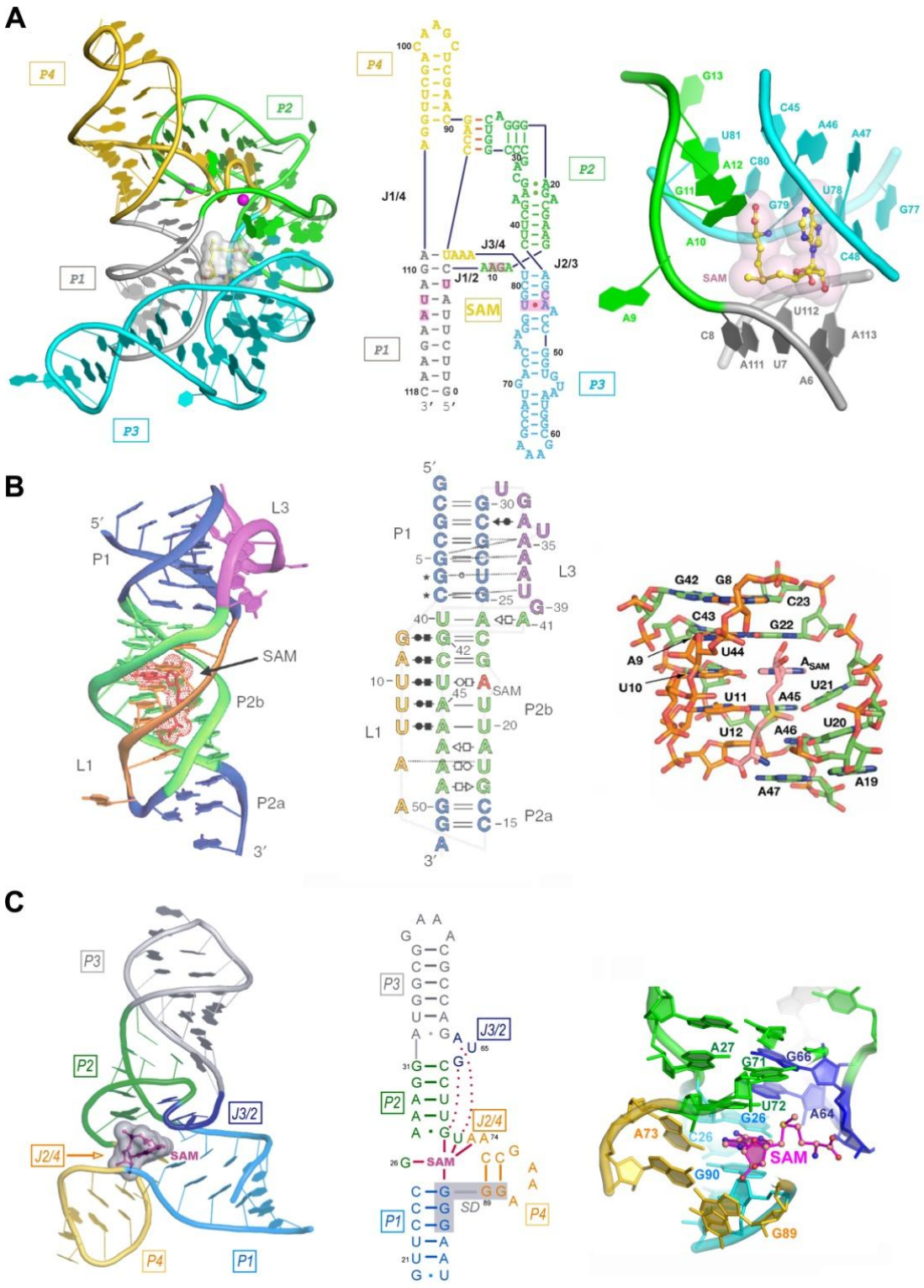
The global architecture of the SAM-II riboswitch is a classic (H-type) pseudoknot (Gilbert et al., 2008). The entire structure bound to SAM resembles a triple helix, with SAM-binding pocket located at the middle of the riboswitch buried by all three RNA strands (Figure 1.5B, left, middle). SAM binds in an extended

configuration, forming making contacts with five successive base pairs and triples along the major groove of the triplex (Figure 1.5B, right). The adenine moiety of SAM saturates its hydrogen bonding in a base triple between U10 and U44. The positively charged sulfonium ion is recognized by electrostatic interaction with the carbonyl oxygen groups of U11 and U21.

The S_{MK} box (SAM-III) riboswitch regulates the translation of SAM synthetase (*metK*) genes in *Enterococcus* bacteria by sequestration of the Shine-Dalgarno (SD) sequence (Fuchs et al., 2006, Lu et al., 2008). Crystal structures by our group have shown that the RNA folds into a "Y" shaped arrangement that organizes an array of conserved nucleotides around a three-way junction for SAM recognition (Figure 1.5C, left, middle). The Shine-Dalgarno sequence, which is sequestered by base pairing with the anti-Shine-Dalgarno sequence in response to SAM binding, also directly participates in SAM recognition (Lu et al., 2008). The adenosine moiety of SAM adopts an energetically unfavorable *syn*-conformation, and intercalates into the three-way-junction similar to a nucleotide, base-pairing with G26 (Figure 1.5C, right). The positive charge on the sulfonium ion of SAM is recognized through favorable electrostatic interactions with the O4 carbonyl of U72 and the 2'-hydroxyl of G71. The methionine end of SAM is exposed to solvent, which provides an easy entry/exit route of SAM, later supported by chemical probing data. Although the SAM ligand adopt different overall conformations in all three riboswitches, the positive charges are all universally stabilized through intramolecular electrostatic interactions with the O4' on the ribose and N3 on the base of SAM Figure 1.6, left). Upon further investigation into the literature, I have discovered that among 30 randomly picked SAM-bound crystal structures, all SAM molecules adopt this 3Å intermolecular interaction (Figure

Figure 1.5. Comparison of SAM-binding riboswitch crystal structures.

S box (SAM-I) riboswitch (Adapted from Lu *et al*, unpublished results). (B) SAM-II riboswitch (PDB code 2QWY) (Adapted from Gilbert et al., 2008)). (C) S_{MK} (SAM-III) riboswitch (PDB code 3E5C) (Adapted from Lu et al., 2008). The coloring schemes of tertiary structure, secondary structure and binding pocket structure are consistent. SAM is shown in surface representation in all three tertiary structures.


A
P4
P2
P1
P3
J1/4
J3/4
J2/3
J1/2
SAM
G13
C45
A46
A47
U81
A12
C80
G11
U78
G79
G77
A10
C48
SAM
U112
A9
C8
A113
A111
U7
A6
B
5′
L3
P1
SAM
P2b
L1
P2a
3′
G42
G8
C23
C43
G22
A9
U44
A_SAM
U10
U21
U11
A45
U12
A46
U20
A47
A19
C
P3
P2
J3/2
J2/4
SAM
P4
P1
A27
G66
G71
U72
G26
A64
A73
C26
SAM
G90
G89

1.6, right)., highly suggesting that this self-recognition mechanism distinguishes SAM from the non-charged SAH molecule (Lu et al., 2008).

Through the above-mentioned examples, one can conclude that RNA molecules are extremely creative in terms of recognizing a variety of metabolites. Despite the little structural or mechanistic similarities across species, the underlying principles of all the recognition are conserved, such as stacking/intercalation and W-C base pairing for purine/pyrimidine bases, metal-coordination for phosphate groups and hydroxyl-mediated electrostatic interactions for sulfonium ion recognition.

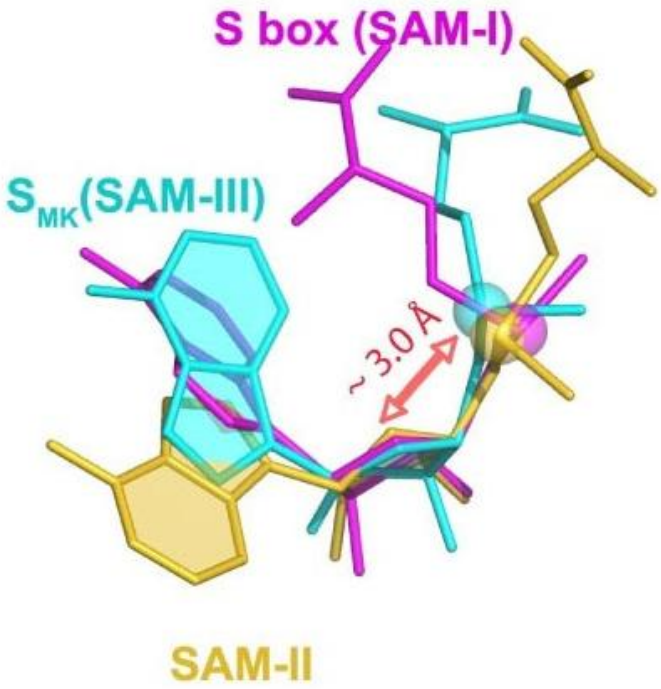
TARGET AS ANTIBIOTICS

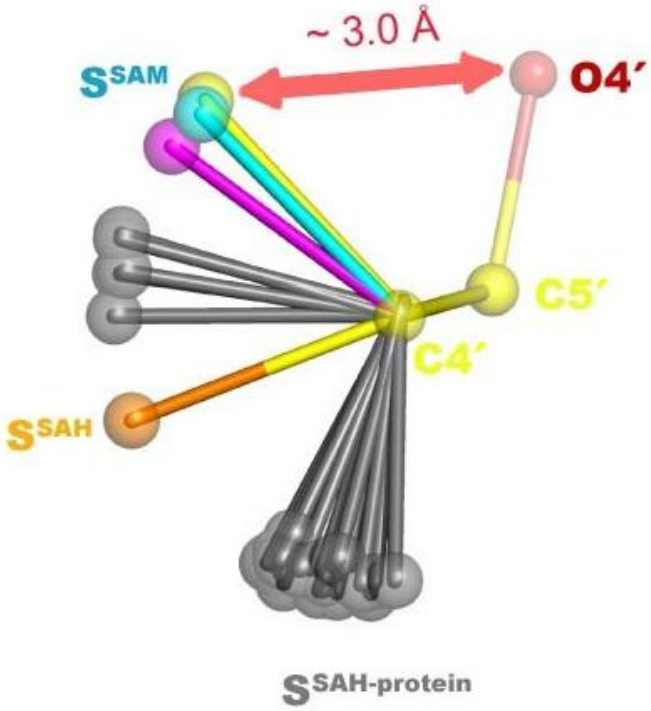
Current development of antibiotics is falling behind the pace of rapidly evolving pathogens (D'Costa et al., 2006). Most current chemical antibiotics target on transcription and translation complexes, which over the years has been countered by rapid bacterial mutagenesis. Raising dosage can only be effective to a certain point when toxicity or side-effects render the drug worthless. The prevalence of bacterial resistance to existing antibiotics requires us to widen our search of novel pathways for antibiotics development. Several properties of the riboswitches attract increasing attention as antibiotic targets:

- They control essential metabolic pathways in many important bacterial pathogens (Figure 1.2), including as *Bacillus anthracis, Staphylococcus, Enterococcus, Streptococcus, Listeria, Clostridium,* and *Mycobacterium.* Recent studies have shown that riboswitches directly control genes that turns on virulence in *L. monocytogenes* (Loh et al., 2009). By repressing biosynthesis of survival-essential pathways in a pathogen, a riboswitch-specific inhibitor can be an ideal agonistic drug that causes either fatality or loss of virulence (Zhang et al., 2004).

Figure 1.6. Conformational comparison between SAM and SAH molecules in various riboswitch and protein structures.

Left, SAM conformation from all three classes of SAM riboswitches, aligned along the ribose portion. Despite conformational differences in the adenosine base and the methionine tail, the sulfur atom makes a constant 3.0 Å polar interaction with the O4' of the ribose. Right, Comparison of the O4'-C4'-C5'-S dihedral angle within the SAM and SAH molecules. Molecules are aligned along the O4'-C4'-C5' linkage. *Gauche* conformation about the C4'-C5' bonds (labeled S^{SAM}) is the predominant conformation of SAM in the riboswitch complexes (the C5'-S bond is colored in yellow, cyan, and magenta for the S box (SAM-I), SAM-II and S_{MK} box [SAM-III] respectively). By contrast, a survey of 30 randomly chosen SAH-bound protein structures revealed that the sulfur ($S^{SAH\text{-}protein}$) does not adopt the same conformation as SAM. SAH in our S_{MK} box structure adopts a near *anti* conformation (the sulfur atom is labeled S^{SAH} and colored in orange). (Adapted from Lu et al., 2008).


S box (SAM-I)
S_MK(SAM-III)
~ 3.0 Å
SAM-II


~ 3.0 Å
S^SAM
O4′
C5′
C4′
S^SAH
S^SAH-protein

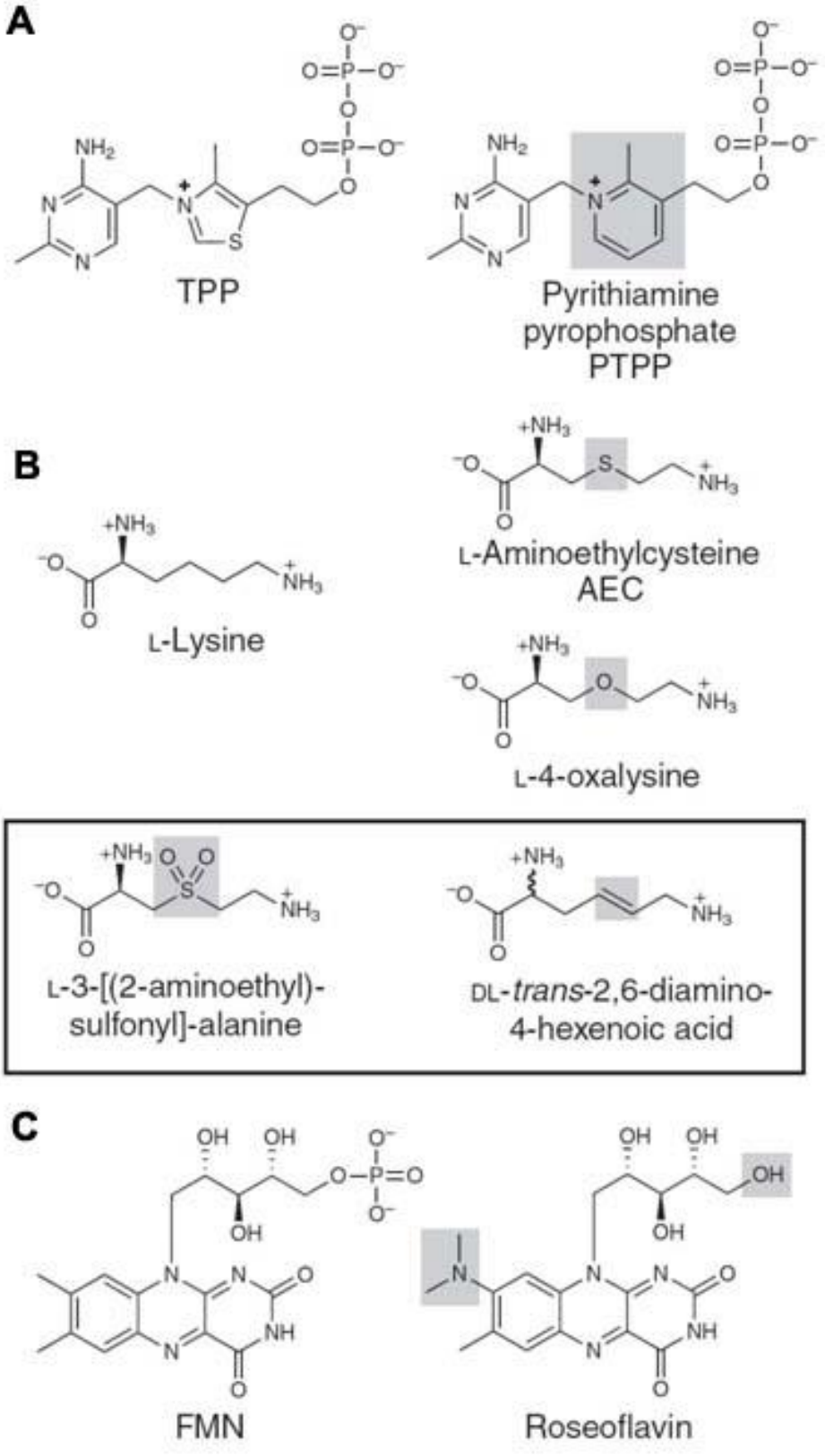
- Their specificity to the ligand is at least comparable to protein–ligand interactions, developing a riboswitch-specific inhibitor is thus a reasonable expectation (Schwalbe et al., 2007).
- Bacteria frequently use "tandem riboswitch architecture" to achieve more precise control of a single operon (Kwon and Strobel, 2008, Sudarsan et al., 2006, Welz and Breaker, 2007), which means more than one mutations are required to develop resistance to a inhibitor that targets that particular riboswitch pair. Recent studies have also shown that a few different classes of riboswitch can cooperatively regulate a single operon (Andre et al., 2008), opening up possibilities to target a single locus with multiple drugs. Conversely, same riboswitch were found to exist in multiple genome locations to control different operons (Welz and Breaker, 2007). These are all unique properties of riboswitch drug targeting that are not observed in traditional pharmacology that focuses on protein and protein complexes.

Recent advances in the field of RNA crystallography and biochemistry has identified historical as well and novel drugs that successfully target riboswitches, as demonstrated for thiamine pyrophosphate (TPP) (Figure 1.7A) (Sudarsan et al., 2005),lysine (Figure 1.7B) (Blount et al., 2007), flavin mononucleotide (FMN) (Figure 1.7C) (Lee et al., 2009) and guanine (Kim et al., 2009) responsive riboswitches. The most recent work on guanine riboswitch has discovered that a pyrimidine compound (PC1) binds guanine riboswitches and reduce of *Staphylococcus aureus* infection in mouse (Mulhbacher et al.).

Although developing new methods for antibiotics discovery can be extremely challenging, so far the progress in targeting bacterial riboswitches has shown many promises and potentials. The concept and feasibility of such approach has already been

Figure 1.7. The chemical structures of riboswitch-binding natural metabolites and antibacterial metabolite analogs.

(A) Compounds that bind TPP riboswitches. (B) Compounds that bind lysine riboswitches. (C) Compounds that bind FMN riboswitches. Shaded regions highlight the functional groups that differ from the natural metabolite. The inset in B shows identified antibacterial lysine analogs that bind to lysine riboswitches. (Adapted from Blount and Breaker, 2006)


A
TPP
Pyrithiamine
pyrophosphate
PTPP
B
L-Lysine
L-Aminoethylcysteine
AEC
L-4-oxalysine
L-3-[(2-aminoethyl)-
sulfonyl]-alanine
DL-trans-2,6-diamino-
4-hexenoic acid
C
FMN
Roseoflavin

proven, and it is just a matter of time before more attention is positioned on developing and screening compounds for drug activities. Essentially, targeting riboswitches require little to no change in the current industry because the underlying technology and principle of drug discovery is essentially the same. In summary, the potential of this novel antimicrobial target offers new hope to our ongoing battle against bacterial pathogens and well-being of the mankind.

CHAPTER 2

CRYSTAL STRUCTURES OF THE S_{MK} BOX (SAM-III) RIBOSWITCH REVEAL THE SAM-DEPENDENT TRANSLATION INHIBITION MECHANISM[1]

SUMMARY

Three distinct classes of SAM-responsive riboswitches have been identified that regulate bacterial gene expression at the levels of transcription attenuation or translation inhibition. The S_{MK} box (SAM-III) translational riboswitch has been identified in the *S*-adenosylmethionine (SAM) synthetase gene in members of the *Lactobacillales*. Here we report the 2.2 Å crystal structure of the *Enterococcus faecalis* S_{MK} box riboswitch. The Y-shaped riboswitch organizes its conserved nucleotides around a three-way junction for SAM recognition. The Shine-Dalgarno sequence, which is sequestered by base pairing with the anti-Shine-Dalgarno sequence in response to SAM binding, also directly participates in SAM recognition. The riboswitch makes extensive interactions with the adenosine and sulfonium moieties of SAM, but completely ignores the tail of the methionine moiety. A structural snapshot of the S_{MK} box riboswitch sampling the near-cognate ligand SAH was captured, in which SAH was found to adopt an alternative conformation and fails to make several key interactions.

---

[1]This work is published in Lu, C., A. M. Smith, et al. (2008). "Crystal structures of the SAM-III/S(MK) riboswitch reveal the SAM-dependent translation inhibition mechanism." Nature Structural & Molecular Biology **15**(10): 1076-83.The text and data was reused here with permission from the publisher.

INTRODUCTION

Recent discoveries of *cis*-acting regulatory RNAs termed riboswitches revealed the novel ability of RNA to directly relay environmental cues to the genetic regulation machinery (for reviews, see (Tucker and Breaker, 2005, Grundy and Henkin, 2006, Vitreschak et al., 2004) and references therein). Bacteria use riboswitches to regulate the metabolism and transport of vitamins (Winkler et al., 2002a, Wickiser et al., 2005, Serganov et al., 2006, Winkler et al., 2002b, Sudarsan et al., 2005, Thore et al., 2006), nucleotides (Christiansen et al., 1997, Mandal and Breaker, 2004a, Mandal et al., 2003, Roth et al., 2007, Kim et al., 2007, Meyer et al., 2008), amino acids (Mandal et al., 2004, Rodionov et al., 2003, Sudarsan et al., 2003b, Grundy et al., 2003), cofactors (Winkler et al., 2003, Wickiser et al., 2005, McDaniel et al., 2003, Epshtein et al., 2003, Corbino et al., 2005, Fuchs et al., 2006) and metal ions (Groisman et al., 2006, Coppins et al., 2007). Riboswitches are especially prevalent in Gram-positive bacteria, as exemplified by *Bacillus subtilis*, where over 4% of its genes are riboswitch-regulated (Winkler, 2005a). A handful of riboswitches were also found in eukaryotic systems, suggesting that such mechanisms may be more widespread than currently appreciated (Sudarsan et al., 2003a, Thore et al., 2006, Cheah et al., 2007).

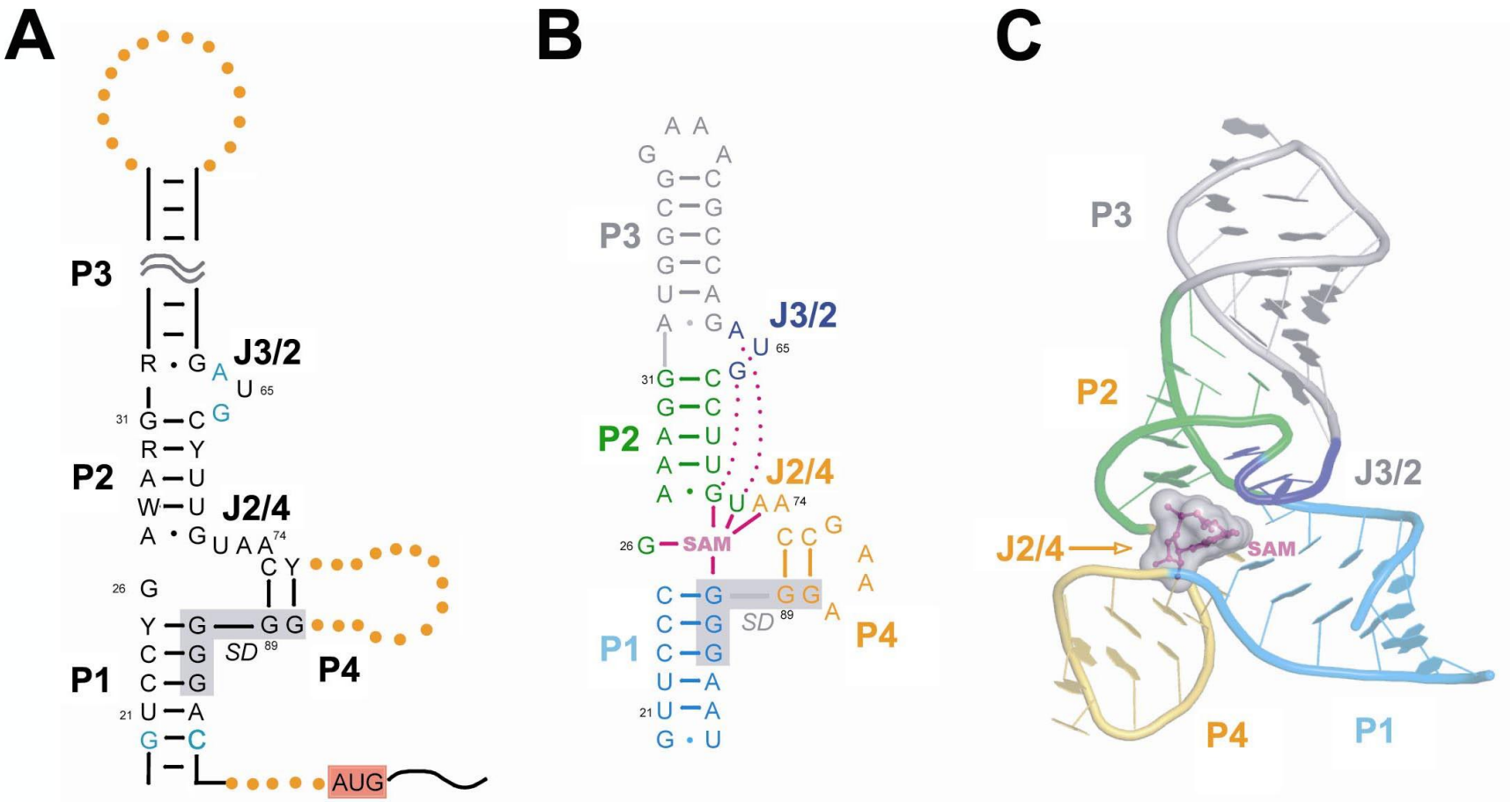
A typical metabolite-binding riboswitch consists of a ligand-sensing "aptamer domain" and an "output domain" that regulates gene expression, generally at the levels of transcription attenuation or translation initiation (Grundy and Henkin, 2006, Winkler et al., 2002a, Henkin and Yanofsky, 2002). Binding of the cognate metabolite to the aptamer domain usually causes conformational changes that result in alternative base pairing in the output domain, which in turn affects expression of the downstream open reading frames. The three classes of *S*-adenosylmethionine (SAM)-responsive

riboswitches, including the S box (SAM-I) (Grundy and Henkin, 1998, Winkler et al., 2003, McDaniel et al., 2003, Epshtein et al., 2003), SAM-II (Corbino et al., 2005), and S_{MK} box (SAM-III) (Fuchs et al., 2006), represent the most commonly found riboswitch families in bacterial genomes (Barrick and Breaker, 2007). All of these riboswitches are highly selective for SAM over its natural analog, *S*-adenosylhomocysteine (SAH), which is generated from utilization of SAM as a methyl donor in enzyme-catalyzed methyltransfer reactions and differs from SAM by the absence of a methyl group and a positive charge on the sulfur atom. Crystal structures of the S box (SAM-I) and SAM-II riboswitches (Montange and Batey, 2006, Gilbert et al., 2008) revealed two completely different RNA folds that utilize distinct mechanisms of SAM recognition.

The S_{MK} box riboswitch (Figure 2.1A) regulates the translation of SAM synthetase (*metK*) genes in lactic acid bacteria by sequestration of the Shine-Dalgarno (SD) sequence that is essential for loading of the 30S ribosomal subunit for translation initiation (Fuchs et al., 2006). Interestingly, the SD sequence within the S_{MK} box family (GGGGG) differs from the consensus SD sequence (GGAGG) at the central position (Figure 2.1A, 2.1B). The three Gs at the 3′ portion of the SD, in conjunction with the following two nucleotides, base pair with the ASD sequence in the presence of SAM (Fuchs et al., 2006) which hinders the binding of 30S ribosomal subunits to the mRNA (Fuchs et al., 2007). Mutational studies demonstrated that the S_{MK} box riboswitch differs from most metabolite binding riboswitches in that the SD-ASD pairing is required for SAM binding, indicating that the output domain (i.e., the SD-ASD pairing) is an intrinsic part of the ligand binding domain (Fuchs et al., 2006). Similar to other SAM riboswitches, the S_{MK} box riboswitch exhibits at least 100-fold preference for SAM over SAH (Fuchs et al., 2006, Fuchs et al., 2007).

Figure 2.1. Structure of the S_{MK} box riboswitch.

(A). Secondary structure of the S_{MK} box riboswitch based on phylogenetic analysis (Fuchs et al., 2006). Black capital letters, residues that are 100% conserved; blue, 50-85% conserved; R, G or A; W, A or U; Y, C or U. Thick dotted lines, hypervariable loops; solid ladders, conserved secondary structure; gray and pink shaded areas, SD sequence and AUG start codon, respectively. (B). Secondary structure of the S_{MK}6 riboswitch RNA based on the crystal structure. Helices P1 through P4 are colored in cyan, green, silver, and yellow, respectively. Grey shade: SD sequence; Solid magenta lines: direct contacts between the RNA and the SAM molecule; dashed magenta lines: tertiary interactions between J3/2 and P2 and J2/4. Numbering is consistent with previous studies (Fuchs et al., 2006, Fuchs et al., 2007). (C). Cartoon representation of the crystal structure of the S_{MK} riboswitch. SAM is shown in overlapping CPK and surface representations in magenta and silver, respectively. The coloring scheme for the RNA is consistent in B and C.


A
P3
J3/2
P2
J2/4
P1
P4
SD
AUG
B
P3
J3/2
P2
J2/4
SAM
P1
P4
SD
C
P3
P2
J3/2
J2/4
SAM
P4
P1

To elucidate the SAM-induced translation inhibition mechanism in the S_{MK} box riboswitch, we determined crystal structures of the *E. faecalis* S_{MK} box riboswitch, including its SD sequence, in the presence of SAM, selenium-derivitized SAM (Se-SAM), or SAH. The RNA was found to organize into a "Y" shaped molecule with SAM intercalated into the three-way helical junction (Figure 2.1C). In agreement with previous genetic and enzymatic probing analyses (Fuchs et al., 2007, Fuchs et al., 2006), the SD sequence contributes to both binding site formation and specific SAM recognition. The near-cognate ligand SAH makes only weak and nonspecific interactions with the riboswitch.

MATERIALS AND METHODS

RNA crystallization and structure determination.

We inserted the *E. faecalis* S_{MK} box riboswitch sequence into a pUC19 plasmid under the control of a T7 RNA polymerase (RNAP) promoter, with the hepatitis delta virus ribozyme placed at the 3′ end of the S_{MK} box sequence to ensure a homogeneous 3′-end. The crystallization construct initiates 1 nt upstream of the ASD sequence and ends 3 nucleotides (nt) downstream of the SD sequence. In the process of searching for better diffracting RNA crystals, the hypervariable P3 stem was replaced with a five bp GC-rich stem capped with a GAAA tetraloop, and the hypervariable linker connecting the SD sequence to the upstream helical domain was systematically shortened. Nine different constructs were generated by combining variations at two hypervariable regions (P3 and P4) (Fuchs et al., 2006). RNA was prepared as previously described (Ke and Doudna, 2004). Se-SAM was a gift from Dr. Booker Squires at the Pennsylvania State University. Its synthesis has been described previously (Iwig and Booker, 2004). SAM and SAH were purchased from Sigma,

dissolved in water and DMSO, respectively, as 100 mM stock solutions (pH < 7) and stored at -80 ℃ until immediately before use.

To ensure conformational homogeneity, the S_{MK} box RNA was heat-refolded as described (Fuchs et al., 2006, Fuchs et al., 2007). The high-resolution S_{MK}6 crystals grew as extremely slim needle crystals of 15x15 μm^2 in cross-section from a solution containing 40 mM sodium cacodylate, pH 7.0, 80 mM strontium chloride, 15% (w/v) 2-methylpentane-2,4-diol (MPD), and 2 mM spermine-HCl. The MPD content was raised step-wise to 25% before the crystals were flash-frozen in liquid nitrogen. Diffraction data were collected using the microcrystallography setup at beamlines APS 24-ID-E and MACCHESS F1 and processed using HKL2000 (Otwinowski and Minor, 1997).

S_{MK} box RNA structure refinement

While most S_{MK} crystals belong to the $I4_122$ space group [one molecule per asymmetric unit (ASU)], a few fall into the $P4_12_12$ space group with similar unit cell dimensions (two molecules per ASU). A 3.8 Å single wavelength anomalous diffraction data set was collected from a $P4_12_12$ crystal soaked in 80 mM iridium hexaammine. The initial phase calculated using program SHELXD (Sheldrick, 2008) from eight iridium hexaammine sites was improved by applying a 2-fold noncrystallographic symmetry (NCS) to average the two molecules in ASU. After solvent-flattening and phase extension to a 2.6 Å native data set, the experimental density map allowed unambiguous model building of 23 nucleotides using program COOT (Emsley and Cowtan, 2004). 20 nucleotides were missing in the initial model, including the entire P3, J3/2, and part of P2. To locate the P3 helix, an ideal RNA model containing a 5-base pair A-form duplex capped by a GAAA tetraloop was

included in the molecular replacement search using program PHASER (Storoni et al., 2004). The correct solution was identified from the top rotation and translation solutions after manual examination of the crystal packing environment. The rest of the missing nucleotides were added one or two bases at a time followed by restrained positional refinement using Refmac5 (Collaborative Computational Project, 1994). The SAM molecule was then real-space fitted into the difference map using COOT (Emsley and Cowtan, 2004). The completed structural model was refined against a 2.2 Å $I4_122$ native data set in iterative cycles of torsion-angle simulated annealing, group/individual B-factor refinement, and Powell energy minimization in CNS (Brunger et al., 1998, Brunger, 2007). The initial model was built using COOT (Emsley and Cowtan, 2004) from experimental phases calculated from SHELXD (Sheldrick, 2008) and refined using Refmac5 (Collaborative Computational Project, 1994) and CNS (Brunger et al., 1998, Brunger, 2007). The final model includes all nucleotides and a total of fifteen strontium ions (Table 2.1) and sixty-nine water molecules. A final round of TLS refinement in Refmac5 was carried out to model domain movements within the RNA (Collaborative Computational Project, 1994). The methionine tail of SAM and the pyrimidine ring of U65 are the only portion of the model without corresponding electron densities. Final R_{work} and R_{free} factors are 22.1% and 22.7%, respectively (Table 2.2). The same set of excluded reflections was used in free R factor calculation between refinement programs.

Structure determination Se-SAM and SAH bound S_{MK} box RNA

A Se-SAM soaking experiment was used to confirm the sulfonium position in the S_{MK} box riboswitch. To introduce Se-SAM and titrate away bound SAM, S_{MK} crystals were exposed to Se-SAM (10 mM final concentration, freshly prepared and confirmed by

Table 2.1. Analysis of Sr^{2+}-RNA interactions in the S_{MK} box riboswitch.

Divalent cation ID	Interacting base number	RNA ligand*	Metal-ligand Distance (Å)
Sr201	A27	O1P	4.7
Sr202	G91	O6	4.5
	G92	O6	4.3
Sr203	G30	O6	4.6
	G31	O6	4.3
Sr204	G88	O6	2.9
		N7	2.9
Sr205	A86	O1P	4.5
	G76	N7	4.4
Sr206	G47	O1P	3.5
	U45	O6	2.8
Sr207	C25	O1P	2.6
	A64	O2P	4.4
Sr208	G71	O6	2.8
	G66	O2P	4.4

Table 2.1 (continued)

	U69	O2P	3.0
	C68	O2P	4.6
Sr209	A73	O1P	2.8
	A27	N3	4.6
		O2'	4.4
Sr210	G71	O1P	4.7
	G88	O2P	4.7
	U72	O1P	4.8
Sr211	G47	O2P	4.7
	U45	N7	2.9
	A44	N7	4.4
Sr212	U69	O2'	3.7
	U70	O4'	4.0
Sr213	U33	O4	2.5
	U34	N6	3.0
	A61	N4	4.2

Table 2.1 (continued)

	U35	O6	4.2
	A60	N4	4.3
Sr214	G66	O6	4.4
	A27	O2P	4.4
	A28	N7	4.0
		O1P	4.5
Sr215	A73	O2P	2.4
	A74	O2P	3.3
		N7	4.6

* Inner-sphere Sr^{2+}-RNA interactions should be around 2.5 Å. In most of identified Sr^{2+} ions, the hydration shell is not resolved in the 2.2 Å electron density map. Thus, we list all RNA functional groups within 4.8 Å of Sr^{2+} ions as potential outer-sphere ligands.

Table 2.2. Crystallographic statistics

	S_{MK} native bound to SAM #1	S_{MK} native bound to SAM #2	S_{MK} Ir derivative	S_{MK} bound to Se-SAM	S_{MK} bound to SAH
Data collection					
Space group	$I4_122$	$P4_12_12$	$P4_12_12$	$I4_122$	$I4_122$
Cell dimensions					
a, *b*, *c* (Å)	97.806, 97.806, 87.138	98.131, 98.131, 86.262	98.180, 98.180, 86.790	96.776, 96.776, 86.765	98.148, 98.148, 86.393
α, β,γ ( )	α=β=γ=90°	α=β=γ=90°	α=β=γ=90°	α=β=γ=90°	α=β=γ=90°
Resolution (Å)	20 (2.2)	20 (2.6)	20 (3.3)	20 (2.7)	20 (2.9)
R_{sym} or R_{merge}	7.1 (30.3)	11.3 (40.1)	13.3 (42.2)	7.6 (28.4)	7.9 (24.6)
I / δ(I)	29.4 (1.2)	12.3 (2.59)	13.2 (5.8)	15.5 (1.7)	80.8 (11.4)
Completeness (%)	95.1 (75.4)	90.4 (85.0)	100 (100)	93.7 (74.0)	100 (100)
Redundancy	9.4 (2.3)	6.3 (4.8)	13.7 (13.6)	3.9 (2.8)	9.3 (8.0)
Refinement					
Resolution (Å)	2.2			2.7	2.9

Table 2.2 (continued)

No. reflections	9147	4980	4694
R_{work} / R_{free}	22.1 / 22.7	22.7 / 25.5	22.2 / 25.9
No. atoms			
RNA	1155	1155	1155
Ligand/ion	42	43	31
Water	69	11	5
B-factors	53	57	56
R.m.s. deviations			
Bond lengths (Å)	0.008	0.005	0.006
Bond angles (°)	1.61	1.29	1.43

*Values in parentheses are for highest-resolution shell.

mass spectrometry to be intact) during the entire cryoprotection procedure over the course of 2 hr. The crystals were first incubated and then sequentially transferred into three fresh Se-SAM solutions with increasing cryoprotectant concentration and incubated for at least 20 min between each transfer before snap-freezing in liquid nitrogen. 2.7 Å diffraction data sets were collected on the APS ID-24E microdiffraction beamline at the K absorption edge of Se. The native $S_{MK}6$ structure (excluding SAM, metal ions, and water coordinates) was used as the starting model for rigid body and B-factor refinement in Refmac5 (Collaborative Computational Project, 1994), followed by simulated annealing refinement in CNS (Brunger, 2007) to remove model bias. The structural model was further refined in CNS (Brunger, 2007) as described for the SAM-bound structure, and the Se-SAM model was introduced at the later stage of the refinement. The selenium atom in Se-SAM was located in the sigmaA-weighted anomalous difference Fourier map, where the only substantial peak overlaps with the selenium/sulfur position in the structure model. The final Se-SAM-bound $S_{MK}6$ structure contains fifteen strontium ions and eleven waters with R_{work}/R_{free} of 22.7%/25.5% (Table 2.2).

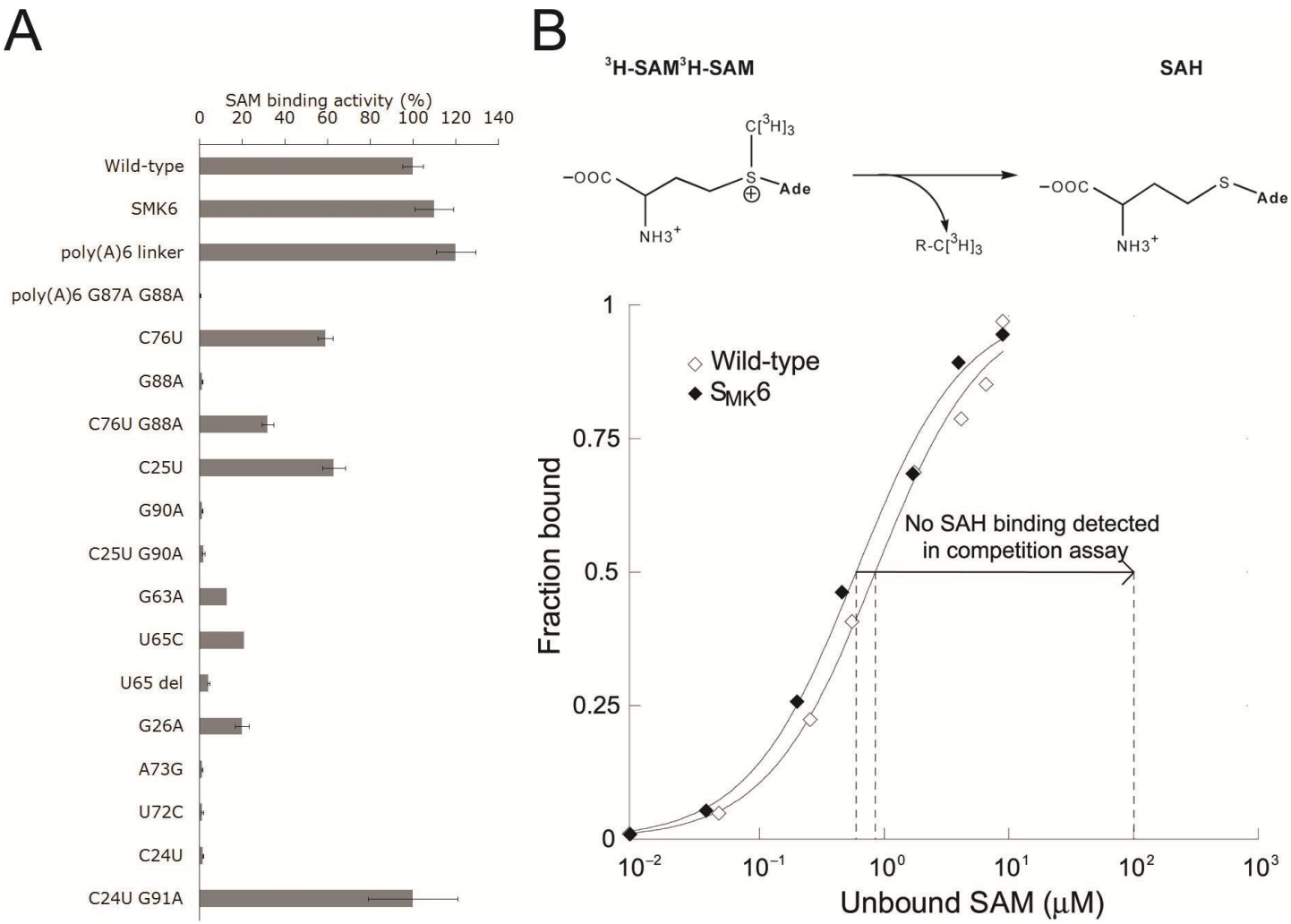
To examine the binding of SAH, the $S_{MK}6$ crystal was repeatedly transferred into solutions containing a saturating amount of SAH (~2 mM) over the course of three days. The structure was solved using a molecular replacement method from a 2.9 Å data set collected at the microdiffraction beamline 24ID-E at APS. The refinement procedure was essentially the same as described for the Se-SAM structure. Final R_{work}/R_{free} is 22.2%/25.9% (Table 2.2). We verified both structures using simulated annealing omit maps by excluding five bases in each calculation.

In vitro SAM-binding assays

The SAM-binding assays were performed as previously described (Fuchs et al., 2006, Fuchs et al., 2007) with minor modifications. DNA templates corresponding to positions 15-118 (relative to the predicted transcription start site) of the *E. faecalis metK* leader were constructed by ligating overlapping pairs of complementary oligonucleotides, including a T7 RNAP promoter sequence, as previously described (Yousef et al., 2003, McDaniel et al., 2005). Ligated products were amplified by PCR and RNAs were synthesized by T7 RNAP transcription using an AmpliScribe T7 High Yield Transcription Kit (Epicentre Biotechnologies). RNAs (3 μM) in 1x transcription buffer (Grundy et al., 2002a) were heated to 65 ℃ for 5 min and slow cooled to 40 ℃ followed by addition of radiolabeled SAM (3 μM [methyl-^{3}H]-SAM, 15Ci/mmol; GE Healthcare) in a total reaction volume of 40 μl. Binding reactions were incubated at room temperature for 15 min followed by passage through a Nanosep 10K Omega filter (Pall Life Sciences) by centrifugation at 14,000 x g for 2.5 min. Filters were washed four times with 40 μl 1x transcription buffer to remove unbound SAM and material retained by the filter was collected, mixed with Packard BioScience Ultima Gold scintillation fluid and counted in a Packard Tri-Carb 2100TR liquid scintillation counter. No SAM was retained by the filter in the absence of RNA, therefore the background for nonspecific binding is negligible. All binding assays were carried out in triplicate and the standard deviations are shown in Figure 2.2.

Figure 2.2. Effects of S_{MK} box riboswitch mutagenesis on SAM binding and apparent SAM binding constant (K_d) determination.

(A). Effects of mutation (with standard deviations) on SAM binding by size-exclusion filtration assay as described (Fuchs et al., 2006, Fuchs et al., 2007). (B). SAM and SAH K_d determination by competition binding assay. Binding curves are shown for the full-length *E. faecalis metK* RNA (open diamonds, K_d = 0.85 M) and the S_{MK}6 crystallization construct (filled diamonds, K_d = 0.57 M). Unlabeled SAH showed no competition at a >100-fold excess of SAH over SAM.


A
SAM binding activity (%)
0 20 40 60 80 100 120 140
Wild-type
SMK6
poly(A)6 linker
poly(A)6 G87A G88A
C76U
G88A
C76U G88A
C25U
G90A
C25U G90A
G63A
U65C
U65 del
G26A
A73G
U72C
C24U
C24U G91A
B
3H-SAM3H-SAM
SAH
C[3H]3
-OOC
S
Ade
NH3+
R-C[3H]3
Wild-type
SMK6
Fraction bound
No SAH binding detected
in competition assay
Unbound SAM (μM)

Apparent K_d determination

The apparent equilibrium dissociation constants (K_d values) for the wild-type 15-118 *metK* leader sequence and the 53-nucleotide crystallization construct were determined using a modified SAM-binding assay (Tomsic et al., 2007). Briefly, T7 RNAP transcribed RNA (1 μM) in 1x transcription buffer (Grundy et al., 2002a) was refolded as described above and incubated with [^{3}H]-SAM ranging in concentration from 0.02 to 10 μM. Binding reactions were loaded onto Nanosep 3K Omega filters (Pall Life Sciences) and centrifuged briefly at 14,000 x g to avoid large volume changes. The concentration of the RNA-[^{3}H]-SAM complex retained by the filter and the unbound SAM present in the flow through were determined by scintillation counting of known volumes and nonlinear regression analyses were performed using KaleidaGraph Version 3.51 (Synergy Software). Apparent K_d values represent the averages of at least two independent experiments for each construct with a margin of error ≤5%.

In vivo lacZ reporter gene assay of S_{MK} box fusions

In vivo lacZ reporter gene assays were performed as described previously (Fuchs et al., 2006, Fuchs et al., 2007). *E. faecalis metK* leader region constructs, which included the first 15 nt of the coding region, were positioned downstream of the highly expressed *B. subtilis glyQS* promoter. The resulting DNA fragment was inserted into a *lacZ* fusion vector (pFG328) (Grundy and Henkin, 1993) to generate an in-frame *metK-lacZ* translational fusion in which the first 5 codons of *metK* were fused to codon 18 of *lacZ*. The constructs were introduced into the chromosome of *B. subtilis* strain BR151 (*metB10 lys-3 trpC2*) by integration into an SPβ prophage. *B. subtilis* strains containing the *lacZ* fusions were grown at 37 °C in Spizizen minimal

medium (Anagnostopoulos and Spizizen, 1961) containing methionine (50 μg ml^{-1}) until early exponential growth. The cells were harvested by centrifugation at 8,000 x g and resuspended in Spizizen medium either with or without methionine, and samples were collected at 1-h intervals and assayed for β-galactosidase activity after toluene permeabilization of the cells (Miller, 1972)**.** All assays were carried out in triplicate and the standard deviations are shown in Table 2.3.

RESULTS

Structure determination

We determined the crystal structure of the *E. faecalis* S_{MK} box riboswitch (Fuchs et al., 2006) to elucidate the SAM-dependent translation inhibition mechanism. The optimized version of the S_{MK} box riboswitch (Figure 2.1B), designated S_{MK}6, allowed structure determination at 2.2 Å and displayed SAM binding activity similar to that of the full-length wild-type construct (Figure 2.2 A,B). The apparent K_d values for SAM determined by size-exclusion filtration were estimated to be 0.85 μM for the full-length transcript (corresponding to positions 15-118 relative to the predicted *E. faecalis metK* transcription start site) (Fuchs et al., 2006) and 0.57 μM for the 53-nt crystallization construct (Figure 2.2B). The K_d value for the full-length S_{MK} transcript was further confirmed by fluorescence quenching assays (Figure 2.3). In a competition assay, the S_{MK} box riboswitch displays at least 100-fold preference for SAM over its near-cognate ligand SAH (Fuchs et al., 2006) (Figure 2.2B).

Table 2.3. Expression of *E. faecalis metK-lac*Z fusions

	-Met	+Met	Ratio[b]
WT	120 +/- 19[a]	24 +/- 5.5	5.0
C24G	110 +/- 14	95 +/- 6.1	1.2
C24U	310 +/- 0.71	200 +/- 22	1.6
C24U+G91A	23 +/- 2.9	15 +/- 2.9	1.5
U22G	220 +/- 27	160 +/- 13	1.4
A93C	77 +/- 8.7	42 +/- 1.0	1.8
U22G+A93C	11 +/- 3.7	2.5 +/- 0.72	4.4
C25U	225 +/- 25	80 +/- 5.0	2.8
C25U+G90A	78 +/- 19	82 +/- 16	0.95

a. β-Galactosidase activities are expressed in Miller units (Miller, 1972).

b. Ratio of expression during growth in the absence of methionine to expression during growth in the presence of methionine.

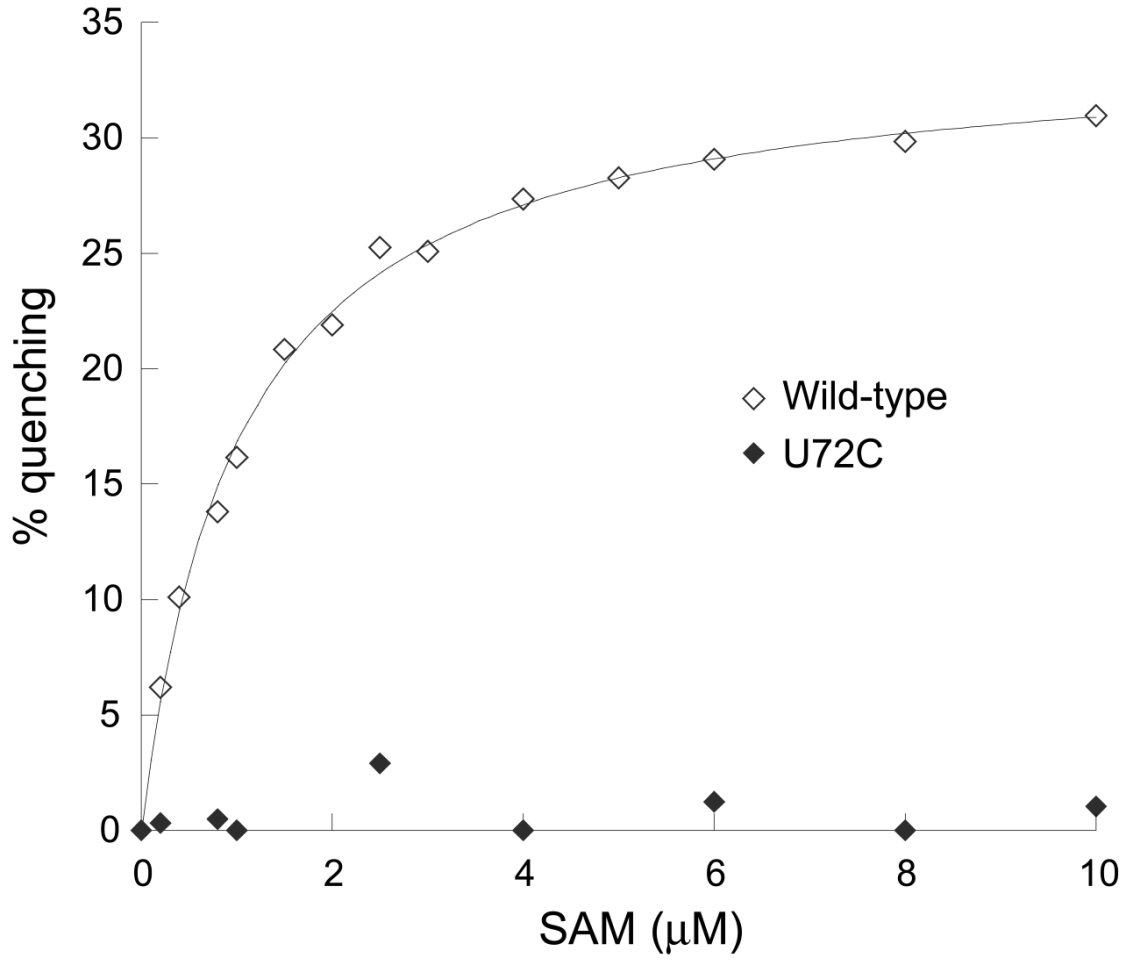


Figure 2.3. 2-AP fluorescence quenching in response to SAM.

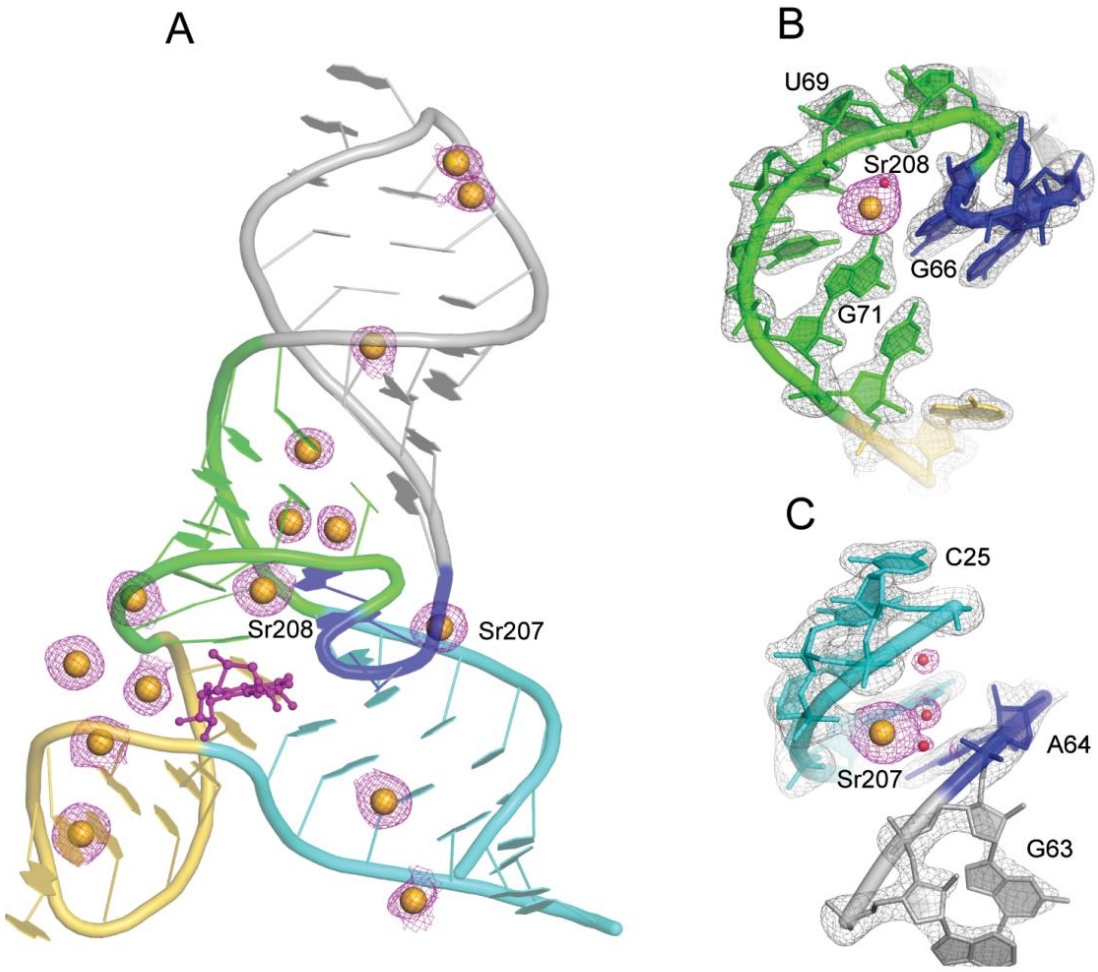
Increasing concentrations of SAM were added to bipartite metK leader RNA containing an internal 2-AP modification at residue A29. 2-AP fluorescence at steady-state was measured at 375 nm with an excitation wavelength of 310 nm. Increasing SAM resulted in quenching of 2-AP fluorescence for the wild-type RNA (open diamonds), but no substantial change in fluorescence was observed with a U72C mutant RNA (closed diamonds) that is predicted to bind SAM very poorly. Percent quenching ($\Delta F/F_0$ x 100) represents the change in fluorescence normalized to the initial fluorescence observed in the absence of SAM. Data were analyzed by nonlinear regression analysis to determine an apparent K_d of 1.0 μM for the wild-type RNA.

Overall structure of the S_{MK} box riboswitch

The 53-nt S_{MK} box riboswitch RNA folds into an inverted "Y" shaped molecule, where helices P1 (SD-ASD helix) and P4 (linker helix) constitute the two short arms, and P3 (top helix) stacks on top of P2 (middle helix) to give rise to the long arm (Figure 2.1C). Most of the secondary structure features match the predictions from the phylogenetic analysis and RNase T1, V1, A, and H probing experiments (Fuchs et al., 2006), including the SAM-dependent formation of the SD-ASD helix and the protection of nucleotides including U69, U70, C75, and G88, although the P4 linker helix was not previously assigned. The quintuple-G SD sequence spans the SD-ASD (P1) and linker (P4) helices (Figure 2.1B), contributing to the overall folding of the RNA as well as the SAM binding site. Among the 23 residues in the S_{MK} box riboswitch that are 100% conserved (Figure 2.1A), 10 participate in the formation of a pocket inside the three-way junction, where the adenosine moiety of SAM intercalates to allow continuous base stacking from P1 to P2. A lid-like structure formed due to a double strand reversal in the J3/2 bulge partially encloses the SAM binding site, and further stabilizes the three-way junction through base triple interactions that enlarge the major groove of P2. A total of 15 divalent metal ions were identified bound to the S_{MK} RNA to shield the unfavorable electrostatic interactions between sugar phosphate backbones as the result of tertiary RNA folding (Table 2.1, Figure. 2.4). The P4 linker helix was further probed by poly(A) scanning mutations. Replacement of the 11-nt hypervariable *E. faecalis* linker region with a poly$(A)_6$ sequence did not affect SAM binding activity; disrupting the linker helix by further mutating G88 and G89 to adenosines, however, dramatically reduced SAM binding (Figure 2.2A), consistent with the participation of these residues in the P4 linker helix. Additional point mutations were generated to confirm the formation of the G88-C76 base pair (Figure 2.2A). Introduction of a C76U wobble mutation in the *E. faecalis* S_{MK} box, which

Figure 2.4. Metal binding sites in the S_{MK} box RNA.

(A). Overall view of the fifteen strontium binding sites on the S_{MK} box RNA. (B, C). Electron density map (grey mesh) showing two strontium ions (Sr208 and Sr207) bound to their respective sites on the S_{MK} box RNA. The magenta mesh shows the simulated composite omit electron density map of strontium contoured at 1 σ. Yellow ball, strontium ion; red ball, water. The coloring scheme for the RNA is consistent with that in Figure 2.1.


A
Sr208
Sr207
B
U69
Sr208
G66
G71
C
C25
A64
Sr207
G63

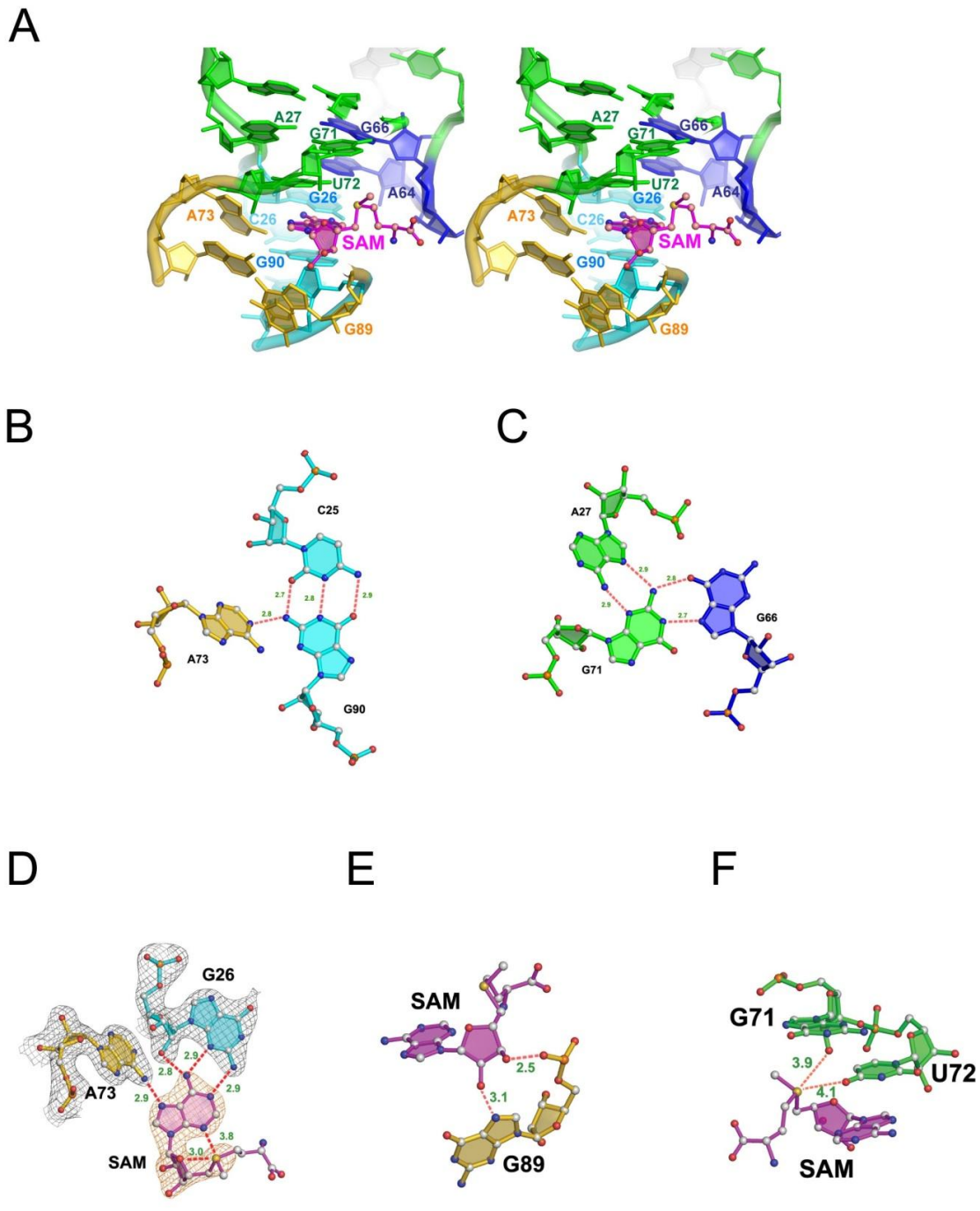
occurs naturally in some S_{MK} box family members when a third 77-87 base pair is maintained (Fuchs et al., 2006), resulted in retention of 59% of the SAM binding activity. By contrast, the G88A mismatch mutation that would severely destabilize the linker helix resulted in a 75-fold reduction in SAM binding activity. Combination of the C76U and G88A mutations, which replaces the G-C pair with a weaker A-U pair, restored the binding to 32% of the wild type level, which is consistent with the sequence alignment results that suggest that an additional base pair is needed in this context to maintain the stability of the linker helix (Fuchs et al., 2006).

Binding pocket formation

The floor of the SAM-binding pocket (Figure 2.5A) is defined by a critical base triple interaction (A73:G90-C25), where N1 of A73 approaches from the minor groove side to accept a hydrogen bond from N2 of G90 (Figure 2.5B). This base triple serves two purposes: first, it ties J2/4 to the P1 SD-ASD helix in conjunction with another interaction between the N1 of A74 and the 2′-hydroxyl of G90; furthermore, it orients the N6 amine of A73 for SAM recognition. Base pairing at C25-G90 is crucial, as a C25·A90 mismatch (G90A mutation) reduced the SAM-binding activity to 1.3% of the wild type, whereas a U25-G90 wobble pair (C25U mutation) exhibited 63% of the SAM binding activity (Figure 2.2A). The importance of the A73:G90-C25 base triple is underlined by the strong, deleterious effect of the C25U/G90A double mutation, which converts the SD element from the quintuple-G sequence found in all S_{MK} box elements to the more commonly found GGAGG sequence, while maintaining the SD-ASD pairing. The loss of SAM-binding activity in this mutant is apparently due to the loss of N2 when G90 is converted to A, which disrupts the A73:G90 side of the base

Figure 2.5. SAM binding pocket and important interactions.

The labeling and base coloring scheme is consistent with that in Figure 2.1. (A). Stereo view of the SAM binding site in the S_{MK} box riboswitch. The adenosine moiety of SAM is shown to base-stack between U72 and G90. (B). A73:G90-C35 base triple paves the "floor" of the SAM binding pocket. The C-G base pair is co-planar, whereas A73 contacts from the minor groove of G90 at a 45 ºtilted angle, which orients N6 of A73 for SAM recognition one base plane above. (C). A27·G71·G66 base triple defines the "ceiling" of the binding pocket where the sheared A27·G71 base pair in P2 is contacted at the major groove side by G66 of the J3/2 bulge. (D). Recognition of the adenosine base of the SAM overlapped with the 2.2 Å experimental electron density contoured at 1.5 σ. Atoms beyond sulfur in SAM do not have corresponding electron density, indicating disorder. The adenine of SAM is extensively recognized through hydrogen bond interactions from A73 and G26. SAM has two intramolecular electrostatic interactions from the sulfonium ion to the O4′ and N3 of SAM. Grey mesh, RNA density; orange mesh, SAM density. (E). The 3′-hydroxyl of the ribose of SAM is recognized by a hydrogen bond to the phosphoryl oxygen of G89, whereas the 2′-OH makes a rare π-hydrogen bond interaction with the N7 of G89. (F). The positively charged sulfonium ion of SAM makes favorable electrostatic interactions with O4 of U72 and 2′-hydroxyl of G71. Distances are given in angstroms. Carbon, oxygen, nitrogen, sulfur, and phosphorus atoms are colored grey, red, blue, yellow, and orange, respectively.


A
A27
G71
G66
U72
G26
A64
A73
C26
SAM
G90
G89
B
C25
A73
G90
C
A27
G66
G71
D
G26
A73
SAM
E
SAM
G89
F
G71
U72
SAM

triple interaction. Thus, the structural and mutational data collectively explain the absolute conservation of an unusual SD sequence in all S_{MK} box riboswitch RNAs.

The ceiling of the SAM binding site is formed from two layers of nonstandard base pairs, both of which involve bases from the J3/2 tri-nucleotide bulge that are highly conserved among the E. faecalis subfamily of the SMK box RNAs (Fuchs et al., 2006). A sheared U72•A64 pair sits immediately on top of the SAM intercalation base plane, which exposes the O4 of U72 for SAM recognition. Further above is the A27•G71•G66 base triple, where G71 mediates extensive hydrogen bond contacts to the Hoogsteen face of both A27 and G66, weaving P2 together with J3/2 (Figure 2.5C). U65 in the J3/2 bulge mediates a U-turn motif and extends toward the opening of the SAM binding pocket; this residue was shown to be more prone to RNase T1 digestion in the presence of SAM (Fuchs et al., 2006), consistent with the proposal that SAM binding reorients this residue. The backbone of the bulge resembles a "lid" that secludes SAM inside the three-way junction, and contributes to the overall electronegative environment inside the pocket that may help attract the SAM molecule. The importance of the J3/2 lid structure is indicated by the U65C and U65 deletion mutations which cause a 5- and 20-fold decrease in SAM binding activity, respectively (Figure 2.2A).

SAM recognition

The universally conserved G26 is left unpaired inside the cavity of the three-way junction. SAM intercalates its adenosine moiety between the P1 and P2 helices from the major groove side, stabilizing the three-way junction through π-stacking interactions (Figure 2.5D). The adenosine moiety of SAM adopts an energetically unfavorable syn-conformation, and presents its Watson-Crick face to form three minor

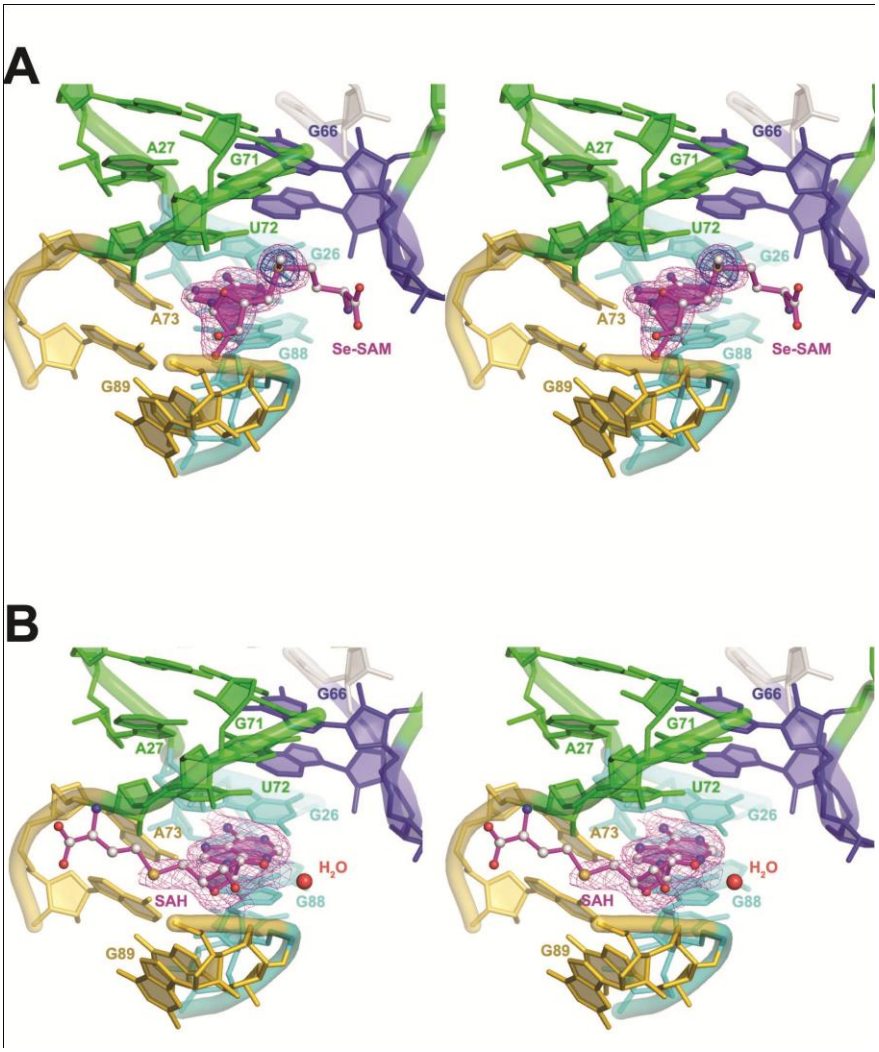
groove contacts (N1-amino, amino-N3, and amino-2′-OH) with the unpaired G26. Even a minor perturbation of the G26-SAM interaction, such as occurs in the G26A mutant that disrupts the N1-amino hydrogen bond, reduces SAM binding by 5-fold (Figure 2.2A). The N7 at the Hoogsteen face accepts a hydrogen bond from the N6 amine of the tilted A73 (Figure 2.5D). Disruption of this hydrogen bond by the A73G mutation caused an 80-fold reduction in SAM binding activity (Figure 2.2A), although we cannot completely separate this effect from the disruption of the A73:G90-C25 base triple. The ribose moiety of SAM adopts a 2′-endo conformation to avoid steric clashes with the base plane of G89 in the SD sequence (Stolowitz and Minch, 1981). The 2′ and 3′-hydroxyl of SAM are recognized by hydrogen bonds to the N7 (π-hydrogen bond) and 2′-hydroxyl of G89, respectively (Figure 2.5E). The 2′-hydroxyl of SAM donates an additional water-mediated hydrogen bond to the N6 of A74.

The positive charge on the sulfonium ion of SAM is recognized through favorable electrostatic interactions with the O4 carbonyl of U72 and the 2′-hydroxyl of G71 (Figure 2.5F). The positive charge is further stabilized through intramolecular electrostatic interactions with the O4′ on the ribose and N3 on the base of SAM (Figure 2.5D). Consistent with the structural observations, the U72C mutation that places a partially positive N4 amine towards the sulfonium decreases the SAM binding activity by 70-fold (Figure 2.2A). A similar charge stabilization scheme involving contacts between the sulfonium ion of SAM and O4 of uracil is used by the other two classes of SAM binding riboswitches to selectively bind the SAM molecule (Montange and Batey, 2006, Gilbert et al., 2008). The neutral, hydrophobic sulfide in SAH is not expected to make these electrostatic interactions with the RNA.

In contrast to the well-defined adenosine moiety, no electron density was observed for functional groups beyond the sulfonium ion in SAM (Figure 2.6A). This

Figure 2.6. Binding of Se-SAM and SAH to the S_{MK} box riboswitch.

(A). Stereo view of the binding pocket in the S_{MK} box riboswitch in complex with Se-SAM. The location of the selenium atom is confirmed by the strong anomalous difference density shown in blue contoured at 8 σ. The rest of the binding pocket in the Se-SAM-bound S_{MK} structure is almost identical to that in the SAM-bound structure. Magenta mesh signifies the simulated composite omit electron density map contoured at 1.5 σ of Se-SAM. (B). Stereo view of the SAH-bound S_{MK} structure from a direction similar to that in A. The simulated annealing omit map contoured at 0.8 σ level clearly shows that the ribose and sulfide moieties rotate 180° to exit the RNA from the linker helix side.


A
G66
A27
G71
U72
G26
A73
G88
Se-SAM
G89
G66
A27
G71
U72
G26
A73
G88
Se-SAM
G89
B
G66
G71
A27
U72
G26
A73
H2O
SAH
G88
G89
G66
G71
A27
U72
G26
A73
H2O
SAH
G88
G89

is a strong indication that the methionine tail (including the main chain and most of the side chain atoms beyond the sulfur) is not specifically recognized by the S_{MK} box riboswitch, which is in sharp contrast to the extensive recognition of the methionine tail observed in the structures of the S box and SAM-II riboswitches (Montange and Batey, 2006, Gilbert et al., 2008). Lack of recognition towards the methionine tail is further supported by binding studies with SAM analogs, where SAM analogs lacking the methionine main chain atoms bind the S_{MK} box riboswitch with affinities similar to that of SAM (A.M.S., F.J. Grundy and T. M. H., manuscript in preparation). The methyl group on the sulfonium ion is not specified by the S_{MK} box riboswitch (Figure 2.5D), since replacing it with an ethyl group does not impair ligand-RNA interactions (A.M.S., F.J. Grundy and T. M. H., manuscript in preparation).

Regulation of E. faecalis metK-lacZ fusions in vivo

The *E. faecalis* metK S_{MK} box was previously shown to confer translational regulation of a *lacZ* reporter gene in which the translation initiation region is replaced by that of the *metK* gene (Fuchs et al., 2006). Integration of the fusion construct into the B. subtilis chromosome results in high expression when cells are grown under conditions where intracellular SAM pools are low, and a 5-fold reduction in expression when cells are grown in the presence of high methionine, conditions that result in high SAM pools (Tomsic et al., 2007). The S_{MK} box riboswitch structure supports the model that SAM-dependent SD-ASD helix formation is responsible for translational inhibition. To verify this hypothesis, we evaluated the functional consequences of three sets of mutations that perturb the stability of the SD-ASD helix. As previously reported (Fuchs et al., 2006), a C24G mutation in the ASD region that is predicted to disrupt the SD-ASD interaction causes loss of repression in vivo (Table

2.3) and loss of SAM binding (Figure 2.2A). A C24U mutation had a similar effect (Table 2.3) although expression was significantly higher and a small response to SAM was retained, presumably because of the maintenance of an U24-G91 wobble pairing. A C24U/G91A doublemutation, which restores pairing at this position, restored SAM binding (Figure 2.2A), and resulted in low expression but only partial restoration of repression by SAM (Table 2.3). The low expression is likely due to disruption of the SD by the G91A mutation, which causes reduced affinity for 30S ribosomal subunits, and therefore obscures the regulatory response in vivo. A similar pattern was observed with the U22G and A93C mutations (Table 2.3), which affect the ASD and SD sequences, respectively, and cause loss of SAM binding in vitro (Fuchs et al., 2007). The U22G/A93C double mutant restores the pairing at this position, and restores SAM binding (Fuchs et al., 2007). This mutant also exhibited SAM-dependent repression in vivo, although expression was reduced relative to that of the wild-type construct (Table 2.2). This reduction in expression may be due to an enhanced SD-ASD interaction from replacing an A-U pair with a G-C pair.

As described above, G90 in the SD sequence plays a pivotal role in the S_{MK} box riboswitch, as it organizes the SAM binding pocket through A73:G90-C25 base triple formation and stacks directly underneath the SAM molecule. The C25U substitution creates a wobble pair with G90. This mutant exhibited partial repression by high SAM *in vivo* (Table 2.2), consistent with a modest reduction of SAM binding in vitro (Figure 2.2A). The C25U/G90A double mutant is predicted to restore Watson-Crick pairing at this position but not the critical base triple interaction with A73. This variant exhibited complete loss of repression in vivo despite maintaining a canonical SD sequence (Table 2.2), in accord with the loss of SAM binding in vitro

(Figure 2.2A). These data confirm the importance of G90 for SAM binding and SAM-dependent regulation.

Se-SAM bound S_{MK} box structure confirms the position of the sulfonium

To confirm the position of the sulfur atom in the SAM molecule, we further determined the structure of the SMK box riboswitch soaked in freshly prepared Se-SAM (where the sulfur atom is substituted with selenium, a heavier chalcogen analog (Iwig and Booker, 2004)). The conformations of the RNA and the Se-SAM molecule are essentially identical to that in the SAM-bound structure, with an r.m.s.d. of 0.3 Å for all-phosphorous-atom alignment. The same set of contacts specifies Se-SAM in the ligand binding site, and the selenomethionine moiety remains unstructured beyond the onium selenium (Figure 2.6A). The location of the selenium atom, which essentially overlaps with the sulfur atom in the omit map, is unambiguously identified by an 8σ anomalous difference signal collected at the absorption edge of Se (Figure 2.6A). The distances of the electrostatic interactions between the RNA and the onium selenium are on average ~ 0.3 Å longer than seen in the SAM-bound structure, reflecting the increased van der Waals radius. Observations from the Se-SAM bound riboswitch structure confirm our assignment of the sulfonium position in SAM, and strengthen our conclusion that the methionine tail of SAM is not specified by the S_{MK} box riboswitch.

SAH makes a minimum set of contacts to the S_{MK} box riboswitch

To investigate how the S_{MK} box riboswitch would respond in the presence of natural near-cognate ligands such as SAH, we determined the S_{MK} box RNA structure

in the presence of SAH at a saturating concentration of 2 mM, about 67-2000 fold higher than the intracellular concentration of SAH (0.5 – 30 μM (Ueland, 1982)). The resulting structure revealed that SAH can displace SAM at saturating concentration. The binding pocket and overall structure of the SMK box riboswitch remains unchanged, presumably due to strong crystal lattice contacts to the P1 and P4 helices. The resulting structure is thus a snapshot of the functional S_{MK} box riboswitch sampling a near-cognate ligand before rejecting it. In the 2.9 Å structure, SAH adopts a conformation very different from SAM in the binding pocket (Figure 2.6B). The ribose moiety of SAH rotates 180 °such that the adenosine moiety now adopts an anti-conformation. 2′- and 3′-hydroxyls in SAH make an alternative set of interactions with the phosphoryl oxygens of G90. More importantly, weak electron density features revealed that the uncharged sulfide in SAH rotates 180° from U72 to the vicinity of G71. As a result, the intramolecular contacts that stabilize the sulfonium ion conformation in SAM and the key electrostatic interactions between the sulfur atom and the RNA are all lost when SAH is in place. It therefore appears that the positive charge on the sulfur atom is crucial in presenting the SAM molecule in the correct conformation for strong ligand-RNA interactions, analogous to the "lock-and-key" mechanism found in many enzyme-substrate complexes. The structural observations are consistent with in vitro assays showing that the S_{MK} box riboswitch has much higher affinity for SAM over SAH (Fuchs et al., 2006, Fuchs et al., 2007), with at least 100-fold preference for SAM.

K_d measurement by 2-AP fluorescence

To verify the apparent K_d value observed in the size-exclusion filtration experiment, a second assay was developed to test the binding affinity of SAM to the

wild-type 15-118 metK leader RNA. For this we used a bipartite RNA comprised of two half RNAs, corresponding to positions 15-46 and 47-118. The two RNA halves are predicted to anneal through extensive base pairing in the P1-P3 helices to form the full-length *metK* leader RNA corresponding to positions 15-118, with a nick immediately upstream of residue G47 at the top of the P3 stem loop. A 2-aminopurine (2-AP) modification was incorporated into the upstream RNA at residue A29 to monitor local conformational changes occurring at this position in response to SAM binding using florescence spectroscopy. In the absence of SAM, sample excitation at 310 nm yielded a fluorescence emission peak at 375 nm. Addition of SAM resulted in a concentration-dependent quenching of 2-AP fluorescence that was not observed with addition of SAH (data not shown). SAM titrations were used to determine an apparent K_d value of 1.03 μM for the wild-type *metK* RNA (Figure 2.2C), which is in close agreement to the value obtained using the size-exclusion filtration assay. No quenching of 2-AP fluorescence was observed with an RNA construct containing a U72C substitution with SAM concentrations up to 100 μM (Figure 2.2C and data not shown), consistent with the loss of SAM binding to the U72C RNA in the filtration assay (Figure 2.3).

DISCUSSION

With the completion of the structure of the S_{MK} box riboswitch, a family portrait of the three known classes of SAM binding riboswitches is now available (Montange and Batey, 2006, Gilbert et al., 2008). These riboswitches adopt completely different RNA folds and most likely have emerged independently during evolution. They nevertheless converged at the functional level to preferentially recognize SAM, an important metabolite inside the cell, and regulate gene expression

in response to the ligand-binding event. We conclude from structural comparisons that while creative ways are used to accommodate the binding of SAM, a conserved mechanism is used by all three classes of SAM riboswitches to distinguish SAM from SAH, which is the biologically relevant SAM analog.

Riboswitches sense both chemical and conformational distinctions in SAM

SAM adopts drastically different conformations among the three classes of SAM riboswitches (Figure 1.6A). The *T. tengcongensis* and *B. subtilis yitJ* S box (SAM-I) riboswitches completely engulf SAM in a pocket between two helical stacks ((Montange and Batey, 2006) and C. Lu el al., manuscript in preparation). The SAM-II riboswitch adopts a classic H-type pseudoknot and encloses SAM inside an RNA triple helix (Gilbert et al., 2008). The S_{MK} box riboswitch allows the intercalation of SAM into a tight pocket in a three-way junction. In each case, SAM mediates important tertiary interactions to stabilize the ligand-bound conformation of the riboswitch. The adenine base is involved in base stacking as well as base triple interactions in all three riboswitches. The positively charged sulfonium moiety is invariably recognized through favorable electrostatic interactions, usually with one or two O4 carbonyl oxygen atoms from uracil residues. This recognition forms the basis for preferential binding of SAM over SAH. The methyl group on the sulfonium moiety is not directly contacted by any of the three riboswitches. Rather, it points toward the solvent region and is recognized collectively with the sulfonium ion, a mechanism perhaps evolved to prevent self-inactivation by spontaneous methylation of the RNA by SAM. Recognition of the methionine tail, however, occurs differently in the three riboswitches. The tail is extensively contacted by the S box riboswitch (Montange and Batey, 2006), less so by the SAM-II riboswitch (Gilbert et al., 2008),

and is completely ignored by the S_{MK} box riboswitch. Previous structural studies seem to suggest that riboswitches are more likely to be identified through bioinformatics rather than SELEX approaches because in each structure the metabolite is completely encapsulated by the RNA. The ability of the biologically active S_{MK} box riboswitch to recognize SAM by specifying only half of the molecule and without engulfing SAM completely suggests that additional classes of SAM riboswitches can potentially be identified by SELEX approaches using a SAM molecule immobilized at the methionine tail as the bait.

In addition to chemical differences between SAM and SAH, SAM also adopts a distinct conformation that is energetically unfavorable for SAH and other non-cognate SAM analogs. Alignment of the conformations of SAM within the three riboswitch structures along the ribose region clearly revealed that while the conformation of the adenine base (*syn*- or *anti*-) and the methionine tail (crouched or extended) varies dramatically among the three riboswitch structures, the sulfonium ion makes an invariable, strong 3 Å electrostatic contact to the O4′ of the SAM ribose (Figure 1.6A,B). This *gauche*-conformation (Figure 1.6B) about the C4′-C5′ bond has been shown by NMR measurement to be the predominant (93%) SAM conformation in solution (Stolowitz and Minch, 1981). The same study showed that SAH, on the other hand, favors the *anti*-conformation (Stolowitz and Minch, 1981), presumably due to its inability to maintain the intramolecular electrostatic interaction. Indeed, in our survey of thirty randomly chosen SAH-bound protein structures, 87% of the SAH molecules adopt the *anti*-conformation (Figure 1.6B). Thus, by simultaneously contacting both the ribose and the sulfonium moieties of SAM, all three riboswitches effectively select for a SAM conformation that is unfavorable for SAH. Consistent with this observation, our structural analysis of the SAH-bound S_{MK} box riboswitch

revealed that while SAH could bind to the crystalline-trapped S_{MK} box riboswitch at a non-physiological high concentration, the sulfide moiety in the riboswitch-bound form of SAH swings 180° away from the sulfonium binding site, severely weakening the ligand-RNA interactions to a level indistinguishable from the binding of other adenosine-containing metabolites such as ATP. Thus, it appears that both chemical and conformational differences at the sugar-sulfur linkage are explored by SAM-binding riboswitches to distinguish between cognate and near-cognate ligands. The recent discovery of a new class of SAH- binding riboswitches upstream of bacterial genes involved in SAM recycling (Wang et al., 2008) demonstrates that RNA is capable of forming a selective binding site that favors either SAM or SAH.

A simple SAM-dependent translation inhibition mechanism

A typical riboswitch is conceptually modular, consisting of a ligand-sensing aptamer domain and a separate output domain whose structure is influenced by the ligand-binding event at the aptamer domain. A much simpler mechanism for translational regulation is found in the S_{MK} box and SAM-II riboswitches, where the SD sequence is an integral part of the SAM-binding aptamer domain, allowing ligand sensing and translational inhibition through SD sequence sequestration take place in a single step. An important feature in the S_{MK} box riboswitch is the direct involvement of the SD sequence in SAM recognition. Although strong crystal packing interactions prevent observation of large ligand-induced conformational changes in crystal structures, mutagenesis and enzymatic probing assays (Fuchs et al., 2006, Fuchs et al., 2007) clearly revealed global conformational changes in the presence of physiologically relevant concentrations of SAM, but not SAH. These SAM-dependent changes are especially evident in the formation of the linker and SD-ASD helices that

sequester the ribosome binding site. Conversely, the ability to form the linker and SD-ASD helices is a prerequisite for SAM binding (Fuchs et al., 2006). The combined structural and biochemical data suggest that SAM shifts the conformational equilibrium toward the ligand-bound state seen in the crystal structure, where the SD sequence required for binding of the 30S ribosomal subunit is sequestered. Consistent with this mechanism, ribosome toeprinting analysis showed that the correct positioning of the 30S ribosomal subunit on the S_{MK} box mRNA is reduced in the presence of SAM, but not SAH (Fuchs et al., 2007). Direct involvement of the SD sequence in the formation of the aptamer domain is also observed in the SAM-II riboswitch structure (Gilbert et al., 2008), although the exact location of the SD sequence is less well defined.

In summary, our combined structural and mutagenesis data on the S_{MK} box riboswitch clearly demonstrate its mechanism for translational inhibition through sequestration of the SD sequence, and the interplay between SD-ASD pairing and SAM-binding activity. The S_{MK} box riboswitch is unique among riboswitch RNAs studied to date in that the same residues of the RNA that are involved in gene regulation (i.e., the SD sequence) are directly involved in specific recognition of the SAM ligand.

Accession codes: Structure factors and coordinates for the SAM, SAH, and Se-SAM bound S_{MK} box riboswitch have been deposited in the Protein Data Bank with accession numbers of 3E5C, 3E5E, and 3E5F, respectively.

CHAPTER 3

FOLDING OF THE SAM-III/S_{MK} BOX RIBOSWITCH IS HYPERSENSITIVE TO SEQUENCES OUTSIDE THE SAM-BINDING CORE

SUMMARY

The S_{MK} box (SAM-III) translational riboswitch has been identified in S-adenosylmethionine (SAM) synthetase *metK* genes in the *Lactobacillales*. It switches between two alternative conformations depending on the presence or absence of SAM and control gene expression level at the translation initiation level. We previously reported crystal structures of the SAM-bound S_{MK} box riboswitch. In this study we used SHAPE chemical probing analysis to characterize the conformational dynamics in S_{MK} RNA. We show during the conformational transition between the non-binding and SAM-bound states, there is a third, intermediate "pre-binding" conformation, where the S_{MK} RNA pre-organizes into a SAM-bound like conformation in the absence of the ligand. Mutagenesis results suggest that an intricate energetic balance is kept between these conformations. In particular, the energetic difference between the non-binding and pre-binding intermediate conformations is as small as forming or breaking two to three hydrogen bonds, and a single nucleotide substitution in a hypervariable region outside the SAM-binding core selectively stabilizes the intermediate information to produce a "constitutively-OFF" riboswitch as a consequence. Furthermore, we showed that SAM's natural analog, *S*-Adenosyl-homocysteine (SAH), cannot induce the "OFF" response in S_{MK} because it cannot switch the RNA conformation even at saturating concentration.

INTRODUCTION

Riboswitches are non-coding RNAs that regulate gene expression via structural changes induced by environmental cues (Tucker and Breaker, 2005, Grundy and Henkin, 2006, Vitreschak et al., 2004). Riboswitch RNAs that regulate a significant number of gene families that are involved in metabolism and are especially prevalent in Gram-positive bacteria and a number of important pathogens (Winkler and Breaker, 2005, Grundy and Henkin, 2006). The SAM-binding S_{MK} riboswitch regulates the translation of SAM synthetase (*metK*) genes in *Enterococcus* bacteria by sequestration of the Shine-Dalgarno (SD) sequence (Fuchs et al., 2006, Lu et al., 2008). It folds into a "Y" shaped arrangement that organizes an array of conserved nucleotides around a three-way junction for SAM recognition (Lu et al., 2008). The Shine-Dalgarno sequence, which is sequestered by base pairing with the anti-Shine-Dalgarno sequence in response to SAM binding, also directly participates in SAM recognition (Lu et al., 2008).

To understand the conformational switching mechanism of the *Enterococcus faecalis* S_{MK} riboswitch, we used **s**elective 2'-**h**ydroxyl **a**cylation **p**robing **e**xperiment (SHAPE) to define its conformation in the presence or absence of SAM. The reactivity profile of the SAM-bound conformation agrees extremely well with the crystal structure. In the absence of SAM, an alternatively base paired structure dominates the *apo* (termed "non-binding" hence forth) S_{MK} conformation. In addition, we discovered an important intermediate (pre-binding form) where the RNA pre-folds into a SAM-bound like conformation in the absence of the ligand. Mutagenesis results suggested that the energetic difference between the unbound and pre-binding conformations is a matter of making or breaking two to three hydrogen bonds, as a single point mutation in a hypervariable region outside the SAM-binding core selectively stabilized the

intermediate conformation and resulted in the formation of a constitutively "OFF" riboswitch. Our results suggested that naturally occurring riboswitches maintain intricate conformational balance between alternative conformational states in order to fulfill their gene regulatory function, and that seemingly hypervariable sequences outside the ligand-sensing core may play important roles in maintaining the conformational balance.

MATERIALS AND METHODS

RNA preparation

The *Enterococcus faecalis* S_{MK} box RNA and all mutation constructs flanked by 5′-T7 promoter and 3′-reversetranscription start site was constructed by overlapping PCR and inserted into pJet cloning vector. The DNA template for *in vitro* transcription was generated by 2 mL PCR reaction with terminal primers. The PCR product was directly used in T7 RNA RNAP run-off transcription as described (Ke and Doudna, 2004, Lu et al., 2008). The RNA product was purified and recovered by denaturing polyacrylamide gel electrophoresis as described (8% polyacrylamide, 8 M urea, 30 W, 3 hrs) (Ke and Doudna, 2004). Eluted RNA was concentrated and buffer exchanged into 0.5X TE using Millipore centrifuge filters with 10 kDa cutoff. The purified RNAs have final concentrations of around 50 μM measured by UV absorbance of 260nm/280nm.

DMS/SHAPE probing analysis

RNA was diluted into a final annealing mix containing 300 nM RNA, 5 mM Tris pH 7.5, 0.5 mM EDTA, 111 mM NaCl, 111 mM Hepes pH 8.0, 11.1 mM $MgCl_2$

(or no $MgCl_2$). The annealing mix was heated at 65 ℃ for 5 min, then added 0.5 mM of SAM (or 0.5 to 5 mM SAH, or nothing), and incubated at 37 ℃ for 3 min, and cooled on ice.

SHAPE was carried out as described (Wilkinson et al., 2006, Mortimer and Weeks, 2007). Every 8 µl of RNA solution was treated with 1 µl 1M7 (100 mM in anhydrous DMSO) at 4 ℃, mix and allowed to react for 10 min or until the solution reaches a deep orange color). The reaction is quenched by the addition of 500 µl precipitation buffer [90 µl sterile ddH_2O, 5 µl NaCl (5 M), 1 µl glycoblue (Ambion) (15 mg/mL), and 400 µl ethanol].

Dimethyl sulphate (DMS) modification for probing the Watson-Crick faces of adenosines was performed as previously described (Gilbert et al., 2008). 10 pmols of annealed RNA mix prepared as described above was added to 200 µl total volume of DMS modification buffer (70 mM HEPES-KOH, pH 8.0, 10 mM $MgCl_2$, 270 mM KCl, and additional 0.1 mM SAM for reactions containing SAM) on ice. The DMS stock is made from mixing one volume of pure DMS (Sigma) with two volumes of ethanol. 1 µl of DMS stock (or ethanol for negative control) was added to every 5 pmol (100 µl) of RNA at 30 °C for 5 min, followed by the immediate addition of 50 µl of DMS termination buffer (1 M Tris-acetate, pH 7.5, 1 M β-mercaptoethanol, 1.5 M NaOAc, 0.1 mM EDTA), 1 µl glycoblue (Ambion) (15 mg/mL) and 750 µl of ethanol at -20 ℃.

Reverse transcription

The modified RNA mix was then stored at -80 ℃ for 30 min and spun in 4 ℃ at 15000 g for 30 min. resulting blue pellet was recovered and resuspended in 10 μl of 0.5X TE.

The reverse transcription and sequencing procedure is generally that outlined by Wilkinson *et al* (Wilkinson et al., 2006). A fluorescently labeled DNA primer (HEX or 6-FAM-labelled GAA CCG GAC CGA AGC CCG; 1 μl, 30 μM) was added to the RNA (10 μl, from the previous step) by heating to 65 ℃ for 2 min, annealed at 35 ℃ for 5 min and cooled to 4 ℃. Reverse transcription buffer [167 mM Tris (pH 8.3), 250 mM KCl, 10 mM $MgCl_2$, 1.67 mM each dNTP] was prepared by mixing four parts SSIII FS buffer, one part 0.1 M DTT and one part 10 mM dNTP mix. After mixing with 6 μl of transcription buffer, the RNA mix was pre-heated to 52 ℃ for 45 sec, before the addition of 1 μl Superscript III (Invitrogen, 200 units), and the mix was incubated at 52 ℃ for 20 min. The reactions were quenched and the RNA template was digested by addition of 0.8 μl of 5M NaOH, followed by heating to 90 ℃ for 4 min. 29 μl of acid stop mix [4:25 (v/v) mixture of 1 M unbuffered Tris-HCl, and stop solution (85% formamide, 0.5X TBE, 50 mM EDTA, pH 8.0)] was added to each reaction, the reaction is allowed additional 4 min at 90 ℃ before cooled to -20 ℃. No internal tracking dye is used in the reaction to avoid interference with fragment analysis.

Sequencing reactions were analyzed by polyacrylamide sequencing gel electrophoresis (12% 29:1 acrylamide:bis-acrylamide, 1X TBE, 8 M urea). Load 5 μl per lane and run at constant 55 watts for 5 hours. For extensions of 100 or fewer nucleotides, perform electrophoresis for 150 min at 70 W.

Data analysis

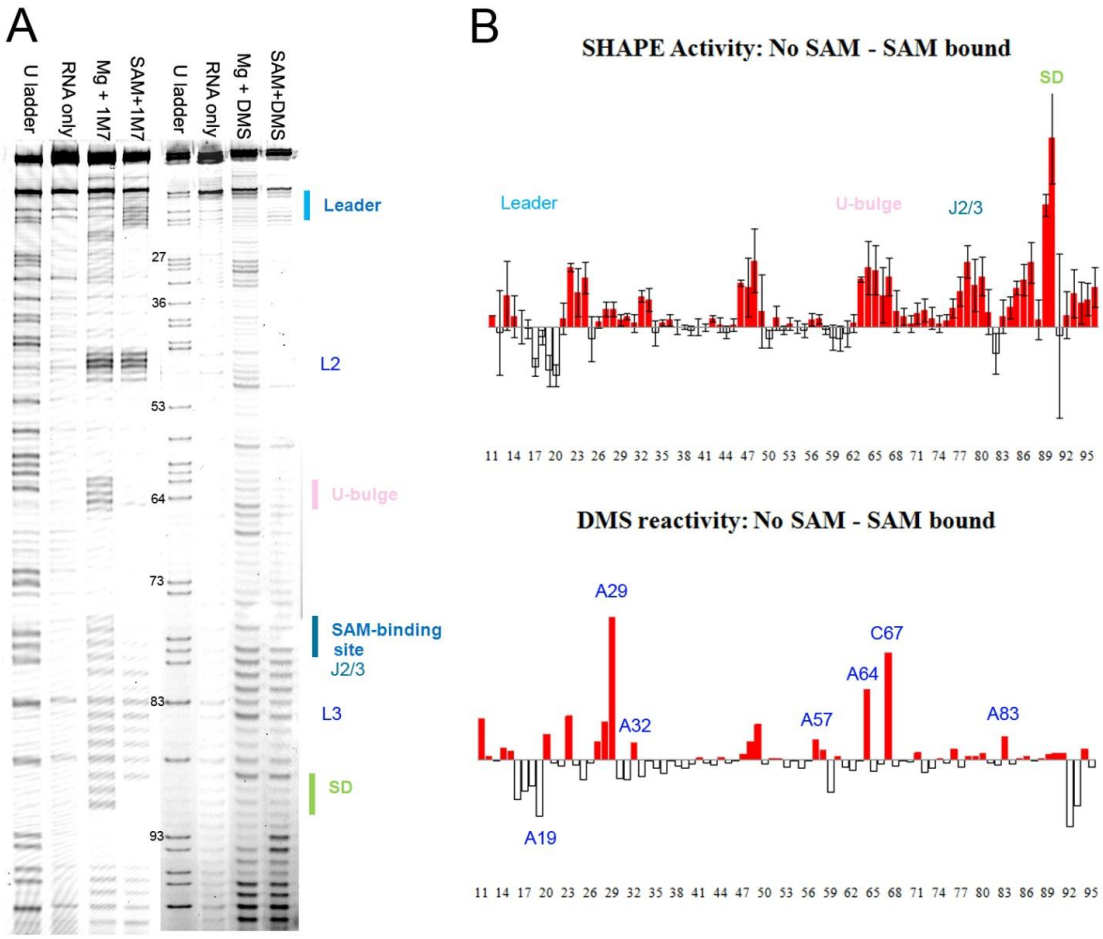
Band intensity from a Typhoon phosphor-imager (Molecular Dynamics) was quantified using the SAFA (Das et al., 2005). Data were averaged across three or four separate experiments run out on two to three gels. A total of 11 bands including base 36-44 (P2 stem), 83 and 87 (flexible linker) that are similarly modified throughout all experiments are used to normalize all intensities. A Microsoft Excel spreadsheet template was used to process and graph all data. After normalization and scaling, we compared 1M7 reactivity among RNA alone, RNA+Mg^{2+}, RNA+Mg^{2+}+SAM, RNA+Mg^{2+}+SAH by subtracting one from another. All individual experiments are processed and graphed separately and then compared with each other before averaged to produce the final result. This procedure ensures the reducibility and consistency of probing experiments, gel analysis and data processing.

RESULTS

SHAPE and DMS probing data agreed very well with the S_{MK} crystal structure.
Our previous structural studies of the S_{MK} box riboswitch provided detailed insights into its ligand-recognition mechanism (Lu et al., 2008). Here we employed chemical probing methods to further characterize the ligand-induced conformational dynamics within the S_{MK} RNA. We used two different chemical probing methods in the initial characterizations to cross-validate the results. The S_{MK} RNA used in this study contains the aptamer domain (U21-A94), the upstream alternative base pairing elements (A11-G20) and the downstream AUG start codon. In SHAPE experiments, the extent of acylation by 2′-hydroxyl alkylation by 1-methyl-7-nitroisatoic anhydride (1M7) was monitored to detect local flexibility in RNA structure. Figure 3.1 shows side-by-side comparisons of single-nucleotide resolution SHAPE probing data of the

Figure 3.1. SHAPE and DMS probing of apo and SAM-bound S_{MK} RNA.

(A). Sequencing gel detects conformational differences between SAM bound and unbound structures. Lane 1and 5, ddUTP sequencing lanes; lanes 2-4, SHAPE analysis lanes; lanes 6-8, DMS probing lanes. Base identities are indicated by numbers next to the ddUTP sequencing lane. Structural motifs are marked on the right. (B). Quantified probing analysis reveals structural elements that respond to SAM binding. Upper panel: SHAPE analysis; lower panel: DMS probing analysis. Base numbers are indicated on the X-axis.


A
U ladder
RNA only
Mg + 1M7
SAM+1M7
U ladder
RNA only
Mg + DMS
SAM+DMS
Leader
27
36
L2
53
U-bulge
64
73
SAM-binding site
J2/3
83
L3
SD
93
B
SHAPE Activity: No SAM - SAM bound
SD
Leader
U-bulge
J2/3
11 14 17 20 23 26 29 32 35 38 41 44 47 50 53 56 59 62 65 68 71 74 77 80 83 86 89 92 95
DMS reactivity: No SAM - SAM bound
A29
C67
A64
A32
A57
A83
A19
11 14 17 20 23 26 29 32 35 38 41 44 47 50 53 56 59 62 65 68 71 74 77 80 83 86 89 92 95

non-binding and SAM-bound S_{MK} RNA samples, together with appropriate background and unfolded RNA negative controls. Sequencing signals were then quantified and normalized against the background to generate SHAPE reactivity profiles. Both the SAM-unbound and SAM-bound S_{MK} adopt defined tertiary structures when compared to the unfolded S_{MK} RNA, although their distinct reactivity profile clearly suggested that S_{MK} adopt different conformations in these two functional states (Figure 3.1, compare lane 3 with lane 4, and lane 7 with lane 8). The SHAPE reactivity profile of the SAM-bound S_{MK} correlated very well with the reported crystal structure (Lu et al., 2008). Residues that were omitted by the reported crystal structure construct showed markedly higher SHAPE reactivity. Within the crystal structure, corresponding residues participated in helix formation (P1 – P4) all showed noticeable reduction in SHAPE reactivity. Similarly, residues involved in SAM-binding and tertiary structure formation (G25, A26, G70-A72, and the AUG tri-nucleotide bulge) were also protected. By contrast, the two hypervariable loops, which were replaced by stable GAAA tetraloops as a necessary step to obtain high-resolution crystal structure, displayed markedly higher SHAPE reactivity in their native form. The 6-nucleotide distal L2 loop was the most reactive region inside the folded structure, and the hypervariable 11- nucleotide loop in linker helix was the next reactive region, suggesting that the potential AT-rich helical structure does not form under physiological conditions (Figure 3.1).

To verify the structure model by SHAPE analysis, we further probed the unbound and SAM-bound S_{MK} using dimethylsulfate (DMS), which probes the solvent accessibility of the Watson-Crick face of adenosine (via N1 modification) and cytosine (via N3 modification). When SAM is absent, A27-A29 is the most flexible region in the RNA, as proposed by secondary structure models (Fuchs et al., 2006,

Fuchs et al., 2007). When SAM is present with the S_{MK} RNA, the leader (shown by A18 and A19), becomes more flexible while A27-29 become protected due to the binding pocket formation and SAM-recognition. Similarly, A64 and C67, which are unprotected in no-SAM condition, forms tertiary interactions (Lu et al., 2008) (termed the "U-bulge"), are protected in the SAM-bound form. Consistent with SHAPE analysis, adenosine and cytosine nucleotides involved in W-C pairing in the no-SAM structure (A16-A19 in the leader sequence) are protected in the absence of SAM, whereas A and Cs that only form base pairs in the SAM-bound structure (A29, 64 and C67) are more protected when SAM is present (Figure 3.1A, 7^{th} and 8^{th} lane, Figure 3.1B, lower panel).

SAM-unbound S_{MK} RNA adopts an alternatively base paired conformation.

SHAPE and DMS probing data clearly support an alternatively base paired S_{MK} conformation in the absence of SAM. The unbound S_{MK} conformation is apparently less stable than the SAM-bound form as reflected by the overall elevated SHAPE reactivity (Figure 3.1A, compare lanes 3 and 4, 7 and 8, 3.1B). SHAPE profile of the SAM-unbound S_{MK} agreed well with the alternatively base paired secondary structure model first proposed using enzymatic probing assay (Fuchs et al., 2006). While P3 in the SAM-bound S_{MK} is preserved, P1 is replaced by a more stable structure between the upstream leader and J3/4 sequences; P2 and P4 are completely disrupted and bases in these regions display significantly higher SHAPE reactivity (Figure 3.1A, B). The protection of leader sequence, together with the de-protection of ASD left a characteristic signature banding pattern to distinguish the SAM-unbound and SAM-bound conformations. The rearrangement of base pairing disrupts structures around the SAM-binding site, and unveiled the sequestered Shine-Delgarno sequence.

Similarly, the DMS probing results revealed similar conformation shift between SAM-bound and unbound forms. The SHAPE and DMS probing data therefore supports the functional data showing that S_{MK} riboswitch controls downstream gene expression at the translation level (Fuchs et al., 2006), through a SAM-dependent translation initiation mechanism. Based on the SHAPE analysis, it is clear that the unbound S_{MK} RNA folds into a different tertiary structure and upon addition of SAM, "switches" to the SAM-bound form.

Destabilization of non-binding structure unveiled a "pre-binding" conformation. How does SAM switch the S_{MK} conformation? We hypothesize that the structure transition traverses through a "pre-binding" intermediate, where S_{MK} pre-organizes into a conformation resembling the SAM-bound state in the absence of the ligand. This state is less stable than the non-binding conformation, and is only infrequently traversed in the absence of SAM. The presence of SAM, however, selectively stabilizes the "pre-binding" conformation through SAM-mediated tertiary interactions, which in turn drives the conformational equilibrium towards the predominant SAM-bound population (Figure 3.2). If the "pre-binding" is a stable conformation, our hypothesis predicts that we can prove its existence by selectively destabilizing the non-binding conformation. This can be done by a poly-(A) substitution of the leader sequence (nt 11-20) to disrupt the base helix in the non-binding structure, while having no effect on SAM-bound structure. When subject to SHAPE analysis in the absence of SAM, this Poly-A construct clearly displayed a SHAPE reaction profile resembling that of the SAM-bound state, but not the non-binding state (Figure 3.3A), consistent with our prediction. Detailed quantitative SHAPE analysis that this pre-binding conformation is largely similar to that of the SAM-bound form, except that

Figure 3.2. Proposed secondary structure changes of S_{MK} RNA binding to SAM.

Flexible regions detected by chemical probing analysis are highlighted in magenta ovals. Five consecutive Gs in solid black boxes are the ribosome binding sites. SAM is shown in red. SAM-mediated tertiary interactions are shown as red dotted lines. The pre-binding form is structurally similar to SAM-bound form, except lacking SAM-mediated tertiary contacts.

Non-binding

Pre-binding

SAM-bound

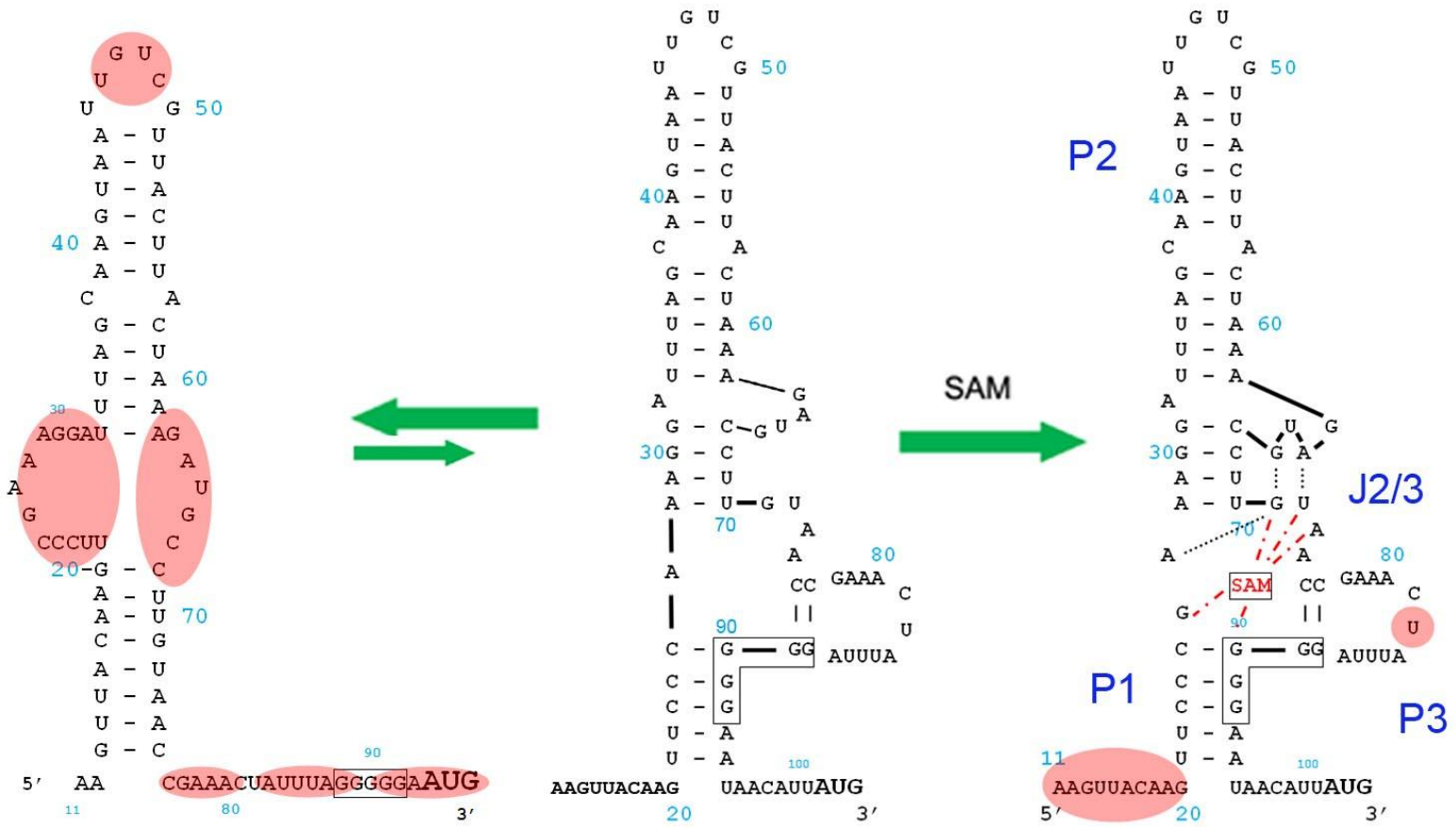
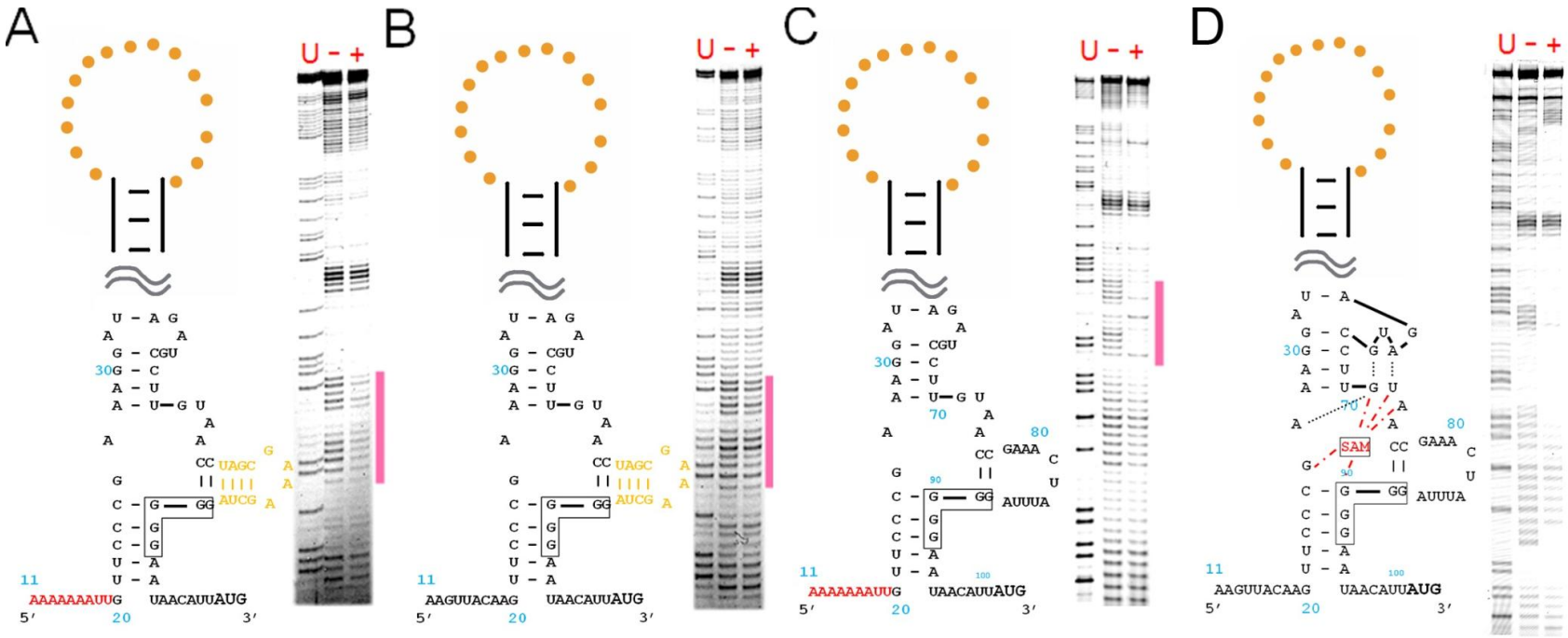


Figure 3.3. Mutations outside SAM-binding domain shift the conformational equilibrium completely to the pre-binding state.

Secondary structures and SHAPE analysis of three mutant S_{MK} RNAs with or without SAM. (A) A construct with Poly-Adenine leader and perfect base-pairing linker helix. (B). A construct with perfect base-pairing linker helix only. (C). A construct with Poly-Adenine leader only. Mutations in the 5′-leader are highlighted in red; mutations in the linker helix are highlighted in gold. SHAPE activity differences in the U-bulge-J2/3 region are highlighted by pink bars. Gel loading sequence, from left to right: ddU ladder, RNA+Mg^{2+}, RNA+Mg^{2+}+SAM. Five consecutive Gs in solid black boxes are the ribosome binding sites.


A
B
C
D
U - +
U - +
U - +
U - +
AAAAAAAUUG
UAACAUUAUG
AAGUUACAAG
UAACAUUAUG
AAAAAAAUUG
UAACAUUAUG
AAGUUACAAG
UAACAUUAUG
SAM
5′
3′

regions involved in SMA-induced tertiary structure formation displayed higher SHAPE reactivation. These include the J2/4 at the SAM-binding site, and the tri-nucleotide bulge (J3/2) that participates in base triple formation immediately atop of SAM-binding site. Our interpretation is that the "pre-binding" conformation resembles the SAM-bound state at the secondary structure level, however, the defined tertiary structure only forms in the presence of SAM, underlining the energetic contribution of SAM-mediated tertiary interactions.

The non-binding state is marginally more stable than the "pre-binding" state

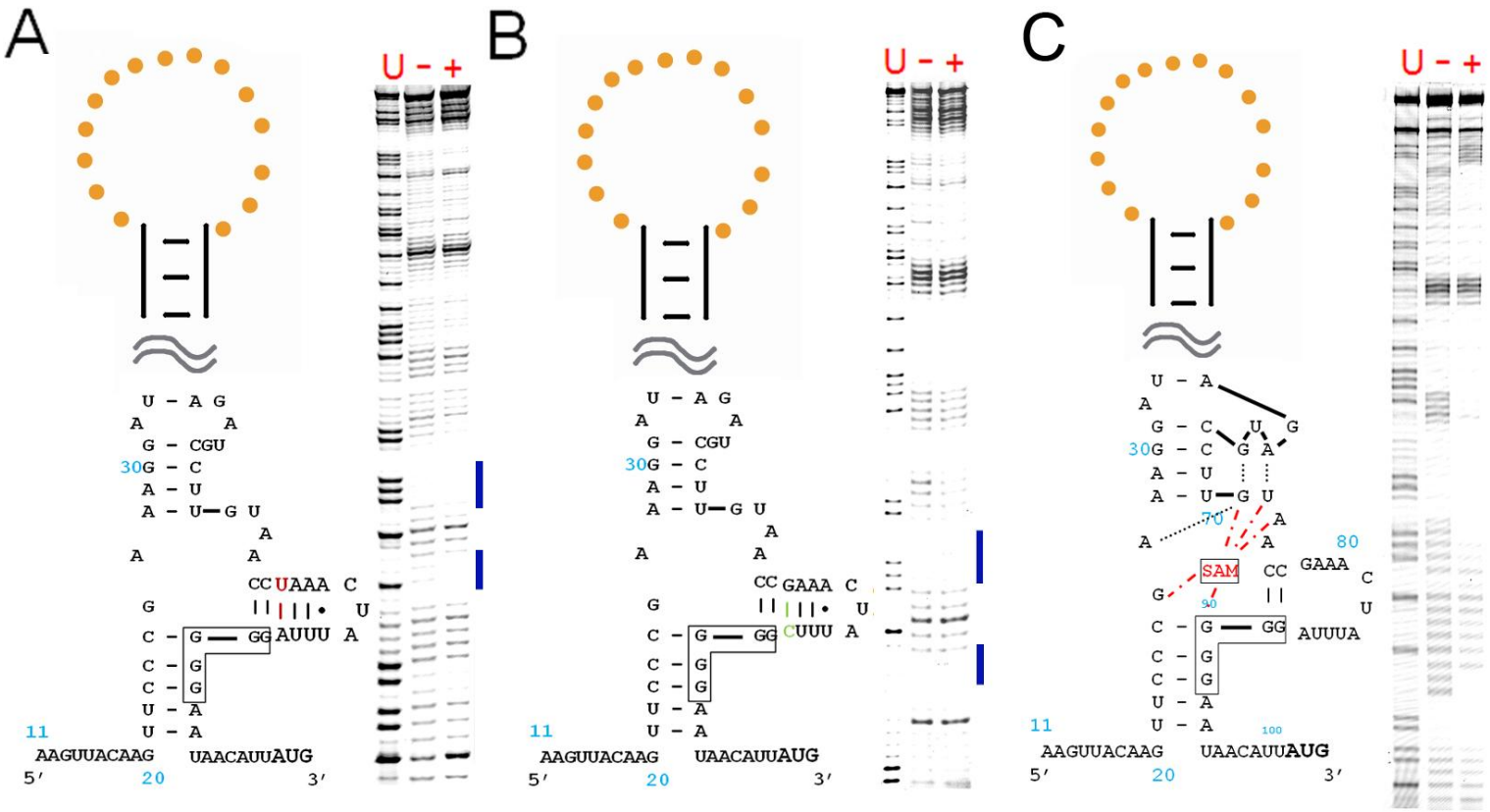
After isolating and stabilizing the "pre-binding" state during the structural transition of the S_{MK} riboswitch, we next investigated the energetic differences among non-binding, pre-binding, and SAM-bound states. To estimate the free energy difference between non-binding and pre-binding states, we again resorted to mutagenesis perturbation of the conformation equilibrium between two states. Here we seek to demonstrate that (i) selective stabilization of the pre-binding state can tip the conformational equilibrium such that S_{MK} predominantly adopts the pre-binding conformation in the absence of SAM, and (ii) further minimization of the perturbation to gauge the energetic difference between non-binding and pre-binding states. To achieve the former, we focused on the 11-nt hypervariable loop on top of the 2-bp P4 linker helix that is near the SAM-binding site and appeared to be poorly structured with elevated SHAPE reactivity, even though theoretically it can fold into an AT-rich stem-loop. We speculate that introducing additional base pairs in this region may stabilize the pre-binding state by reducing its conformational entropy. We tested this idea by replacing the linker helix with a perfected base paired stem-loop, capped by a GAAA tetraloop. SHAPE analysis showed that similar to the poly-(A) mutant, this

construct adopt the same pre-binding instead of non-binding conformation in the absence of SAM, supporting the notion that high conformational entropy in the linker region was evolved to maintain the energetic balance between non-binding and pre-binding states. Not surprisingly, S_{MK} mutant containing both poly-(A) leader and a perfectly paired linker helix folds into the pre-binding conformation in the absence of SAM as well.

To gauge the energetic difference between non-binding and pre-binding states, we seek to identify a minimum perturbation that leads to pre-binding conformation. A total of 5 single nucleotide substitutions were systematically introduced at G77•A87 mismatch. Among them, G77U and A87C (Figure 3.4A, B) introduce a G-C or A-U pair in place of the G•A mismatch. SHAPE analysis not only confirmed the intended base pair formation between position 77 and 87, but also a dramatic stabilization of the linker helix region as a whole, where two to three additional pairs (A78-80 with U83-U85) were observed. This is accompanied by the adoption of predominant pre-binding instead of non-binding conformation in the absence of ligand in these two mutants. On the contrary, when G77•A87 is replaced by a G77•U87 wobble pair, SHAPE analysis showed a mixture of non-binding and pre-binding conformation in the absence of SAM, suggesting that energetic gain from forming a G•U wobble in this position is not sufficient to shift the conformational equilibrium (data not shown). As controls, a G•G mismatch introduced by A87G mutation did not perturb the conformational equilibrium in any significant way (data not shown). All together, our data suggested that the free energy difference between the non-binding and pre-binding states is within the making and breaking of one base pair. Replacing a hypervariable mismatched base pair outside the SAM-binding core region with a normal W-C pair, but not a wobble pair, is sufficient to drastically stabilize the linker helix by reducing

Figure 3.4. Free energy difference between non-binding and pre-binding state diminished by a point mutation in the linker helix.

Secondary structures and SHAPE analysis of two point mutant S_{MK} RNAs with or without SAM. Point mutations in the secondary structure are highlighted in red and green respectively. Protection of linker helix stem is marked by blue bars and the right. Gel loading sequence, from left to right: ddU ladder, RNA+Mg^{2+}, RNA+Mg^{2+}+SAM. Five consecutive Gs in solid black boxes are the ribosome binding sites.


A
B
C
U - +
U - +
U - +
5′
3′
11
20
30
80
90
100
SAM
AAGUUACAAG
UAACAUUAUG
CCUAAA C
CCGAAA C
CC GAAA C
AUUU A
CUUU A
AUUUA

its conformational entropy. The subtle change selectively stabilized the pre-binding conformation and making it the predominant conformation.

Selective stabilization of "pre-binding" state improves S_{MK}'s affinity for SAM.

One prediction from the model of S_{MK} folding energy diagram is that there would be bigger free energy gain if S_{MK} exists as pre-binding instead of the non-binding state before SAM is introduced. Previous studies (Winkler et al., 2003, Winkler et al., 2002a) have shown that peripheral sequences can affect the disassociation constant (K_d) of riboswitches, by eliminating the adjoining sequences that permits alternative conformations. Therefore, it is expected that our mutations on the S_{MK} RNA will improve binding of SAM by favoring the pre-binding form. We performed equilibrium dialysis assays to measure the disassociation constants of our constructs to further characterize the effects locking the RNA into the pre-binding form. As expected, our assays showed that compared to the wild-type RNA, all of the mutations that stabilizes the pre-binding state lowered the K_d of SAM-binding, by up to five fold (Table 3.1).

At saturating concentration, SAH induces pre-binding form but unable to completely shift the conformation equilibrium towards SAM-bound state

Previous X-ray crystal structure by Lu and co-workers showed that at high concentrations, SAH partially fits into the binding pocket of the S_{MK} RNA by making three hydrogen bonds with G26 (Lu et al., 2008). We therefore speculated that at saturating concentrations, SAH might affect the equilibrium between non-binding and pre-binding form of the S_{MK} RNA. SHAPE analysis on the S_{MK} RNA in the presence

Table 3.1.Disassociation constant (K_d) measurements for the S_{MK} RNA constructs by equilibrium dialysis.

Construct	K_d (μM)
Wild-type	0.9 ±0.2
Perfect linker helix + GAAA tetraloop	0.3 ±0.1
GC-rich perfect linker helix + GAAA tetraloop	0.3 ±0.2
Poly-A leader	0.5 ±0.1
Poly-A leader + GC-rich perfect linker helix + tetraloop	0.2 ±0.1

high SAH concentration (5 mM) showed that SAH shifted the RNA to the pre-binding form (but not SAM-bound form), while low SAH concentration (50 μM) has no such effect and the RNA adopts the non-binding form (Figure 3.5A). Further quantification, normalization and comparison of SHAPE reactivity at four domains on the S_{MK} RNA, we clearly demonstrated that even at saturating concentrations, SAH cannot completely induce the SAM-bound form (Figure 3.5B). This result also suggests that the free energy difference between non-binding and pre-binding form of the S_{MK} RNA is small enough that the equilibrium can be perturbed by a non-specific SAH incorporation into the binding pocket, which is in accord with our point-mutagenesis SHAPE results.

DISCUSSION

Our SHAPE analysis and SAM binding assays clearly demonstrated that the non-binding S_{MK} RNA adopts a distinct alternative conformation in solution. In order to bind SAM, the RNA has to unzip partially and reform a pre-binding conformation, an intermediate state between non-binding and SAM-bound form. The first evidence of pre-binding form was the S_{MK} riboswitch crystal structure bound to SAH (PDB code: 3e5e) (Lu et al., 2008). The crystallization construct was essentially locked into the pre-binding form by eliminating the leader sequence and stabilizing the linker helix, similar to our poly-A + perfect-linker construct (Figure 3.3A). In addition to SAM, the pre-binding form also associates with high concentration of SAH (Figure 3.5), in accord with previous crystallographic observations) (Lu et al., 2008).

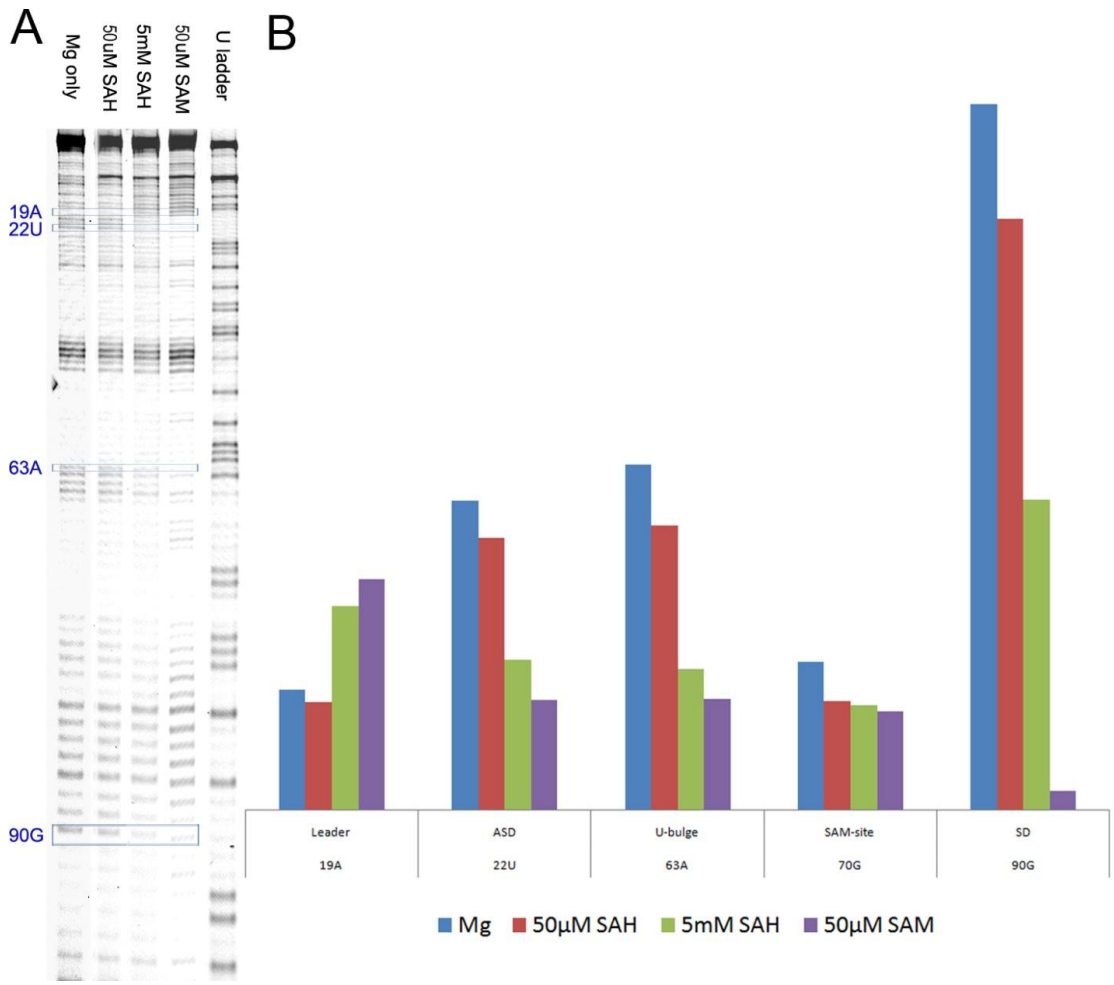
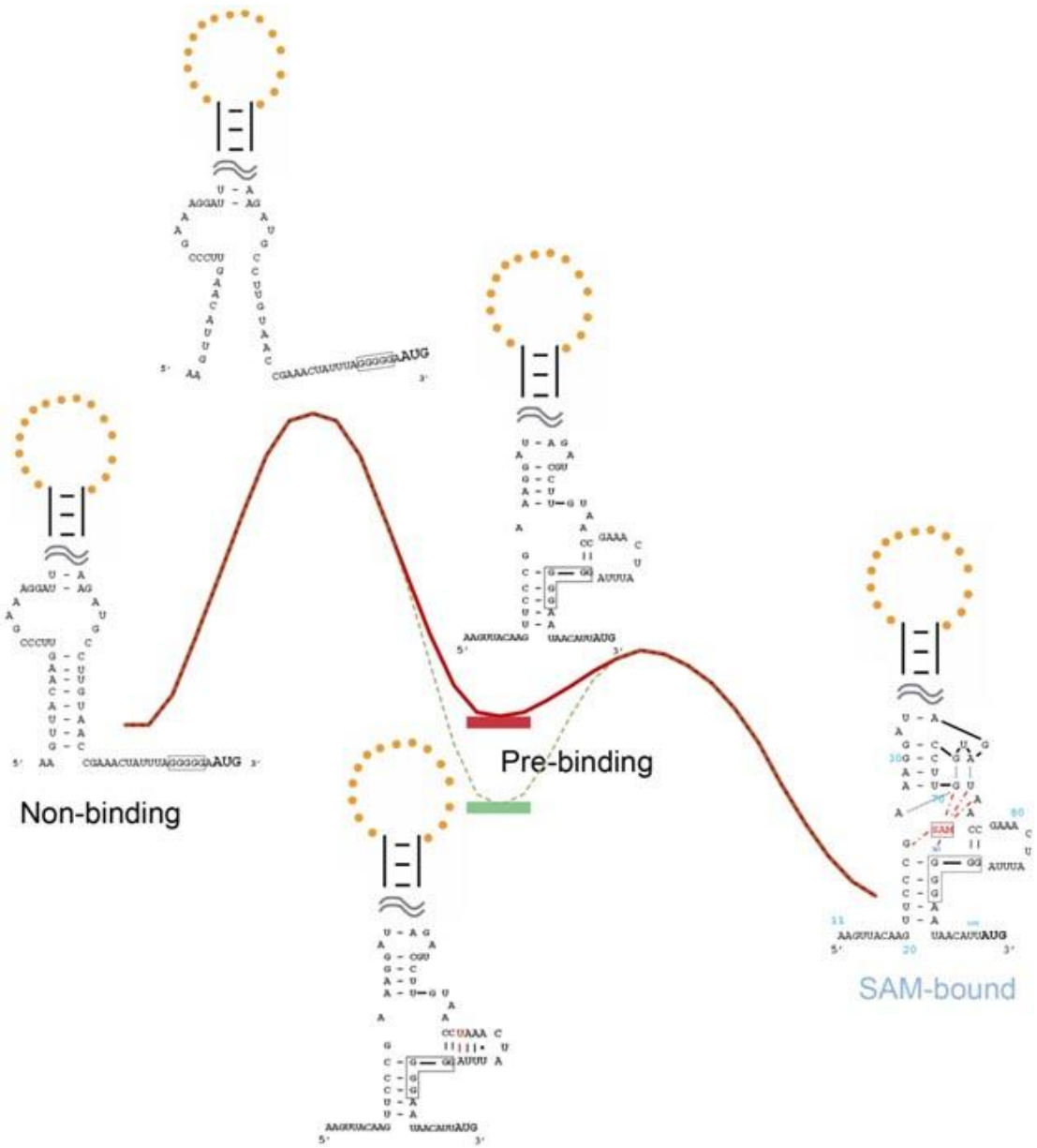


Figure 3.5. SAH induced pre-binding form at saturating concentration.

Left: Gel electrophoresis of SHAPE reactivity of the S_{MK} RNA in the presence of Mg2+ only, 50μM SAH, 5mM SAH and 50μM SAM; right: quantification of band intensity at five structural motifs show that SAH can induce the pre-binding form, but not the SAM-bound form, at saturating concentration. Gel loading sequence, from left to right: RNA+Mg^{2+}, RNA+Mg^{2+}+50 μM SAH, RNA+Mg^{2+}+5mM SAH, RNA+Mg^{2+}+50 μM SAM, ddU ladder

Figure 3.6. Proposed free energy landscape of the S_{MK} riboswitch associating with SAM.

Red trajectory: folding path of wild-type S_{MK} RNA with SAM; green dotted trajectory: G77U mutant RNA folding pathway with SAM. Proposed secondary structures of non-binding, pre-binding, SAM-bound and first transition state are shown. Five consecutive Gs in solid black boxes are the ribosome binding sites.


Non-binding
Pre-binding
SAM-bound

Apparently, intercalating the adenosine moiety of SAH into the SAM-binding site gains enough free energy to drive the equilibrium to pre-binding state, but not SAM-bound state (Figure 3.5), suggesting the free energy difference between the two forms is relatively small. Our model suggests that in solution, the non-binding S_{MK} RNA exist predominately as the non-binding form because the non-binding form is slightly more energetically favorable than the pre-binding form. Upon SAM introduction, the free energy of the SAM-RNA complex is much lower, driving the equilibrium away from the non-binding form. The ligand binding process of the S_{MK} RNA can be described as a double transition mechanism (Figure 3.6, red trajectory). We also speculate that the transition from non-binding to pre-binding is the rate-limiting step because it involves unzipping the leader helix (Figure 3.6).

Our results showed that by lowering the free energy of the pre-binding form below the non-binding form, we effectively shifted equilibrium towards the pre-binding form (Figure 3.6, green trajectory). The energy difference between wild-type pre-binding form and non-binding form is on the order of one base pair formation (two to three hydrogen bonds plus energy gain from correcting a mismatch), as demonstrated by our point mutation experiments (Figure 3.4). I believe that though evolution, nature has used this intricate balance to tweak the S_{MK} riboswitch sensitivities.

CHAPTER 4
THE HIERARCHICAL, SAM-DEPENDENT AND TEMPERATURE SENSITIVE STRUCTURAL TRANSITION IN THE S_{MK} RIBOSWITCH

SUMMARY

The S_{MK} box (SAM-III) RNA is a translational riboswitch identified in *S*-adenosylmethionine (SAM) synthetase *metK* genes in the Lactobacillales. It switches between two alternative conformations depending on the presence (translation "off") or absence (translation "on") of SAM and control gene expression level at the translation initiation level. We previously reported crystal structures of the S_{MK} box riboswitch and identified its folding pathway and intermediate. In this study we used SHAPE chemical probing analysis to characterize temperature effects on the conformational dynamics in S_{MK} RNA. We show that the S_{MK} RNA loses its ability to switch conformations rapidly at lower temperatures, suggesting that the RNA has evolved to operate rapidly only when metabolism is at optimum levels. Subsequent thermal melting experiments have shown that the S_{MK} RNA – SAM complex undergoes a two-step (ligand disassociation + overall structure melting), non-hierarchical unfolding process. Contrast to its SAM-I (S box) riboswitch counterpart, the S_{MK} RNA's relatively easier release of ligand might have physiological importance due to its translational regulatory role.

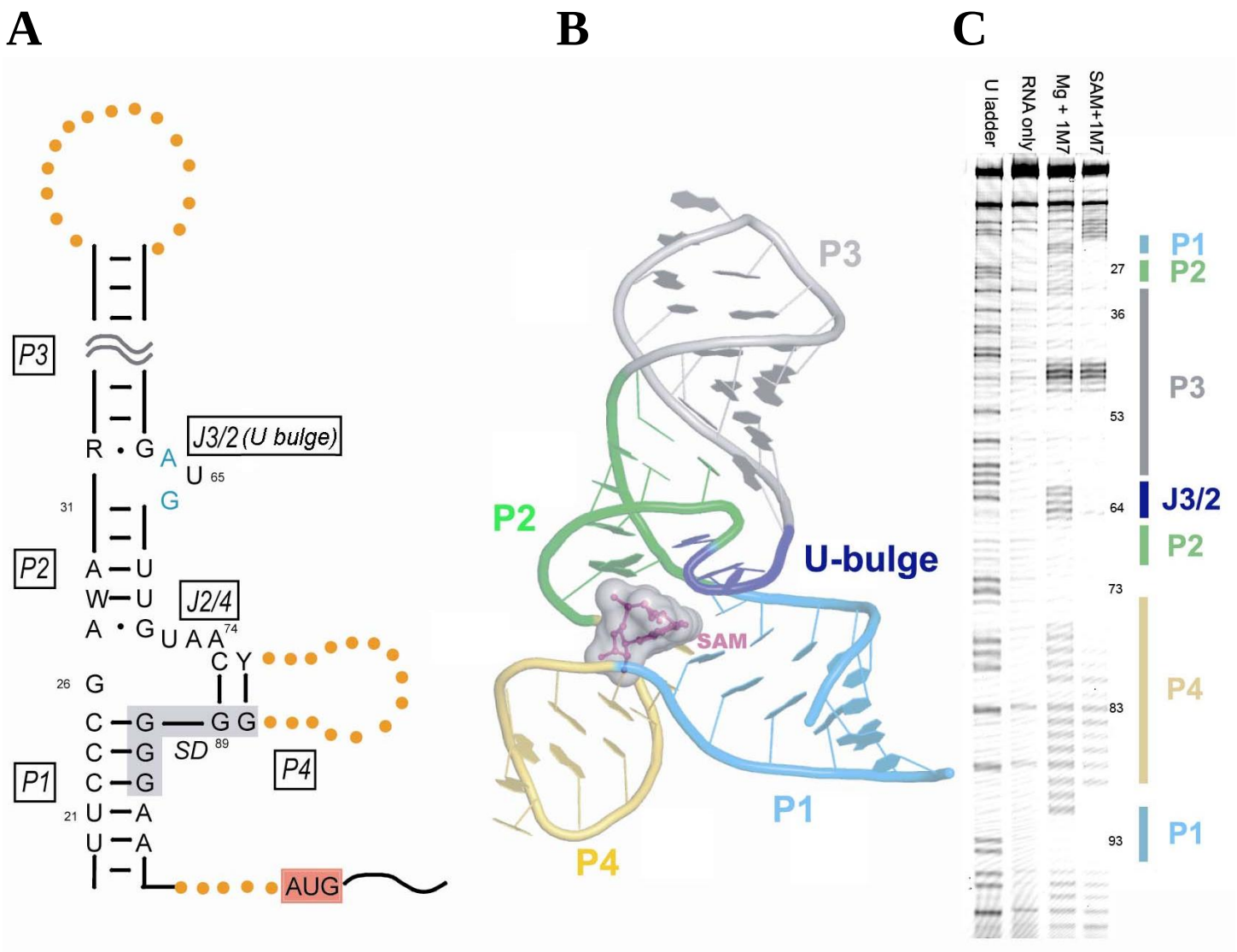
INTRODUCTION

Through structural changes induced by ligand binding, non-coding Riboswitch RNAs regulate a significant number of gene families that are involved in metabolism. They are especially prevalent in Gram-positive bacteria and a number of important pathogens (Tucker and Breaker, 2005, Grundy and Henkin, 2006, Vitreschak et al., 2004, Winkler and Breaker, 2005), among which is the SAM-binding S_{MK} riboswitch found in Lactobacillales. The S_{MK} box riboswitch regulates the translation of SAM synthetase (*metK*) genes in *Enterococcus* bacteria by sequestration of the Shine-Dalgarno (SD) sequence (Fuchs et al., 2006, Lu et al., 2008). Previous structural studies have shown that the RNA folds into a "Y" shaped arrangement (Figure 4.1A, B) that organizes an array of conserved nucleotides around a three-way junction for SAM recognition against SAH (Lu et al., 2008). SHAPE analysis that probes flexible RNA structures at single nucleotide resolution (Wilkinson et al., 2006) have provided additional information on the switching dynamics and completely agrees with the crystal structure (Figure 4.1C) (see Chapter 3 for details). Further chemical probing and mutation analysis have shown that and the RNA maintains its intricate balance of conformation switching through base-pairing schemes in the hypervariable sequences in the P3 stem, outside the SAM-binding domain (see Chapter 3).

After determining the mechanism through which the RNA controls its on/off equilibrium, the next logical step is to investigate environmental effects on the dynamics and stability of the riboswitch switch. Our temperature dependent SHAPE analyses have shown that the RNA has a narrow range of operational temperatures. Lowering the temperature reduces the RNA's ability to switch conformation between the unbound and SAM-bound form despite the presence of SAM, and increasing the temperature causes the RNA to quickly lose tertiary interactions. This result highly

Figure 4.1.Overall crystal structure of S_{MK} RNA bound to SAM.

(A) Secondary structure of the S_{MK} box riboswitch based on previous studies. Black capital letters, residues that are 100% conserved phylogenetically; blue, 50-85% conserved; R, G or A; W, A or U; Y, C or U. Beads, hypervariable loops; solid ladders, conserved secondary structure; gray and pink shaded areas, SD sequence and AUG start codon, respectively. (B) Cartoon representation of the crystal structure of the S_{MK} riboswitch. SAM is shown in overlapping CPK and surface representations in magenta and silver, respectively. The coloring scheme for the RNA is consistent within B, C and rest of the figures in this chapter. Sequencing gel detects conformational differences between SAM bound and unbound structures. Lane 1, ddUTP sequencing lanes; lanes 2-4, SHAPE analysis lanes; Base identities and structural motifs are marked on the right


A
B
C
P3
J3/2 (U bulge)
R • G
A
U 65
G
31
P2
A—U
W—U
A • G
J2/4
U A A 74
C Y
26 G
C—G — G G
C—G
SD 89
C—G
P1
21 U—A
U—A
P4
AUG
P3
P2
U-bulge
SAM
P1
P4
U ladder
RNA only
Mg + 1M7
SAM+1M7
27
36
53
64
73
83
93
P1
P2
P3
J3/2
P2
P4
P1

suggests that the S_{MK} RNA is fine-tuned to function within optimum growth temperatures.

MATERIALS AND METHODS

RNA Preparation for SHAPE analysis

The *Enterococcus faecalis* S_{MK} RNA with and without 3′ anti-terminator sequence flanked by 5′-T7 RNAP promoter and 3′-reversetranscription start site was constructed by overlapping PCR and inserted into pJet cloning vector. The DNA template for *in vitro* transcription was generated by 2 mL PCR reaction with terminal primers. The PCR product was directly used in T7 RNA RNAP run-off transcription as described (Ke and Doudna, 2004, Lu et al., 2008). The RNA product was purified and recovered by denaturing polyacrylamide gel electrophoresis as described (8% polyacrylamide, 8 M urea, 30 W, 3 hrs) (Ke and Doudna, 2004). Eluted RNA was concentrated and buffer exchanged into 0.5X TE using Millipore centrifuge filters with 10 kDa cutoff. All RNAs are subsequently diluted into RNA folding buffer (111 mM HEPES pH 8, 6.7 mM $MgCl_2$, 111 mM NaCl) with final RNA concentrations around 0.5 μM.

Chemical Probing

SHAPE procedure was similar as that outlined previously (Wilkinson et al., 2006, Wilkinson et al., 2005b, Merino et al., 2005). A BioRad thermocycler was used to incubate samples at specific temperatures. SAM-free RNA samples where first pre-heated to the desired temperature (4, 20 and 37 ℃) for one minute, proceeded by the addition and mixing of 1 mM (final) SAM for specified time (0, 5, 10, 15 and 30

minutes) before the addition of 1.5µl 100mM electrophile 1M7 to allow 2'-ribose hydroxyl alkylation. After the reaction is complete after 5 minutes on ice, manifested by the orange color of reaction contents, the reaction is then quenched with 900 ul precipitation buffer containing 80% ETOH, 45µM NaCl, 0.45µM EDTA, 2 µl glycoblue (15mg/mL, Ambion, Inc). The RNA in the reaction mixture was precipitated by incubation at –80 ℃ for 30 minutes, followed by centrifugation at 16,000 g at 4 ℃ for 45 minutes. The pellet was air dried and re-suspended in 12 µl of 0.5X TE buffer.

Primer Extension

The primer extension protocol is similar to that described previously (Wilkinson et al., 2006). Briefly, a fluorescently labeled DNA primer (HEX-GAACCGGACCGAAGCCCG; 1 µl, 50 µM) was annealed to 12 µl of the recovered RNA by incubation at 65 ℃ for 2 min then 35 ℃ for 5 min, followed by snap cooling to 4 °C for another 2 minutes. The resulting 13 µl RNA-primer mixture was mixed with 6 µl Reverse transcription buffer [167 mM Tris (pH 8.3), 250 mM KCl, 10 mM $MgCl_2$, 1.67 mM each dNTP] and pre-heated to 52 ℃ for 45 sec, then incubated with 1 µl of Superscript III (Invitrogen, 200 units) at 52 °C for 20 min. The reactions were quenched and the RNA template was digested by addition of 0.8 µl of 5M NaOH, followed by heating to 90 °C for 4 min. 29 µl of acid stop mix [4:25 (v/v) mixture of 1 M unbuffered Tris-HCl, and stop solution (85% formamide, 0.5X TBE, 50 mM EDTA, pH 8.0)] was added to each reaction, the reaction was allowed additional 4 min at 90 ℃ before cooled to -20 ℃. No internal tracking dye was used in the reaction to avoid interference with fragment analysis.

SHAPE Data Analysis

Data analysis was performed similar to previous studies (Watts et al., 2009). Primer extension reactions were visualized by polyacrylamide gel electrophoresis (PAGE) and SHAPE activity was quantified by capillary electrophoresis. For sequencing gels, 5-10 μl of reverse transcription reaction was loaded per lane onto denaturing acrylamide gels (10% acrylamide, 0.5X TBE, 8 M urea) and run at constant 60 watts for four hours. We used the program ShapeFinder (Vasa et al., 2008) to quantify the trace from fragment analysis to determine differences of intensities at regions of interest. Normalization of nucleotide reactivities consists of dividing the RNA into three segments, and each segment was normalized to its respective uniformly reactive positions (A12-U15 for nucleotides A11-C39; G47-U48 for A40-G77; G91-U95 for A78-G104). Intensities were rescaled to a unit (0 to 1) and plotted against increasing temperature.

RESULTS

The S_{MK} RNA acts as a Ribo-"switch" only at temperatures above 20 degrees Celsius

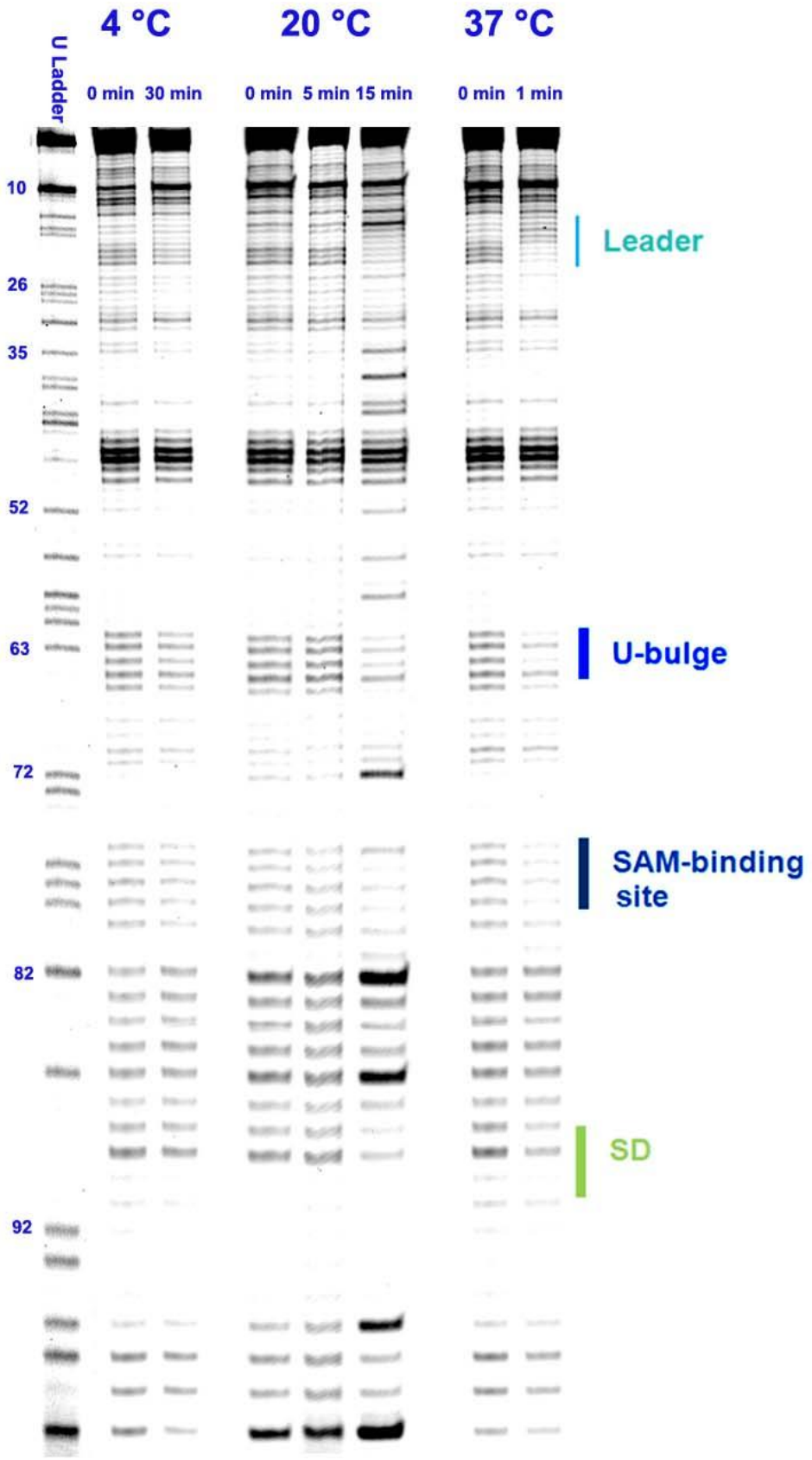
Previous biochemical studies and crystal structures (Fuchs et al., 2007, Fuchs et al., 2006, Lu et al., 2008) have shown that the apo S_{MK} RNA has to overcome the alternative base-pairing scheme to fold into the final SAM-bound conformation. This SAM-independent transition involves unzipping of the leader helix and the formation of the three-way-junction, which involves a significant energy barrier and may be the rate-limiting step in the entire SAM-binding process (see Chapter 3). To study the effects and implications of this energy barrier in the transition process, we performed time and temperature-dependent SHAPE experiments to investigate the effect of temperatures on the conformational dynamics of the S_{MK} RNA.

We first performed SHAPE probing at 37 ℃ to establish a baseline for the riboswitch. The RNA switched conformation from non-binding to SAM-bound within one minute of SAM exposure (Figure 4.2, right panel), manifested by change in base-pairing schemes at the leader sequence, and protection of U-bulge, SAM-binding site and SD sequence, consistent with previous results (see Chapter 3). The switching process can potentially be much faster, since one minute is the detection limit of 1M7(Mortimer and Weeks, 2007, Mortimer and Weeks, 2009). This result indicates that organism that utilizes the S_{MK} riboswitch requires near-instant response from SAM stimulation. Combined with the fact that the S_{MK} RNA is translational riboswitch, its fast response to SAM concentration increase at physiological temperature make it capable to shuts down SAM biosynthesis minutes after the desired SAM concentration has been reached.

On the other hand, at 4 ℃, our SHAPE experiments showed that the S_{MK} RNA is almost non-responsive to SAM, as it remained at the non-binding conformation even after 30 minutes of exposure to 1mM SAM. The conformations between 0 minutes and 30 minutes after SAM addition are identical as revealed by SHAPE activities (Figure 4.2, left panel). This means that the energy barrier involved in switching is big enough that at 4 ℃, the population of RNA molecules that possess the activation energy is minimal.

Figure 4.2. SHAPE analyses show that S_{MK} RNA respond to SAM only at physiological temperatures.

Base identities are labeled next to the sequencing lane on the left. Important structural motifs are marked on the right. Temperatures and time points are marked on the top


4 °C
20 °C
37 °C
U Ladder
0 min
30 min
0 min
5 min
15 min
0 min
1 min
10
26
35
52
63
72
82
92
Leader
U-bulge
SAM-binding site
SD

We proceeded to find a temperature between 4 ℃ and 37 ℃ at which the S_{MK} RNA switches conformation slowly enough to be observed. At temperatures below 15 ℃, no switching action can be observed even after one hour of incubation with SAM (data not shown). At 20 ℃, roughly 80% of the S_{MK} RNA partially folded into the SAM-bound form after 15 minutes of exposure to SAM, while having no visible change at T = 5 minutes mark (Figure 4.2, middle panel). The 20 ℃, 15-minute lane displayed changes in leader base-pairing, and reduced activities in the U-bulge, SAM-binding site and SD sequence, similar to the 37 ℃, 1-minute lane (compare Figure 4.2, left and right panels). However, at 20 ℃, the RNA has developed many additional bands in helical regions, compared with its 37 ℃ counterpart (Figure 4.2, right panel). This is an indication that the RNA is at least partially mis-folded, leaving many bases flipped out or unpaired. Nevertheless, most of the RNA did overcome the activation energy barrier and is partially functional at 20 ℃, which can be physiologically important for the organism's metabolism at that temperature.

In summary, the S_{MK} riboswitch is a temperature-dependent switch, and the conformational dynamics is extremely sensitive to temperature change. The riboswitch appears to be tuned to only work efficiently above room temperature with optimum temperature at physiological 37 ℃.

The S_{MK} RNA unfolds in a non-hierarchical, two-step process with increasing temperatures

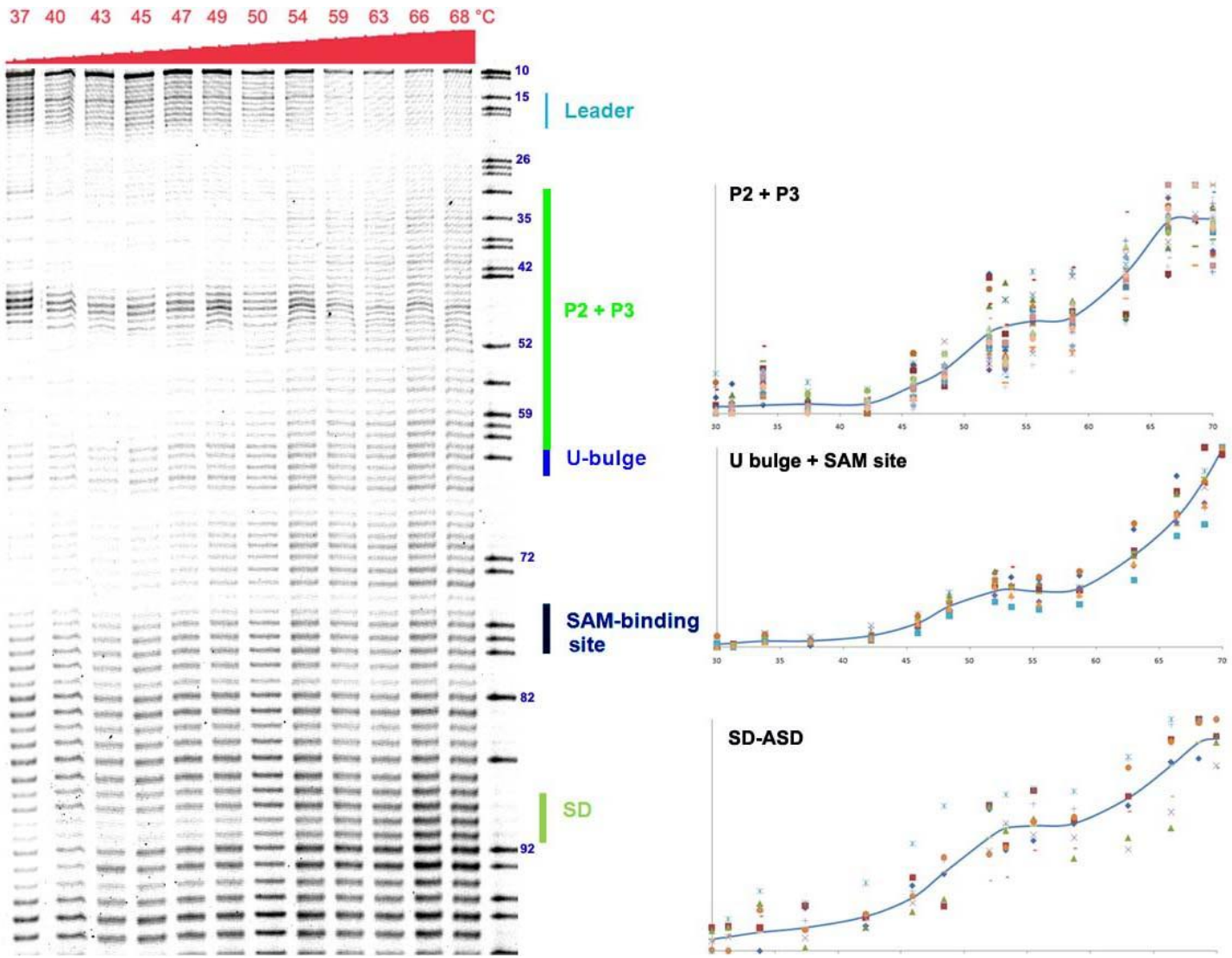
Crystallographic analysis (Lu et al., 2008) describes the SAM-bound S_{MK} RNA as an SAM-stabilized three-way junction. Part of SAM, namely the methionine tail, is not enclosed in the RNA and is exposed to the solvent. This organization can potentially play an important in the stability and dynamics of the RNA-SAM complex.

In order to investigate the unfolding/disassociation process of the S_{MK}-SAM complex, we employed the temperature melting SHAPE technique to study the unfolding process of the S_{MK} RNA complex with SAM at nucleotide resolution ((Wilkinson et al., 2005a) and Chapter 3 in this thesis).

Compared with the thermal melting process of other organized RNA structures, such as t-RNA (Wilkinson et al., 2005a) and the SAM-I (S box) RNA (see Chapter 5), the S_{MK} RNA-SAM complex displays a 2-staged unfolding pattern that has not been demonstrated previously (Figure 4.3). Quantification of SHAPE activities show that, as the temperature increases from 37 ℃ to 70 ℃, the entire RNA structure unfolds at roughly the same rate, but the unfolding curve resembles a double-sigmoidal curve. The first transition occurs at T_m=47 ℃ for SAM-binding site and U-bulge (~50 ℃ for P2/P3 and SD-Anti-SD helix), while the second and final transition is at T_m~65 ℃ when almost all secondary structures unfold (Figure 4.3). The individual structure elements, such as the SAM-binding site, P2/P3, U-bulge and SD-ASD helix exhibit some but limited difference in stability, since SAM-binding site and U-bulge melt 2~3 ℃ earlier than the rest of the structure. The distal P3 exhibits similar melting behavior as P2, albeit making no direct contact with SAM. This is probably because distal P3 stem is stabilized by the U-bulge, which forms SAM-dependent tertiary H-bonds with G71 and U72. Therefore, the stability of both P2 and P3 are directly related to SAM-binding in the same fashion. The above results suggest that the melting the S_{MK} RNA is non-hierarchical, which is also distinct from previous observations on SAM-binding RNA structures (see Chapter 5).

Figure 4.3. SHAPE detects a two-step S_{MK} thermal melting process.

Left, temperature-dependent SHAPE gel analysis of the S_{MK} aptamer domain bound to SAM. RNA structures are indicated at right. The dideoxy-U sequencing ladder is the left-most lane. Right, nucleotide unfolding profiles computed by quantifying intensities shown in gel by fragment analysis and processed by program ShapeFinder (Vasa et al., 2008). Data are plotted on a linear scale; solid blue line indicates trends of average behavior.


37 40 43 45 47 49 50 54 59 63 66 68 °C
10
15
26
35
42
52
59
72
82
92
Leader
P2 + P3
U-bulge
SAM-binding site
SD
P2 + P3
U bulge + SAM site
SD-ASD

Considering the three-dimensional structure and unique ligand binding mechanism of the S_{MK} riboswitch, it is distinctly possible that upon increasing temperatures, tertiary interactions mediated by SAM, such as hydrogen bonds that recognize SAM and electrostatic interactions that mediates the sulfonium ion and the U-bulge, are disrupted at 47 ℃. At that temperature, the RNA structure still possess the three-way-junction overall fold, but no longer has any affinity to SAM (Figure 4.4). As the temperature continues to rise, the helices that hold the three-way junction begin to break down at 65 ℃ (Figure 4.4). This two-staged melting of the S_{MK} RNA bound to SAM suggests that the riboswitch is probably not designed to hold the ligand tightly, and that the complex can easily reform as temperature decreases to physiological levels.

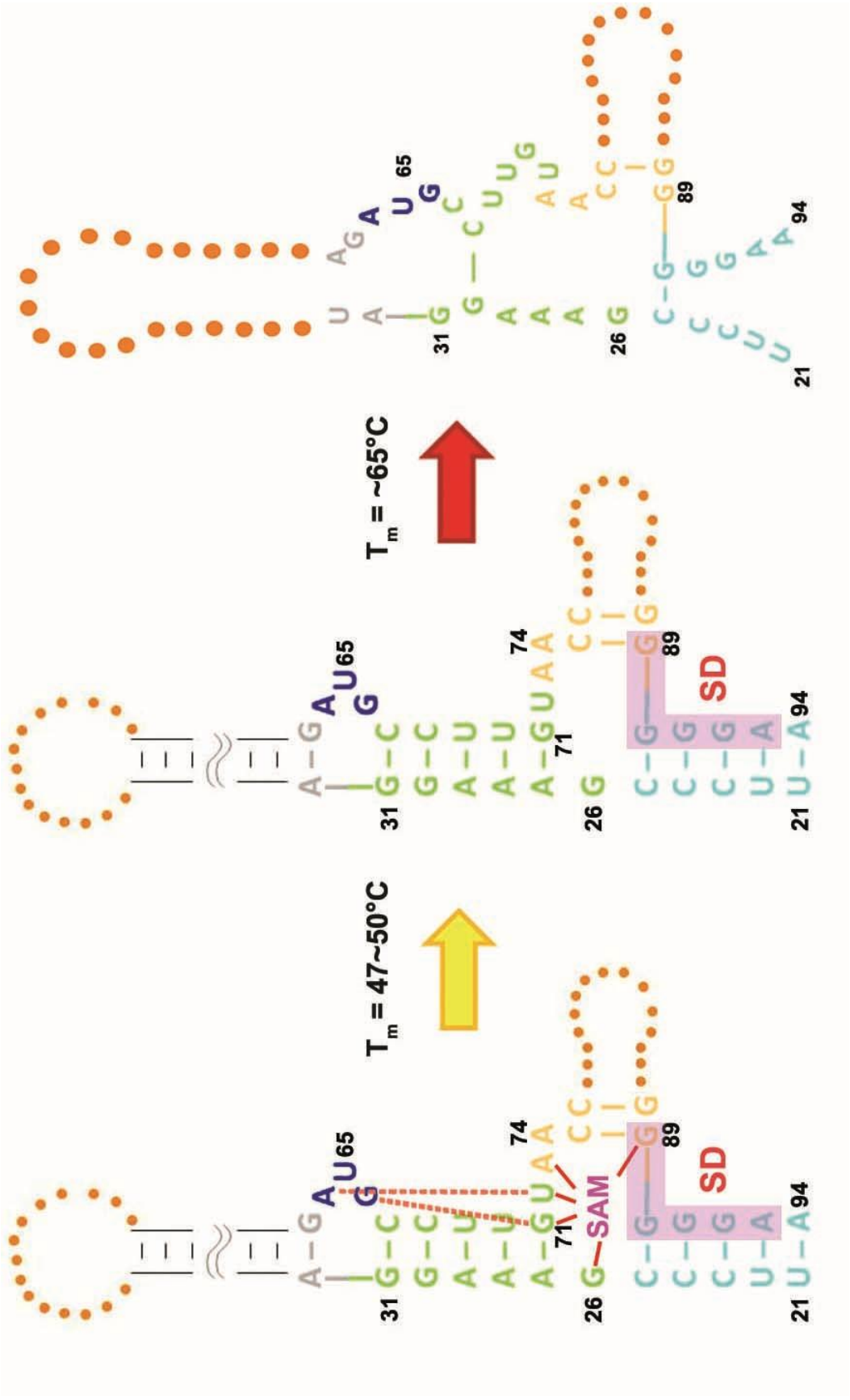
DISCUSSION

S_{MK} RNA provides rapid response to changing SAM concentration

Biochemical and structural studies on the S_{MK} RNA have provided many insights on how the RNA recognizes the ligand and many details on the mechanism and dynamics of the "switching" process. The free RNA samples two different conformations, "non-binding" and "pre-binding" forms, and the "pre-binding" form binds SAM at the three-way junction. Being a riboswitch that operates at translational level, the S_{MK} RNA is designed to provide rapid response to environmental cues. Structures of the S_{MK} RNA have demonstrated that SAM has a relatively direct route to enter and leave the RNA, which is the first clue that the riboswitch utilizes a lock and key mechanism in ligand recognition, and has relatively fast association/disassociation rates. The second clue comes from a previous discovery and isolation of a "pre-binding" S_{MK} RNA intermediate from mutagenesis and SHAPE

Figure 4.4. Proposed model of S_{MK} RNA/SAM complex thermal unfolding pathway.

First transition, loss of tertiary interactions and ligand binding. Second transition, loss of nearly all secondary structures.


SAM
SD
SD
T_m = 47~50°C
T_m = ~65°C

analyses (Chapter 3), in which the RNA does not adopt the alternative base-pairing described earlier (Fuchs et al., 2007, Fuchs et al., 2006), and folds into a SAM-like overall structure that is ready to accept SAM binding. The third evidence is that at physiological temperature, the activation energy between the "non-binding" and "pre-binding" form is easily overcome at physiological temperatures. Combined with its ability to instantly start the translation process by releasing the mRNA for ribosomal binding, the S_{MK} RNA has been evolved to rapidly respond to changes in SAM concentrations.

The S_{MK} RNA is a fine-tuned temperature switch

In our studies, the S_{MK} riboswitch displayed temperature sensitivity through its ability to switch conformation as well as stability at higher temperatures. The conformational switch between "on" and "off" is not observed at temperatures below 15 ℃, and the riboswitch starts to lose structural integrity above 45 ℃. Therefore, the S_{MK} RNA is not only SAM-riboswitch, but also a temperature-switch that has different behaviors at low and high temperatures. At below 15 ℃, the S_{MK} riboswitch delays the inhibition of SAM-biosynthetic pathway by not rapidly respond to increasing SAM concentration. The delayed response is in accord with the relatively low rate of metabolism at that temperature, and a fast response would be obviously unnecessary. At temperatures above 45 ℃, the RNA partially loses its tertiary structure, reducing its ability to inhibit translation, perhaps as part of the response to heat-shock. Throughout evolution, the S_{MK} riboswitch has been fine-tuned to respond differently according to temperature as well as SAM concentrations.

CHAPTER 5

SAM RECOGNITION MECHANISM REVEALED BY THE CRYSTAL STRUCTURE OF THE *BACILLUS SUBTILIS YitJ* SAM-I/S BOX RIBOSWITCH

SUMMARY

S box (SAM-I) riboswitches are the most widespread class of riboswitches involved in regulating methionine or cysteine biosynthesis in Gram-positive bacteria. Here we report the 3.0 Å crystal structure of the aptamer domain of the *Bacillus subtilis YitJ* S box (SAM-I) riboswitch bound to *S*-Adenosyl-*L*-methionine (SAM)**.** The RNA folds into two sets of helical stacks spatially arranged by tertiary interactions including a K-turn and a pseudoknot at the four-way junction. The tertiary structure is further stabilized by extensive ribose zipper H-bonds between helices and metal coordination. Despite structural differences on the periphery, the SAM binding core of the *YitJ* S box, located at the interface of P1/P4 and P2/P3 coaxial stacks, is virtually superimposable with the previously determined *T. tengcongensis YitJ* riboswitch structure. An identical set of interactions specify the binding of a U-shaped SAM, but not *S*-Adenosyl-homocysteine (SAH), suggesting that a highly conserved ligand recognition mechanism is utilized by all SAM-I riboswitches, including those found in the human pathogens such as *Bacillus anthracis*, *Staphylococcus*, *Enterococcus*, *Streptococcus*, *Listeria*, *Clostridium*, and *Mycobacterium*. Our structural and chemical probing studies show that SAM-independent interactions are responsible for the switching dynamics and the hierarchical stability among different S box RNA structural elements..

INTRODUCTION

Riboswitches are short *cis*-acting RNAs that regulate gene expression in response to environmental cues (reviewed in (Tucker and Breaker, 2005, Grundy and Henkin, 2006, Vitreschak et al., 2004)). They usually reside in the 5' untranslated region (UTR) of bacterial genes/operons and directly interact with small molecule metabolites to regulate downstream gene expression at either the transcriptional or translational level. Bacteria use riboswitches to control vitamin metabolism, nucleotide biosynthesis, amino acids, sulfur metabolism, and metal transport (reviewed in (Serganov, 2009, Henkin, 2008, Mandal and Breaker, 2004b, Nudler and Mironov, 2004))**.** Two examples of riboswitches were found in eukaryotes to regulate alternative splicing and the stability of the host gene (Cheah et al., 2007, Wachter et al., 2007). Some riboswitches, such as the T-box riboswitch that recognizes both uncharged and charged tRNA (Grundy et al., 2005, Yousef et al., 2005, Grundy and Henkin, 2003, Yousef et al., 2003, Grundy et al., 2002b), monitors the substrate:product ratio (Yousef et al., 2005). Other metabolite-sensing riboswitches only specifically respond to the substrate, detecting the absolute amount of the metabolite.

Among more than a dozen distinct riboswitch classes that respond to various metabolites, the *S*-adenosyl-*L*-methionine (SAM)-recognizing SAM-I (S box) riboswitch class is arguably the most prevalent class (Barrick and Breaker, 2007, Blount and Breaker, 2006), especially in Gram-positive bacteria genomes. S box riboswitches were originally identified and extensively studied in *Bacillus subtilis* (Grundy and Henkin, 1998, McDaniel et al., 2003, McDaniel et al., 2005), where it controls more than 26 genes involved in methionine import, synthesis and recycling

system (Grundy and Henkin, 1998, Tomsic et al., 2007, Winkler et al., 2003). The S box (SAM-I) riboswitch is composed of a metabolite-sensing (in this case the SAM molecule) "aptamer domain" followed by an "expression platform" that regulates downstream gene expression mainly using transcription attenuation mechanisms (Grundy and Henkin, 1998, Tomsic et al., 2007, Winkler et al., 2003), and translational control in some cases (Weinberg et al., 2008, Weinberg et al., 2007). An alternative base pairing scheme between the aptamer domain and expression platform is regulated by the intracellular concentration of SAM. High SAM concentration induces a tightly folded structure in the aptamer domain which in turn allows the downstream formation of a transcription terminator to inhibit transcription (Grundy and Henkin, 2006, Henkin and Yanofsky, 2002, Montange and Batey, 2006, Tomsic et al., 2007, Tomsic et al., 2008, Winkler, 2005b, Winkler et al., 2003). When intracellular concentration of SAM is low, the aptamer domain is destabilized, which allows the formation of an antiterminator stem loop, allowing the RNA polymerase to read-though and produce full-length functional transcripts (Epshtein et al., 2003, Winkler et al., 2003, Grundy and Henkin, 1998, McDaniel et al., 2005). The *B. subtilis yitJ* S-box riboswitch exhibits high affinity for SAM (Kd of ~ 10 nM) and strongly discriminates against closely related natural analogs such as *S*-Adenosyl-*L*-homocysteine (SAH, 100-fold lower) and *S*-Adenosyl-*L*-cysteine (SAC, nearly 10,000-fold lower) (Winkler et al., 2003, McDaniel et al., 2003). Phylogenetic and mutagenesis analyses revealed that the S box aptamer domain consists four helical segments organized around a four-way junction, among them P2 and P4 were shown to participate in a pseudoknot formation (Winkler et al., 2003, Grundy and Henkin, 1998).

The crystal structure of a 94-nt S box (SAM-I) riboswitch from a hemophilic Gram-negative bacteria *Thermoanaerobacter tengcongensis* has been reported (Montange and Batey, 2006, Montange et al., 2009). The structure shows that the phylogenetically conserved helices organize into two sets of coaxial helices packing against each other at a tilted angle, wrapping SAM at helical packing interface (Montange and Batey, 2006, Montange et al., 2009). Mutagenesis studies on the *T. tengcongensis YitJ* SAM-I riboswitch further confirmed that the riboswitch distinguishes SAM from SAH by specifying the positively charged sulfonium ion in SAM (Montange et al., 2009). Here we report the 3.0 Å, 119-nt *Bacillus subtilis YitJ* S box aptamer RNA, the S box riboswitch that has been most carefully characterized in both biochemical and genetic studies. We show that while these two SAM-I riboswitches differ in their peripheral regions, their SAM-binding sites are almost superimposable, suggesting that the S box/ SAM-I riboswitch family very likely uses a conserved ligand-recognition mechanism. The presence of extensive SAM-independent tertiary interactions and the SHAPE chemical probing results of the apo aptamer domain points to a transient "pre-binding" conformation, which subsequently recruits a SAM molecule through J1/2 and J3/4 to arrive at the final SAM-bound conformation. Our temperature-dependent SHAPE analysis revealed an extremely stable SAM-binding core structure as the result of synergistic effects of both SAM-dependent and -independent tertiary interactions.

MATERIALS AND METHODS

RNA preparation for crystallization

The aptamer domain of the *Bacillus subtilis YitJ* S-box riboswitch gene (nt 29-146) was inserted into a pUC19 plasmid under the control of the T7 RNAP promoter, and flanked at the 3'-end by the hepatitis delta virus ribozyme to ensure a homogeneous 3'-end after ribozyme cleavage. RNA was prepared by the standard T7 polymerase run-off transcription using linearized plasmid as template and purified by denaturing gel electrophoresis as described (Ke and Doudna, 2004). SAM was purchased from Sigma and stored as 100 mM stock solution (pH < 7) at -80 ℃ until immediately before usage.

Before crystallization, the S box RNA was refolded by heating up to 65 ℃ for 10 minutes in a buffer containing 25 mM Tris-HCl pH 7.5, 40 mM NaCl, and 5 mM $MgCl_2$, at which point SAM was added to the final concentration of 1 mM, and the solution was allowed to cool gradually to 20 ℃ over the course of an hour and stored frozen at -80 ℃. We minimized the effect of spontaneous SAM degradation by adding 1mM final SAM concentration in all steps during the crystal cryoprotection and freezing process. The wild-type S box aptamer domain constructs produced crystals that only diffracted to ~6 Å resolution. The distal P3 loop was found to have undergone spontaneous cleavage during crystallization process (data not shown). This observation led to the design of three constructs in which the flexible P3 loop was replaced with a stable GAAA tetraloop and the base pairs below were systematically shortened; these construct retained wild type SAM-binding activity. One of the resulting constructs, designated TL5, produced a rare crystal form in 40 mM sodium cacodylate, pH 7.0, 80 mM Strontium chloride, 15% MPD, and 2 mM spermine after

prolonged incubation, which diffracted X-ray to ~ 3 Å (Table 5.1). The crystals were soaked in increasing concentrations of MPD (up to 25%) before flash-frozen in liquid nitrogen.

Molecular replacement and structure refinement

A 3 Å data set was collected from a TL5 S-box crystal using beamline F1 at Cornell High Energy Synchrotron Source (CHESS). Data were processed in hexagonal space group $P3_121$ with unit cell dimensions of 58.732 Å by 58.732 Å by 204.091 Å, and γ=120° (Table 5.1). We solved the S-box structure by molecular replacement using PHASER (Storoni et al., 2004). First, the core *T. tengcongensis* SAM-I structure coordinates (Montange and Batey, 2006) (PDB code: 2GIS), excluding SAM and the distal portion of P3 helix region, was used successfully in rotation and translation searches using PHASER (Storoni et al., 2004). A 17-nt ideal A-form stem loop capped by GAAA tetraloop was then positioned at the location of the distal P3 helix. The initial electron density map was poor due to large phase errors and relatively high R/R_{free} (48%/44%). Repeated rounds of density modification with model phase inputs, rigid body refinement, and harmonically–restrained simulated annealing refinement procedures in CNS (Brunger et al., 1998, Brunger, 2007) reduced the R_{work}/R_{free} to 40%/44% and allowed tracing of most of the sugar-phosphate backbone. Phylogenetically conserved base pairing was reinforced by specifying Watson-Crick base pairs throughout CNS refinement (Brunger, 2007, Brunger et al., 1998). To further reduce the model bias and phase errors, the original search model was refined for several rounds as a poly-cytosine model, which produced an electron density map of sufficient quality to distinguish purines from pyrimidines. This map allows accurate

Table 5.1. Crystallographic statistics

Data collection and refinement statistics	
Data collection	CHESS F1
Space group	$P3_121$
Unit cell parameters	
a, b, c (Å)	a = 58.732, b = 58.732, c = 204.091
α, β, γ (°)	α = 90, β = 90, γ = 120
Wavelength (Å)	
Resolution (Å)	20–3.0
Average I/σ(I)	
Reflections	8152 (931)
Completeness (%)	93.5
R_{merge} (%)	
Refinement	
Resolution (Å)	20–3.0
Number of reflections used in the working set	7350

Table 5.1 (continued)

Number of reflections used in the testing set	802
Average B-factor (Å^2)	72.4
R-factor (%)[c]	23.6
Free R-factor (%)	28.6
r.m.s.d.	
r.m.s.d. bond lengths (Å)	0.007
r.m.s.d. bond angles ( )	1.57

* $R_{sym} = \sum| I - \langle I \rangle | / \sum I$, where I is the observed intensity and $\langle I \rangle$ is the statistically weighted absolute intensity of multiple measurements of symmetry related reflections.

§ $R = \sum|F_o - k|F_c|| / \sum|F_o|$, R from the working set and R_{free} from the test set.

‡ $R_{cullis} = \sum||F_{PH} \pm F_P| - F_H(calc)| / \sum|F_{PH}|$ reported for all centric reflections.

ζ Mean figure of merit = $\langle \sum P(\alpha) e^{i\alpha} / \sum P(\alpha) \rangle$, where α is the phase and $P(\alpha)$ is the phase probability distribution.

modeling of bases in COOT (Emsley and Cowtan, 2004) and restrained refinement in Refmac5 (Collaborative Computational Project, 1994) until the model is ~93% complete and R_{free} is below 35%. The final model (R_{work}/R_{free} to 29%/33%) was generated after iterative cycles of group B-factor refinement, energy minimization and simulated annealing in CNS (Brunger et al., 1998, Brunger, 2007). Although the coordinates of SAM, and metal ions were never included in molecular replacement and structure refinement, f_o-f_c difference density corresponding to the SAM molecule and two ordered magnesium ions were clearly identified and refined using real-space fitting in COOT (Emsley and Cowtan, 2004) after the completion of the 118-nucleotide RNA model. Electron densities revealed all but four flipped out nucleotides in P3 stem and the P4 loop. The final refinement statistics is summarized in Table 5.1 with R_{work}/R_{free} of 0.236/0.286. Eight water molecules were located and refined by CNS (Brunger et al., 1998, Brunger, 2007). The entire structure was verified using simulated annealing composite-omit maps by excluding 5% of the model in each calculation (Brunger, 2007, Brunger et al., 1998).

RNA Preparation for SHAPE analysis

The *Bacillus Subtilis YitJ* S box RNA with or without the 3′ expression platform (containing the anti-terminator sequence) were inserted into SHAPE cassette containing a 5’ T7 promoter and a 3′-reverse transcription (RT) primer site was constructed by overlapping PCR and inserted into a pJet cloning vector. The SHAPE cassette DNA template were PCR amplified using terminal primers and used directly to produce RNAs for SHAPE analysis in *in vitro* transcription as described (Ke and Doudna, 2004, Lu et al., 2008). The RNA product was purified and recovered by denaturing polyacrylamide gel electrophoresis as described (Ke and Doudna, 2004).

All RNAs are subsequently diluted into RNA folding buffer (111 mM HEPES pH8, 6.7mM $MgCl_2$, 111mM NaCl) with final RNA concentrations around 0.5 μM. Immediately prior to chemical modification, the diluted RNA is refolded as previously described (Lu et al., 2008). For –SAM conditions, no SAM is added during the annealing process.

Chemical Probing

SHAPE procedure was similar as that outlined previously (Wilkinson et al., 2006, Wilkinson et al., 2005b, Merino et al., 2005). Briefly, 15ul of the refolded S box RNA (+/- SAM) from was equilibrated at a specific temperature (20 ℃, 30.0 ℃, 33.8 ℃, 43.9 ℃, 48.4 ℃, 50.0 ℃, 52.0 ℃, 53.3 ℃, 58.7 ℃, 58.9 ℃, 63.0 ℃, 66.4 ℃, 68.6 ℃) for one minute, and subsequently mixed with 1.5 μl 100mM electrophile 1M7 to allow 2'-ribose hydroxyl alkylation. After the reaction is complete (related to half life of 1M7 at different temperatures, see (Wilkinson et al., 2005b, Wilkinson et al., 2009) for reference), manifested by the orange color of reaction contents, the reaction is then quenched with 900 ul precipitation buffer containing 80% ETOH, 45 μM NaCl, 0.45 μM EDTA, 2 μl glycoblue (15mg/mL, Ambion, Inc). The RNA in the reaction mixture was precipitated by incubation at –80 ℃ for 30 minutes, followed by centrifugation at 16,000 g at 4 ℃ for 45 minutes. The pellet was air dried and re-suspended in 12 μl of 0.5X TE buffer.

Primer Extension

The primer extension protocol is similar to that described previously (Wilkinson et al., 2006). A fluorescently labeled DNA primer (HEX-

GAACCGGACCGAAGCCCG; 1 μL, 50 μM) was annealed to 12 μL of the recovered RNA by incubation at 65 ℃ for 2 min then 35 ℃ for 5 min, followed by snap cooling to 4 °C for another 2 minutes. The resulting 13 μL RNA-primer mixture was mixed with 6 μL Reverse transcription buffer [167 mM Tris (pH 8.3), 250 mM KCl, 10 mM $MgCl_2$, 1.67 mM each dNTP] and pre-heated to 52 ℃ for 45 sec, then incubated with 1 μL of Superscript III (Invitrogen, 200 units) at 52 °C for 20 min. The reactions were quenched and the RNA template was digested by addition of 0.8 μL of 5M NaOH, followed by heating to 90 °C for 4 min. 29 μL of acid stop mix [4:25 (v/v) mixture of 1 M unbuffered Tris-HCl, and stop solution (85% formamide, 0.5X TBE, 50 mM EDTA, pH 8.0)] was added to each reaction, the reaction was allowed additional 4 min at 90 ℃ before cooled to -20 ℃. No internal tracking dye was used in the reaction to avoid interference with fragment analysis.

SHAPE Data Analysis

Each result of primer extension reactions was evaluated by both polyacrylamide sequencing gel electrophoresis and capillary electrophoresis/fragment analysis. To visualize band intensities, 5-10 μL of each reverse transcription reaction was separately loaded onto lanes of 10% denaturing sequencing gels (10% 29:1 acrylamide:bis acrylamide, 1X TBE, 8 M urea) and run at constant 55 watts. In order to obtain best separation for both high and low molecular weight frangmenst, we perform two scans from a Typhoon phosphor-imager (Molecular Dynamics) for every gel image. First image to see low molecular weight bands was scanned after 2.5 hours of running, and a second image for higher molecular weight fragments was obtained after 5 hours of running.

Our quantification analysis used results from fragment analysis (Applied BioSystems 3730xl DNA Analyzer) to determine differences of intensities at regions of interest. Traces from fragment analysis were quantified by program ShapeFinder (Vasa et al., 2008). Nucleotide reactivities were scaled to uniformly reactive loop positions (G61-A63 for nucleotides 1-C66; A96-C100 for A68-C118). Strong signals caused by pausing are manually eliminated from the data. For temperature melting data, intensities were rescaled to a unit (0 to 1) and fitted to the following equation assuming a unimolecular transition (John and Weeks, 2000) using the program Origin (Origin Lab):

$$I = A\frac{1}{1+\left[\exp\left[\frac{\Delta H_{\mathrm{vH}}}{R}\left(\frac{1}{T_{\mathrm{m}}}-\frac{1}{T}\right)\right]\right]^{-1}} + b$$

where I is SHAPE activity at a given temperature (T); R is the gas constant, and A and b are the maximum and initial intensity, respectively.

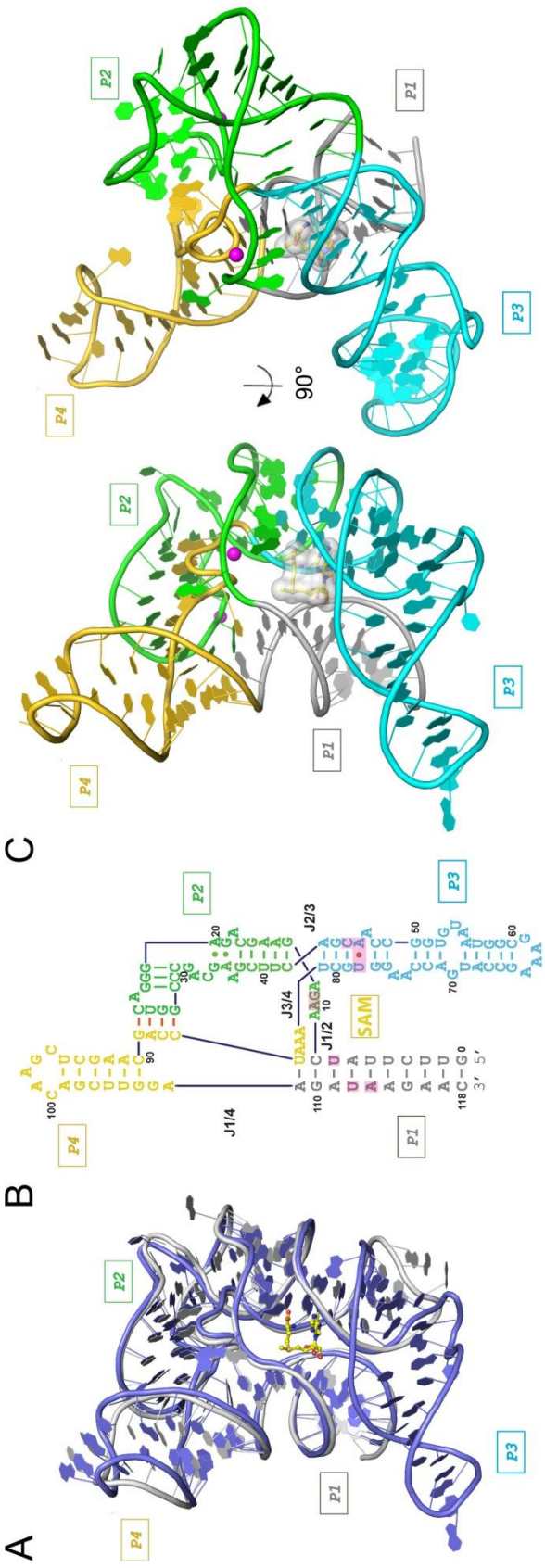
RESULTS

Comparison of the B. subtilis and T. tengcongensis S box riboswitch structures

The overall architecture of the 119-nt *B. subtilis YitJ* S box RNA is similar to that of the 94-nt *T. tengcongensis YitJ* SAM-I riboswitch (PDB code: 2GIS, 3GX5) (Montange and Batey, 2006, Montange et al., 2009), despite a large deviation in the equivalent phosphorous atoms (r.m.s.d. = 2.5 Å) (Figure 5.1A). The four helices defined by secondary structure prediction programs assemble into *two sets of* coaxial

Figure 5.1. Overall crystal structure of S box RNA bound to SAM.

(A) Cartoon diagrams of structure alignment between *Bacillus subtilis YitJ* S box (SAM-I) riboswitch (grey) and *T. tengcongensis YitJ* riboswitch (blue) (Montange and Batey, 2006). (Equivalent phosphate r.m.s.d. = 2.5 Å) (B) Secondary structure of the *Bacillus subtilis YitJ* S box (SAM-I) riboswitch. Helices P1 through P4 are colored in gray, green, cyan, and yellow, respectively. Red dashes are pseudoknot basepairs. Magenta shades label SAM-contacting bases. Red dot labels the sheared A-U basepair which recognizes adenosine face of SAM. (C) Cartoon representations of the S box RNA structure. The two views represent a front view (left) and a 90 °rotation (right). The labeling and base coloring schemes are consistent with that in Figure 5.1(B). SAM is shown as ball-and-stick model in overlapping CPK and surface representations in grey. Two magnesium metal ions are shown in pink.

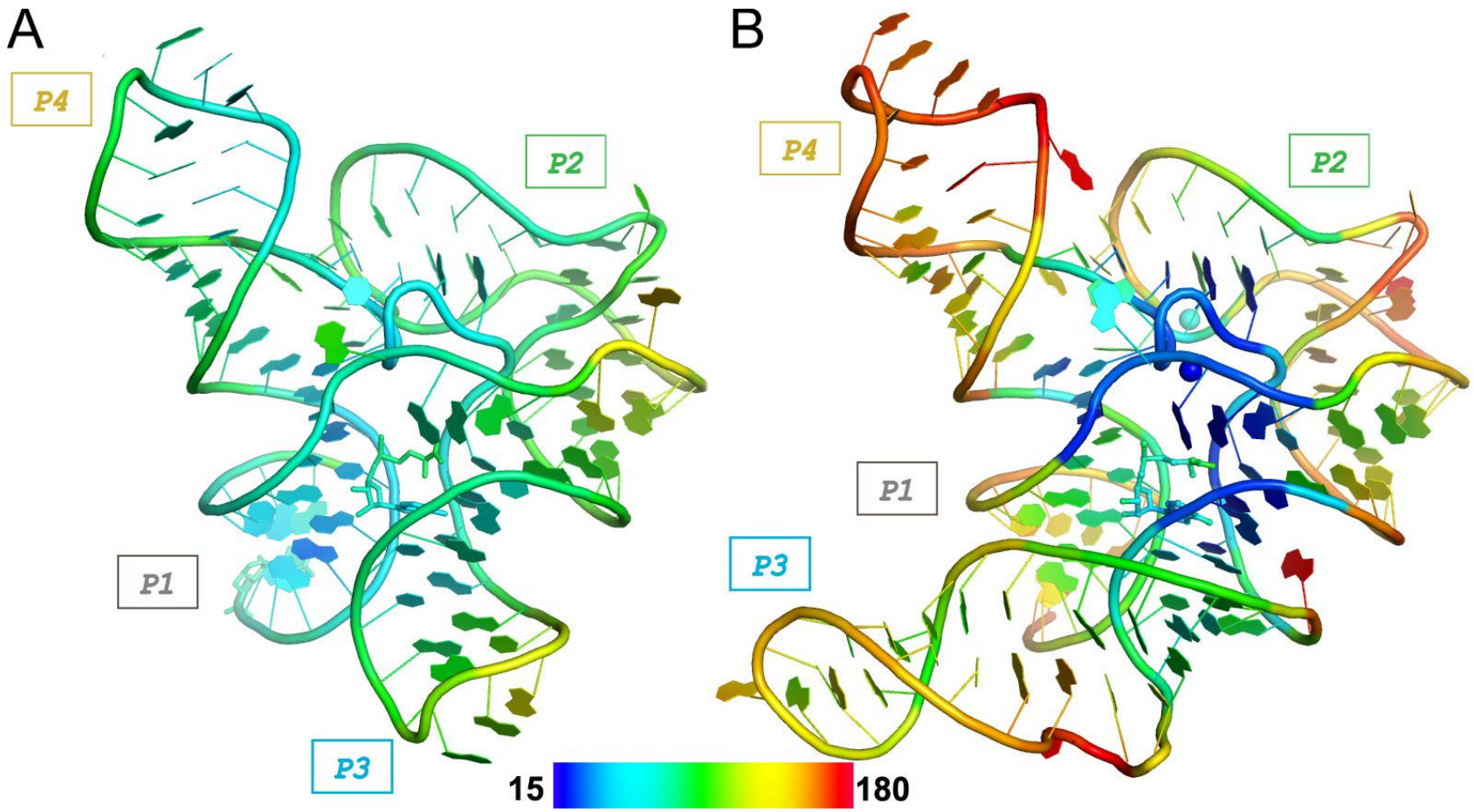

A
B
C
P1
P2
P3
P4
J1/4
J1/2
J2/3
J3/4
SAM
90°
5′
3′

helices, P1/P3 and P2/P4, which then pack against each other at a ~70-degree tilted angle through extensive minor-groove ribose zipper interactions (Figure 5.1B). Additional tertiary interactions orient the two coaxial helices and anchor them together at this configuration. These include a K-turn in the middle of P2 that bends its distal portion ~120 degrees to wrap around P1/P3, leading to a pseudoknot between the distal loop of P2 and J3/4 that weaves the P1/P3 and P2/P4 into a rigid conformation. The distal portion of the P3 helix, which is significantly longer in our structure than that of the *T. tengcongensis YitJ* SAM-I riboswitch (Montange and Batey, 2006, Montange et al., 2009), makes van der Waals minor groove interactions with the P4 helix and wraps around P4 helix in the *B. subtilis YitJ* structure (Figure 5.1B).

Our *YitJ* S box structure has a wider distribution of temperature factors for each nucleotide, which is a good indicator their thermal motion, in contrast to the thermophilic homolog structures (Montange and Batey, 2006, Montange et al., 2009) (Figure 5.2). Nucleotides surrounding the SAM binding site and the SAM molecule itself exhibit a much lower average temperature factor (B=35) than that of the overall structure (B=87) (Figure 5.2). Certain peripheral regions of the *YitJ* structure, such as the P4 loop, and J2/3 display significantly higher temperature factors (>150) (Figure 5.2). While the absolute temperature factor is complicated by both thermal motion and the order of the crystal lattice, the relative distribution of the B factors is a good indicator of the thermal motions within a structure. It is thus apparent that in our *YitJ* S box structure, the peripheral P4 terminal loop, distal P1, and the K-turn region in P2 possess much more flexibility than the central SAM binding site, whereas the *T. tengcongensis* SAM-I riboswitch is uniformly more rigid (Figure 5.2). This is unsurprising since in a thermophilic environment, the *T. tengcongensis YitJ* riboswitch requires more stability and temperature tolerance. Higher overall B factors in the

Figure 5.2. B-factor analysis of two S box (SAM-I) riboswitches.

Comparison of B-factors between (A) *T. tengcongensis* SAM-I riboswitch (Montange and Batey, 2006) and (B) *Bacillus subtilis YitJ* S box (SAM-I) riboswitch. The same B-factor coloring scheme (scale shown) is applied to both molecules


A
P4
P2
P1
P3
B
P4
P2
P1
P3
15
180

peripheral P4 terminal loop, distal P1, and the K-turn region in P2 therefore suggests that these regions possess much more flexibility than the SAM binding core. This is in good agreement with the temperature-dependent SHAPE analysis shown later.

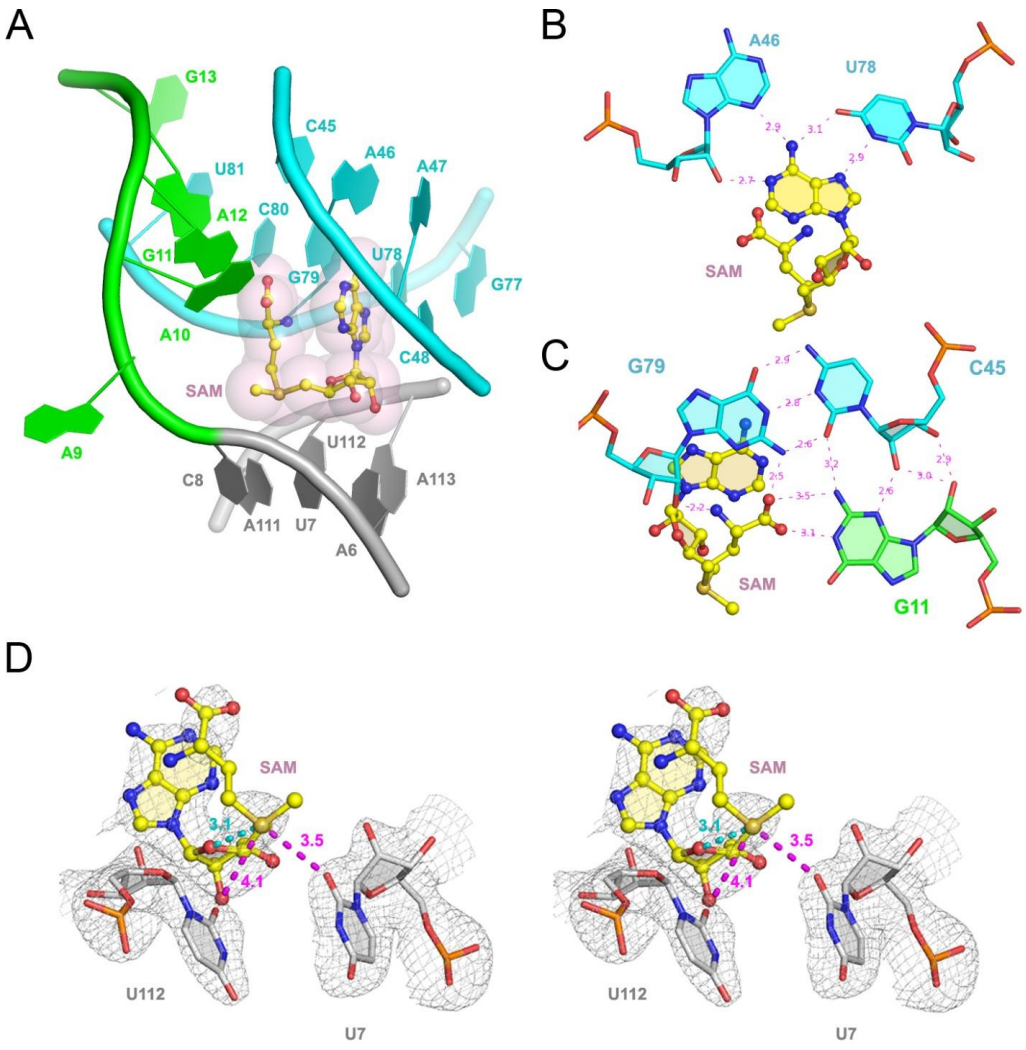
The SAM binding pocket is almost superimposable with that of the T. tengcongensis YitJ riboswitch

While some structural differences reside on the surface and periphery regions of the two S box riboswitches, their SAM-binding sites, however, bear striking similarity (r.m.s.d. of 0.37 Å for all phosphorous atoms), suggesting that the S box family of riboswitches use a conserved mechanism to achieve ligand recognition. Similar to seen in the *T. tencongensis YitJ* SAM-I structure, SAM adopts a compact "U" conformation that are rarely seen in the protein-bound structures to allow its methionine moiety to stack on top of the adenine base at a distance of 4 ~ 5 Å (Figure 5.3A). The Watson-Crick and Hoogsteen faces of the adenosine base of SAM are collectively recognized by four hydrogen bonds in a universally conserved, sheared A46 ŀU78 pair (Figure 5.3B). The methionine tail of SAM makes hydrogen bond contacts with G79-C45•G11 triple, which are responsible for tying J1/2 and P3 together (Figure 5.3C).

The positively charged sulfonium group in SAM is specified by two ~ 4Å favorable electrostatic interactions to the O4 carbonyl oxygens of U7 and U112 (Figure 5.3D, magental distances). The sulfonium ion also makes a very strong ~ 3.0 Å intramolecular electrostatic interaction with the O4′ of its ribose (Figure 5.3D, cyan distance). By specifying a SAM conformation allowing this favorable intramolecular interaction, the S box family of riboswitches further discourages the binding of SAH. The methyl group on SAM is not directly in contact with the riboswitch and orients

Figure 5.3. Overview of SAM binding pocket and key interactions between the RNA and SAM.

(A) The SAM molecule sits at the three-way junction between J2/3 and J1/2, and folds into a "U"-shape at the binding site of S box RNA. (B) A sheared A46•U78 pair recognizes the Watson-Crick and Hoogsteen faces of the adenosine moiety of SAM. (C) The G11•C45-G79 base triple recognizes the methionine tail of SAM. (D) (Stereoview) of U7 and U112 interacting with the sulfonium ion, and the ~ 3.0 Å intramolecular electro static interaction with O4′ of SAM's ribose moiety. The labeling and base coloring schemes are consistent with that in Figure 5.1B and C. The 3.0 Å̊ simulated annealing omit map is contoured at 1 sigma. Distances are given in angstroms in magenta. Carbon, oxygen, nitrogen, sulfur, and phosphorus atoms are colored grey, red, blue, gold, and orange, respectively.


A
G13
C45
U81
A46
A47
A12
C80
G11
U78
G77
G79
A10
C48
SAM
A9
U112
C8
A113
A111
U7
A6
B
A46
U78
SAM
C
G79
C45
SAM
G11
D
SAM
3.1
3.5
4.1
U112
U7
SAM
3.1
3.5
4.1
U112
U7

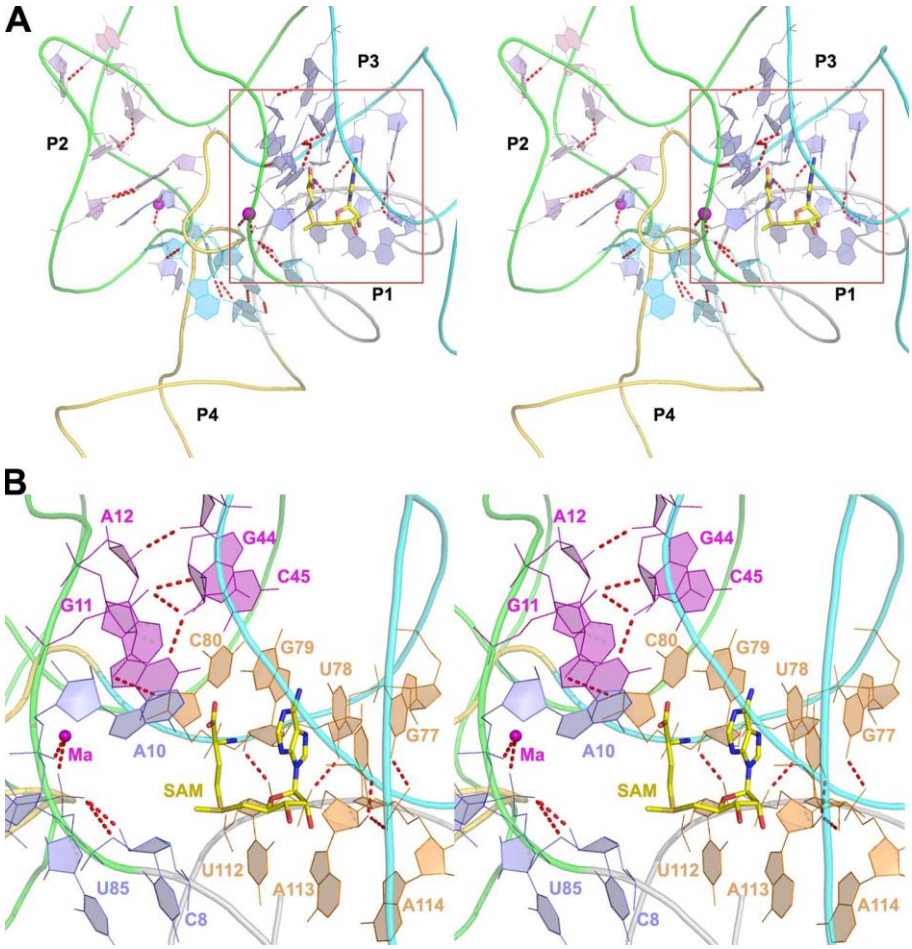
towards the solvent. Favorable inter- and intra-molecular electrostatic interactions with the sulfonium ion, and the lack of direct contact with the S-methyl group, is a general theme in all three classes of SAM riboswitches and forms the basis for strong discrimination against the near-cognate ligand SAH (~100-fold in the case of the *YitJ* S box riboswitch), which bears a neutral sulfide in place of sulfonium ion. This is consistent with our previous observation (Lu et al., 2008) based on the SAM-III riboswitch structure. Upon the determination of all three classes of SAM riboswitches(Gilbert et al., 2008, Lu et al., 2008, Montange and Batey, 2006, Montange et al., 2009) and this study, it is now clear that all the SAM riboswitches use similar mechanisms and exact same principles to distinguish cognate and near-cognate ligands. There is no conserved specific contact for the methionine carbon backbone, and their corresponding electron densities were significantly weaker, likely caused by more thermal motions than the rest of the SAM molecule. The amino and carboxyl terminal of SAM makes hydrogen bond contacts with the J1/2 in the S box RNA (Figure 5.3B, C). Similar interactions are observed in the *T. tencongensis* SAM-I RNA (Montange and Batey, 2006, Montange et al., 2009). These favorable electrostatic interactions account for about two orders of magnitude drop in affinity of the RNA to SAH.

A set of conserved tertiary interactions zip up the SAM binding site

In addition to the pseudoknot interactions previously reported in the *T. tencongensis* SAM-I structure, we report additional sets of SAM-independent interactions that stabilize the tertiary structure of the *YitJ* S box riboswitch (Figure 5.4A, 4B). Surrounding the SAM binding pocket are several clusters of conserved ribose zippers around the binding pocket (Figure 5.4A, red bases). One set of such interactions forms

Figure 5.4. Stereoview of important SAM-independent tertiary interactions within the S box RNA.

The orientation of the S box RNA structure is consistent with that in Figure 5.3, by positioning SAM so that the adenosine is symbolized by the right arm of the "U" and the methionine is the left arm. (A) Distribution of ribose zippers in the S box RNA. Blue, interaction around the binding pocket; pink, interaction important to the K-turn in P2; cyan, interaction orienting the P2 pseudoknot. Boxed interactions around the binding pocket are shown in detail in B. (B) Zoom-in on the SAM binding pocket. The interactions are sub-divided into three groups: slate, orange and purple based on their respective locations. Red dashes are the important interactions involved. Backbone coloring is consistent with that in Figure 5.1A and B.


A
P3
P2
P1
P4
P3
P2
P1
P4
B
A12
G44
C45
G11
C80
G79
U78
G77
Ma
A10
SAM
U112
A113
A114
U85
C8
A12
G44
C45
G11
C80
G79
U78
G77
Ma
A10
SAM
U112
A113
A114
U85
C8

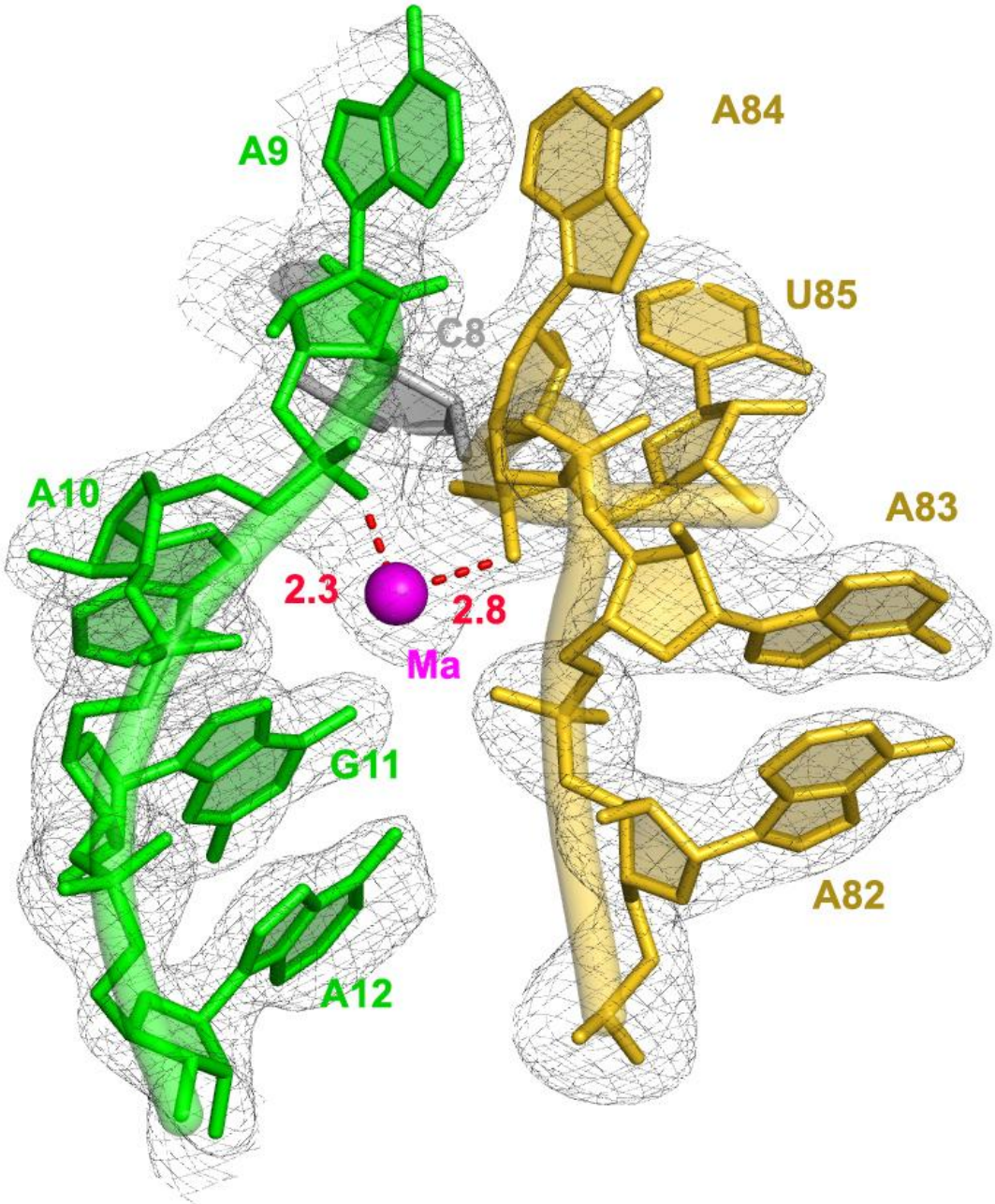
between nt 112-114 of P1 and 77-80 of P3, which is responsible for adenine Hoogsteen face recognition and defines the back and right side of the SAM binding pocket (Figure 5.4B, blue strands). Another set of a similar ribose zipper is above the binding pocket for the methionine tail of SAM (Figure 5.4B, red strands), where G44, C45, G11 and A12 bring J1/2 with J2/3 together and position G11 for hydrogen bond interactions with the carboxyl tail of SAM. Last but not least, a conserved C48 makes two hydrogen bonds through its O2 and ribose O2′ with A113, closing off the adenosine side of the binding pocket. These networks of ribose-mediated hydrogen bonds seals the back, top, right and left portion of the binding pocket in an apparent SAM-independent fashion. These interactions highly suggest that at least part of the S box RNA structure forms in a SAM-independent manner.

A magnesium ion mediates key ligand-induced conformational changes

Two inner-sphere coordinated magnesium ions were assigned with confidence in our 3.0 Å resolution YitJ S box structure. One of the magnesium ions, Ma, is located near the methionine tail side of the SAM binding pocket and makes two inner-spherical coordinations with phosphoryl oxygens from A10 and U65, as well as four outer spherical interactions with N7 of A10, O6 and N7 of G11, and the phosphoryl oxygen of A83 (Figure 5.5). Ma serves the important function of tying J3/4 U-turn and J1/2 together to stabilize the helical stacking of P1 and P4, and facilitates the pseudoknot formation by precisely orienting A82, A83 and C86 to base-pair with C30:G22, C29:G23 and G26 respectively. A single divalent metal ion was assigned in the T. tengcongensis YitJ SAM-I riboswitch structure (Montange and Batey, 2006), which binds to the equivalent location as Ma and makes identical interactions with surrounding RNA including the conserved A10 and G11 residues (Figure 5.5).

Figure 5.5. A Magnesium ion, Ma, mediates tertiary folding of the S box RNA.

Ma ties the J3/4 U-turn and J1/2 together, sealing the SAM-binding site at the methionine tail end. The labeling and base coloring scheme is consistent with that in Figure 5.1B, C. The 3.0 Å simulated annealing omit map is contoured at 1 sigma. Distances are given in angstroms in red.


A84
A9
U85
C8
A10
A83
2.3
2.8
Ma
G11
A82
A12

Interestingly, A10 appears to be conserved solely for the purpose of participating in the Ma binding pocket. This magnesium mediated junction also contains an A9/A84 di-nucleotide stack, which is not present in the first *T. tengcongensis YitJ* SAM-I riboswitch (Montange and Batey, 2006), but later identified in the A94G/U34 SAM-I variant (Montange et al., 2009). Our chemical probing data suggests that this metal ion site could potentially serve as a SAM entry site, and secure J1/2 and J3/4 with together with SAM in a cooperative manner.

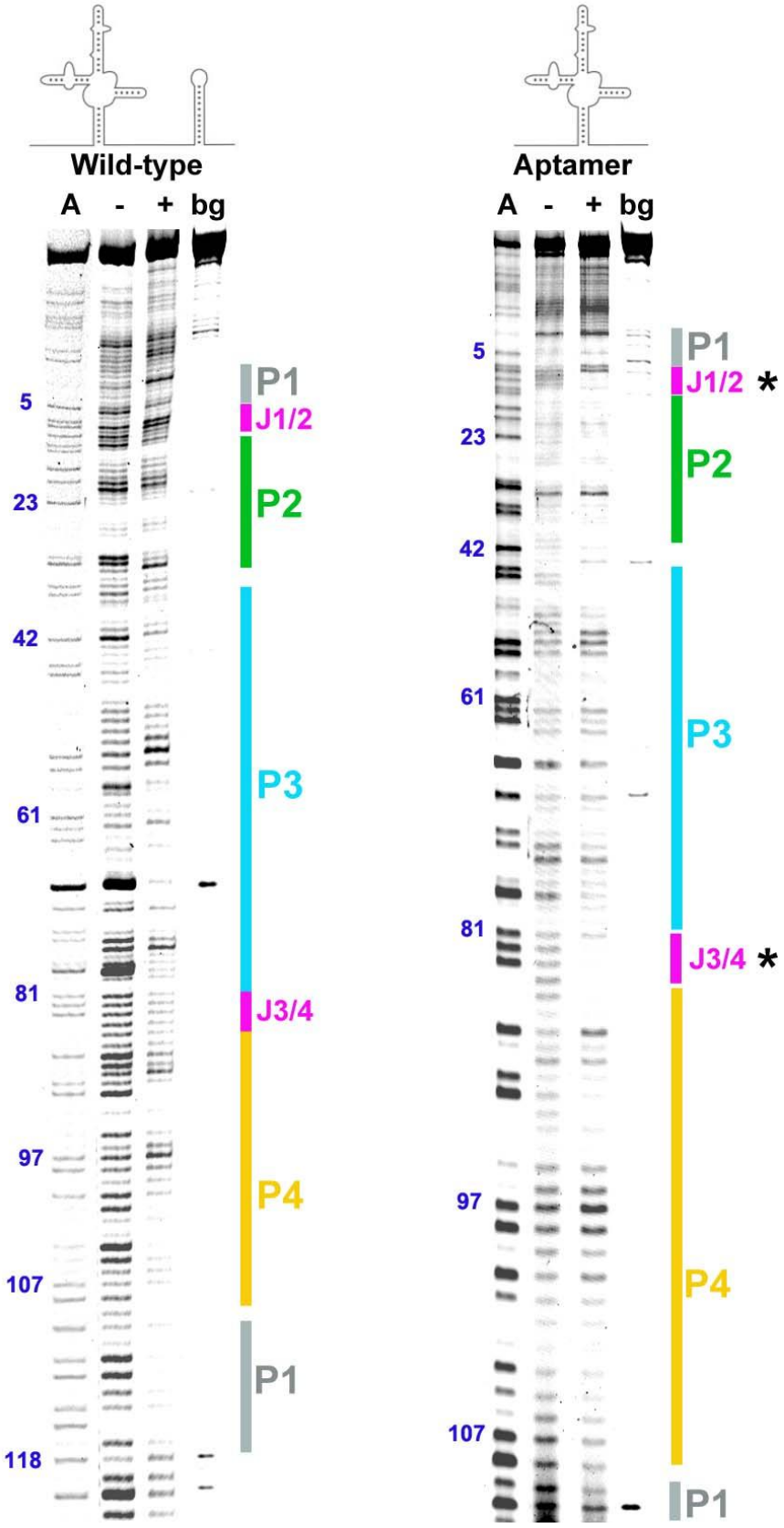
Expression platform is required for conformational dynamics of the S box riboswitch

To evaluate changes in conformations of the S box RNA upon SAM binding, we employed SHAPE analysis, which detects flexibilities in 2′-hydroxyl groups on the RNA ribose of the riboswitch. In order to investigate the dynamics of SAM-dependent conformational switching, we performed SHAPE analysis on TL5 RNA constructs with or without the 5′ transcription terminator/anti-terminator.

SHAPE analysis of the full length S box RNA reveals that SAM is required for the tertiary structure formation for the wild-type RNA. Reactivity patterns with SAM present show that only single-stranded loops and junctions, such as single-stranded P2 K-turn (G19, A20, A33), P3 bulge (G53-A55) and P4 terminal loop (C96) all exhibited higher than average flexibility (Figure 5.6, left panel, lane 3), which correlates well with our crystal structure. Meanwhile, without SAM, the full length S box riboswitch displayed a different and drastically less protected pattern (Figure 5.6, left panel, lane 2), especially at P1 stem, J1/2, P2 K-turn, J2/3, J3/4, part of P3, and P4 pseudoknot. Downstream at the expression platform, the formation of the transcription terminator when SAM is present is obvious, but the terminator is not generally more flexible without SAM (data not shown), which could result from multiple alternative

Figure 5.6. Chemical probing gel electrophoresis of the structural difference between unliganded and liganded S box riboswitches.

Left panel: full-length RNA construct with both the aptamer domain and the expression platform; right panel: S box TL5 aptamer domain only RNA. Both panels, lane 1: A ladder, lane 2: RNA + Mg^{2+}, lane 3: RNA + Mg^{2+}+SAM, lane 4: RNA background. Nucleotide identities are marked at the left and structural motifs are marked at the right. Coloring of secondary structure elements are consistent with that of Figure 5.1. Important activity differences between SAM-bound and –unbound aptamer structures are highlighted by asterisks (*) on the right.


Wild-type
A - + bg
5
23
42
61
81
97
107
118
P1
J1/2
P2
P3
J3/4
P4
P1
Aptamer
A - + bg
5
23
42
61
81
97
107
P1
J1/2 *
P2
P3
J3/4 *
P4
P1

base-pairing schemes and a relatively large terminator hairpin loop. This result indicates that the *apo* wild-type S box RNA adopts an alternative flexible conformation very different from the SAM-bound form. SAM is essential for the integrity of both the secondary and tertiary structure of the aptamer, and that the anti-terminator formation has lower free energy than the folding of the aptamer domain without SAM.

In order to understand the folding dynamics of the S box aptamer domain and the role of the expression platform, we also performed SHAPE analysis on our crystallization construct, which does not contain the downstream terminator sequence. Without the possibility of alternative base-pairing at the expression platform, the aptamer domain of S box folds into a compact structure very similar to that of the SAM-bound form (Figure 5.6, right panel, lane 2-3). This result strongly indicates that the overall folding of the aptamer domain without SAM is very similar to that of SAM-bound form, except at critical SAM-dependent tertiary contacts at J1/2, J3/4 and nt 46-47 (Figure 5.6, right panel), which agrees well with SAX data published earlier (Baird and Ferre-D'Amare). The P2-P4 pseudoknot region (25-28, 86-90) does not exhibit noticeable increase in flexibility, suggesting that the pseudoknot formation is SAM-independent in the presence of Mg ions. This is consistent with previous 2AP fluorescence experiment observations (Heppell and Lafontaine, 2008). Three areas of the S box RNA displayed noticeable conformational difference between the SAM-bound and unbound form: J1/2 (A10, G11 and A12), J3/4 (A82—U85) and A46-47. A46 makes direct hydrogen bonds with the adenosine face of SAM, while A10 and G11 interact with SAM's methionine tail and A12 is responsible for tethering P3 to enclose the binding pocket via contact with J3/4 (U81 and A82) (Figure 5.3A, C). Nucleotides A82-U85 are involved in coordinating the Mg ion Ma to stabilize the

electronegative repulsion between J1/2 and J3/4 (Figure 5.5) (Montange et al., 2009) and this study. The overall similarity between liganded and un-liganded structure strongly suggests that by removing the expression platform, the folding equilibrium shifts towards the formation of a "pre-binding" aptamer structure that has the binding pocket largely pre-formed but accessible to SAM through the opening between J1/2 and J3/4. After SAM docks into the binding pocket, A10, G10 form contact with the methionine moiety, and A11 finally brings P3 to close the opening. Our result strongly points towards a synergistic relationship between SAM-binding and Mg^{2+}-mediated J1/2-J3/4 coupling. This experimental result agrees with previous MD predictions performed on the *T. tencongensis* SAM-I riboswitch (Huang et al., 2009).

The above result confirms that with no SAM present, the full length wild type S box favors the unfolded/anti-terminator form against the pre-binding/terminator form, while addition of SAM shifts the equilibrium towards the SAM-bound/terminiator form by lowering its free energy. This conformation dynamics does not exist if the expression platform is removed from the riboswitch.

Thermal melting of the S box RNA bound to SAM reveals differential stability in structural elements

The SAM-bound S box RNA is a compact, well-ordered structure that encompasses many tertiary interactions such as a pseudoknot, ribose zippers, Mg-mediated strand couplings, and of course the SAM-dependent binding pocket formation. Therefore, it is highly likely that the folding/refolding process of the SAM-bound S box RNA is highly hierarchical. To dissect the differential stability of folded S box structural elements, we employed temperature dependent SHAPE analysis to obtain thermal stability information at nucleotide-resolutions.

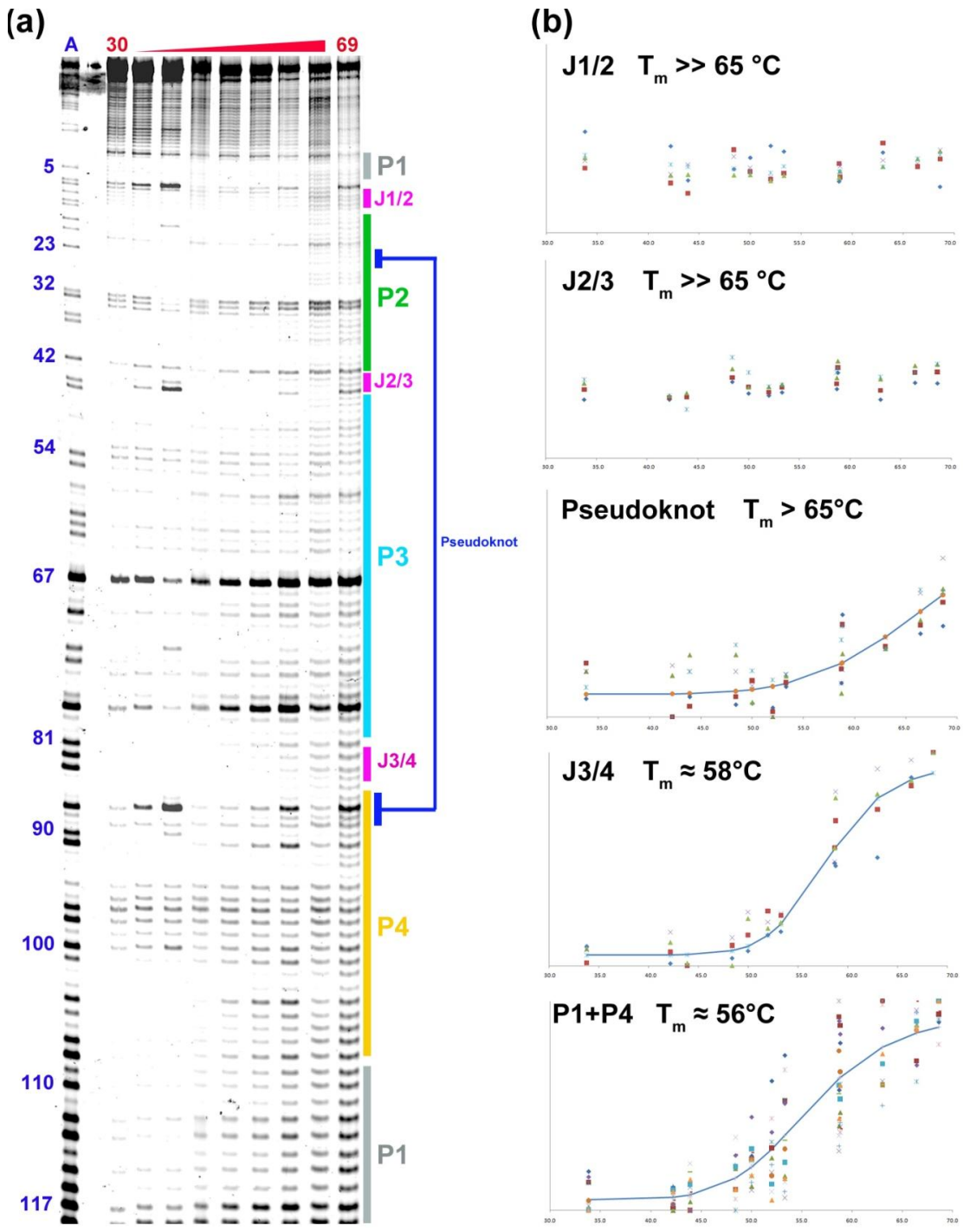
The SAM-bound S box aptamer domain displays remarkable thermal stability, as almost all the tertiary interactions stay intact even at 50 degrees Celsius, which is well above physiological temperature. As temperature increases even higher, about 80% of the structure eventually becomes more flexible at 70 degrees (Figure 5. 7A). Throughout the temperature gradient, >95% of the structure displayed increase in SHAPE reactivity (Figure 5.7A). However, the changes in reactivities are not uniform. P1, J1/4 and J3/4 showed T_m between 55 and 60 ℃ (Figure 5.7B, lower panels) while J1/2, J2/3 and P2-P4 pseudoknot areas have surprisingly higher T_m (>65 ℃) (Figure 5.7B, middle panel). This hierarchical melting pathway of the S box riboswitch suggests strong tertiary interactions, both from ligand-induced (e.g. J1/2, J2/3) and ligand-independent (e.g. pseudoknot). These interactions stabilize core of the riboswitch, which holds the SAM-binding pocket intact even after P1 helix has disintegrated. This is probably because the S box RNA completely "enclosed" the ligand at its core, where the structure is most stable, as demonstrated by its lowest temperature factors in our crystal structure (Figure 5.2B, 5.7B upper panels).

DISCUSSION

Through structure comparison of the two S box (SAM-I) riboswitch structures, we concluded that despite differences on their surface and peripheral regions, they utilize a virtually identical mechanism to recognize SAM molecule, because SAM conformation, as well as RNA residues involved in SAM recognition, are virtually superimposable. We also revealed that conserved SAM-dependent and -independent tertiary interactions likely play important roles in stabilizing the structural integrity as well as the conformational dynamics. The wild-type riboswitch, containing the aptamer domain and the expression platform, exhibits clear conformational dynamics

Figure 5.7. Hierarchical melting of the S box aptamer domain.

(A). Temperature-dependent SHAPE gel analysis of S box aptamer domain bound to SAM. RNA structures are indicated at right. Lane 1: the dideoxy-U sequencing ladder. (B). Nucleotide unfolding profiles. Profiles were computed by quantifying intensities shown in A by fragment analysis and processed by program ShapeFinder (Vasa et al., 2008). Data are plotted on a linear scale; solid blue line indicates average behavior. Nucleotides at pseudoknot, J3/4 and P1+P4 were fit to a unimolecular transition equation.


(a)
A
30
69
5
23
32
42
54
67
81
90
100
110
117
P1
J1/2
P2
J2/3
P3
J3/4
P4
P1
Pseudoknot
(b)
J1/2 T_m >> 65 °C
J2/3 T_m >> 65 °C
Pseudoknot T_m > 65°C
J3/4 T_m ≈ 58°C
P1+P4 T_m ≈ 56°C

in response to SAM. This switching behavior is completely abolished once the terminator/anti-terminator sequence is removed, which suggests that the SAM-independent interactions within the aptamer domain is strong enough to hold the structure together. It is highly likely that this "pre-binding" S box structure has a completely pre-formed binding pocket, with the opening at the methionine recognition site, which closes once SAM and a divalent ion synergistically bind to J1/2 and J3/4. This hypothesis is supported by both crystallographic observations as well as by SHAPE analysis: we observed extensive ribose-mediated hydrogen bonds around the SAM-binding pocket and the P2/P4 pseudoknot, both of which stay intact even without SAM present in the SHAPE analysis of the aptamer domain, except for J1/2 and J3/4.

These results provide further evidence towards the mode of SAM binding to the S box riboswitch, likely through the opening of J1/2 and J3/4 in the absence of SAM, as these are the only regions that showed disorder in the SHAPE analysis, compared with that of the SAM-bound structure. The entry and binding of SAM through this region of RNA lowers the free energy of the binding pocket and facilitates the incorporation of a magnesium ion (Ma) and the stabilization of J1/2 and J3/4 to the final SAM-bound structure (Figure 5.8).

Once SAM is secured into the binding site, the core of the S box aptamer displayed remarkable thermal stability. Part of the SAM binding pocket (J1/2, J2/3) and pseudoknot region showed very little perturbation even at 60 ℃. This means that these tertiary interaction-mediated single-stranded regions are more stable than helical duplexes such as P1, which has a melting point (Tm) around 55 ℃. This hierarchical unfolding pattern of the S box RNA bound to SAM suggests that the riboswitch very tightly encloses SAM in a thermal stable and solvent-inaccessible pocket. Such

Figure 5.8. Proposed mode of SAM-entry to S box aptamer RNA based on the crystal structure and SHAPE analysis.

Left, “open” structure with flexible without SAM bound; right, “closed” structure when SAM and an Mg ion seal the opening. Regions of flexibility detected by SHAPE analysis are colored red. The flexible J1/2 and J3/4 on the left are arbitrary idealized structure generated by Coot(Emsley and Cowtan, 2004). The coloring scheme for the rest of the structure is consistent with that in Figure 5.1B, C. Blue moieties on the surface are water molecules.

J3/4
J1/2
SAM
J3/4
J1/2

configuration of SAM-enclosing also explains the RNA's high affinity to SAM (sub-micromolar K_d, compared with sub- millmolar for SAM-II and S_{MK} (SAM-III)) riboswitches, by having a relatively slow off-rate.

Based on these evidence, we propose a transcription-dependent conformational switching model for the *B. subtilis yitJ* S box riboswitch The RNAP first transcribes the aptamer domain of the S box RNA, as it immediately folds into the pre-binding conformation before the transcription of downstream terminator sequence (RNA folding completes less than 100 milliseconds vs. bacterial transcription 42nt per second). This pre-binding aptamer conformation has two alternative fates: 1). If intracellular SAM level is high, the aptamer structure is greatly stabilized by SAM-mediated tertiary interactions and cannot be competed away by the subsequent emergence of the expression platform sequence, leading to terminator hairpin formation and transcription termination; or 2). If intracellular SAM level is low, the aptamer structure collapses due to competing base pairing formation from the expression platform, anti-terminator hairpin forms, and transcription continues to produce the full-length transcript. Either decision by the RNAP is irreversible, therefore SAM-I/S box class of transcription riboswitches are not under selection pressure to be able to switch from the SAM-bound to the unfolded conformation. Our "window-of-opportunity" model requires that the rate of RNA transcription to be much slower than the rate of S box riboswitch folding.

APPENDIX 1

CRYSTAL STRUCTURE OF THE *S. SOLFATARICUS* ARCHAEAL EXOSOME REVEALS CONFORMATIONAL FLEXIBILITY IN THE RNA-BINDING RING[2]

SUMMARY

Background: The exosome complex is an essential RNA 3'-end processing and degradation machinery. In archaeal organisms, the exosome consists of a catalytic ring and an RNA-binding ring, both of which were previously reported to assume three-fold symmetry.

Methodology/Principle Findings: Here we report an asymmetric 2.9 Å *Sulfolobus solfataricus* archaeal exosome structure in which the three-fold symmetry is broken due to combined rigid body and thermal motions mainly within the RNA-binding ring. Since increased conformational flexibility was also observed in the RNA-binding ring of the related bacterial PNPase, we speculate that this may reflect an evolutionarily conserved mechanism to accommodate diverse RNA substrates for degradation.

Conclusion/Significance: This study clearly shows the dynamic structures within the RNA-binding domains, which provides additional insights on mechanism of asymmetric RNA binding and processing.

This work is published in Lu, C., F. Ding, et al. (2010). "Crystal Structure of the *S. solfataricus* Archaeal Exosome Reveals Conformational Flexibility in the RNA-Binding Ring." PLoS ONE 5(1): e8739. doi:10.1371/journal.pone.0008739.The text and data was reused here with permission from the publisher.

INTRODUCTION

The exosome and its related bacterial polynucleotide phosphorylase (PNPase) represent a class of conserved multi-subunit protein complexes responsible for the 3′-to-5′ processing and degradation of RNA (Butler, 2002, Houseley et al., 2006, Parker and Song, 2004). In the nucleus of eukaryotic cells, the exosome is responsible for the 3′-end trimming of rRNA, snRNA, and snoRNA, but also plays a major role in the degradation of the spliced introns and pre-mRNAs that fail the quality control processes (van Hoof et al., 2000a, van Hoof et al., 2000b, Allmang et al., 2000, Bousquet-Antonelli et al., 2000, Allmang et al., 1999a, Allmang et al., 1999b, Mitchell et al., 1997, Kadaba et al., 2004, Milligan et al., 2005, Burkard and Butler, 2000). In addition, the nuclear exosome participates in nuclear surveillance pathways to degrade aberrant rRNA, tRNA, and mRNA transcripts (Hilleren et al., 2001, Kadaba et al., 2004, Vasudevan and Peltz, 2003, Lee et al., 2005, Hieronymus et al., 2004, Allmang et al., 2000, Milligan et al., 2005). In the cytoplasm, the exosome is a key player during mRNA turnover, and is responsible for the 3′-to-5′ degradation of normal mRNA after deadenylation as well as for the products of endonucleolytic events (Gatfield and Izaurralde, 2004), including those of the RNAi pathway (Orban and Izaurralde, 2005). Like the nuclear exosome, the cytoplasmic exosome also participates in mRNA surveillance pathways such as nonsense-mediated decay, non-stop decay, no-go decay, and ARE-mediated decay to eliminate aberrant mRNAs (Hilleren and Parker, 2003, Chen et al., 2001, Doma and Parker, 2006, van Hoof et al., 2002, Mitchell et al., 2003).

The function and architecture of the exosome is conserved among all three kingdoms of life. Its bacterial counterpart, PNPase, is an 81 KDa polypeptide that trimerizes to form a ~ 240 KDa complex. The primary sequence of the PNPase

contains two domains homologous to RNase PH, followed by RNA-binding KH and S1 domains (Symmons et al., 2000). Architecturally, the ~250 KDa prototypic archaeal exosome is similar to the bacterial PNPase, consisting of a catalytic ring containing a trimer of Rrp41/Rrp42 heterodimers, and with a trimetric Rrp4 or Csl4 RNA-binding ring stacked on the top. Both Rrp41 and Rrp42 are homologous to RNase PH, but only Rrp41 possesses the phosphorolytic 3'-to-5' exoribonuclease activity (Lorentzen and Conti, 2005). The core eukaryotic exosome consists of nine different subunits (Liu et al., 2006). Six of them classify into two RNase PH-like groups: Rrp41-like (Rrp41, Rrp46, and Mtr3) and Rrp42-like (Rrp42, Rrp43, and Rrp45), however, in yeast and human none of these subunits possess any exoribonuclease activity (Dziembowski et al., 2007, Liu et al., 2006). The rest of the three subunits (Rrp4, Rrp40, and Csl4) are predicted to be RNA binding proteins (Mitchell et al., 1997). The eukaryotic core exosome interacts with many protein cofactors including 3′-to-5′ exoribonuclease, RNA helicase, and polyadenylase to carry out its RNA processing and degradation functions. Crystal structures of the PNPase (Symmons et al., 2000, Shi et al., 2008), archaeal (Lorentzen et al., 2007, Buttner et al., 2005, Navarro et al., 2008), and eukaryotic (Liu et al., 2006) exosomes have been determined, from which it has become clear that these complexes adopt a very similar doughnut-shaped architecture.

The structure and enzymatic mechanism of the archaeal exosome have been extensively studied. Published structural work revealed that the archaeal exosomes adopt a three-fold symmetric structure, where a trimeric RNA-binding ring packs on top of the catalytic ring consisting of a trimer of Rrp41/Rrp42 heterodimers. A large "processing chamber" was found inside the catalytic ring, where three identical phosphorolytic active sites reside. RNA substrate likely gains access to the

"processing chamber" from the side of the RNA-binding ring through a narrow pore at the neck region, which only allows the passage of a single-stranded (ss) RNA (Buttner et al., 2005, Lorentzen and Conti, 2006, Navarro et al., 2008). The proposed recruitment pathway is consistent with the observation that the archaeal exosome is stalled by strong RNA secondary structures, leaving RNA products with 7-9 nucleotides-long 3′ overhangs (Lorentzen and Conti, 2005).

To further understand the RNA recruitment mechanism of the exosome family of macromolecular machines, we determined the crystal structure of the archaeal *Sulfolobus solfataricus* exosome. We observed that while the catalytic ring of the *S. solfataricus* exosome adopts a three-fold symmetric structure, its RNA-binding ring displays considerable rigid body and thermal motions, which broke its internal symmetry. Since conformational flexibility was also observed in the RNA-binding ring of the bacterial PNPase (Symmons et al., 2000, Shi et al., 2008), we propose that the conformational dynamics may be harnessed by the archaeal exosome, and perhaps by all exosome family members, to accommodate the binding of various RNA substrates and to disrupt their structures.

MATERIALS AND METHODS

Expression, and purification of the 4-subunit Sulfolobus solfataricus exosome

Genes encoding Rrp4, Rrp41, Rrp42, and Csl4 have been cloned from the genomic DNA of the archaeal hyperthermophile *Sulfolobus solfataricus*. We designed a polycistronic construct to promote exosome assembly by co-expressing four subunits in *E. coli* simultaneously. This was inspired by the observation that the three Rrp genes are arranged as a superoperon (in the order of Rrp4-Rrp41-Rrp42) in the

archaeal genomes (Koonin et al., 2001). The order of the genes was preserved in the polycistronic vector construction, followed by the insertion of the Csl4 gene (Tan et al., 2005, Selleck and Tan, 2008). A His_6-tag and a Tobacco Etch Virus (TEV) protease cleavage site were engineered into the N-terminus of Rrp4 to facilitate purification.

After overexpression in *E.coli*, the supernatant of the cell lysate was incubated at 75° for half an hour and precipitated using centrifugation to efficiently denature and remove *E. coli* proteins. Target proteins were then purified using nickel affinity chromatography using supernatant from the heat treatment. All four exosome subunits were present in the eluate in roughly predicted stoichiometry. Purified proteins were then dialyzed to a buffer containing 25 mM Tris-HCl pH 8.0 and 100 mM NaCl, incubated with TEV protease to remove the His_6-tag, and further purified to homogeneity on an anion exchange chromatography (Mono Q column). Alternatively, the complex can by purified using size exclusion chromatography (Suprose 6 column). We confirmed the identity of the four archaeal proteins in the purified exosome by separation on SDS polyacrylamide gel (SDS-PAGE), followed by in-gel protease digestion, electrospray mass spectrometry, and data base matching (data not shown).

RNase assay

The 254-nt RNA substrate containing the hepatitis delta virus ribozyme followed by a 161-nt 3′-tail was generated from *in vitro* T7 RNA polymerase transcription reaction and purified following standard protocols as previously described (Ke and Doudna, 2004). The RNA was incubated with the purified archaeal exosome was incubated at 37 °C under multiple turnover conditions (~10 pmol enzyme, ~40 pmol substrate) in a buffer containing 25 mM Tris-HCl pH 7.5, 100 mM

NaCl, and10 mM inorganic phosphate (PO_4^{3-}). Control experiments were carried out in the same buffer in the absence of PO_4^{3-} or in the presence of 20 mM SO_4^{3-}.

Crystallization and structure determination of the Rrp4-exosome

Our Rrp4-exosome isoform was concentrated to ~10 mg/mL in a buffer containing 100 mM Tris-HCl pH 8.6, 200 mM $Mg(Ac)_2$, and 25 mM $(NH_4)_2SO_4$), and crystallized by hanging vapor diffusion against a well solution containing 30% PEG4000 with 100 mM Tris-HCl pH 8.6, 200 mM $Mg(Ac)_2$. Crystals appeared after two months incubation at room temperature among heavy precipitations. The crystals were soaked in the well solution plus 10 mM $NaSO_4$ for 10 minutes before flash-frozen in liquid nitrogen. A complete data set to 2.9 Å was collected at 90K from CHESS beamline A1. The crystal belongs to C2 space group with unit cell dimensions of 151 Å x 145 Å x 97 Å, and γ=93.8°, as indexed and scaled by HKL2000 (Otwinowski and Minor, 1997). The apo Rrp4-exosome structure was solved by molecular replacement using PHASER (Storoni et al., 2004) from the structure of the catalytic core of the *S. solfataricus* exosome (PDB code: 2BR2) (Lorentzen et al., 2005) as the search model. Initial search using the nine-subunit complex resulted in very poor and untraceable maps in seven out of the nine subunits, while searching attempt with three copies each of individual Rrp41or Rrp42 subunits resulted in more than 100 steric clashes. Finally, a search using three pairs of Rrp41-42 heterodimer produced a reasonable solution and electrondensity map revealing traceable difference between the search model and our structure. Rigid body refinement by Refmac (Collaborative Computational Project, 1994) further improved the overall electron density and revealed extra electron densities corresponding to Rrp4 in the F_o-F_c difference map. Initially, strict three-fold non-crystallographic symmetry (NCS) was

applied to refine the catalytic ring structure (Rrp41/42 heterotrimer) but was dropped when refinement can no longer improve the model and R_{free}. Only two of the three Rrp4 subunits can be correctly placed by molecular replacement in PHASER using the published Rrp4 structure (PDB code 2JE6) (Lorentzen et al., 2007), while attempting to position the third subunit resulted in broken densities. Therefore the third Rrp4 subunit was manually traced using COOT (Emsley and Cowtan, 2004) and built from three- to ten-amino acid-long peptide segments of Rrp4. At this stage the strict NCS matrix was removed from subsequent refinements. Successive steps of manual refinement of all nine subunits using COOT (Emsley and Cowtan, 2004) allowed tracing of more flexible regions as well as locating three sulfate ions at catalytic core domains., Multiple rounds of manual fitting followed by restrained and rigid body refinement in Refmac5 (Collaborative Computational Project, 1994) reduced R_{work}/R_{free} to 0.31/0.34 and positioned all three Rrp4 proteins into density map. At this stage the, R_{work}/R_{free} stopped dropping due to over refinement in some of the subunits and under refinement in others. Therefore, we treated each of the nine subunits separately and independently subject them to energy minimization, simulated annealing, grouped B-factor refinement in CNS (Brunger et al., 1998) while holding remaining eight subunits in position, which reduced R_{work}/R_{free} to 0.29/0.31. We then defined a total of 15 translation-liberation-screw (TLS) groups in the entire structure: three of each Rrp41, Rrp42, N-ter, S1 and KH domain of Rrp4 subunit, to be used in TLS refinement in Refmac5 which ultimately reduced refinement statistics to the final R_{work}/R_{free}=0.268/0.289 (Table A1.1). The TLS refinement result is visualized by Raster3D (Merritt, 1999). The structure is verified by simulated annealing omit maps by excluding 7.5 percent of structure in each calculation.

Table A 1.1. Crystallographic statistics

Resolution (Å)	20 - 2.9
Space Group	C2
Unit Cell Dimensions	
a (Å)	151.1
b (Å)	145.5
c (Å)	97.2
α (°)	90
β ( )	93.8
γ (°)	90
Completeness (%)	99.1(95.2)
Beamline	CHESS A1
Redundancy	4.0
I/s	10.1(3.0)
R_{sym} (%)	7.8(31.5)
R_{free}/R_{work}	28.9/26.7
r.m.s.d. bonds (Å)	0.007
r.m.s.d. angles ( )	1.2
Average B ($Å^2$)	32.9
Coordinate error (Å)	0.38

RESULTS

Reconstitution and phosphorolytic RNase activity in the four-subunit archaeal exosome.

When we started the work, studies had shown that all four eukaryotic exosome homologous proteins, namely Rrp41, Rrp42, Rrp4, and Csl4, are present in the purified archaeal exosome complex (Evguenieva-Hackenberg et al., 2003, Buttner et al., 2005, Lorentzen et al., 2005, Makarova and Koonin, 2005, Portnoy et al., 2005, Walter et al., 2006). We therefore attempted to reconstitute the archaeal *Sulfolobus solfataricus* exosome by co-expressing all four exosome subunits from a polycistronic construct in *E. coli* (Tan et al., 2005, Barrios et al., 2007, Selleck and Tan, 2008). As shown in the results, while Rrp4, Rrp41, and Rrp42 were present at stoichiometric amounts in the purified complex, the Csl4 protein appeared sub-stoichiometric (Figure A1.1A). The purified exosome complex was shown to be monodispersed and homogeneous under negative staining EM (data not shown).

To demonstrate the phosphorylytic RNase activity in our purified archaeal exosome, we incubated the purified exosome with a 254-nt RNA substrate flanked by a strong 5′-tertiary RNA structure (the hepatitis delta virus ribozyme) and a 161-nt flexible 3′ tail under multiple turnover conditions. As shown in Figure A1.1B, the archaeal exosome is highly processive in degrading the RNA substrate. The strong tertiary structure in the HDV ribozyme effectively blocks the action of the exosome, resulting in accumulation of a procession intermediate. The enzymatic activity is strictly phosphate-dependent as the activity decreased to background level in the absence of PO_4^{3-} or in the presence of 20 mM SO_4^{2-} (data not shown).

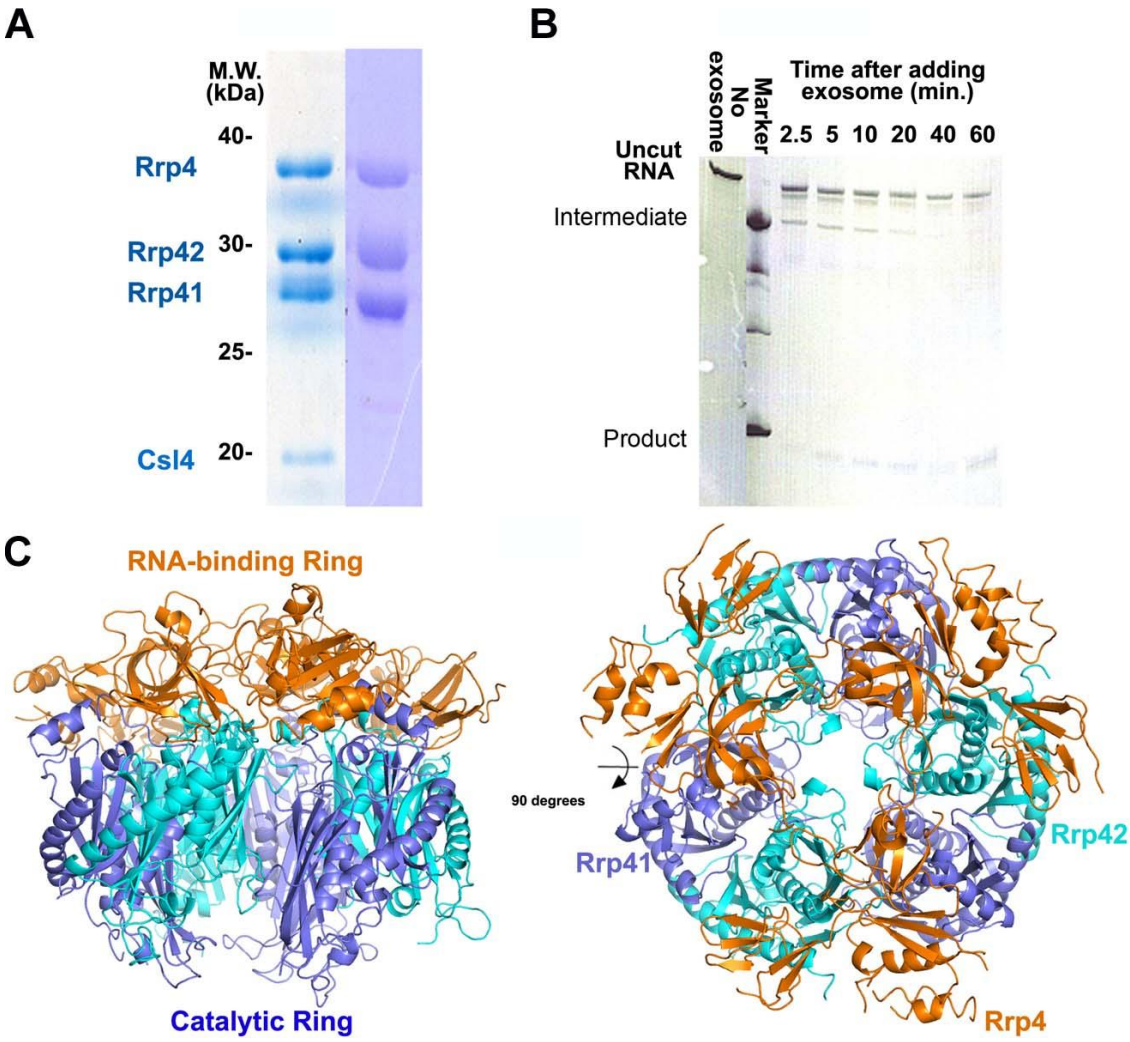


Figure A 1.1. Purification and activity assay of the purified *S. solfataricus* full exosome.

(A) SDS-PAGE of the purified *S. solfataricus* full exosome (left) and the Rrp4-exosome isoform (right). (B) RNase activity assay for the intact *S. solfataricus* exosome. The exosome is stalled by the HDV ribozyme sequence. (C) Side and top-down view of the 2.9 Å *S. solfataricus* exosome structure. Gold, trimeric Rrp4 RNA-binding ring; blue/cyan, Rrp41/Rrp42 catalytic ring

We obtained the crystals of the four-protein component archaeal exosome complex grown in ammonium sulfate and polyethylene glycol 4000 (PEG 4000). The crystals diffracted X-rays to 2.4 Å resolution at a synchrotron radiation source. Data processing by HKL2000 (Otwinowski and Minor, 1997), MOSFLM (Leslie, 2006) was problematic as a large fraction of data were rejected in all possible space groups (the most likely being P2/P21). We speculate this is due to epitaxial twinning of Rrp4-exosome and Csl4-exosome crystal lattices. Not surprisingly, no structure solution was obtained from the four-subunit *Sulfolobus solfataricus* exosome crystals.

Crystallization and structure determination of the Rrp4-exosome isoform.

We reasoned that the problematic diffraction pattern from the four-subunit exosome crystals was due to the presence of a sub-stoichiometric amount of Csl4 protein. To generate homogeneous exosome samples, the Csl4 gene was deleted from the co-expression cassette and biochemically, the resulting Rrp4-exosome isoform behaved similarly as the four-subunit complex. We independently crystallized the complex under high PEG conditions similar to that reported by the Conti group (Lorentzen et al., 2007, Lorentzen and Conti, 2008). A complete data set was collected to 2.9 Å showing that the crystals belong to C2 space group (Table A1.1).

We solved the *S. solfataricus* exosome Rrp4-isoform structure by molecular replacement using the Rrp41/Rrp42 heterodimer in the *S. solfataricus* exosome catalytic core (PDB code: 2BR2) as the search model (Lorentzen et al., 2005). A single exosome core complex was reconstructed in the asymmetric unit after the rotation and translation searches successfully located each of the three heterodimers. Although no model corresponding to the Rrp4 protein was included in the search model, the resulting molecular replacement phases allowed the initial tracing of Rrp4

protein backbone. Despite similarities in the crystallization condition, unit cell dimensions and packing arrangement between our crystal structure and that by the Conti group, we noticed early on in our refinement that applying strict three-fold non-crystallographic symmetry (NCS), as the Conti group did (Lorentzen and Conti, 2005, Lorentzen et al., 2005, Lorentzen et al., 2007, Lorentzen and Conti, 2008), caused considerable fragmentation of the electron densities in the RNA-binding ring where the trimeric Rrp4 proteins are located. We therefore continued structure refinement without the three-fold NCS, re-traced one of the three copies of Rrp4, and continued refinement until the R_{work}/R_{free} were within the acceptable range of 26.7%/28.9%.

Conformation flexibility mainly within the RNA-binding ring

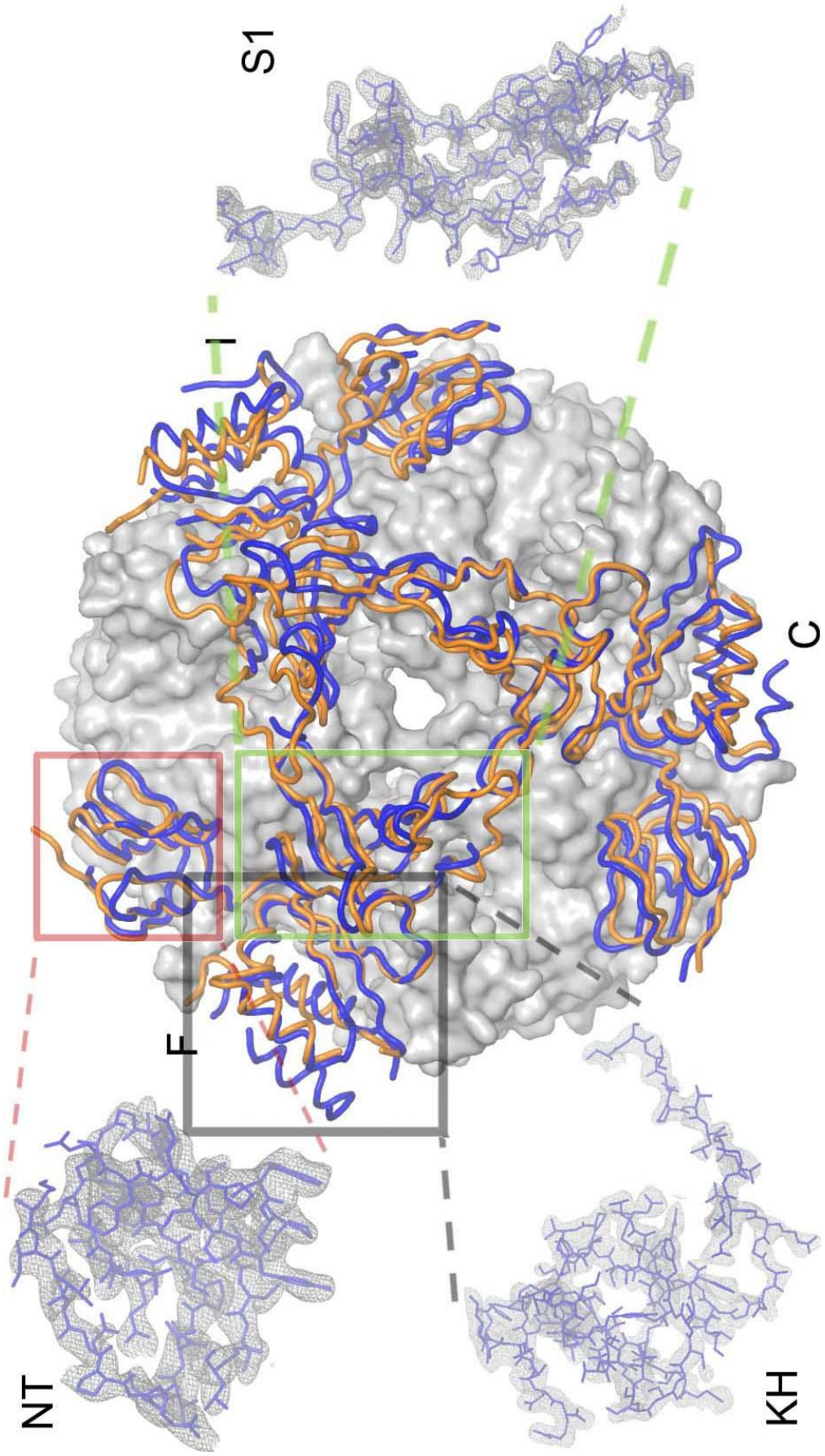
The catalytic ring in our *S. solfataricus* Rrp4-exosome crystal structure (Figure A1.1C) is very similar to that of the catalytic mutant structure from the same organism (PDB code 2JE6) previously reported (Lorentzen et al., 2007). The three copies of Rrp41/Rrp42 heterodimers within the catalytic ring align extremely well with their counterparts in the previous structure (Lorentzen et al., 2007), even though the three-fold NCS was not enforced in our structure, the three heterodimers does not deviate notably from their symmetric positions. The r.m.s.d. of the Cα atom alignment is 0.6 Å for the whole catalytic ring (1529 aa) alignment, and between 0.2 to 0.4 Å for pair-wise Rrp41/Rrp45 heterodimer (501 aa) alignment within our structure. This result suggests that the positions of the heterodimers deviate slightly (on average ~ 0.3 Å) from the symmetric positions.

The three Rrp4 subunits (designated Chains C, F, and I in our structure, respectively) in the RNA-binding ring of our structure, however, deviate significantly from their three-fold symmetric positions as reported previously (Figure A1.2

(Lorentzen et al., 2007)). When one of the Rrp4 subunits (i.e. Chain C) is superimposed with that in the Conti structure, the other two Rrp4 proteins deviate significantly from their symmetric counterparts (Figure A1.2 and Figure A1.3A). In addition to rigid body shifts, we observed conformational heterogeneity among Rrp4 subunits. While two of the three copies (Chain C and I) overlap reasonably well with their counterparts in the previous structure (Lorentzen et al., 2007), with the r.m.s.d of the Cα alignment at 0.5 Å and 0.6 Å, respectively, and at 0.8 Å with each other, the third Rrp4 copy (Chain F) adopts a significantly different conformation (the r.m.s.d. of 250 Cα alignment is 1.5 Å with the previously published structure (Lorentzen et al., 2007), Figure A1.3 B-D). Compared with the published structure by Lorentzen *et al*. (Lorentzen et al., 2007), almost identical sets of intramolecular interactions were observed between the Rrp4 trimeric cap and the Rrp41/42 hetertodimers. The conformational changes are not due to changes in the crystal packing environment, as minimal crystal contacts were observed around the Rrp4 ring. The exceptionally large geometric deviation in Chain F is due to a rigid body hinge motion at the linker region between the N-terminal domain and the S1/KH domains, because when aligned separately, each domain agrees well with their counterparts in other two Rrp4 subunits (r.m.s.d. of Cα alignments for the N-ter, S1, and KH domains are at 0.3 Å (50 aa), 0.3 Å (70 aa), and 0.7 Å (115 aa), respectively, Figure A1.3 B-D). The hinge motion causes the S1 and KH domains in chain F to move away from the central channel and thus increases the diameter of the pore opening.

Figure A 1.2. Comparison of Rrp4 trimeric caps.

Rrp4 trimeric cap of our *S. solfataricus* exosome (blue) deviates from perfect three-fold symmetry as compared with the structure by Lorentzen *et al* (Lorentzen et al., 2007) (orange). Subunit F and I considerably deviates (2 to 3 Å at the periphery) from Lorentzen *et al* previously report when aligning subunit C. Inlets: 2.9 Å experimental electron density map of Rrp4 subunit F contoured at 1.0 σ.


S1
I
C
F
NT
KH

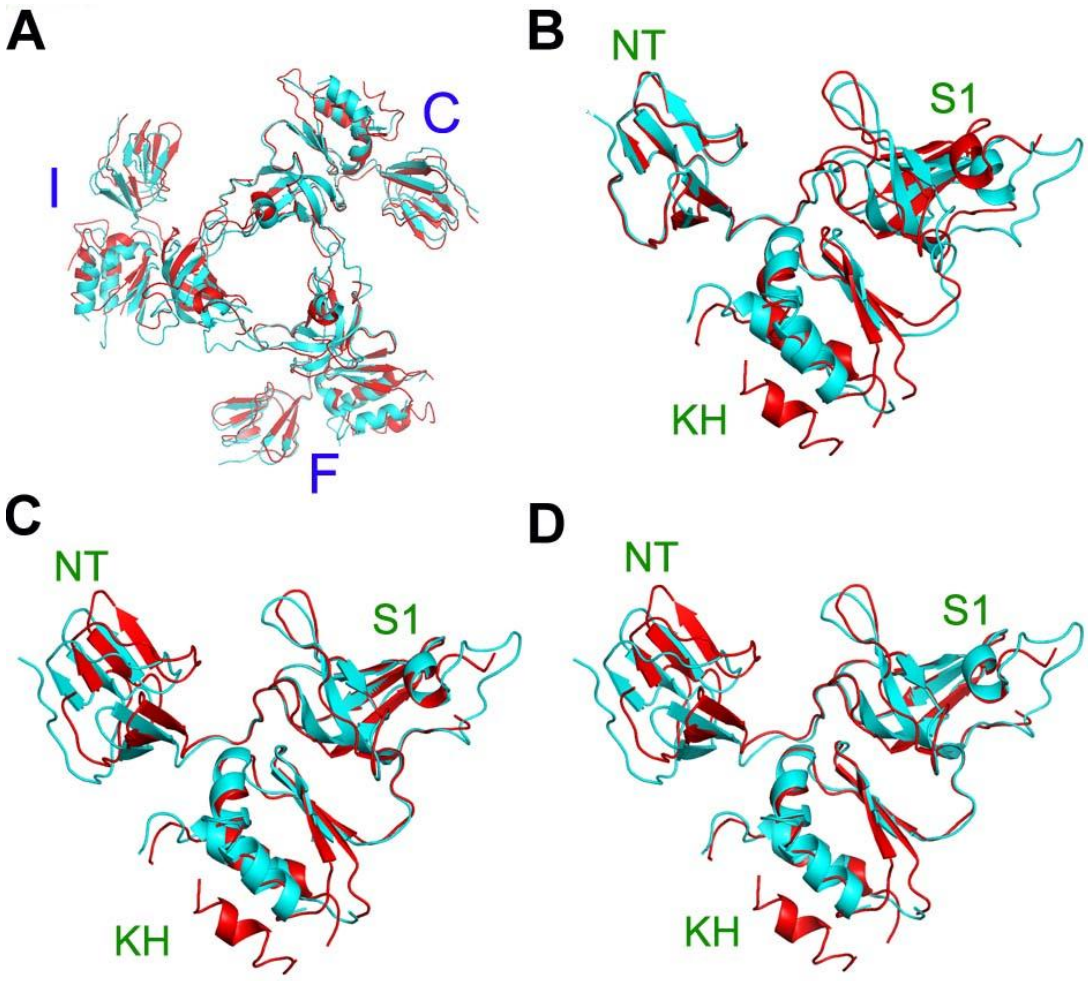


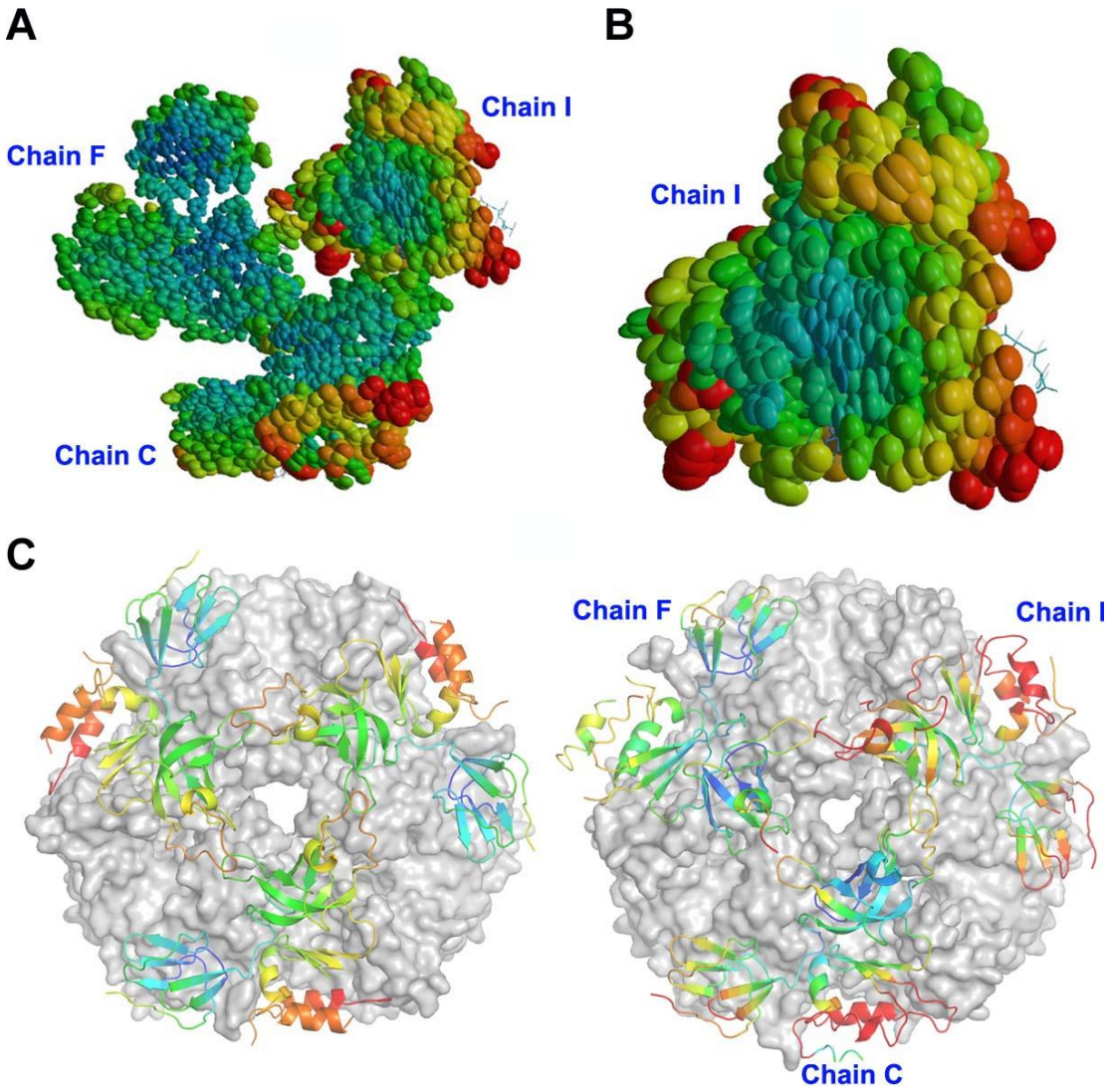
Figure A 1.3. Structure alignment of Rrp4 trimeric caps.

Alignment of Rrp4 trimeric cap between our *S. solfataricus* exosome (red) and that Lorentzen *et al* previously reported (Lorentzen et al., 2007) (cyan). (A) Overall alignment of the 750- aa trimeric cap (r.m.s.d. =1.5 Å). (B) Alignment of the most flexible Rrp4 subunit (Chain F) by N-ter domain, aa 1-50 (r.m.s.d. =0.4 Å). (C) Alignment of Chain F subunit by S1 domain, aa 56-126 (r.m.s.d. =0.3 Å). (D) Alignment of Chain F subunit by KH domain, aa 135-250 (r.m.s.d. =0.7 Å).

Besides conformational heterogeneity caused by the hinge motion, the Rrp4 trimer also exhibits distinct temperature factor distributions, which reflect the RNA binding ring's inherent conformational flexibilities. TLS refinement followed by thermal ellipsoids analysis revealed unique thermal motions in Chain I copy of Rrp4 (Figure A1.4A, B). Although Chain C and Chain I of Rrp4 adopt very similar conformations, Chain I displays considerably higher thermal motion than Chain C, manifested by a higher average temperature factor of 81 versus 61 (Figure A1.4A, C), as well as wider thermal ellipsoids in Chain I (Figure A1.4B). Each Rrp4 subunit has distinct temperature factor distributions within the polypeptide (Figure A1.4A, C). The shape of the thermal ellipsoids revealed that Chain I wobbles to a larger extent on top of the catalytic ring, and in a concentric motion relative to the central RNA-processing chamber. All above observations indicate that the Rrp4 trimeric cap is intrinsically flexible, despite bound to the relatively rigid catalytic ring. The observed conformational heterogeneity agrees perfectly with the function of the RNA-binding ring of the exosome to accommodate diverse RNA substrates.

Figure A 1.4. Detecting thermal motions in Rrp4 RNA-binding ring using thermal ellipsoid and B-factor analyses.

(A) Overall analysis of the trimeric cap. (B) According to the thermal ellipsoid analysis, the most thermal flexible Rrp4 subunit is Chain I, but not Chain F, which displays the largest rigidbody motion. TLS sensors were obtained from TLS refinement in Refmac5 (Collaborative Computational Project, 1994) (see Methods section for details) and plotted using Raster3D (Merritt, 1999).(C) B-factor comparison of the Rrp4 between our *S. solfataricus* exosome (left) and that Lorentzen *et al* previously reported (Lorentzen et al., 2007) (right). B-factor coloring: blue, 30 and below; red, 100 and above.


A
Chain I
Chain F
Chain C
B
Chain I
C
Chain F
Chain I
Chain C

DISCUSSION

Previous structural studies of the archaeal exosome from three different species all reported perfect three-fold symmetric subunit arrangements within this RNA processing machine. An interesting observation from these studies was that the RNA-binding ring displays large conformational differences among different species. For example, the conformation of the *S. solfataricus* exosome RNA-binding ring differs dramatically from what was seen in the *A. fulgidus* structure mostly due to rigid body rearrangement of the KH, S1, and NT domain locations (Buttner et al., 2005) (r.m.s.d. of the Cα atom alignment in the 750-aminoacid Rrp4 region is 5.2 Å). This either suggests a faster evolution rate within the RNA-binding part of the exosome, or hints at inherent conformational flexibilities in this region. Through careful analyses of the conformation and thermal motion of each subunit, we provide evidence to support the higher-flexibility hypothesis. We showed that each Rrp4 subunit within the *S. solfataricus* RNA-binding ring adopts a distinct conformation and thermal motion distribution pattern. As a result, the previously assumed three-fold symmetric subunit arrangement is broken. Interestingly, it was found that significantly higher conformational flexibility and thermal motions are present in the RNA-binding ring of the crystal structures of the exosome-related PNPases from mesophilic bacteria, *Escherichia coli* (Shi et al., 2008) and *Streptomyces antibioticus* (Symmons et al., 2000). The RNA-binding ring (occupying 23% of the entire PNPase) in *E. coli* PNPase is so flexible that it could not be successfully traced from the electron density map (Shi et al., 2008). It is therefore likely the RNA-binding ring may become even more dynamic in its native thermophilic environment.

What is the significance for maintaining the evolutionarily conserved conformation flexibility in the RNA-binding ring of the archaeal exosome and

PNPase? We speculate two functional roles: (1) The conformational flexibility may play a passive role in accommodating the binding of diverse RNA substrates before threading the RNA into the catalytic ring. The fact that each of the three Rrp4 subunits possesses distinct thermal and conformational characteristics suggests that the exosome might respond to the substrate by activating one of the three Rrp4 subunits. It is also possible that the substrate binding is a cooperative event that involves more than one Rrp4 subunit. (2) On the other hand, the increased thermal motion and flexibility in this region may further play an active role in unzipping the secondary structures of the RNA substrates so that the RNA is 'ready' to pass through the narrow neck region of the exosome, which can only accommodate a single-stranded RNA. The RNA-binding ring therefore resembles an RNA helicase in its RNA-unwinding function. Consistent with this hypothesis, we observed that a strong tertiary structure, such as the HDV ribozyme, initially stalls the action of the *S. solfataricus* exosome, but can later be gradually degraded by the exosome, albeit at much slower rate (Figure A1.1B).

REFERENCES

ALLMANG, C., KUFEL, J., CHANFREAU, G., MITCHELL, P., PETFALSKI, E. & TOLLERVEY, D. (1999a) Functions of the exosome in rRNA, snoRNA and snRNA synthesis

10.1093/emboj/18.19.5399 [doi]. *EMBO J,* 18**,** 5399-410.

ALLMANG, C., MITCHELL, P., PETFALSKI, E. & TOLLERVEY, D. (2000) Degradation of ribosomal RNA precursors by the exosome

gkd297 [pii]. *Nucleic Acids Res,* 28**,** 1684-91.

ALLMANG, C., PETFALSKI, E., PODTELEJNIKOV, A., MANN, M., TOLLERVEY, D. & MITCHELL, P. (1999b) The yeast exosome and human PM-Scl are related complexes of 3' --> 5' exonucleases. *Genes Dev,* 13**,** 2148-58.

ALTSCHUL, S. F., MADDEN, T. L., SCHAFFER, A. A., ZHANG, J., ZHANG, Z., MILLER, W. & LIPMAN, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research,* 25**,** 3389-402.

ANAGNOSTOPOULOS, C. & SPIZIZEN, J. (1961) Requirements for transformation in Bacillus subtilis. *J Bacteriol,* 81**,** 741-746.

ANDRE, G., EVEN, S., PUTZER, H., BURGUIERE, P., CROUX, C., DANCHIN, A., MARTIN-VERSTRAETE, I. & SOUTOURINA, O. (2008) S-box and T-box riboswitches and antisense RNA control a sulfur metabolic operon of Clostridium acetobutylicum. *Nucleic Acids Research,* 36**,** 5955-69.

BAIRD, N. J. & FERRE-D'AMARE, A. R. Idiosyncratically tuned switching behavior of riboswitch aptamer domains revealed by comparative small-angle X-ray scattering analysis. *RNA (New York, N.Y.),* 16**,** 598-609.

BARRICK, J. E. & BREAKER, R. R. (2007) The distributions, mechanisms, and structures of metabolite-binding riboswitches. *Genome biology,* 8**,** R239.

BARRICK, J. E., CORBINO, K. A., WINKLER, W. C., NAHVI, A., MANDAL, M., COLLINS, J., LEE, M., ROTH, A., SUDARSAN, N., JONA, I., WICKISER, J. K. & BREAKER, R. R. (2004) New RNA motifs suggest an expanded scope for riboswitches in bacterial genetic control. *Proceedings of the National Academy of Sciences of the United States of America,* 101**,** 6421-6426.

BARRIOS, A., SELLECK, W., HNATKOVICH, B., KRAMER, R., SERMWITTAYAWONG, D. & TAN, S. (2007) Expression and purification of recombinant yeast Ada2/Ada3/Gcn5 and Piccolo NuA4 histone acetyltransferase complexes. *Methods (San Diego, Calif.),* 41**,** 271-7.

BATEY, R. T., GILBERT, S. D. & MONTANGE, R. K. (2004) Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine. *Nature,* 432**,** 411-5.

BLOUNT, K., PUSKARZ, I., PENCHOVSKY, R. & BREAKER, R. (2006) Development and application of a high-throughput assay for glmS riboswitch activators. *RNA biology,* 3**,** 77-81.

BLOUNT, K. F. & BREAKER, R. R. (2006) Riboswitches as antibacterial drug targets. *Nature Biotechnology,* 24**,** 1558-64.

BLOUNT, K. F., WANG, J. X., LIM, J., SUDARSAN, N. & BREAKER, R. R. (2007) Antibacterial lysine analogs that target lysine riboswitches. *Nature chemical biology,* 3**,** 44-49.

BOCOBZA, S., ADATO, A., MANDEL, T., SHAPIRA, M., NUDLER, E. & AHARONI, A. (2007) Riboswitch-dependent gene regulation and its evolution in the plant kingdom. *Genes & Development,* 21**,** 2874-9.

BOUSQUET-ANTONELLI, C., PRESUTTI, C. & TOLLERVEY, D. (2000) Identification of a regulated pathway for nuclear pre-mRNA turnover S0092-8674(00)00065-9 [pii]. *Cell,* 102**,** 765-75.

BRUNGER, A. T. (2007) Version 1.2 of the crystallography and NMR system. *Nature protocols,* 2**,** 2728-2733.

BRUNGER, A. T., ADAMS, P. D., CLORE, G. M., DELANO, W. L., GROS, P., GROSSE-KUNSTLEVE, R. W., JIANG, J. S., KUSZEWSKI, J., NILGES, M., PANNU, N. S., READ, R. J., RICE, L. M., SIMONSON, T. & WARREN, G. L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta crystallographica.Section D, Biological crystallography,* 54**,** 905-921.

BURKARD, K. T. & BUTLER, J. S. (2000) A nuclear 3'-5' exonuclease involved in mRNA degradation interacts with Poly(A) polymerase and the hnRNA protein Npl3p. *Molecular and Cellular Biology,* 20**,** 604-16.

BUTLER, J. S. (2002) The yin and yang of the exosome S0962892401022255 [pii]. *Trends Cell Biol,* 12**,** 90-6.

BUTTNER, K., WENIG, K. & HOPFNER, K. P. (2005) Structural framework for the mechanism of archaeal exosomes in RNA processing S1097-2765(05)01713-2 [pii] 10.1016/j.molcel.2005.10.018 [doi]. *Mol Cell,* 20**,** 461-71.

CHEAH, M. T., WACHTER, A., SUDARSAN, N. & BREAKER, R. R. (2007) Control of alternative RNA splicing and gene expression by eukaryotic riboswitches. *Nature,* 447**,** 497-500.

CHEN, C. Y., GHERZI, R., ONG, S. E., CHAN, E. L., RAIJMAKERS, R., PRUIJN, G. J., STOECKLIN, G., MORONI, C., MANN, M. & KARIN, M. (2001) AU binding proteins recruit the exosome to degrade ARE-containing mRNAs

S0092-8674(01)00578-5 [pii]. *Cell,* 107**,** 451-64.

CHRISTIANSEN, L. C., SCHOU, S., NYGAARD, P. & SAXILD, H. H. (1997) Xanthine metabolism in Bacillus subtilis: characterization of the xpt-pbuX operon and evidence for purine- and nitrogen-controlled expression of genes involved in xanthine salvage and catabolism. *Journal of Bacteriology,* 179**,** 2540-2550.

COCHRANE, J. C., LIPCHOCK, S. V. & STROBEL, S. A. (2007) Structural investigation of the GlmS ribozyme bound to Its catalytic cofactor. *Chemistry & biology,* 14**,** 97-105.

COCHRANE, J. C. & STROBEL, S. A. (2008) Riboswitch effectors as protein enzyme cofactors. *RNA (New York, N.Y.),* 14**,** 993-1002.

COLLABORATIVE COMPUTATIONAL PROJECT, N. (1994) The CCP4 suite: programs for protein crystallography. *Acta crystallographica.Section D, Biological crystallography,* 50**,** 760-763.

COLLINS, J. A., IRNOV, I., BAKER, S. & WINKLER, W. C. (2007) Mechanism of mRNA destabilization by the glmS ribozyme. *Genes & Development,* 21**,** 3356-68.

COPPINS, R. L., HALL, K. B. & GROISMAN, E. A. (2007) The intricate world of riboswitches. *Current Opinion in Microbiology,* 10**,** 176-81.

CORBINO, K. A., BARRICK, J. E., LIM, J., WELZ, R., TUCKER, B. J., PUSKARZ, I., MANDAL, M., RUDNICK, N. D. & BREAKER, R. R. (2005) Evidence for a second class of S-adenosylmethionine riboswitches and other regulatory RNA motifs in alpha-proteobacteria. *Genome biology,* 6**,** R70.

D'COSTA, V. M., MCGRANN, K. M., HUGHES, D. W. & WRIGHT, G. D. (2006) Sampling the antibiotic resistome. *Science (New York, N.Y.),* 311**,** 374-377.

DAMBACH, M. D. & WINKLER, W. C. (2009) Expanding roles for metabolite-sensing regulatory RNAs. *Current Opinion in Microbiology,* 12**,** 161-9.

DAS, R., LAEDERACH, A., PEARLMAN, S. M., HERSCHLAG, D. & ALTMAN, R. B. (2005) SAFA: semi-automated footprinting analysis software for high-throughput quantification of nucleic acid footprinting experiments. *RNA (New York, N.Y.),* 11**,** 344-54.

DOMA, M. K. & PARKER, R. (2006) Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation. *Nature,* 440**,** 561-4.

DZIEMBOWSKI, A., LORENTZEN, E., CONTI, E. & SERAPHIN, B. (2007) A single subunit, Dis3, is essentially responsible for yeast exosome core activity nsmb1184 [pii] 10.1038/nsmb1184 [doi]. *Nat Struct Mol Biol,* 14**,** 15-22.

EDDY, S. R. & DURBIN, R. (1994) RNA sequence analysis using covariance models. *Nucleic Acids Research,* 22**,** 2079-88.

EDWARDS, T. E. & FERRE-D'AMARE, A. R. (2006) Crystal structures of the thi-box riboswitch bound to thiamine pyrophosphate analogs reveal adaptive RNA-small molecule recognition. *Structure (London, England : 1993),* 14**,** 1459-1468.

EDWARDS, T. E., KLEIN, D. J. & FERRE-D'AMARE, A. R. (2007) Riboswitches: small-molecule recognition by gene regulatory RNAs. *Current opinion in structural biology,* 17**,** 273-279.

EMSLEY, P. & COWTAN, K. (2004) Coot: Model-building tools for molecular graphics. *Acta crystallographica.Section D, Biological crystallography,* 60**,** 2126-2132.

EPSHTEIN, V., MIRONOV, A. S. & NUDLER, E. (2003) The riboswitch-mediated control of sulfur metabolism in bacteria. *Proceedings of the National Academy of Sciences of the United States of America,* 100**,** 5052-6.

EVGUENIEVA-HACKENBERG, E., WALTER, P., HOCHLEITNER, E., LOTTSPEICH, F. & KLUG, G. (2003) An exosome-like complex in Sulfolobus solfataricus
10.1038/sj.embor.embor929 [doi] embor929 [pii]. *EMBO Rep,* 4**,** 889-93.

FUCHS, R. T., GRUNDY, F. J. & HENKIN, T. M. (2006) The SMK box is a new SAM-binding RNA for translational regulation of SAM synthetase. *Nature structural & molecular biology,* 13**,** 226-233.

FUCHS, R. T., GRUNDY, F. J. & HENKIN, T. M. (2007) S-adenosylmethionine directly inhibits binding of 30S ribosomal subunits to the SMK box translational riboswitch RNA. *Proceedings of the National Academy of Sciences of the United States of America,* 104**,** 4876-4880.

GARST, A. D. & BATEY, R. T. (2009) A switch in time: detailing the life of a riboswitch. *Biochimica et Biophysica Acta,* 1789**,** 584-91.

GATFIELD, D. & IZAURRALDE, E. (2004) Nonsense-mediated messenger RNA decay is initiated by endonucleolytic cleavage in Drosophila
10.1038/nature02559 [doi] nature02559 [pii]. *Nature,* 429**,** 575-8.

GILBERT, S. D., RAMBO, R. P., VAN TYNE, D. & BATEY, R. T. (2008) Structure of the SAM-II riboswitch bound to S-adenosylmethionine. *Nature Structural & Molecular Biology,* 15**,** 177-82.

GOTTESMAN, S. (2004) The small RNA regulators of Escherichia coli: roles and mechanisms*. *Annual Review of Microbiology,* 58**,** 303-28.

GROISMAN, E. A., CROMIE, M. J., SHI, Y. & LATIFI, T. (2006) A Mg2+-responding RNA that controls the expression of a Mg2+ transporter. *Cold Spring Harb Symp Quant Biol,* 71**,** 251-8.

GRUNDY, F. J. & HENKIN, T. M. (1993) tRNA as a positive regulator of transcription antitermination in B. subtilis. *Cell,* 74**,** 475-482.

GRUNDY, F. J. & HENKIN, T. M. (1998) The S box regulon: a new global transcription termination control system for methionine and cysteine biosynthesis genes in gram-positive bacteria. *Molecular microbiology,* 30**,** 737-749.

GRUNDY, F. J. & HENKIN, T. M. (2003) The T box and S box transcription termination control systems. *Frontiers in bioscience : a journal and virtual library,* 8**,** d20-31.

GRUNDY, F. J. & HENKIN, T. M. (2004) Regulation of gene expression by effectors that bind to RNA. *Current opinion in microbiology,* 7**,** 126-131.

GRUNDY, F. J. & HENKIN, T. M. (2006) From ribosome to riboswitch: control of gene expression in bacteria by RNA structural rearrangements. *Critical Reviews in Biochemistry and Molecular Biology,* 41**,** 329-38.

GRUNDY, F. J., LEHMAN, S. C. & HENKIN, T. M. (2003) The L box regulon: lysine sensing by leader RNAs of bacterial lysine biosynthesis genes. *Proceedings of the National Academy of Sciences of the United States of America,* 100**,** 12057-12062.

GRUNDY, F. J., MOIR, T. R., HALDEMAN, M. T. & HENKIN, T. M. (2002a) Sequence requirements for terminators and antiterminators in the T box transcription antitermination system: disparity between conservation and functional requirements. *Nucleic acids research,* 30**,** 1646-1655.

GRUNDY, F. J., WINKLER, W. C. & HENKIN, T. M. (2002b) tRNA-mediated transcription antitermination in vitro: codon-anticodon pairing independent of the ribosome. *Proceedings of the National Academy of Sciences of the United States of America,* 99**,** 11121-11126.

GRUNDY, F. J., YOUSEF, M. R. & HENKIN, T. M. (2005) Monitoring uncharged tRNA during transcription of the Bacillus subtilis glyQS gene. *Journal of Molecular Biology,* 346**,** 73-81.

HAMPEL, K. J. & TINSLEY, M. M. (2006) Evidence for preorganization of the glmS ribozyme ligand binding pocket. *Biochemistry,* 45**,** 7861-71.

HENKIN, T. M. (2008) Riboswitch RNAs: using RNA to sense cellular metabolism. *Genes & Development,* 22**,** 3383-90.

HENKIN, T. M. & YANOFSKY, C. (2002) Regulation by transcription attenuation in bacteria: how RNA provides instructions for transcription termination/antitermination decisions. *BioEssays : news and reviews in molecular, cellular and developmental biology,* 24**,** 700-707.

HEPPELL, B. & LAFONTAINE, D. A. (2008) Folding of the SAM aptamer is determined by the formation of a K-turn-dependent pseudoknot. *Biochemistry,* 47**,** 1490-9.

HIERONYMUS, H., YU, M. C. & SILVER, P. A. (2004) Genome-wide mRNA surveillance is coupled to mRNA export

gad.1241204 [pii] 10.1101/gad.1241204 [doi]. *Genes Dev,* 18**,** 2652-62.

HILLEREN, P., MCCARTHY, T., ROSBASH, M., PARKER, R. & JENSEN, T. H. (2001) Quality control of mRNA 3'-end processing is linked to the nuclear exosome

10.1038/35097110 [doi] 35097110 [pii]. *Nature,* 413**,** 538-42.

HILLEREN, P. J. & PARKER, R. (2003) Cytoplasmic degradation of splice-defective pre-mRNAs and intermediates

S109727650300488X [pii]. *Mol Cell,* 12**,** 1453-65.

HOUSELEY, J., LACAVA, J. & TOLLERVEY, D. (2006) RNA-quality control by the exosome

nrm1964 [pii] 10.1038/nrm1964 [doi]. *Nat Rev Mol Cell Biol,* 7**,** 529-39.

HUANG, W., KIM, J., JHA, S. & ABOUL-ELA, F. (2009) A mechanism for S-adenosyl methionine assisted formation of a riboswitch conformation: a small molecule with a strong arm. *Nucleic Acids Research,* 37**,** 6528-39.

IWIG, D. F. & BOOKER, S. J. (2004) Insight into the polar reactivity of the onium chalcogen analogues of S-adenosyl-L-methionine. *Biochemistry,* 43**,** 13496-13509.

JANSEN, J. A., MCCARTHY, T. J., SOUKUP, G. A. & SOUKUP, J. K. (2006) Backbone and nucleobase contacts to glucosamine-6-phosphate in the glmS ribozyme. *Nature Structural & Molecular Biology,* 13**,** 517-23.

JOHN, D. M. & WEEKS, K. M. (2000) van't Hoff enthalpies without baselines. *Protein Science,* 9**,** 1416-9.

KADABA, S., KRUEGER, A., TRICE, T., KRECIC, A. M., HINNEBUSCH, A. G. & ANDERSON, J. (2004) Nuclear surveillance and degradation of hypomodified initiator tRNAMet in S. cerevisiae

10.1101/gad.1183804 [doi] 1183804 [pii]. *Genes Dev,* 18**,** 1227-40.

KE, A. & DOUDNA, J. A. (2004) Crystallization of RNA and RNA-protein complexes. *Methods (San Diego, Calif.),* 34**,** 408-414.

KIM, J. N., BLOUNT, K. F., PUSKARZ, I., LIM, J., LINK, K. H. & BREAKER, R. R. (2009) Design and antimicrobial action of purine analogues that bind Guanine riboswitches. *ACS Chem Biol,* 4**,** 915-27.

KIM, J. N., ROTH, A. & BREAKER, R. R. (2007) Guanine riboswitch variants from Mesoplasma florum selectively recognize 2'-deoxyguanosine. *Proceedings of the National Academy of Sciences of the United States of America,* 104**,** 16092-16097.

KLEIN, D. J., BEEN, M. D. & FERRE-D'AMARE, A. R. (2007) Essential role of an active-site guanine in glmS ribozyme catalysis. *Journal of the American Chemical Society,* 129**,** 14858-9.

KLEIN, D. J., EDWARDS, T. E. & FERRE-D'AMARE, A. R. (2009) Cocrystal structure of a class I preQ1 riboswitch reveals a pseudoknot recognizing an essential hypermodified nucleobase. *Nature Structural & Molecular Biology,* 16**,** 343-4.

KLEIN, D. J. & FERRE-D'AMARE, A. R. (2006) Structural basis of glmS ribozyme activation by glucosamine-6-phosphate. *Science (New York, N.Y.),* 313**,** 1752-1756.

KLEIN, D. J. & FERRE-D'AMARE, A. R. (2009) Crystallization of the glmS ribozyme-riboswitch. *Methods in Molecular Biology,* 540**,** 129-39.

KOONIN, E. V., WOLF, Y. I. & ARAVIND, L. (2001) Prediction of the archaeal exosome and its connections with the proteasome and the translation and transcription machineries by a comparative-genomic approach

10.1101/gr.162001 [doi]. *Genome Res,* 11**,** 240-52.

KWON, M. & STROBEL, S. A. (2008) Chemical basis of glycine riboswitch cooperativity. *RNA (New York, N.Y.),* 14**,** 25-34.

LEE, A., HENRAS, A. K. & CHANFREAU, G. (2005) Multiple RNA surveillance pathways limit aberrant expression of iron uptake mRNAs and prevent iron toxicity in S. cerevisiae

S1097-2765(05)01348-1 [pii] 10.1016/j.molcel.2005.05.021 [doi]. *Mol Cell,* 19**,** 39-51.

LEE, E. R., BLOUNT, K. F. & BREAKER, R. R. (2009) Roseoflavin is a natural antibacterial compound that binds to FMN riboswitches and regulates gene expression. *RNA Biol,* 6**,** 187-94.

LESLIE, A. G. (2006) The integration of macromolecular diffraction data. *Acta Crystallogr D Biol Crystallogr,* 62**,** 48-57.

LIU, Q., GREIMANN, J. C. & LIMA, C. D. (2006) Reconstitution, activities, and structure of the eukaryotic RNA exosome

S0092-8674(06)01427-9 [pii] 10.1016/j.cell.2006.10.037 [doi]. *Cell,* 127**,** 1223-37.

LOH, E., DUSSURGET, O., GRIPENLAND, J., VAITKEVICIUS, K., TIENSUU, T., MANDIN, P., REPOILA, F., BUCHRIESER, C., COSSART, P. & JOHANSSON, J. (2009) A trans-acting riboswitch controls expression of the virulence regulator PrfA in Listeria monocytogenes. *Cell,* 139**,** 770-9.

LORENTZEN, E. & CONTI, E. (2005) Structural basis of 3' end RNA recognition and exoribonucleolytic cleavage by an exosome RNase PH core

S1097-2765(05)01715-6 [pii] 10.1016/j.molcel.2005.10.020 [doi]. *Mol Cell,* 20**,** 473-81.

LORENTZEN, E. & CONTI, E. (2006) The exosome and the proteasome: nano-compartments for degradation

S0092-8674(06)00567-8 [pii] 10.1016/j.cell.2006.05.002 [doi]. *Cell,* 125**,** 651-4.

LORENTZEN, E. & CONTI, E. (2008) Expression, reconstitution, and structure of an archaeal RNA degrading exosome. *Methods in Enzymology,* 447**,** 417-35.

LORENTZEN, E., DZIEMBOWSKI, A., LINDNER, D., SERAPHIN, B. & CONTI, E. (2007) RNA channelling by the archaeal exosome

7400945 [pii] 10.1038/sj.embor.7400945 [doi]. *EMBO Rep,* 8**,** 470-6.

LORENTZEN, E., WALTER, P., FRIBOURG, S., EVGUENIEVA-HACKENBERG, E., KLUG, G. & CONTI, E. (2005) The archaeal exosome core is a hexameric ring structure with three catalytic subunits. *Nature Structural & Molecular Biology,* 12**,** 575-81.

LU, C., SMITH, A. M., FUCHS, R. T., DING, F., RAJASHANKAR, K., HENKIN, T. M. & KE, A. (2008) Crystal structures of the SAM-III/S(MK) riboswitch reveal the SAM-dependent translation inhibition mechanism. *Nature Structural & Molecular Biology*.

MAKAROVA, K. S. & KOONIN, E. V. (2005) Evolutionary and functional genomics of the Archaea

S1369-5274(05)00119-0 [pii] 10.1016/j.mib.2005.08.003 [doi]. *Curr Opin Microbiol,* 8**,** 586-94.

MANDAL, M., BOESE, B., BARRICK, J. E., WINKLER, W. C. & BREAKER, R. R. (2003) Riboswitches control fundamental biochemical pathways in Bacillus subtilis and other bacteria. *Cell,* 113**,** 577-586.

MANDAL, M. & BREAKER, R. R. (2004a) Adenine riboswitches and gene activation by disruption of a transcription terminator. *Nature structural & molecular biology,* 11**,** 29-35.

MANDAL, M. & BREAKER, R. R. (2004b) Gene regulation by riboswitches. *Nature reviews.Molecular cell biology,* 5**,** 451-463.

MANDAL, M., LEE, M., BARRICK, J. E., WEINBERG, Z., EMILSSON, G. M., RUZZO, W. L. & BREAKER, R. R. (2004) A glycine-dependent riboswitch that uses cooperative binding to control gene expression. *Science (New York, N.Y.),* 306**,** 275-9.

MCCARTHY, T. J., PLOG, M. A., FLOY, S. A., JANSEN, J. A., SOUKUP, J. K. & SOUKUP, G. A. (2005) Ligand requirements for glmS ribozyme self-cleavage. *Chemistry & biology,* 12**,** 1221-6.

MCDANIEL, B. A., GRUNDY, F. J., ARTSIMOVITCH, I. & HENKIN, T. M. (2003) Transcription termination control of the S box system: direct measurement of S-adenosylmethionine by the leader RNA. *Proceedings of the*

National Academy of Sciences of the United States of America, 100**,** 3083-3088.

MCDANIEL, B. A., GRUNDY, F. J. & HENKIN, T. M. (2005) A tertiary structural element in S box leader RNAs is required for S-adenosylmethionine-directed transcription termination. *Molecular microbiology,* 57**,** 1008-1021.

MERINO, E. J., WILKINSON, K. A., COUGHLAN, J. L. & WEEKS, K. M. (2005) RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). *J Am Chem Soc,* 127**,** 4223-31.

MERRITT, E. A. (1999) Expanding the model: anisotropic displacement parameters in protein structure refinement. *Acta Crystallogr D Biol Crystallogr,* 55**,** 1109-17.

MEYER, M. M., ROTH, A., CHERVIN, S. M., GARCIA, G. A. & BREAKER, R. R. (2008) Confirmation of a second natural preQ1 aptamer class in Streptococcaceae bacteria. *RNA (New York, N.Y.),* 14**,** 685-95.

MILLER, J. H. (Ed.) (1972) *Experiments in molecular genetics.*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

MILLIGAN, L., TORCHET, C., ALLMANG, C., SHIPMAN, T. & TOLLERVEY, D. (2005) A nuclear surveillance pathway for mRNAs with defective polyadenylation

25/22/9996 [pii] 10.1128/MCB.25.22.9996-10004.2005 [doi]. *Mol Cell Biol,* 25**,** 9996-10004.

MITCHELL, P., PETFALSKI, E., HOUALLA, R., PODTELEJNIKOV, A., MANN, M. & TOLLERVEY, D. (2003) Rrp47p is an exosome-associated protein required for the 3' processing of stable RNAs. *Mol Cell Biol,* 23**,** 6982-92.

MITCHELL, P., PETFALSKI, E., SHEVCHENKO, A., MANN, M. & TOLLERVEY, D. (1997) The exosome: a conserved eukaryotic RNA processing complex containing multiple 3'-->5' exoribonucleases S0092-8674(00)80432-8 [pii]. *Cell,* 91**,** 457-66.

MONTANGE, R. K. & BATEY, R. T. (2006) Structure of the S-adenosylmethionine riboswitch regulatory mRNA element. *Nature,* 441**,** 1172-1175.

MONTANGE, R. K. & BATEY, R. T. (2008) Riboswitches: emerging themes in RNA structure and function. *Annu Rev Biophys,* 37**,** 117-33.

MONTANGE, R. K., MONDRAGON, E., VAN TYNE, D., GARST, A. D., CERES, P. & BATEY, R. T. (2009) Discrimination between Closely Related Cellular Metabolites by the SAM-I Riboswitch. *Journal of Molecular Biology*.

MORITA, M. T., TANAKA, Y., KODAMA, T. S., KYOGOKU, Y., YANAGI, H. & YURA, T. (1999) Translational induction of heat shock transcription factor sigma32: evidence for a built-in RNA thermosensor. *Genes & Development,* 13**,** 655-65.

MORTIMER, S. A. & WEEKS, K. M. (2007) A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. *J Am Chem Soc,* 129**,** 4144-5.

MORTIMER, S. A. & WEEKS, K. M. (2009) Time-resolved RNA SHAPE chemistry: quantitative RNA structure analysis in one-second snapshots and at single-nucleotide resolution. *Nat Protoc,* 4**,** 1413-21.

MULHBACHER, J., BROUILLETTE, E., ALLARD, M., FORTIER, L. C., MALOUIN, F. & LAFONTAINE, D. A. (2010) Novel riboswitch ligand analogs as selective inhibitors of guanine-related metabolic pathways. *PLoS Pathog,* 6**,** e1000865.

NAHVI, A., SUDARSAN, N., EBERT, M. S., ZOU, X., BROWN, K. L. & BREAKER, R. R. (2002) Genetic control by a metabolite binding mRNA. *Chemistry & biology,* 9**,** 1043.

NAVARRO, M. V., OLIVEIRA, C. C., ZANCHIN, N. I. & GUIMARAES, B. G. (2008) Insights into the mechanism of progressive RNA degradation by the archaeal exosome

M801005200 [pii] 10.1074/jbc.M801005200 [doi]. *J Biol Chem*.

NUDLER, E. & MIRONOV, A. S. (2004) The riboswitch control of bacterial metabolism. *Trends in Biochemical Sciences,* 29**,** 11-7.

ORBAN, T. I. & IZAURRALDE, E. (2005) Decay of mRNAs targeted by RISC requires XRN1, the Ski complex, and the exosome

rna.7231505 [pii] 10.1261/rna.7231505 [doi]. *RNA,* 11**,** 459-69.

OTWINOWSKI, Z. & MINOR, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods in enzymology,* 276**,** 307-326.

PARKER, R. & SONG, H. (2004) The enzymes and control of eukaryotic mRNA turnover. *Nature Structural & Molecular Biology,* 11**,** 121-7.

PORTNOY, V., EVGUENIEVA-HACKENBERG, E., KLEIN, F., WALTER, P., LORENTZEN, E., KLUG, G. & SCHUSTER, G. (2005) RNA polyadenylation in Archaea: not observed in Haloferax while the exosome polynucleotidylates RNA in Sulfolobus

7400571 [pii] 10.1038/sj.embor.7400571 [doi]. *EMBO Rep,* 6**,** 1188-93.

REICHOW, S. & VARANI, G. (2006) Structural biology: RNA switches function. *Nature,* 441**,** 1054-5.

RODIONOV, D. A., VITRESCHAK, A. G., MIRONOV, A. A. & GELFAND, M. S. (2003) Regulation of lysine biosynthesis and transport genes in bacteria: yet another RNA riboswitch? *Nucleic Acids Research,* 31**,** 6748-57.

ROTH, A., WINKLER, W. C., REGULSKI, E. E., LEE, B. W., LIM, J., JONA, I., BARRICK, J. E., RITWIK, A., KIM, J. N., WELZ, R., IWATA-REUYL, D. & BREAKER, R. R. (2007) A riboswitch selective for the queuosine precursor preQ1 contains an unusually small aptamer domain. *Nature structural & molecular biology,* 14**,** 308-317.

SCHWALBE, H., BUCK, J., FURTIG, B., NOESKE, J. & WOHNERT, J. (2007) Structures of RNA switches: insight into molecular recognition and tertiary structure. *Angewandte Chemie (International ed.in English),* 46**,** 1212-1219.

SELLECK, W. & TAN, S. (2008) Recombinant protein complex expression in E. coli. *Curr Protoc Protein Sci,* Chapter 5**,** Unit 5 21.

SERGANOV, A. (2009) The long and the short of riboswitches. *Current Opinion in Structural Biology,* 19**,** 251-9.

SERGANOV, A. & PATEL, D. J. (2009) Amino acid recognition and gene regulation by riboswitches. *Biochimica et Biophysica Acta,* 1789**,** 592-611.

SERGANOV, A., POLONSKAIA, A., PHAN, A. T., BREAKER, R. R. & PATEL, D. J. (2006) Structural basis for gene regulation by a thiamine pyrophosphate-sensing riboswitch. *Nature,* 441**,** 1167-1171.

SERGANOV, A., YUAN, Y. R., PIKOVSKAYA, O., POLONSKAIA, A., MALININA, L., PHAN, A. T., HOBARTNER, C., MICURA, R., BREAKER, R. R. & PATEL, D. J. (2004) Structural basis for discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs. *Chemistry & biology,* 11**,** 1729-1741.

SHELDRICK, G. M. (2008) A short history of SHELX. *Acta Crystallogr A,* 64**,** 112-22.

SHI, Z., YANG, W. Z., LIN-CHAO, S., CHAK, K. F. & YUAN, H. S. (2008) Crystal structure of Escherichia coli PNPase: central channel residues are involved in processive RNA degradation. *RNA (New York, N.Y.),* 14**,** 2361-71.

STOLOWITZ, M. L. & MINCH, M. J. (1981) S- Adenosyl-L-methionine and 5'-adenosyl-L-homocysteine, an NMR Study. *J. Am. Chem. Soc.,* 103**,** 6015-6019.

STORONI, L. C., MCCOY, A. J. & READ, R. J. (2004) Likelihood-enhanced fast rotation functions. *Acta crystallographica.Section D, Biological crystallography,* 60**,** 432-438.

STORZ, G., ALTUVIA, S. & WASSARMAN, K. M. (2005) An abundance of RNA regulators. *Annual Review of Biochemistry,* 74**,** 199-217.

SUDARSAN, N., BARRICK, J. E. & BREAKER, R. R. (2003a) Metabolite-binding RNA domains are present in the genes of eukaryotes. *RNA (New York, N.Y.),* 9**,** 644-647.

SUDARSAN, N., COHEN-CHALAMISH, S., NAKAMURA, S., EMILSSON, G. M. & BREAKER, R. R. (2005) Thiamine pyrophosphate riboswitches are targets for the antimicrobial compound pyrithiamine. *Chemistry & biology,* 12**,** 1325-1335.

SUDARSAN, N., HAMMOND, M. C., BLOCK, K. F., WELZ, R., BARRICK, J. E., ROTH, A. & BREAKER, R. R. (2006) Tandem riboswitch architectures exhibit complex gene control functions. *Science (New York, N.Y.),* 314**,** 300-304.

SUDARSAN, N., WICKISER, J. K., NAKAMURA, S., EBERT, M. S. & BREAKER, R. R. (2003b) An mRNA structure in bacteria that controls gene expression by binding lysine. *Genes & Development,* 17**,** 2688-97.

SYMMONS, M. F., JONES, G. H. & LUISI, B. F. (2000) A duplicated fold is the structural basis for polynucleotide phosphorylase catalytic activity, processivity, and regulation. *Structure,* 8**,** 1215-26.

TAN, S., KERN, R. C. & SELLECK, W. (2005) The pST44 polycistronic expression system for producing protein complexes in Escherichia coli. *Protein Expression and Purification,* 40**,** 385-95.

THORE, S., LEIBUNDGUT, M. & BAN, N. (2006) Structure of the eukaryotic thiamine pyrophosphate riboswitch with its regulatory ligand. *Science (New York, N.Y.),* 312**,** 1208-11.

TOMSIC, J., MCDANIEL, B. A., GRUNDY, F. J. & HENKIN, T. M. (2007) Natural variability in SAM-dependent riboswitches: S box elements in Bacillus subtilis exhibit differential sensitivity to SAM in vivo and in vitro. *Journal of Bacteriology,* 190**,** 822-833.

TOMSIC, J., MCDANIEL, B. A., GRUNDY, F. J. & HENKIN, T. M. (2008) Natural variability in S-adenosylmethionine (SAM)-dependent riboswitches: S-box elements in bacillus subtilis exhibit differential sensitivity to SAM In vivo and in vitro. *Journal of Bacteriology,* 190**,** 823-33.

TUCKER, B. J. & BREAKER, R. R. (2005) Riboswitches as versatile gene control elements. *Current Opinion in Structural Biology,* 15**,** 342-8.

UELAND, P. M. (1982) Pharmacological and biochemical aspects of S-adenosylhomocysteine and S-adenosylhomocysteine hydrolase. *Pharmacol Rev,* 34**,** 223-253.

VAN HOOF, A., FRISCHMEYER, P. A., DIETZ, H. C. & PARKER, R. (2002) Exosome-mediated recognition and degradation of mRNAs lacking a termination codon

10.1126/science.1067272 [doi] 295/5563/2262 [pii]. *Science,* 295**,** 2262-4.

VAN HOOF, A., LENNERTZ, P. & PARKER, R. (2000a) Yeast exosome mutants accumulate 3'-extended polyadenylated forms of U4 small nuclear RNA and small nucleolar RNAs. *Mol Cell Biol,* 20**,** 441-52.

VAN HOOF, A., STAPLES, R. R., BAKER, R. E. & PARKER, R. (2000b) Function of the ski4p (Csl4p) and Ski7p proteins in 3'-to-5' degradation of mRNA. *Mol Cell Biol,* 20**,** 8230-43.

VASA, S. M., GUEX, N., WILKINSON, K. A., WEEKS, K. M. & GIDDINGS, M. C. (2008) ShapeFinder: a software system for high-throughput quantitative analysis of nucleic acid reactivity information resolved by capillary electrophoresis. *RNA (New York, N.Y.),* 14**,** 1979-90.

VASUDEVAN, S. & PELTZ, S. W. (2003) Nuclear mRNA surveillance S0955067403000516 [pii]. *Curr Opin Cell Biol,* 15**,** 332-7.

VITRESCHAK, A. G., RODIONOV, D. A., MIRONOV, A. A. & GELFAND, M. S. (2004) Riboswitches: the oldest mechanism for the regulation of gene expression? *Trends in genetics : TIG,* 20**,** 44-50.

WACHTER, A., TUNC-OZDEMIR, M., GROVE, B. C., GREEN, P. J., SHINTANI, D. K. & BREAKER, R. R. (2007) Riboswitch Control of Gene Expression in Plants by Splicing and Alternative 3' End Processing of mRNAs. *The Plant Cell*.

WALTER, P., KLEIN, F., LORENTZEN, E., ILCHMANN, A., KLUG, G. & EVGUENIEVA-HACKENBERG, E. (2006) Characterization of native and reconstituted exosome complexes from the hyperthermophilic archaeon Sulfolobus solfataricus MMI5393 [pii] 10.1111/j.1365-2958.2006.05393.x [doi]. *Mol Microbiol,* 62**,** 1076-89.

WANG, J. X., LEE, E. R., MORALES, D. R., LIM, J. & BREAKER, R. R. (2008) Riboswitches that sense S-adenosylhomocysteine and activate genes involved in coenzyme recycling. *Molecular Cell,* 29**,** 691-702.

WATTS, J. M., DANG, K. K., GORELICK, R. J., LEONARD, C. W., BESS, J. W., JR., SWANSTROM, R., BURCH, C. L. & WEEKS, K. M. (2009) Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature,* 460**,** 711-6.

WEINBERG, Z., BARRICK, J. E., YAO, Z., ROTH, A., KIM, J. N., GORE, J., WANG, J. X., LEE, E. R., BLOCK, K. F., SUDARSAN, N., NEPH, S., TOMPA, M., RUZZO, W. L. & BREAKER, R. R. (2007) Identification of 22 candidate structured RNAs in bacteria using the CMfinder comparative genomics pipeline. *Nucleic acids research,* 35**,** 4809-4819.

WEINBERG, Z., REGULSKI, E. E., HAMMOND, M. C., BARRICK, J. E., YAO, Z., RUZZO, W. L. & BREAKER, R. R. (2008) The aptamer core of SAM-IV riboswitches mimics the ligand-binding site of SAM-I riboswitches. *RNA (New York, N.Y.),* 14**,** 822-8.

WEINBERG, Z., WANG, J. X., BOGUE, J., YANG, J., CORBINO, K., MOY, R. H. & BREAKER, R. R. Comparative genomics reveals 104 candidate structured RNAs from bacteria, archaea, and their metagenomes. *Genome biology,* 11**,** R31.

WELZ, R. & BREAKER, R. R. (2007) Ligand binding and gene control characteristics of tandem riboswitches in Bacillus anthracis. *RNA (New York, N.Y.),* 13**,** 573-82.

WHITE, H. B., 3RD (1976) Coenzymes as fossils of an earlier metabolic state. *Journal of Molecular Evolution,* 7**,** 101-4.

WICKISER, J. K., WINKLER, W. C., BREAKER, R. R. & CROTHERS, D. M. (2005) The speed of RNA transcription and metabolite binding kinetics operate an FMN riboswitch. *Molecular Cell,* 18**,** 49-60.

WILKINSON, K. A., MERINO, E. J. & WEEKS, K. M. (2005a) RNA SHAPE chemistry reveals nonhierarchical interactions dominate equilibrium structural transitions in tRNA(Asp) transcripts. *Journal of the American Chemical Society,* 127**,** 4659-67.

WILKINSON, K. A., MERINO, E. J. & WEEKS, K. M. (2005b) RNA SHAPE chemistry reveals nonhierarchical interactions dominate equilibrium structural transitions in tRNA(Asp) transcripts. *J Am Chem Soc,* 127**,** 4659-67.

WILKINSON, K. A., MERINO, E. J. & WEEKS, K. M. (2006) Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nat Protoc,* 1**,** 1610-6.

WILKINSON, K. A., VASA, S. M., DEIGAN, K. E., MORTIMER, S. A., GIDDINGS, M. C. & WEEKS, K. M. (2009) Influence of nucleotide identity on ribose 2'-hydroxyl reactivity in RNA. *RNA (New York, N.Y.),* 15**,** 1314-21.

WILKINSON, S. R. & BEEN, M. D. (2005) A pseudoknot in the 3' non-core region of the glmS ribozyme enhances self-cleavage activity. *RNA (New York, N.Y.),* 11**,** 1788-94.

WINKLER, W., NAHVI, A. & BREAKER, R. R. (2002a) Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature,* 419**,** 952-956.

WINKLER, W. C. (2005a) Metabolic monitoring by bacterial mRNAs. *Arch Microbiol,* 183**,** 151-9.

WINKLER, W. C. (2005b) Riboswitches and the role of noncoding RNAs in bacterial metabolic control. *Current opinion in chemical biology,* 9**,** 594-602.

WINKLER, W. C. & BREAKER, R. R. (2005) Regulation of bacterial gene expression by riboswitches. *Annual Review of Microbiology,* 59**,** 487-517.

WINKLER, W. C., COHEN-CHALAMISH, S. & BREAKER, R. R. (2002b) An mRNA structure that controls gene expression by binding FMN. *Proceedings of the National Academy of Sciences of the United States of America,* 99**,** 15908-15913.

WINKLER, W. C., NAHVI, A., ROTH, A., COLLINS, J. A. & BREAKER, R. R. (2004) Control of gene expression by a natural metabolite-responsive ribozyme. *Nature,* 428**,** 281-286.

WINKLER, W. C., NAHVI, A., SUDARSAN, N., BARRICK, J. E. & BREAKER, R. R. (2003) An mRNA structure that controls gene expression by binding S-adenosylmethionine. *Nature structural biology,* 10**,** 701-707.

YAO, Z., WEINBERG, Z. & RUZZO, W. L. (2006) CMfinder--a covariance model based RNA motif finding algorithm. *Bioinformatics,* 22**,** 445-52.

YOUSEF, M. R., GRUNDY, F. J. & HENKIN, T. M. (2003) tRNA requirements for glyQS antitermination: a new twist on tRNA. *RNA (New York, N.Y.),* 9**,** 1148-1156.

YOUSEF, M. R., GRUNDY, F. J. & HENKIN, T. M. (2005) Structural transitions induced by the interaction between tRNA(Gly) and the Bacillus subtilis glyQS T box leader RNA. *Journal of Molecular Biology,* 349**,** 273-287.

ZHANG, R., OU, H. Y. & ZHANG, C. T. (2004) DEG: a database of essential genes. *Nucleic acids research,* 32**,** D271-2.