Theses and Dissertations

Graduate School

2007

The KsgA methyltransferase: Characterization of a universally conserved protein involved in robosome biogenesis

Heather Colleen O'Farrell
Virginia Commonwealth University

THE KSGA METHYLTRANSFERASE: CHARACTERIZATION OF A

UNIVERSALLY CONSERVED PROTEIN INVOLVED IN RIBOSOME BIOGENESIS

A Dissertation submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy at Virginia Commonwealth University

by

HEATHER COLLEEN O'FARRELL
Bachelor of Arts, University of Virginia, 1996

Director: JASON P. RIFE, PH.D.
PROFESSOR, DEPARTMENT OF MEDICINAL CHEMISTRY

Virginia Commonwealth University
Richmond, Virginia
May 2007

Acknowledgement

My first and loudest thank you goes to Jason Rife, without whose encouragement I would still be a lab tech. Jason has a nearly infinite supply of patience, which even I couldn't quite exhaust, and he managed to turn me into a scientist when I wasn't even looking.

I would also like to thank my committee members, who were all helpful in various ways, and also the faculty members who served as temporary committee members at various times in the last five years.

Finally, I would like to thank my family and friends, who helped keep me (mostly) sane over the last several years. This dissertation is dedicated to my grandfather, the first "Doc" O'Farrell.

Table of Contents

List of Tables

List of Figures

Abstract

The KsgA methyltransferase: Characterization of a universally conserved protein involved

in ribosome biogenesis

By Heather Colleen O'Farrell, Ph.D.

A Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University.

Virginia Commonwealth University, 2007

Major Director: Jason P. Rife
Professor, Department of Medicinal Chemistry

The KsgA enzymes comprise an ancient family of methyltransferases that are intimately involved in ribosome biogenesis. Ribosome biogenesis is a complicated process, involving numerous cleavage, base modification and assembly steps. All ribosomes share the same general architecture, with small and large subunits made up of roughly similar rRNA species and a variety of ribosomal proteins. However, the fundamental assembly process differs significantly between eukaryotes and eubacteria, not only in distribution and mechanism of modifications but also in organization of assembly steps. Despite these

differences, members of the KsgA/Dim1 methyltransferase family and their resultant modification of small-subunit rRNA are found throughout evolution, and therefore were present in the last common ancestor.

The first member of the family to be described, KsgA from *Escherichia coli*, was initially shown to be the determining factor for resistance/sensitivity to the antibiotic kasugamycin and was subsequently found to dimethylate two adenosines in 16S rRNA during maturation of the 30S subunit. Since then, numerous other members of the family have been characterized in eubacteria, eukaryotes, archaea and in eukaryotic organelles. The eukaryotic ortholog, Dim1, is essential for proper processing of the pre-rRNA, in addition to and separate from its methyltransferase function. The KsgA/Dim1 family bears sequence and structural similarity to a larger group of S-adenosyl-L-methionine dependent methyltransferases, which includes both DNA and RNA methyltransferases.

In this document we report that KsgA orthologs from archaea and eukaryotes are able to complement for KsgA function in bacteria, both *in vivo* and *in vitro*. This indicates that all of these enzymes can recognize a common ribosomal substrate, and that the recognition elements must be largely unchanged since the evolutionary split between the three domains of life. We have characterized KsgA structurally, and discuss aspects of KsgA's activity in light of the structural data. We also propose a model for KsgA binding to the 30S subunit, based on solution probing data. This model sheds light on KsgA's unusual regulation and on the dual function of the Dim1 enzymes.

Introduction

Ribosome biogenesis

Ribosome biogenesis is one of the fundamental processes of the cellular machinery, and growing cells devote an extraordinary amount of energy to the production of ribosomes. This requires precise and coordinated regulation of rRNA transcription, production of ribosomal proteins, and production of extra-ribosomal factors that are involved in ribosome maturation. In addition, the pre-rRNA transcript must be co- and post-transcriptionally processed and modified to produce the mature RNA species.

Given the ribosome's importance, our understanding of it has to some extent lagged behind other fields. A major milestone in ribosome research was the *in vitro* assembly of bacterial ribosomes using only the component rRNAs and ribosomal proteins[1-3]. Subsequently, 30S subunits were reconstituted using *in vitro* transcribed 16S rRNA, which lacked modifications[4], and using purified recombinant proteins[5]. Although ribosomes assembled *in vitro* are less active in *in vitro* assays than natural purified ribosomes[3, 5], given the size and intricacy of the structures, and their crucial role in the cell, *in vitro* assembly is a remarkable and important achievement. Another milestone in our knowledge of the ribosome was the first high-resolution crystal structures of ribosomal subunits, the large subunit from the archaeon *Haloarcula marismortui*[6] and the small subunit from the eubacterium *Thermus thermophilus*[7]. These structures were followed by that of the large

1

subunit from the eubacterium *Deinococcus radiodurans*[8] and of 70S ribosomes from *T. thermophilus*[9] and *E. coli*[10]. More recently, Selmer *et al.* have solved the structure of the bacterial ribosome with bound mRNA and tRNA[11]. These structures served as definitive proof that the ribosome is, indeed, a ribozyme, and have proven invaluable in our understanding of the structure and function of this complex molecule.

Although functional bacterial ribosomes can be assembled *in vitro* in the absence of accessory proteins[1-3], additional factors are required *in vivo* for optimal assembly[12]. With respect to the 30S subunit, a diverse group of proteins are indirectly necessary for proper processing of the pre-16S rRNA, apart from the nucleases that are directly responsible for rRNA cleavage: Era[13] and RsgA[14], which are small GTPases; and RbfA[15], RimM[15], and RimN[16]. Depletion of any of these proteins results in accumulation of immature forms of 16S rRNA. These pre-16S species are similar to each other and resemble the 17S precursor rRNA, which is trimmed at the 5' and 3' ends to produce the mature 16S. Era, RbfA, and RimN seem to be important for ribosome assembly at low temperatures, and RbfA is considered a cold-shock protein. Era and RimN are both essential, even at normal temperatures; RimM, RbfA, and RgsA are not essential, at least under optimal conditions. In wild-type cells the 17S 5' and 3' processing occurs in the context of a ribosomal particle[17]. Therefore, depletion of these five proteins may impair 17S processing by interfering with 30S assembly. Era, RbfA and RimM seem to have some degree of overlap in function, as measured by complementation; overexpression of Era complements RbfA knockout, and RbfA compensates for lack of RimM.

In vitro assembly experiments have allowed construction of bacterial 30S subunit assembly maps[18]. Small-subunit ribosomal proteins assemble onto the pre-16S rRNA in a cooperative manner in a roughly 5' to 3' order, corresponding to co-transcriptional assembly *in vivo* (Figure 1a). This means that the body (5' region) of the ribosome is assembled first, followed by platform (central region) assembly, and the head (3' major region) is formed last. The final two helices, helix 44 and helix 45, are referred to as the 3' minor region; they assemble along the body and platform regions (Figure 1b). *In vitro* assembly occurs through two defined reconstitution intermediates, RI and RI*[19] (Figure 2). RI consists of 16S rRNA plus a subset of ribosomal proteins corresponding to the earlier binding proteins in the assembly map. Unimolecular rearrangement of RI yields RI*, which is able to bind the remaining proteins and form the 30S subunit. The RI to RI* transition is the rate-limiting step in assembly, and requires either heat (42°) or the presence of the DnaK chaperone system[20]. Although RI and RI* are seen in *in vitro* assembly, they strongly resemble intermediate particles seen *in vivo* in certain cold-sensitive mutants[21] and in conditional DnaK mutants[22] at restrictive temperatures.

Eukaryotic ribosome assembly is more complex than that of prokaryotes. Eukaryotic ribosome biogenesis requires a significant number of trans-acting factors, including snoRNPs, helicases, GTPases, nucleases, transport proteins, chaperones, and many more proteins whose function has not been elucidated. In addition, ribosome assembly in eukaryotes is spatially and temporally organized, with early processing and assembly steps taking place in the nucleolus and the final steps being completed after transport to the cytoplasm[23, 24] (Figure 3). As the pre-rRNA is transcribed, a complex is
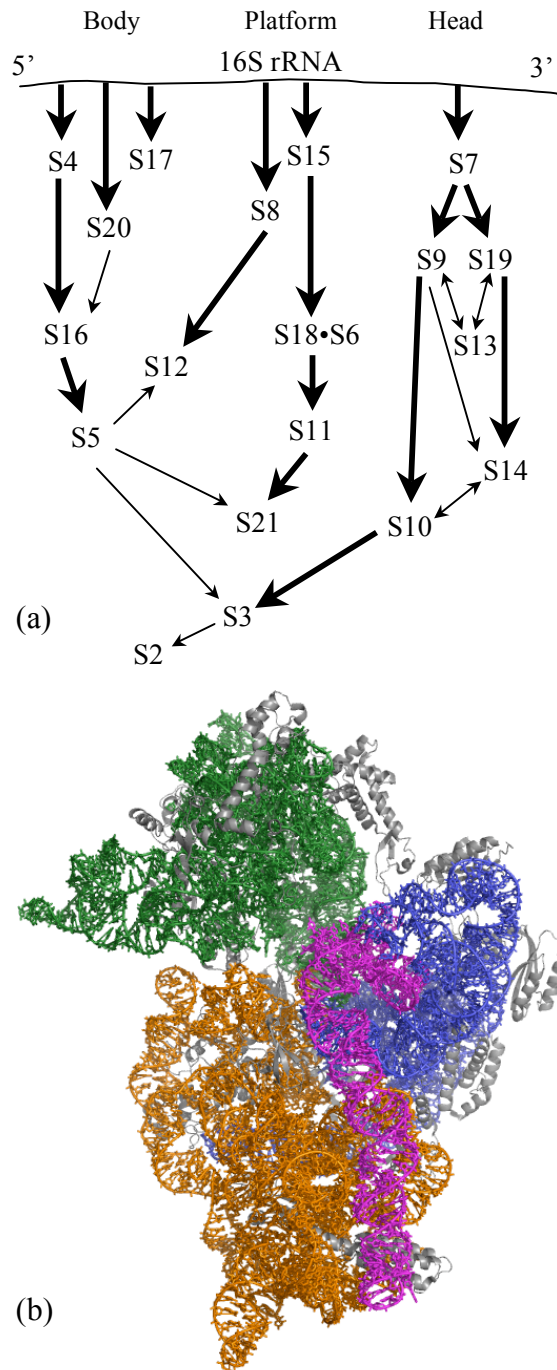
Figure 1. Assembly of the 30S subunit. (a) Assembly map. (b) Domains of the subunit. rRNA in the body is colored orange, the platform is blue, the head is green, and the 3' minor domain is magenta. Ribosomal proteins are gray.
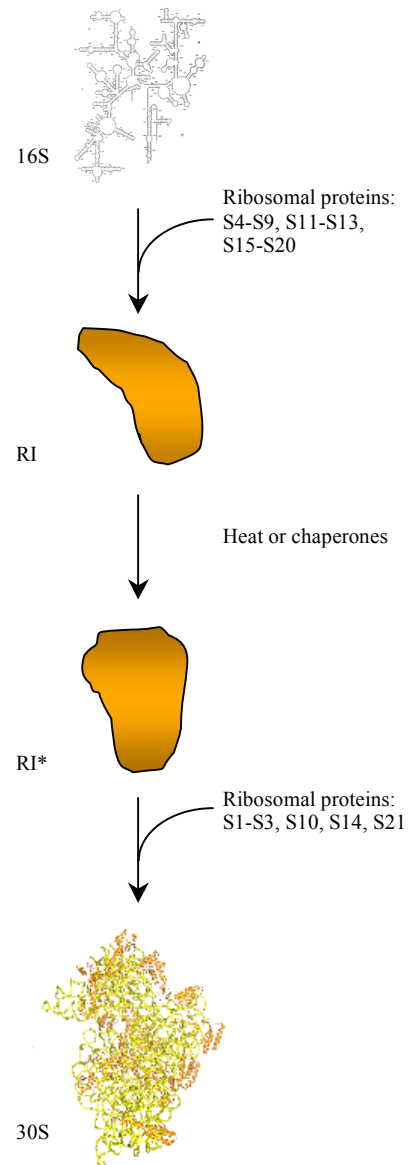
Figure 2. Prokaryotic ribosome assembly.
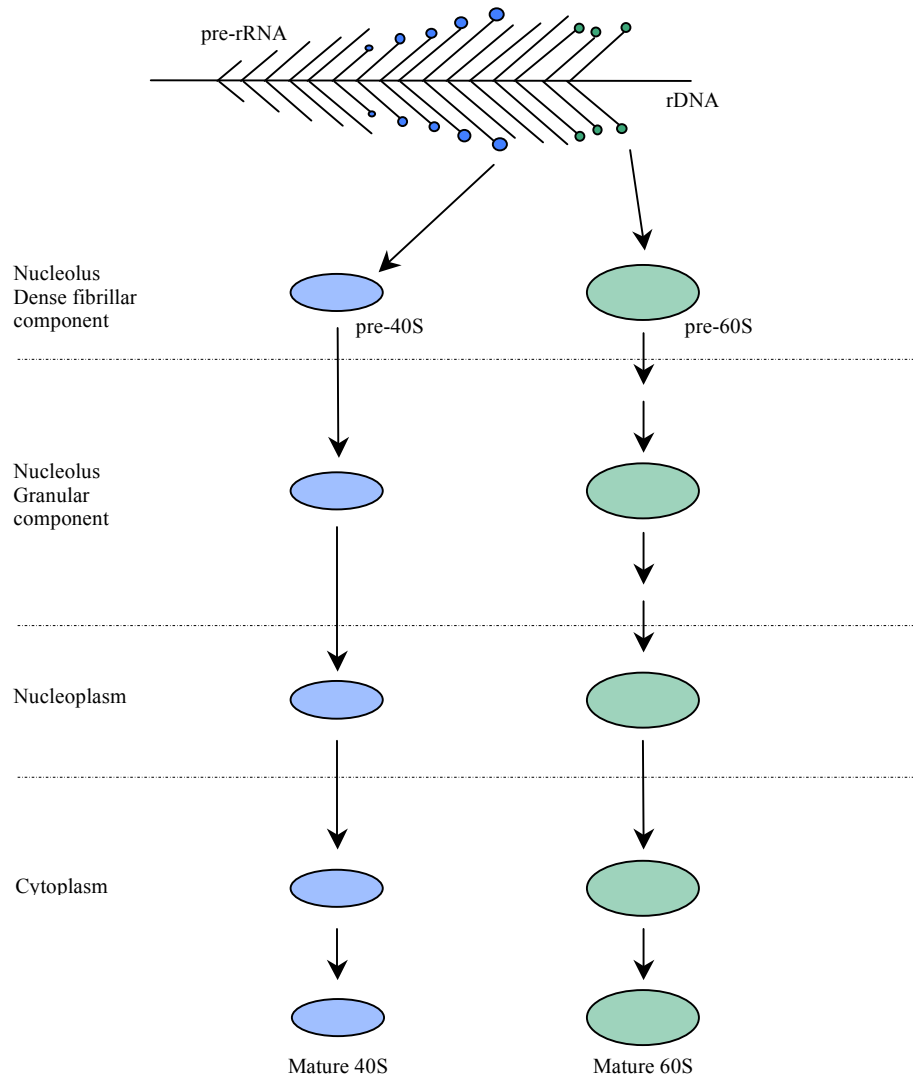
Figure 3. Eukaryotic ribosome assembly. Terminal knobs are shown as blue spheres on the transcribing rRNA.

formed consisting of the U3 snoRNP, small subunit r-proteins and other factors, and the nascent rRNA[25]. This complex corresponds to the terminal knobs seen in Miller chromatin spreads of actively transcribing rDNA. Cleavage events separate the pre-40S particle and the pre-60S particle. The pre-40S particle is exported from the nucleolus through the nucleoplasm to the cytoplasm, where a few final processing steps yield the mature 40S particle. The pre-60S is subject to additional processing steps in the nucleolus and nucleoplasm before export and final maturation steps. Although eukaryotic ribosomes have not been assembled *in vitro* from only component rRNAs and proteins, functional *Dictyostelium discoideum* ribosomes have been assembled from immature rRNA species and purified proteins in the presence of unidentified small nucleolar RNA(s)[26].

There is a dearth of information on ribosome assembly in archaeal systems. Although archaeal cells lack a nuclear membrane, at least one archaeal genome contains putative homologs to nuclear and nucleolar structural genes from eukaryotes[27, 28]. Spatio-temporal control of archaeal ribosome synthesis is therefore a possibility, although highly speculative. Ribosomes from at least one archaeon, the extreme halophile *Haloferax mediterranei*, have been assembled *in vitro* from purified components[29-31].

Common rRNA modifications include conversion of uridine to pseudouridine, methylation on the 2'-O of the ribose moiety, and base methylation[32] (Figure 4). The most frequent modification in bacterial rRNA is base methylation. Modifications in bacteria are performed by individual enzymes; each enzyme catalyzes a single modification or, in the case of some pseudouridylases, modifies up to three specific bases. Eukaryotic rRNAs are more highly modified than bacterial rRNAs, and the modifications are mainly
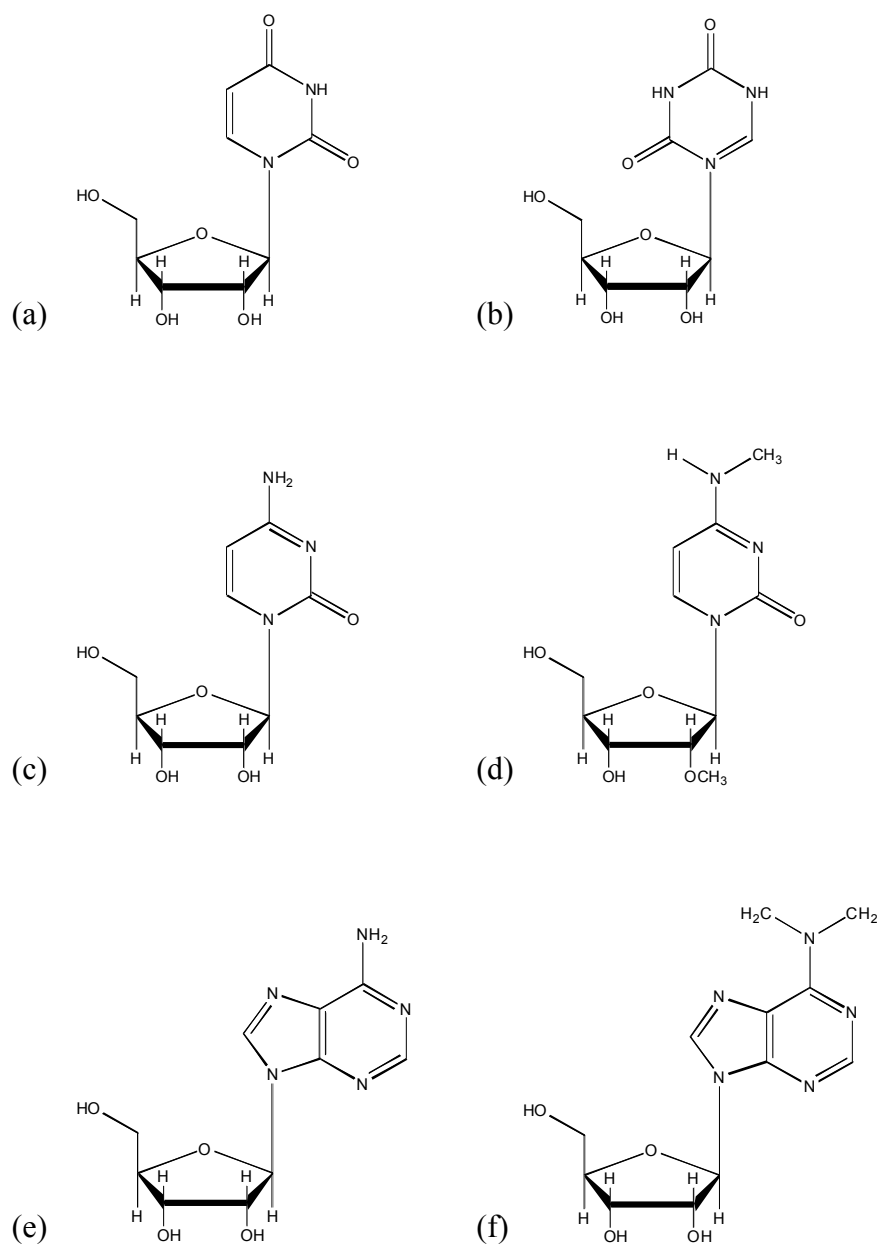
Figure 4. Examples of modified nucleosides. (a) Uridine. (b) Pseudouridine. (c) Cytidine. (d) N^4, 2'-O dimethylcytidine. (e) Adenosine. (f). N^6,N^6-dimethyladenosine.

pseudouridines and 2'-O methyls. In contrast to bacteria, eukaryotic cells generally utilize a common enzyme for each general type of modification, using sequence-specific snoRNAs as guides to direct the modifications to particular bases[33-35]. Archaeal rRNA modification is less well-defined, and there is a great deal of variation between species, but in general the amount of modification falls somewhere between bacteria and eukaryotes, with most of the modifications being 2'-O methyls and relatively few pseudouridines[36]. Archaeal rRNA modification seems to be more similar to that in eukaryotic organisms, with snoRNAs guiding pseudouridylation and ribose methylation[37-39]. Indeed, at least one archaeal snoRNA can direct 2'-O methylation when microinjected into the nucleus of a eukaryotic cell[40].

While post-transcriptional modification of rRNA is common to all life, in most cases the modifications are not very well characterized or understood, and very few specific modifications have been conserved in evolution. A notable exception to this general lack of conservation is the dimethylation of two adjacent adenosines in the 3'-terminal helix of small-subunit rRNA, A1518 and A1519 in helix 45 by *E. coli* numbering (Figure 5). Helix 45 is one of the most highly conserved sequences in small subunit rRNA[41, 42], and the presence of two dimethylated adenosines in the loop of the helix is equally conserved. These residues are thus a rare example of an evolutionarily conserved post-transcriptional rRNA modification. The enzyme responsible for dimethylation of these two adenosines is also universally conserved; in all known instances it has been found to be a member of the KsgA/Dim1 family of methyltransferases.
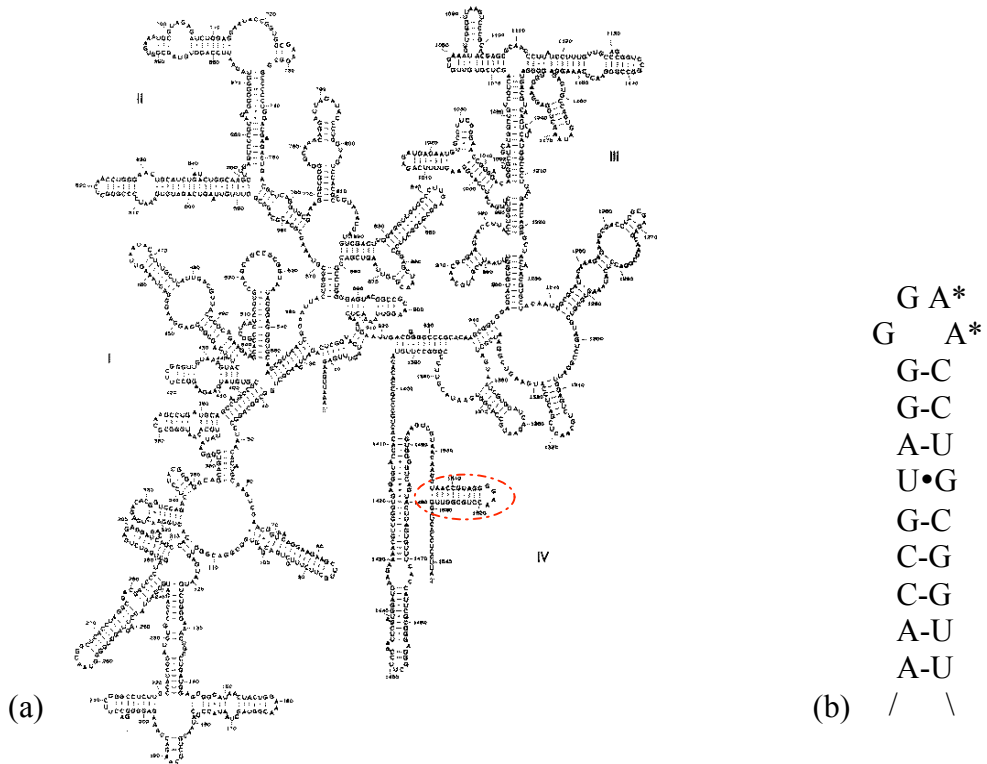
Figure 5. Dimethylated adenosines in 16S rRNA. (a) 16S. Roman numerals refer to the body (I), platform (II), head (III), and 3'-minor (IV) domains. Helix 45 is circled in red. (b) Helix 45. Stars denote modified adenosines.

KsgA

The *ksgA* gene was first described as a locus that determined resistance to the antibiotic kasugamycin (ksg)[43]. Since 16S rRNA from the resistant mutant lacked dimethylation of two adenosines near the 3' end, it was proposed that the *ksgA* locus encoded an rRNA methyltransferase. Subsequent work showed this to be the case, as a methylating activity could be purified from the sensitive strain, but not the resistant strain, and this activity could be used to restore ksg sensitivity to ribosomes from the resistant strain[44]. Methylation was assayed on a variety of substrates; isolated 16S was not a substrate but 21S core particles were. 21S particles consist of 16S plus a subset of ribosomal proteins (Table 1); interestingly, the protein composition of these particles is very similar to the *in vitro* RI and RI*assembly intermediates.

Further study of the substrate requirements for methyltransferase activity defined a minimal substrate as consisting of 16S plus eight ribosomal proteins: S4, S6, S8, S11, and S15-S18[45] (Table 1). Five proteins, S3, S9, S10, S14 and S21, were found by this group to have inhibitory effects on the methyltransferase reaction. In the crystal structure of the 30S subunit, all thirteen of these proteins are on the opposite side of the subunit from the target adenosines and are not in a position to directly interact with KsgA (Figure 6). Therefore, these proteins must regulate KsgA's activity in an indirect manner. All eight of the required proteins reside in the body and platform of the subunit, while all of the inhibitory proteins are found in the head region, with the exception of S21, which was not present in any crystal structure. Additionally, all of the required proteins are present in the *in vitro* RI and RI* intermediates as well as the *in vivo* 21S core particle. None of the

Table 1. Composition of various sub-30S particles.

	21S core particle	RI intermediate	Minimal substrate
S1			
S2			
S3			
S4	+	+	+
S5		+	
S6	+	+	+
S7	+	+	
S8	+	+	+
S9		+	
S10			
S11	+	+	+
S12	+	+	
S13	+	+	
S14			
S15	+	+	+
S16	+	+	+
S17	+	+	+
S18	+	+	+
S19	+	+	
S20	+	+	
S21	+		
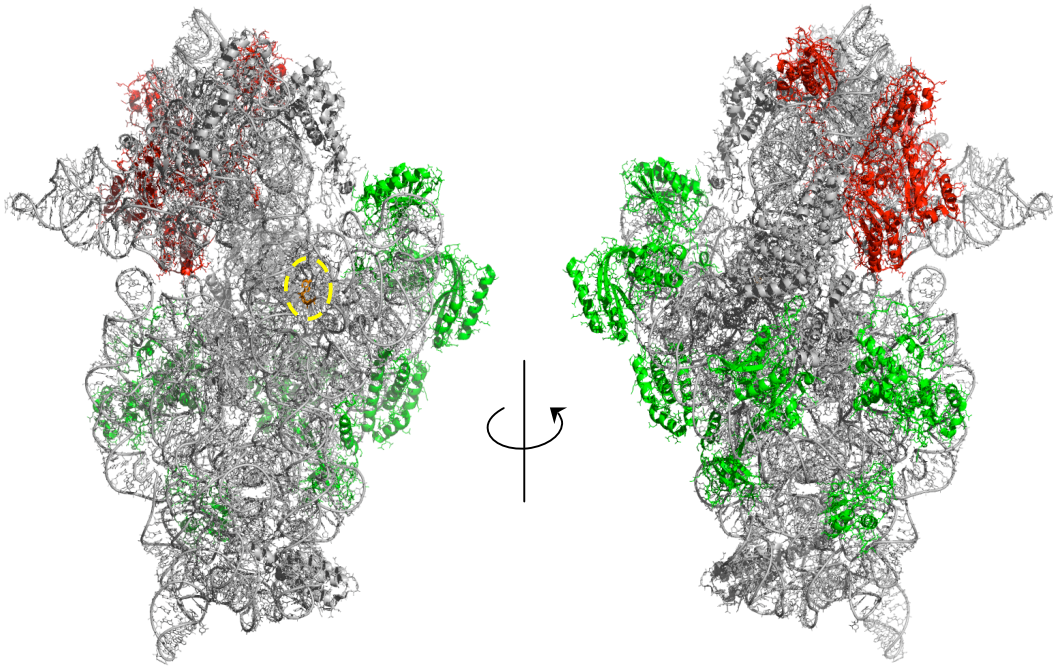
Figure 6. Proteins required for and inhibitory to KsgA activity, according to Thammana and Held. Required proteins are green, and inhibitory proteins are red; all other proteins are gray. Helix 45 is circled in yellow; the target adenosines are shown in orange.

inhibitory proteins are present in these particles, again with the exception of S21, which is found in the 21S core particle but not the RI intermediate. It should be noted that the RI intermediate formed by incubation of 16S with the ribosomal proteins at low temperature is not a substrate for KsgA[46], although it is possible that RI* could serve as a substrate. It seems likely that KsgA is not able to methylate the target adenosines until some level of assembly of the 30S subunit has occurred; this minimum structure includes the body and platform regions of the forming subunit. Once the final head proteins bind and the subunit is fully formed, KsgA is no longer able to methylate the target adenosines. According to Thammana and Held, fully formed 30S are not competent methylation substrates for this reason[45].

Poldermans *et al.*, on the other hand, found that 30S subunits could be methylated, and that of the five "inhibitory" proteins only S21 inhibits KsgA activity[47]. A resolution to this contradiction has recently been proposed by Desai and Rife[48]. 30S subunits purified from bacteria have been found to exist in two distinct conformational states, one that is active in translation assays and one that is translationally inactive[49, 50]. The two conformations are compositionally identical, except that inactive subunits contain more S1 than active subunits[51], and they can be interconverted by varying salt and temperature conditions. Chemical modification experiments reveal only slight variations in rRNA conformation between the two structures, most of which occur near the site of modification by KsgA[52]. In fact, antibodies raised against the dimethylated adenosines bind to inactive subunits but not to active subunits, suggesting that the adenosines are less accessible in the translationally active conformation[53]. In the crystal structures of small subunits from *T.*

thermophilus[7, 9] and *E. coli*[10], which represent active forms, the loop of helix 45 is tucked

into the minor groove of helix 44, confirming the inaccessibility of the target adenosines in

active 30S subunits. Along with this, Desai and Rife found that KsgA is only able to

methylate the translationally inactive form of 30S. They hypothesize that, under the salt

conditions used by Thammana and Held, addition of the so-called inhibitory proteins

resulted in formation of the active form of 30S, which is not a competent substrate for

KsgA. In the experiments performed by Poldermans *et al.*, on the other hand, the salt

conditions did not allow conversion of 30S to the active form, even upon addition of all of

the component proteins, thus these particles were able to be methylated by KsgA (Figure

7). S21, which was found to be inhibitory by both groups, has a strong influence on the

conformation of the 3' end of the 16S rRNA[54]. The conformational change promoted by

S21 has parallels in the inactive to active conformational change. Therefore, it seems likely

that binding of S21 may promote a conformation in the 3' region of the rRNA which does

not allow methylation.

The question of temporal regulation of KsgA activity may prove to be crucial to

our understanding of small subunit maturation. Of the ten methylated bases in *E. coli* 16S

rRNA, six are clustered in the same immediate area (Figure 8). m^2G1516 is located in the

loop of helix 45 along with the dimethyladenosines at 1518 and 1519; $m^4Cm1402$,

m^5C1407 and m^3U1498 are located at the base of helix 44. RsmE and RsmF catalyze the

methylations at U1498 and C1407 respectively[55, 56]; the methyltransferase responsible for

methylation of G1516 has not yet been identified. C1402 is methylated both on the

exocyclic amine of its base and on the 2'-O of the ribose, presumably by two different
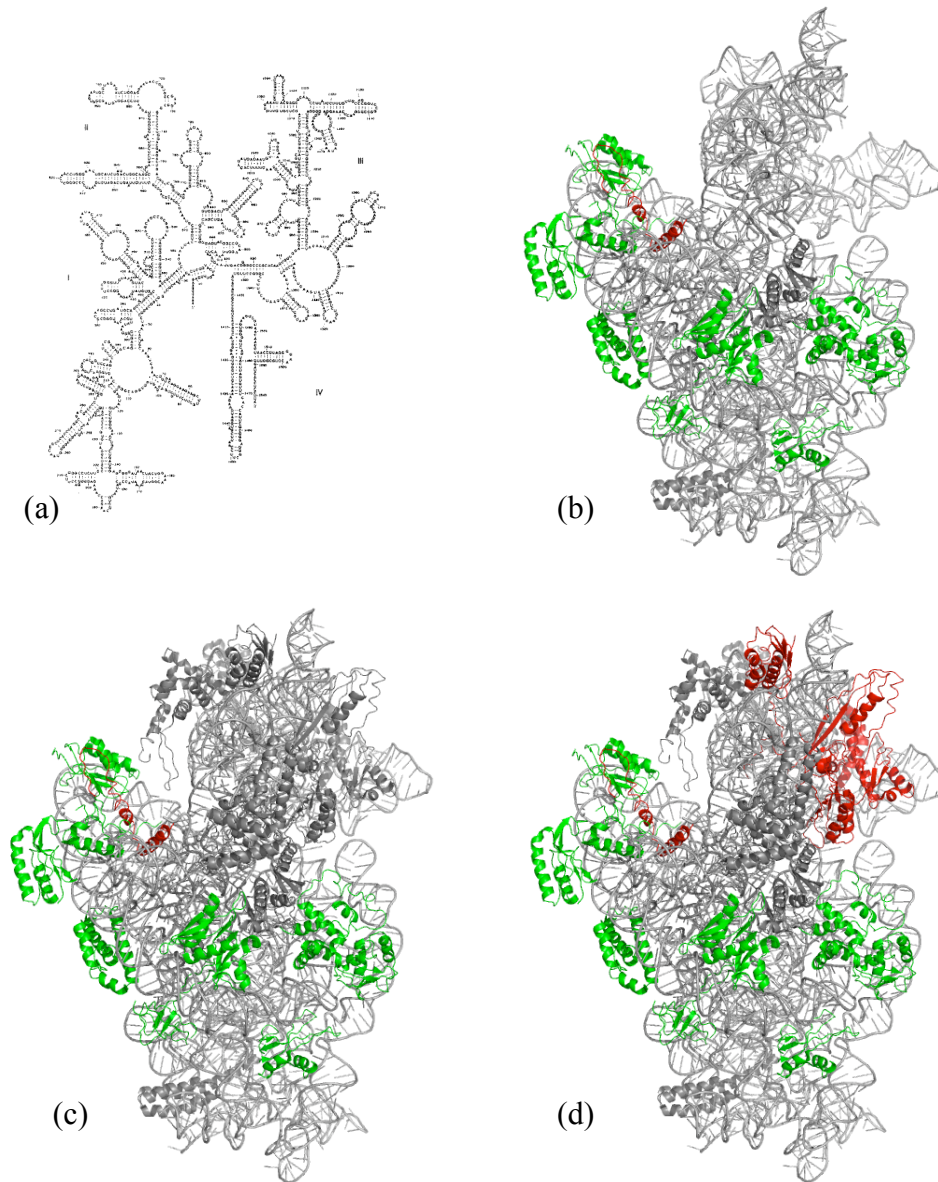
Figure 7. Substrate competence. Naked 16S rRNA (a) is not a substrate. Addition of the eight required proteins (green) allows assembly of the body and platform (b); this particle is a substrate. Addition of the remaining proteins allows head formation (c); this particle is a substrate under conditions which promote the inactive conformation. Under conditions that promote the translationally active conformation, addition of the five inhibitory proteins (red) and subsequent head formation results in a particle (d) which is not a substrate. S21, shown in red in panels b-d, is inhibitory under all conditions.
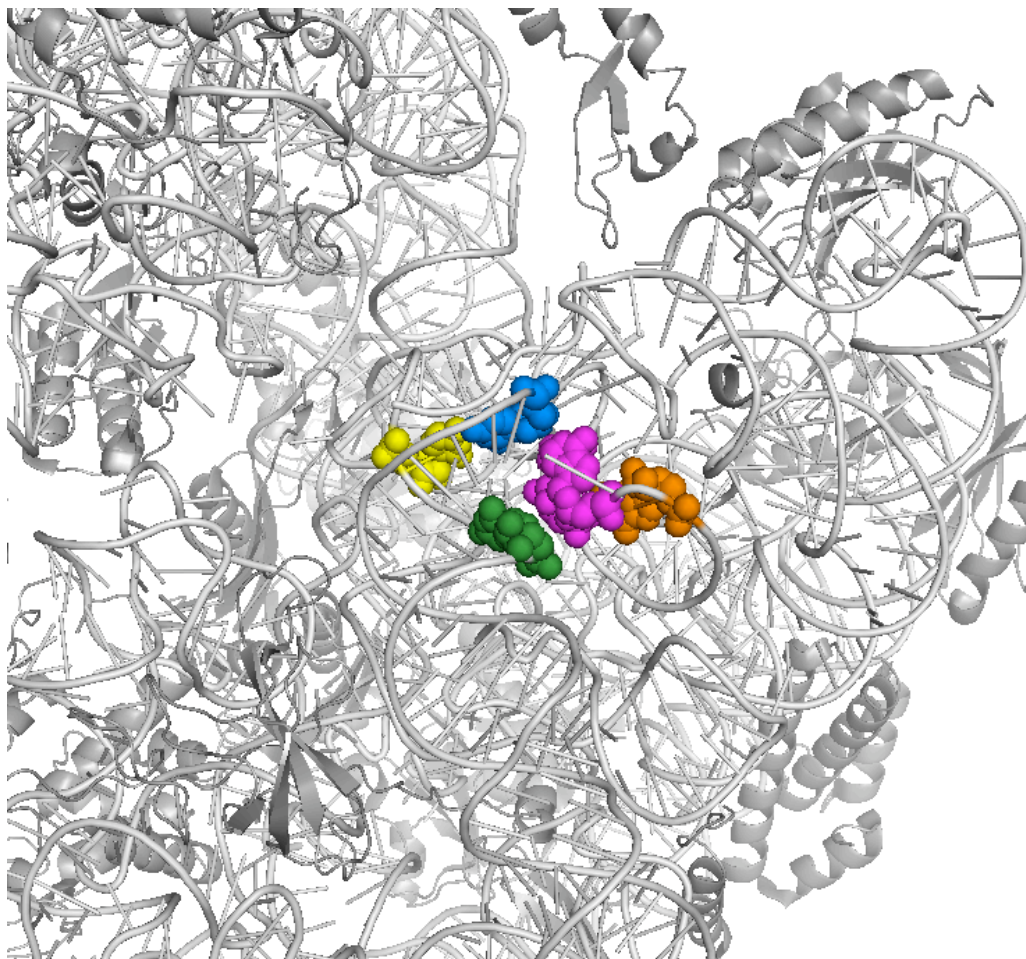
Figure 8. Methylated nucleotides in the area of A1518 and A1519 (magenta). G1516 is shown in orange, U1498 is shown in green, C1407 is shown in blue, and C1402 is shown in yellow.

methyltransferases; these enzymes are also as yet unidentified. RsmE and RsmF both methylate 30S subunits but not naked 16S. Neither of these enzymes has specifically been tested for activity against active vs. inactive subunits, but assays for both were performed under conditions that would favor the active conformation. So KsgA does not methylate until the body and platform are mostly formed, but it must methylate before the head forms and the subunit becomes active. During this window of time between KsgA binding and involvement of the 30S subunits in translation, at least two, and possibly five other methyltransferases bind to their respective target sites, which must overlap the KsgA binding site to some extent. These processing events must require exquisite timing to ensure that all necessary modifications are carried out without the various enzymes interfering with one another.

ksgA knockout and kasugamycin resistance

Although knockout of KsgA is tolerated in bacteria[57], there are consequences to loss of the modifications. Poldermans *et al.* observed that, in the absence of IF-1, more IF-3 is required for binding of fMet-tRNA to unmethylated 30S subunits than to methylated subunits[58]. This difference could be eliminated by methylating the subunits, showing that the methyl groups were solely responsible for the difference in factor requirements in the two types of subunits. An explanation for this requirement may be found in structural studies that described the effect of the four methyl groups on the local structure of the rRNA. As shown in Figure 5 the modified adenosines lie in a hairpin loop. Unsurprisingly, in an oligonucleotide containing the hairpin, the methylation state of the adenosines has a strong impact on formation of the loop[59]. A variety of evidence points to destabilization of

helix 45 by dimethylation of the two adenosines[60], both by increased stacking of the

modified adenosines[61] and by reduction of the hydrogen bonding potential within the

loop[62]. Wickstrom et al. showed that IF-3 binding to 30S disturbed the secondary structure

of the 3'-OH end of the 16S rRNA; the authors hypothesized that increased stability of the

unmethylated relative to the methylated loop made it more difficult for IF-3 to disrupt the

local structure, and therefore more IF-3 would be needed to promote translation

initiation[63].

Igarishi *et al*. described a kasugamycin resistant strain which grew slower than the

wild type strain, but only in certain culture media[64]. *In vitro*, ribosomes from the resistant

strain also showed slower polypeptide synthesis than wild-type ribosomes. However, this

translational impairment was not directly a result of the undermethylation of 16S.

Ribosomes from the resistant strain contained lower amounts of ribosomal protein S1 than

those from the wild-type strain. Since S1 is essential for normal translation in *E. coli*[65], this

was likely the reason for slower polypeptide synthesis. In this case, loss of the

modification affects assembly of the ribosomal proteins onto the 30S subunit, and thereby

indirectly affects translation. This effect of methylation status of the 30S on S1 content

may help explain the ksg[R] phenotype in cells lacking functional KsgA. Ribosomes lacking

protein S1 cannot translate mRNA containing leader sequences, but are able to translate

leaderless mRNAs[66]. Kasugamycin acts in part by blocking translation initiation. It has

been shown that this inhibition is not as pronounced on mRNAs lacking a leader sequence,

allowing translation of these mRNAs[67]. If ribosomes in resistant strains contain less S1

than wild-type ribosomes, one hypothesis is that lower amounts of S1 bound to 30S in

resistant cells could bias these ribosomes toward translation of leaderless mRNAs; therefore, these cells would be less sensitive to ksg inhibition of initiation. Also, Poldermans *et al*. showed that ksg is able to displace fMet-tRNA from 30S subunits, regardless of methylation state, but not from 70S ribosomes[68]. Leaderless mRNAs may bind directly to 70S ribosomes, as opposed to mRNAs with leader sequences, which bind to 30S[69]. This would further protect the mutant cells from the effects of kasugamycin on translation.

van Buul *et al.* described a *ksgA* mutant strain which displayed ribosomal ambiguity in *in vitro* translation assays[70], providing further evidence of the importance of the methyl groups for proper ribosome function. Ribosomes purified from the mutant strain allowed leakiness of nonsense and frameshift mutations in *in vitro* translation assays. This group did not examine the protein composition of 30S purified from wild-type and *ksgA* mutant strains. However, given the proximity of the methylated adenosines to the functional center of the ribosome, it seems likely that the presence or absence of the methyl groups could have a direct effect on translational fidelity. The same study showed that kasugamycin causes translation to become hyper-stringent, and the two effects work against each other; ribosomes from the ksg[R] strain translated normally in the presence of kasugamycin. These results offer further explanation of how loss of the methyl groups leads to ksg sensitivity.

Another indication that the methyl groups are important, if not essential, comes from a study that found a relationship between S20 and KsgA activity[71]. In an S20 knockout strain a portion of the 16S rRNA was found to be submethylated at

A1518/A1519. 30S subunits containing 16S that was unmethylated at 1518/1519 were not able to join with 50S to form 70S. Since functional 30S can be assembled with T7-transcribed 16S[72], which is completely unmodified, the phenotype of the S20 knockout strain must be caused by loss of modifications combined with absence of S20.

The *apaH* gene lies downstream of *ksgA,* in the same operon[73]. ApaH is responsible for the hydrolysis of Ap_4N nucleotides such as diadenosine tetraphosphate[74]; knockout of *apaH* results in a sharp increase in the levels of Ap_4A[57, 75]. Curiously, knockout of both *ksgA* and *apaH* simultaneously results in restoration of the kasugamycin sensitive phenotype[57]. The basis for this is not known, although the authors hypothesize that the resulting Ap_4A binds to the 3' region of the 30S and causes it to adopt a conformation more similar to the methylated state. An even more astonishing finding in the same study is that overexpression of the chaperone DnaK in the *ksgA⁻/apaH⁻* background renders the bacteria once again resistant to ksg. This may be explained by the finding that Ap_4A binds to DnaK[76]; in this case overexpression of DnaK may act to sequester the increased amounts of Ap_4A produced by *apaH* knockout, allowing the *ksgA⁻* resistant phenotype to dominate. This also suggests an alternate mechanism for the ksg sensitive phenotype in the *ksgA⁻/apaH⁻* cells. DnaK is known to be important in ribosome synthesis; it is possible that elevated levels of Ap_4A could interfere with DnaK in this process, resulting in ribosomes that are sensitive to kasugamycin despite lacking the methyl groups in helix 45.

KsgA in evolution

Since the first description of KsgA in *E. coli*, orthologs have been characterized in a variety of organisms, including KsgA in *Bacillus stearothermophilus*[77], Dim1 in

Saccharomyces cerevisiae[78] and *Kluyveromyces lactis*[79], Pfc1 in the chloroplasts of

Arabidopsis thaliana[80], h-mtTFB in human mitochondria[81], and MjDim1 in

Methanocaldococcus jannaschii[82]. During evolution the KsgA proteins have been recruited

to play other roles in the cell. Dim1 is essential for proper processing of pre-rRNA at

cleavage sites A_1 and A_2[83] (Figure 9), and knockout of Dim1 in yeast is lethal[78]. Depletion

of Dim1 in *S. cerevisiae* leads to buildup of an aberrant 22S rRNA species as a result of

misprocessing; 5.8S and 25S rRNA can still be processed normally. Pfc1 is important for

chloroplast development under chilling conditions[80]; accumulation of misprocessed rRNA

in chloroplasts upon Pfc1 depletion indicates that Pfc1 may play a role in chloroplast

rRNA processing similar to Dim1 in yeast[84]. The methyltransferase function of Dim1 is

not required for its RNA processing activity; Lafontaine *et al*. described a *dim1* mutant

which contained properly processed, but unmethylated, rRNA[85]. Although this strain grew

normally compared to wild-type, indicating that the unmethylated ribosomes were not

noticeably impaired *in vivo*, cell extract from the mutant strain did not support translation

in vitro. It is possible that the methylations are required for ribosome function in the less

favorable *in vitro* conditions; alternately, lack of the modifications could affect ribosomal

protein assembly in a manner similar to that seen in *ksgA* mutation in bacteria.

mtTFB, which is encoded in the nucleus, serves as a transcription factor in

mitochondria[86]. h-mtTFB from human mitochondria is able to methylate bacterial 30S *in*

vitro, and is presumed to be the methyltransferase for the mitochondrial small subunit

rRNA, although this RNA seems to be submethylated relative to bacteria [86]. Interestingly,

in some mitochondria, there are two separate mtTFB proteins, mtTFB1 and mtTFB2,
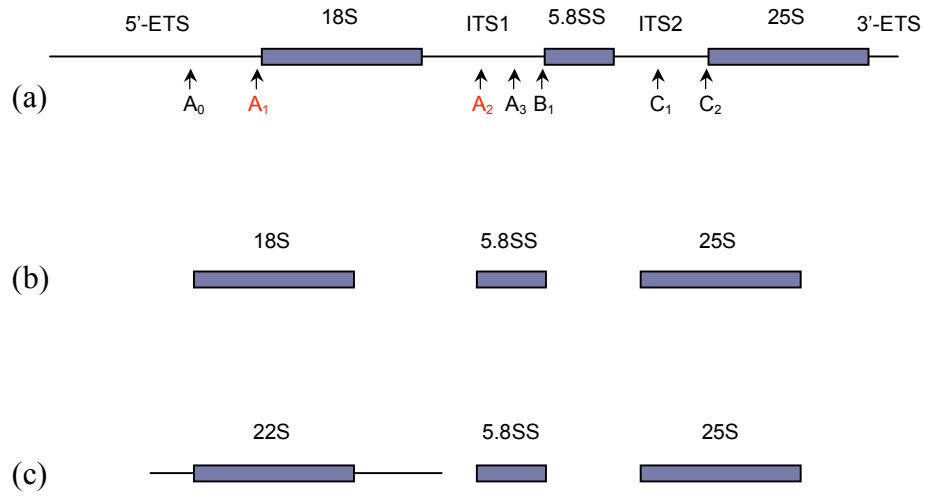
Fgure 9. rRNA processing in yeast. (a) The 35S pre-rRNA transcript. Cleavage sites are noted; the cleavage sites which require Dim1 activity are shown in red. (b) Properly processed rRNA. (c) Improper processing resulting from depletion of Dim1.

which are proposed to have arisen from a gene duplication event[87, 88]. There is evidence that mtTFB1 has retained stronger methyltransferase activity, while mtTFB2 is more active as a transcription factor[88-91]. The fungi have only a single mtTFB, suggesting loss of one of the paralogs in this lineage. The *S. cerevisiae* ortholog, sc-mtTFB, serves as a transcription factor but does not have methyltransferase activity. sc-mtTFB lacks significant sequence homology to any of the KsgA/Dim1 enzymes; yeast mtTFBs are generally poorly conserved and difficult to identify via sequence homology[92]. It is worth noting that the mtTFB2 variants, which have evolved to be stronger transcription factors at the expense of the methyltransferase activity, are more closely related to fungal mtTFBs than to mtTFB1s[88, 93].

The Erm family of methyltransferases mediates resistance to the macrolide-lincosamide-streptogramin B (MLS-B) group of antibiotics[94]. These enzymes share high sequence homology with the KsgA family[95] and are likely descended from one or more bacterial *ksgA* genes. The Erms mono- or dimethylate a single adenosine of 23S rRNA, near the peptidyl transferase site of the 50S subunit[96]. Unlike KsgA, which requires at least a partially formed 30S subunit for enzymatic activity, ErmC' from *Bacillus subtilis* can methylate a fragment of 23S as small as 32 nucleotides[97].

The KsgA family belongs to a well-characterized group of S-adenosyl-L-methionine (SAM) dependent methyltransferases which includes RNA methyltransferases, DNA methyltransferases, protein methyltransferases and small molecule methyltransferases[98]. All of the enzymes in this group contain a Rossman-like fold, which consists of a central β-sheet surrounded by a variable number of α-helices, and they share

several conserved motifs[99]. Many of these Rossman-like methyltransferases have been well characterized structurally and biochemically, and their mechanisms have been explored. In the case of nucleic acid methyltransferases, the target base is often involved in secondary structure and must be "flipped out" of a stable conformation in order to be methylated. This process has been extensively studied in DNA methyltransferases, whose targets reside within the double helix. KsgA is unique among this class of enzymes in that it must catalyze the transfer of four methyl groups onto two adjacent bases.

Unanswered questions

In the three decades since its discovery a picture of KsgA's role in the cell has slowly emerged. However, a plethora of questions still remain and more recent studies on KsgA's orthologs have opened up intriguing lines of inquiry. There is a glaring lack of information about the mechanism catalyzed by KsgA, which is no surprise given the convoluted reaction involved. Methyl groups from four separate SAM molecules must be transferred to two adenosines in the 16S rRNA. Are all of the transfers accomplished in a single binding event, or does the enzyme:substrate complex dissociate and reassociate between transfers? In the case of ErmC, there is strong evidence that dimethylation of the target adenosine occurs in two separate events; the base is monomethylated, the enzyme dissociates, and the monomethylated RNA serves as a substrate for the second methylation[100]. In a related question, is there a preferred order of methyl transfer, or are the methyl groups transferred randomly to one or the other adenosine? It has been reported that, under stringent conditions (low SAM concentration and low temperature), KsgA preferentially dimethylates the 3' proximal adenosine, suggesting some preference for

ordered methylation[101]. However, if either adenosine is mutated the other is still able to be methylated, indicating that any such preference is not obligatory[102].

Next, the exact nature of KsgA's intricate substrate requirements is not fully understood. KsgA can bind to 16S rRNA, but does not methylate it until certain proteins have assembled onto the RNA. One possible explanation is that KsgA is allosterically regulated. Free KsgA could exist in a conformation that is competent for RNA binding but not methylation; at some point during subunit assembly a conformational change could be triggered which allows KsgA to catalyze the methyl transfers. This hypothesis is supported by the observation that free KsgA does not bind SAM appreciably[47], suggesting that binding of the RNA and SAM substrates may occur in an ordered manner. A similar ordered binding mechanism is seen in the DNA methyltransferase HhaI, which does not bind SAM unless the enzyme is first bound to its target DNA[103]. Another possibility is that a conformational change in the substrate is required. KsgA can bind to naked 16S, but in this structural context the target adenosines may not be available to the enzyme. Some assembly of the subunit could be required to engender a substrate conformation which supports methylation. It should be noted that these two scenarios are not mutually exclusive.

Another intriguing aspect of the KsgA/Dim1 family is their diversity of function. KsgA has been recruited to play an essential role in pre-rRNA processing in eukaryotes, and serves as a transcription factor in mitochondria. The structural basis for these "new" functions is unknown. It is feasible that the basic structure of KsgA is sufficient to allow function as a processing factor or transcription factor, in which case a bacterial KsgA could

complement for Dim1 or mtTFB knockout. However, it is also possible that structural and/or sequence evolution of KsgA was required before KsgA could acquire functions other than rRNA methylation. Extensive sequence alignments of KsgA, Dim1 and mtTFB enzymes identify elements unique to each group, suggesting possibilities for domains that could be involved in these other functions. Notably, the eukaryotic Dim1 enzymes have a 25-30 residue insert in the C-terminal domain relative to bacterial KsgAs. It is conceivable that this insert could mediate RNA or protein interactions important for Dim1's processing function. The mtTFB enzymes have three inserts that are not found in KsgA or Dim1, which could be related to their transcription activity. These inserts will be discussed in more detail in Chapter 3.

Given the range of functions displayed by KsgA orthologs, it is tempting to speculate that KsgA may play a larger role in ribosome biogenesis, beyond methylation. This possibility was hinted at by the discovery of a functional link between KsgA and Era. The Inouye group isolated a cold-sensitive Era mutation which resulted in defects in rRNA processing and also in cell division[13, 104, 105]. The cold-sensitive phenotype was complemented by overexpression of KsgA[105]. Although the effects of KsgA overexpression on rRNA processing in the Era cold-sensitive mutant background were not specifically addressed, it is certainly feasible that KsgA has a role in ribosome assembly similar to that of Era, RbfA, RimM, RimN, and RsgA. Although KsgA is not essential for viability under normal conditions, it could certainly play a more pronounced role in ribosome assembly under non-optimal conditions such as cold shock.

Scope and objectives

KsgA and its orthologs play an integral role in the biosynthesis of functional ribosomes. The studies presented here were undertaken in order to better understand the structure and function of these proteins, in order to lay a foundation for future investigations. In a first step toward understanding this fascinating family of proteins, we have performed a more in-depth characterization of the functional conservation of the KsgA/Dim1 enzymes. We have also solved the crystal structure of KsgA from *E. coli*, the first structure of a KsgA/Dim1 enzyme. This structure will serve as a point of reference for other work on the KsgA/Dim1 family. The structures of sc-mtTFB[106] and ErmC'[107] have been solved by X-ray crystallography, and a second Erm enzyme, ErmAM from *Streptococcus pneumoniae*, has been solved by NMR[108]. Since this work was completed, the crystal structure of the human Dim1 enzyme (hDim1) has been solved[109]. Comparison of these structures yields information about their common methyltransferase function and also hints at such issues as substrate recognition, mechanism, and evolution of function. Finally, using biochemical experiments in concert with structural modeling, we have characterized the interaction of KsgA with the 30S subunit. This model may help explain the substrate requirements for KsgA's methyltransferase activity.

Functional Conservation

All ribosomes share the same general architecture, with two subunits made up of rRNA and ribosomal proteins. Mears *et al.* modeled a minimal ribosome structure based on conservation of both RNA and protein elements in the three kingdoms and in organellar ribosomes[110]. This model shows high conservation in the functional core of the ribosome, indicating that critical regions of the ribosome have changed very little throughout evolution. Indeed, hybrid ribosome can be engineered combining RNA and protein elements from different kingdoms. Such combinations include ribosomes with small and large subunits coming from different sources[111-115], ribosomes containing a hybrid rRNA[116], and ribosomes from one organism containing selected proteins from another organism[117-121]. A surprising number of these combinations retain some functionality. Clearly, there is a high degree of structural and functional conservation in the ribosome, which is not really unexpected given its central role in the workings of all cells. Despite this conservation, however, there is an astonishing degree of divergence in ribosome assembly pathways from the different kingdoms.

As cells grew more complex with the transition from prokaryotic to eukaryotic organisms, the ribosome assembly process became correspondingly more elaborate. Eukaryotic rRNA is transcribed in the nucleolus; ribosomal proteins and other factors assemble onto the nascent rRNA co-transcriptionally[23, 24]. Early processing and

29

modification steps occur on the pre-40S and pre-60S particles in the nucleolus, then the particles travel into the nucleoplasm, where the pre-60S is further processed, and finally the particles are exported into the cytoplasm, where final maturation steps take place. This compartmentalization was accompanied by a large increase in the amount of extra-ribosomal factors required to produce functional ribosomes. At least 170 factors are involved in eukaryotic ribosome production, many of which are indispensable *in vivo* [35]; this number continues to grow and the final count could be well over 200 proteins and RNPs. This is in contrast to bacteria, in which only a few extra-ribosomal assembly factors have been identified [12, 122]. While some of these factors are essential for ribosome formation under non-optimal conditions such as cold-shock[13, 16], none of them are essential under normal conditions. In fact, prokaryotic ribosomes can be assembled *in vitro* from rRNA and ribosomal proteins, without any accessory factors present[1-3].

In addition to compartmentalization and increased involvement of assembly factors, billions of years of evolution have resulted in a shift in rRNA modification patterns, as well as the way in which nucleotide modification is carried out. In bacteria, base methylation is the predominant modification, with few pseudouridines and ribose methylations. Each base methylation and pseudouridylation is carried out by a distinct enzyme, although in the case of pseudouridylation, the same enzyme may have up to three specific target residues. In contrast, eukaryotes contain mostly pseudouridines and ribose methylations, in roughly equal amounts; eukaryotic rRNA contains few methylated bases. Eukaryotes, as well as archaea, use a snoRNA guide system to carry out most modifications[36, 123]. A single enzyme, Nop1p/fibrillarin, carries out all of the ribosme

methylations; box C/D snoRNAs target the snoRNP complex to specific ribose moieties. Similarly, Cbf5p/dyskerin catalyzes all of the pseudouridylations, guided by box H/ACA snoRNAs. Little is known about the mechanism of base methylation in eukaryotes, and ribose methylation in bacteria is likewise poorly understood.

In comparing prokaryotic and eukaryotic ribosome biogenesis, the conservation of the KsgA/Dim1 family is unique. The presence and function of this enzyme has been maintained in every evolutionary lineage, including eukaryotic organelles. Exceptions are found in only a few disparate organisms[101, 124-129]. This high degree of conservation begs the question of why the protein has been kept around. Although KsgA is dispensable in bacteria[57], it must perform a function important enough to justify its preservation throughout evolution. Dim1 and h-mtTFB have been shown to complement for KsgA function in bacteria[78, 86]. This indicates that the bacterial, eukaryotic, and organellar enzymes can recognize a common bacterial substrate, and suggests that structural cues that mediate this recognition have been conserved throughout ribosomal evolution. Given the high degree of rRNA conservation in the region of the methylated adenosines, this may not seem so remarkable. However, 16S rRNA structure alone is not enough to allow KsgA activity; the ribosomal proteins S4, S6, S8, S11 and S15-S18 must be assembled onto the RNA in order to form a competent substrate conformation. Since these proteins are not in position to directly contact helix 45 or KsgA, their requirement for KsgA activity must depend on the rRNA adopting a particular conformation upon binding of the proteins, which are all located in the body and platform of the 30S subunit. Although these eight proteins are sufficient and necessary to form a competent substrate structure, three out of

the eight, S6, S16 and S18 are not conserved outside of bacteria[110]. That means that the local 16S conformation which allows methylation must be achieved with a different subset of proteins in other organisms. Given this, it is more striking that KsgA orthologs from yeast and from mitochondria can recognize their substrate in bacterial ribosomes *in vivo*. We have extended previous complementation studies to include an ortholog from an archaeal organism, *M. jannaschii*, and have also examined *in vitro* activity and mono- and dimethyladenosine profiles produced by the eukaryotic and archaeal orthologs.

Cloning, expression and purification of proteins

The KsgA gene was amplified by PCR of genomic DNA from *E. coli* XL1-Blue cells, using primers designed from the coding sequence (accession number M11054), and cloned into the pET15b vector, in-frame with the vector-encoded N-terminal poly-histidine tag. Cloning was confirmed by sequencing. The plasmid was transformed into BL-21 (DE3) cells for protein production.

To identify the archaeal ortholog of the KsgA/Dim1 family we used NCBI's genomic BLAST tool to search the *M. jannaschii* sequence, using the *E. coli* KsgA protein sequence (accession number P06992) as the query sequence. This search identified a putative ortholog, accession number NP_248023, which was annotated as "dimethyladenosine transferase (ksgA)" in the Entrez protein record. Alignment of this protein with other KsgA/Dim1 family members revealed a high degree of similarity (Figure 10), including the presence of motifs common to SAM-dependent methyltransferases. We followed the eukaryotic nomenclature and designated the protein as MjDim1 (for <u>M</u>. *jannaschii* <u>dim</u>ethyltransferase <u>1</u>). We obtained *M. jannaschii* genomic

```
KsgA      MNNRVHQGHLARKR-------------------FGQNFLNDQFVIDSIV
HsDim1    -MPKVKSGAIGRRRGRQEQRRELKSAGGLMFNTGIGQHILKNPLIINSII
ScDim1    -MGKAAKKKYSGATSSKQVSAEKHLSSVFKFNTDLGQHILKNPLVAQGIV
MjDim1    -------------------------MFKPKKKLGQCFLIDKNFVNKAV
                                              :** :* :   . : :
```

```
                       I                        II
                 <----------->              <----->
KsgA      SAINPQKGQAMVEIGPGLAALTEPVGERLDQLTVIELDRDLAARLQTHPF
HsDim1    DKAALRPTDVVLEVGPGTGNMTVKLLEKAKKVVACELDPRLVAELHKRVQ
ScDim1    DKAQIRPSDVVLEVGPGTGNLTVRILEQAKNVVAVEMDPRMAAELTKRVR
MjDim1    ESANLTKDDVVLEIGLGKGILTEELAKNAKKVYVIEIDKSLEPYANKLKE
            .         :.::*:* * . :*   : :. .:: . *:*  :  .   .
```

```
                     III                        IV           V
                 <----------->              <----->    <------------
KsgA      ---LGPKLTIYQQDAMTFNFGELAEKMGQPLRVFGNLPYNISTPLMFHLF
HsDim1    GTPVASKLQVLVGDVLKTDLPFF-------DTCVANLPYQISSPFVFKLL
ScDim1    GTPVEKKLEIMLGDFMKTELPYF-------DICISNTPYQISSPLVFKLI
MjDim1    ---LYNNIEIIWGDALKVDLNKLDF-----NKVVANLPYQISSPITFKLI
            :   :: :   * :. ::  :        ..* **:**:*: *:*:
```

```
          -->             VI                            VII
                       <-->                       <----->     <-
KsgA      SYTDAIADMHFMLQKEVVNRLVAGPNSKAYGRLSVMAQYYCNVIPVLEVP
HsDim1    LHRPFFRCAILMFQREFALRLVAKPGDKLYCRLSINTQLLARVDHLMKVG
ScDim1    NQPRPPRVSILMFQREFALRLLARPGDSLYCRLSANVQMWANVTHIMKVG
MjDim1    K--RGFDLAVLMYQYEFAKRMVAKEGTKDYGRLSVAVQSRADVEIVAKVP
                   :* * *.. *::*  . . * ***  .*  . *   : :*
```

```
          VIII
          -------->
KsgA      PSAFTPPPKVDSAVVRLVPHATMPHPVK-DVRVLSRITTEAFNQRRKTIR
HsDim1    KNNFRPPPKVESSVVRIEPKNPPP--PI-NFQEWDGLVRITFVRKNKTLS
ScDim1    KNNFRPPPQVESSVVRLEIKNPRP--QV-DYNEWDGLLRIVFVRKNRTIS
MjDim1    PSAFYPKPKVYSAIVKIKPNKGKY--HIENENFFDDFLRAIFQHRNKSVR
           . * *  *:*  *::*::  :       :   . . :    *  ::.:::
```

```
KsgA      NSLGNLF---------SVEVLTG--------------------------
HsDim1    AAFKSSA---------VQQLLEKNYRIHCSVHNIIIPEDFS-----IADK
ScDim1    AGFKSTT---------VMDILEKNYKTFLAMNNEMVDDTKGSMHDVVKEK
MjDim1    KALIDSSKELNYNKDEMKKILEDFLNTNSEIKNLINEKVFK-----LSVK
           .:  .                 .:*
```

```
KsgA      ------MGIDPAMRAENISVAQYCQMANYLAENA-PLQES
HsDim1    IQQILTSTGFSDKRARSMDIDDFIRLLHGFNAEGIHFS--
ScDim1    IDTVLKETDLGDKRAGKCDQNDFLRLLYAFHQVGIHFS--
MjDim1    DIVNLSN-------------EFYRFLQNRGRL-------
                          :: ::
```

Figure 10. Structure-based sequence alignment of KsgA (P06992), putative Dim from *Homo sapiens* (HsDim1, Q9UNQ2), ScDim1 (P41819), and MjDim1 (NP_248023). The alignment was generated using the TCoffee web server[130]. Identical residues are denoted with a star, and strongly conserved residues with a colon; weakly conserved residues are marked with a period. Double-headed arrows indicate motifs common to SAM-dependent methyltransferases. Structures used for alignment were those of KsgA[131] (PDB ID 1QYR) and HsDim1[109] (PDB ID 1ZQ9).

DNA from ATCC and amplified the MjDim1 gene by PCR, using primers designed from the coding sequence (accession number NC_000909). The gene was cloned into the pET15b vector, in-frame with the vector-encoded N-terminal poly-histidine tag, and was confirmed by sequencing. The plasmid was transformed into BL-21 (DE3) cells for protein production.

The identification and cloning of ScDim1 was described by Lafontaine *et al*. (1994), and ScDim1 was provided to us by the same group as a pET15b construct engineered to contain an N-terminal poly-histidine tag. Analysis of the ScDim1 gene sequence revealed the use of codons that are rare in *E. coli*; therefore, the plasmid was transformed into BL21-CodonPlus (DE3)-RIL cells, which contain extra copies of rare codons. Protein purification was made difficult by insolubility of the overexpressed protein; expression conditions were optimized to produce the majority of protein in a soluble form.

All three proteins were overexpressed with 1mM IPTG and purified by nickel chelate chromatography. SDS-PAGE analysis estimated proteins to be >95% pure.

In vivo analysis

ScDim1 has been shown to complement for KsgA function in *ksgA⁻ E. coli* cells[78], demonstrating functional equivalence of the two proteins. We asked whether MjDim1 could also complement KsgA knockout. *In vivo* activities of both ScDim1 and MjDim1 were assessed using a modified minimal inhibitory concentration (MIC) assay, which takes advantage of the fact that loss of KsgA function renders bacteria resistant to kasugamycin[44, 77]. Plasmids containing the two proteins were transformed into a ksg[R] strain

of *E. coli*, which lacks endogenous KsgA activity. This strain was constructed from BL-21 (DE3) cells; this allows leaky expression from the pET15b T7 promoter. Growth on ksg was compared to cells transformed with pET15b-KsgA plasmid (positive control) and cells transformed with empty vector (negative control). Unlike in a traditional MIC, untransformed cells are naturally resistant to the antibiotic and become sensitive when transformed with a functional dimethyltransferase. Therefore, cells transformed with KsgA have a low MIC of 400 μg/ml ksg, while cells transformed with empty vector have a high MIC, greater than 3000 μg/ml ksg. As shown in Figure 11, MjDim1 is fully functional in this *in vivo* system, with an MIC of 400 μg/ml. ScDim1, on the other hand, shows partial activity on bacterial ribosomes *in vivo*, with an MIC of 1200 μg/ml. While ScDim1 doesn't restore full sensitivity to the antibiotic, it does show increased sensitivity, indicating that the enzyme is able to recognize the small subunit as a substrate. Lack of full complementation may correlate with slower and/or incomplete methylation of 30S as compared to the other two enzymes (see nucleotide analysis).

In vitro analysis

We next asked how efficiently ScDim1 and MjDim1 were able to methylate *E. coli* 30S in an *in vitro* assay. Unmethylated 30S subunits were prepared from the ksg^R strain described above. Incorporation of ^{3}H-methyl from labeled SAM by each enzyme was followed at discrete time points over an interval of two hours. Control experiments were performed with 30S subunits purified from wild-type *E. coli* cells, which are methylated by endogenous KsgA and thus do not serve as substrates. Initial experiments were performed with 10 pmol of 30S substrate and 1 pmol of enzyme. This amount of protein did not allow
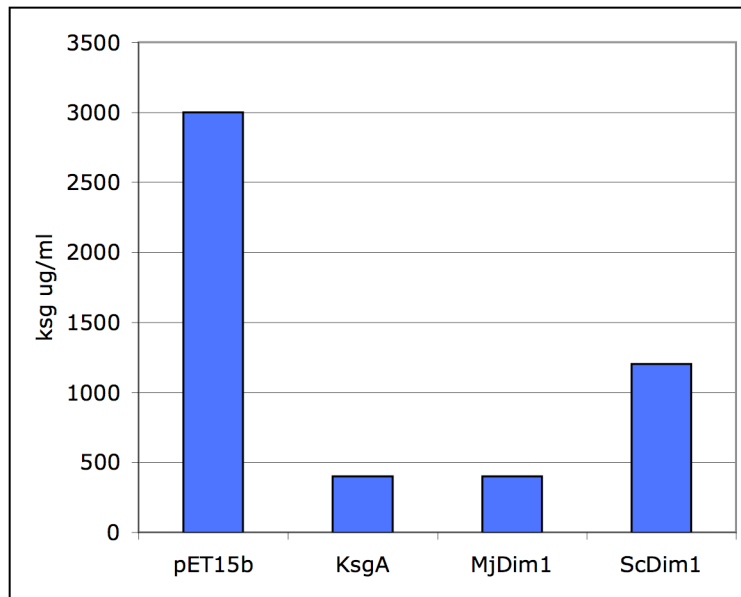
Figure 11. *In vivo* activity of KsgA orthologs.

for completion of the reaction within two hours, so experiments were also performed using 10 pmol each of 30S and enzyme.

Figures 12a, 12b, and 12c show the time-course of methylation for KsgA, ScDim1, and MjDim1 respectively. Methylation of *E. coli* 30S by ScDim1 and MjDim1 closely followed the KsgA time-course, both in rate of incorporation and final level of methylation. In reactions with 1 pmol of protein, MjDim1 showed a slightly higher rate of ^{3}H incorporation than KsgA and ScDim1 at later time points. With stoichiometric amounts of protein relative to 30S, the time-course of methylation was essentially indistinguishable between the three proteins, confirming the ability of the enzymes from archaea and eukaryotes to recognize bacterial 30S subunits as substrates.

We then estimated the amount of methyl groups transferred at the two-hour time point by each enzyme, with both 1 pmol and 10 pmol amounts of enzyme, by constructing a standard curve of cpm vs. concentration of ^{3}H-methyl-SAM. With 10 pmol of 30S per reaction, and 4 methylation sites per 30S molecule, we would expect to see transfer of 40 pmol methyl groups if the reactions have gone to completion. As shown in Figure 13, our calculations lead to slight overestimation of methyl group transfer, probably due to error in ^{3}H counting. Nevertheless, the reactions performed with 10 pmol enzyme appear to be more or less complete after 2 hours. Reactions performed with only 1 pmol enzyme are approximately halfway completed after 2 hours.

Nucleoside analysis

In vitro assays of the three proteins, performed as described above for the time-course, were incubated for two hours and analyzed to determine relative amounts of m^{6}A
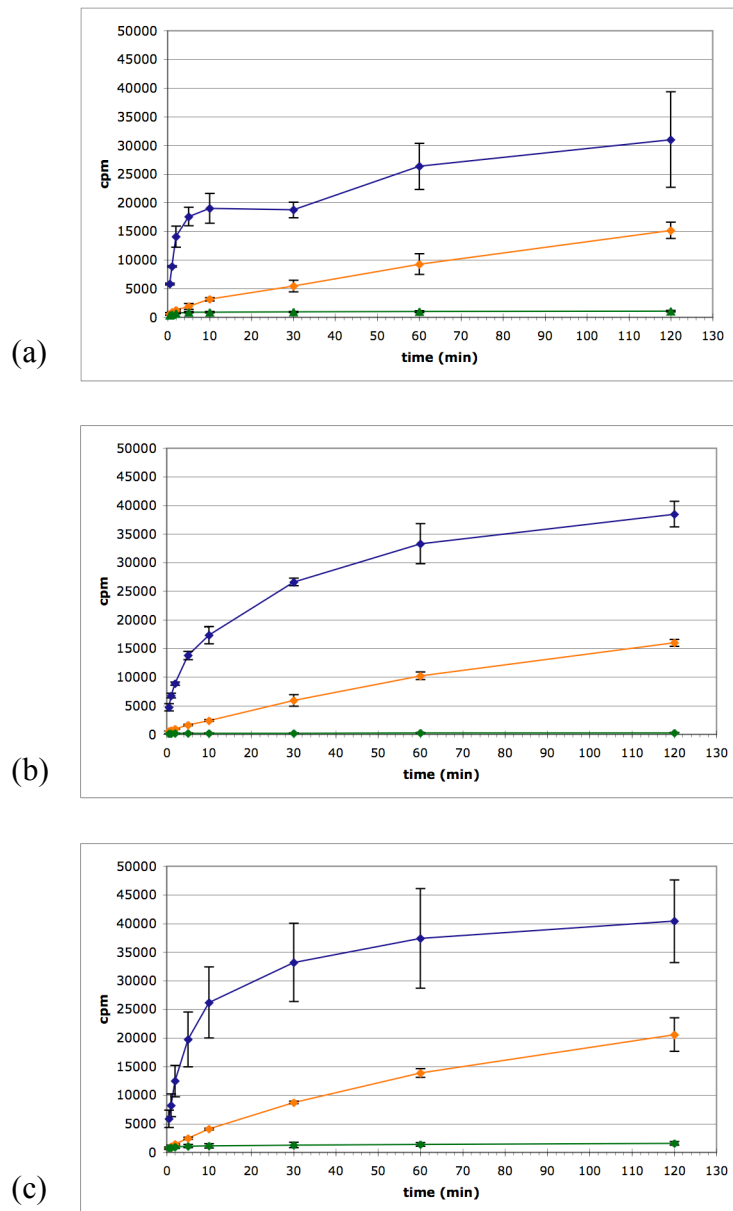
Figure 12. *In vitro* methylation of 30S. Time-course assays for KsgA (a), ScDim1 (b), and MjDim1 (c). Blue lines indicate assays containing 10 pmol ksg^R 30S, 10 pmol enzyme; orange indicates assays containing 10 pmol ksg^R 30S, 1 pmol enzyme; green indicates control assays containing 10 pmol wt 30S subunits, 10 pmol enzyme. Assays were performed in triplicate; error bars represent standard deviation.
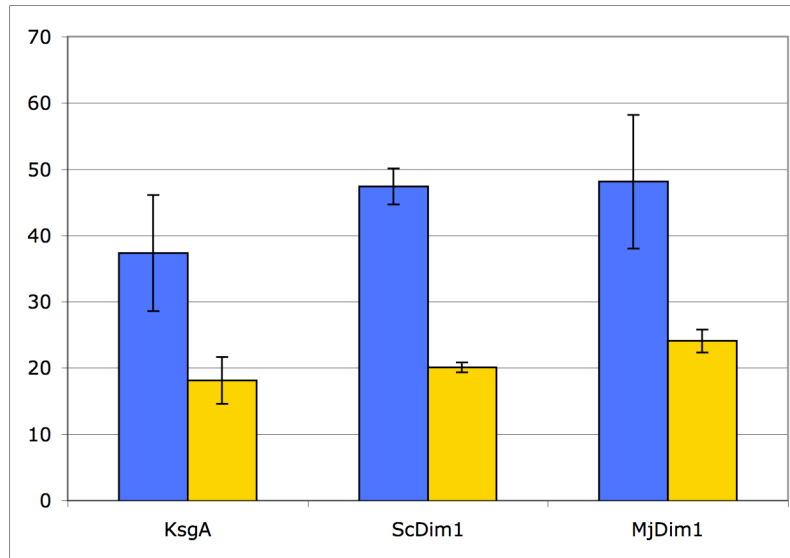
Figure 13. Quantitation of methyl groups transferred after two hours. Blue bars represent assays performed with 10 pmol enzyme; yellow bars represent assays using 1 pmol enzyme. Assays were performed in triplicate; error bars represent standard deviation.

and m^6_2A (Figure 14, Table 2). 16S rRNA isolated from 30S subunits methylated by 10 pmol of either KsgA or MjDim1 contained no detectable labeled m^6A; radioactive incorporation was seen only in the dimethyladenosine peak. This agrees with the *in vitro* data suggesting that these reactions have gone to completion (Figure 13). ScDim1, on the other hand, produced a mixture of m^6A and m^6_2A; approximately 28% of the incorporated radiolabel was found on monomethylated adenosine. This indicates that at most 80% of the potential sites were methylated after two hours.

Partially methylated 30S from reactions using 1 pmol of enzyme showed radioactive peaks at both m^6A and m^6_2A. Surprisingly, although the total level of methylation was similar for all three enzymes (Figure 13), rRNA methylated by the different enzymes showed different ratios of m^6A to m^6_2A. ScDim1 produced approximately 1.4 times as much m^6A as m^6_2A. MjDim1, conversely, produced almost no m^6A; only about 1% of the incorporated methyl groups were found on m^6A. KsgA fell somewhere between the other two, producing both m^6A and m^6_2A, with only 0.8 times as much m^6A as m^6_2A. These results could be a result of assaying enzymes from different species on the bacterial substrate, or they could reflect a difference in reaction mechanism.

The KsgA/Dim1 enzymes transfer a total of four methyl groups from four SAM molecules to two adenosines. The exact mechanism of transfer has not yet been established; questions remain as to order of addition, if any, and the number of binding events required for the four methylations. The above data begin to address the question of the multiple methyl group transfers. Partially methylated 30S were produced in reactions containing 10 pmol 30S subunits and 1 pmol enzyme. Therefore, m^6A produced in excess
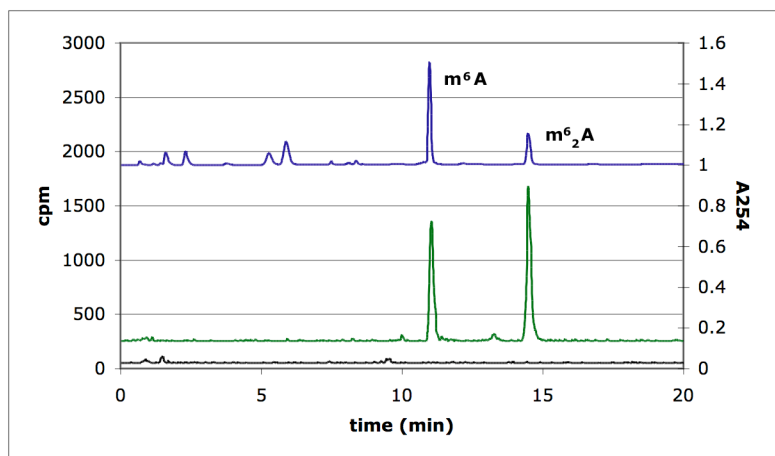
Figure 14. Representative HPLC trace. The bottom trace (shifted upward by 50 cpm) is data from a control assay using wt 30S; the middle trace (shifted upward by 250 cpm) represents an assay using 1 pmol ScDim1 and 10pmol ksg[R] 30S; the top trace (shifted upward by 1 OD unit) represents reference nucleotides, with m^6A and m^6_2A peaks labeled.

Table 2. Quantitation of methylated adenosine species.

		total CH$_3$ (pmol)[a]	ratio m^{6}A:m$^6_2$A	m^{6}A (pmol)	m$^6_2$A (pmol)[b]
KsgA	1 pmol	18.1 ± 1.7	0.8:1	5.2	6.5
	10 pmol	37.4 ± 10.1	-	ND[c]	37.4
ScDim1	1 pmol	20.1 ± 0.7	1.4:1	8.3	5.9
	10 pmol	47.4 ± 2.7	0.8:1	13.5	16.9
MjDim1	1 pmol	24.1 ± 3.6	0.02:1	0.2	11.9
	10 pmol	48.1 ± 8.8	-	ND	48.1

[a]Data from Figure 13.
[b]1 pmol m$^6_2$A corresponds to 2 pmol methyl groups.
[c]ND=none detected.

of 2 pmol (corresponding to two adenosines available for methylation per subunit) will only be seen if the enzyme releases the substrate after monomethylation and rebinds to a new substrate. With 20.1 pmol of methyl groups incorporated by 1 pmol ScDim1, the 1.4:1 $m^6A:m^6_2A$ ratio represents approximately 8.3 pmol labeled m^6A and 5.9 pmol labeled m^6_2A (1 pmol m^6_2A represents 2 pmol incorporated methyl groups). Therefore, under our assay conditions, ScDim1 clearly releases the m^6A intermediate, which is subsequently converted to the m^6_2A product after an additional binding event.

In contrast, there is no indication that MjDim1 produces m^6A as anything but a transient intermediate. Of the 24.1 pmol methyl groups transferred, only 0.3 pmol were found on m^6A. These results suggest that the archaeal enzyme preferentially forms dimethyladenosine, without release of the monomethyl intermediate. While release of a monomethyl intermediate and subsequent re-binding and addition of the second methyl cannot be ruled out, such a model requires that MjDim1 prefer the monomethylated substrate to the unmethylated substrate to a large degree.

In terms of m^6A vs. m^6_2A production, KsgA falls somewhere in between ScDim1 and MjDim1. Unlike ScDim1, KsgA produces less m^6A than m^6_2A; however, of the 18.1 pmol of methyl groups transferred, 5.2 pmol are found on m^6A, which is still indicative of a released intermediate. Although it is possible that these differences are a result of suboptimal assay conditions, these results also allow the possibility of distinct mechanisms for the three enzymes, thus demonstrating a need for future analysis to dissect the exact scheme of methyl transfer.

Conclusions

Substrate recognition by the KsgA/Dim1 methyltransferases is complex. KsgA is able to methylate 30S subunits, but it can also methylate a pre-ribosomal particle containing 16S and a partial complement of ribosomal proteins[45]. Dim1 is essential for early processing of the pre-18S rRNA, but doesn't methylate 18S until very late in the 40S maturation process[83]. Despite evolutionary divergence of ribosomal assembly and processing pathways, eukaryotic and archaeal KsgA orthologs are able to methylate *E. coli* 30S both *in vivo* and *in vitro*. This requires the conservation of similar structural cues in small ribosomal subunits across evolution. Also complex is the mechanism of the modification performed by these enzymes. A total of four methyl groups are transferred, from four SAM molecules, to two separate adenosines. It is clear from the crystal structure of KsgA[131] that only one SAM molecule is bound at a time, and that the adenosines enter the active site separately. It has not been determined in what order, if any, the methyl groups are transferred, or if all four of the transfers take place within a single or multiple binding events. The work presented here demonstrates clear differences in the reaction intermediate profiles produced *in vitro* by KsgA enzymes from bacteria, archaea and yeast, despite the fact that the three enzymes methylate bacterial 30S to a similar extent and at similar rates in the assay used. However, we cannot exclude the possibility that the differences in the respective rates of m^6A and m^6_2A production seen here are a result of suboptimal substrate or assay conditions rather than a reflection of true differences in mechanism. For example, the yeast and archaeal enzymes may show different activity if assayed on their respective small ribosomal subunits rather than on bacterial 30S. Our

results demonstrate the remarkable cross-recognition of a complex substrate by evolutionarily distant members of an enzyme family and emphasize the need to further investigate the multi-step reaction mechanism.

Experimental

Cloning

E. coli genomic DNA was provided by Dennis J. Wilson. The 819 base-pair KsgA gene was amplified by PCR using the following primers, purchased from Integrated DNA Technologies: 5'-ATC GCC CAT ATG ATG AAT AAT CGA GTC CAC CAG G-3' and 5'-ATT ATG CAC GAG TTA ACT CTC CTG CAA AGG CG-3'. The gene was cloned into the pET15b vector (Novagen) as an NdeI/XhoI fragment for expression as a His-tagged fusion construct. The recombinant plasmid was sequenced (Nucleic Acids Research Facilities, Virginia Commonwealth University) to confirm the presence and correct sequence of the insert.

M. jannaschii genomic DNA was obtained from ATCC. The MjDim1 gene was amplified with the following primers, purchased from Integrated DNA Technologies: 5'-GCC GCA CCA TAT GTT CAA ACC AAA GAA AAA ATT AGG-3' and 5'-GCT ACT CGA GCT ATA ACC TAC CCC TAT TTT GCA G-3'. The amplicon was then inserted into pET15b as an NdeI-XhoI fragment for expression as a His-tagged fusion construct. The correct clone was confirmed by sequencing.

Protein expression and purification

pET15b-KsgA and pET15b-MjDim1 plasmids were transformed into BL-21 (DE3) cells for overexpression. Cell cultures were grown to an OD_{600} of 0.6 in the presence of

ampicillin and induced with 1 mM IPTG (Sigma-Aldrich). After 4 hours at 37ºC, cells were harvested by centrifugation. Pellets were resuspended in lysis buffer (50 mM NaPO$_4$, 300 mM NaCl, 10 mM imidazole, pH 8.0), broken with two passages through an Emulsiflex cell breaker (Avestin), and centrifuged to remove cell debris. Cleared lysate was loaded onto a HiTrap Chelating column (Amersham) equilibrated with 0.1M NiSO$_4$, washed twice with increasing amounts of imidazole (wash buffer 1: 50 mM NaPO$_4$, 300 mM NaCl, 20 mM imidazole, pH 8.0; wash buffer 2: 50 mM NaPO$_4$, 300 mM NaCl, 50 mM imidazole, pH 8.0), and the protein eluted with elution buffer (50 mM NaPO$_4$, 300 mM NaCl, 250 mM imidazole, pH 8.0).

The pET15b-ScDim1 construct was provided by Dr. Jean Vandenhaute and was confirmed by sequencing. Protein was expressed in BL21-CodonPlus (DE3)-RIL cells (Stratagene). The cells were grown at 37ºC to an OD$_{600}$ of 1.2 in the presence of ampicillin. Then the protein was induced under mild conditions with 0.1 mM IPTG and transferred to 25ºC for 4 hours. Cells were harvested and broken as for KsgA and MjDim1. Purification was carried out by affinity chromatography using a Ni^{2+} column; buffers were as above with the addition of 15 % glycerol and 3 mM 2-mercaptoethanol. To increase the stability of the protein, glycerol and 2-mercaptoethanol were added to the purified protein to final concentrations of 25% and 6 mM respectively.

Proteins were estimated to be >95% pure by SDS-PAGE analysis. Protein concentration was measured using the Bio-Rad Protein Assay.

30S purification

An *E. coli* strain lacking functional KsgA was constructed by growing BL-21 (DE3) cells on kasugamycin (ksg) to select for loss of the dimethylations. 30S ribosomes from this ksg^R strain were used in an *in vitro* assay to confirm that the adenosines were able to be methylated and therefore that the resistance to ksg was due to lack of KsgA activity. 30S subunits from both the ksg^R strain and the wild-type strain were prepared as previously described[132], except that cells were broken as described above. Purified subunits were dialyzed into reaction buffer (40 mM Tris, pH 7.4; 40 mM NH$_4$Cl; 4 mM MgOAc; 6 mM 2-mercaptoethanol) and stored at –80ºC in single-use aliquots. 30S concentration was estimated by measuring the absorbance at 260 nm and using a relationship of 67 pmol 30S per 1 unit of optical density.

In vivo assay

ksg^R cells were transformed with the pET15b constructs and selected on LB plates containing ampicillin. Transformed colonies were picked into liquid media and grown in overnight culture. These cultures were diluted 1:25 in fresh LB containing 50 μg/ml ampicillin and grown to OD$_{600}$ of 0.7-0.8, diluted 1:100 in fresh LB, and plated onto LB/ampicillin containing increasing amounts of ksg, from 0 to 3000 μg/ml. Plates were incubated at 37º overnight and visually inspected for colony formation.

In vitro assay

The *in vitro* assay was adapted from Poldermans *et al.*[47]. Time-course reactions were performed in 500 ul volumes containing 40 mM Tris, pH 7.4, 40 mM NH$_4$Cl, 4 mM MgOAc, 6 mM 2-mercaptoethanol, 0.02 mM ^{3}H-methyl-SAM (780 cpm/pmol; MP Biomedicals), 100 pmol 30S subunits, and 10 or 100 pmol enzyme; volume and

components were sufficient for 10 reactions. Buffer and reagents were pre-warmed to 37º

and added into pre-warmed tubes to minimize any lag in the reaction start. At each of eight

designated time points 50 ul was removed and added to a pre-chilled tube containing 10 ul

of 100 mM unlabeled SAM (Sigma-Aldrich) to quench the reaction; the remaining 100 ul

was stored at -20º and used for HPLC analysis (see below). The quenched reactions were

deposited onto DE81 filter paper (Whatman), washed twice with ice-cold 5% TCA, and

rinsed briefly with ethanol. Filters were air-dried for one hour, placed into scintillation

fluid, and counted.

HPLC analysis

Labeled 16S rRNA was extracted from 30S subunits with

phenol/chloroform/isoamyl alcohol. 16S was digested and dephosphorylated as described

by Gehrke and Kuo[133], and subjected to nucleoside analysis by reversed-phase HPLC.

Nuclease P1 was obtained from USBiological, shrimp alkaline phosphatase was from MBI

Fermentas. HPLC analysis was performed on a Polaris C-18 column (Varian). The HPLC

system used consisted of a Waters 600 Controller, a Waters 2487 Dual λ Absorbance

Detector, and Waters Empower software. Radioactivity was monitored with a Packard

150TR Flow Scintillation Analyzer. Buffer A was 20 mM NaH_2PO_4, pH 5.1. Buffer B was

20 mM NaH_2PO_4, pH 5.1:acetonitrile 70:30. Separation was performed at room

temperature using a linear gradient from 100%A-100%B over twenty minutes, at a flow

rate of 1.0 ml/min. Nucleoside standards used were N^6-methyladenosine (Sigma-Aldrich)

and $N^6 N^6$-dimethyladenosine, synthesized as in Rife et al.[134]. Peak integration was

calculated by the Empower software and used to determine ratios of m^6A:m^6_2A.

Crystal Structure

A significant step towards understanding an enzyme's function is often to obtain information about its structure. Knowledge of the structure can help guide predictions of areas important for the protein's function, including interactions with substrate and cofactor. A structure can also direct selection of amino acid residues for mutational analysis and biochemical characterization. Although structures of related methyltransferases have been solved, the KsgA/Dim1 family has not been explored structurally.

KsgA belongs to a large group of related SAM-dependent methyltransferases[98]. This group is responsible for methylating a diverse array of target molecules, including proteins, nucleic acids, and small molecules. At least fifty of these methyltransferases have been characterized structurally[135, 136]. Despite generally low sequence conservation, these structures share a remarkable degree of structural conservation. When SAM is present in the crystal structures, it is found in very similar orientations in each protein active site, which goes along with the strong structural conservation in this region. The conserved core structure consists of a central beta sheet and bears a strong resemblance to the well-defined Rossman fold characteristic of proteins which bind NAD(P)[98] (Figure 15). The Rossman fold proteins contain a six-stranded parallel sheet; the methyltransferase fold is characterized by an additional strand inserted antiparallel between strands five and six. The
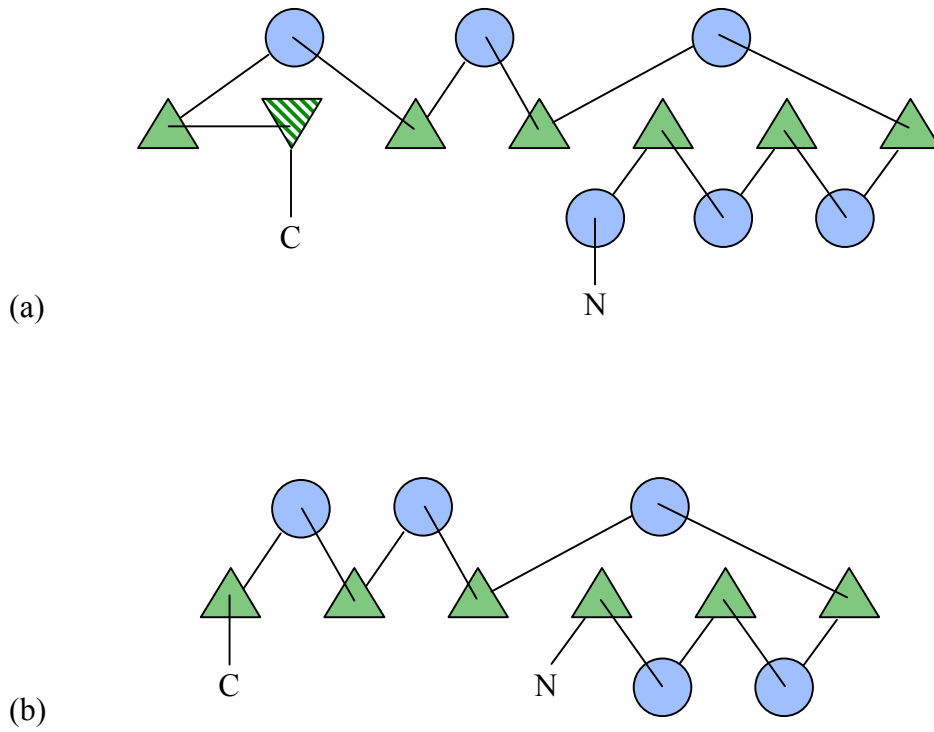
Figure 15. Topology diagrams of the consensus SAM-dependent methyltransferase core structure (a) and the Rossman fold (b). Blue circles represent α-helices; green triangles represent β-sheets. The hatched green triangle denotes the antiparallel β-strand found in the methyltransferase fold.

β-strands alternate with α-helices, which are not as well conserved in the secondary structure as the β-strands, to form an α/β sandwich. The SAM-binding site lies at the edge of the sheet, analogous to the NAD(P) binding site in Rossman fold proteins. The adenosine moieties of both molecules are bound in a virtually identical manner. Given this high degree of structural conservation, a reasonable model of KsgA's N-terminal catalytic domain can be constructed using known structures of other methyltransferases. Sequence similarity with the most closely related enzymes can aid predictions of catalytically important residues. However, homology models of KsgA's structure are ultimately inadequate. Sequence conservation, even among closely related enzymes, is limited, and there are many sequence elements that lie outside of the central β-sheet that cannot be modeled in a satisfactory way.

Outside of the catalytic center, the methyltransferase structures are not conserved. This is probably due to the wide range of substrate molecules targeted by this enzyme family. Of the structurally characterized SAM-dependent methyltransferases, ErmC' and ErmAM are the most closely related to KsgA. The two families share essentially no sequence homology in the C-terminal domains, which are thought to be important in substrate binding. Therefore, this region of KsgA cannot be modeled in a satisfactory way, and predictions about substrate binding are tenuous at best.

We have crystallized KsgA from *E. coli* and solved its structure at 2.1 Å resolution and we discuss implications of the structure for substrate binding and activity. This is the first structural characterization of a protein from this remarkable family. Recently the human cytoplasmic ortholog, hDim1, has also been crystallized and its structure

determined[109]; comparisons of these two closely related proteins may shed light on the

acquisition of a second function in the Dim1 proteins. Extensive biochemical and structural

analyses have been carried out on other SAM dependent methyltransferases. These data,

combined with the KsgA crystal structure, have allowed prediction of residues important

for KsgA function. We have performed mutagenic analysis of selected residues, testing the

mutant proteins for *in vivo* and *in vitro* activity.

Crystallization and refinement

KsgA-His fusion protein was overexpressed and purified as described in Chapter 2.

Purified protein was dialyzed into Buffer A (50mM Tris, pH 7.4, 400mM NH_4Cl, 6mM β-

mercaptoethanol) and stored in 10% glycerol at 253 K. Prior to crystallization, protein was

simultaneously concentrated and exchanged into Buffer B (50mM Tris, pH 7.4, 50mM

NH_4Cl, 6mM β-mercaptoethanol). For crystallization experiments, equal amounts of

protein and precipitant solution were mixed in hanging drops. High quality crystals were

obtained with 18-20% PEG 4000 and 0.2M calcium acetate (Figure 16). The best crystals

were obtained by using 4 μl of protein sample in Buffer B mixed with 4 μl of reservoir

solution. Under these conditions, protein readily formed irregular monoclinic/pseudo-

orthorhombic crystals as clusters from which single crystals could be obtained.

Crystals were of the C2 space group with unit cell dimensions of a=173.9 Å,

b=38.4 Å, c=83.0 Å, β=90.0°. The crystals diffracted X-rays to 1.9 Å; however, due to

limitations imposed by a long unit cell axis and a finite detector area, data were truncated

at 2.1 Å. Initial single isomorphous replacement with anomalous scattering (SIRAS)

phases proved sufficient to permit automated tracing of approximately 80% of the

Figure 16. A representative crystal of KsgA. Its approximate dimensions are 0.5 x 0.1 x 0.05 mm.

backbone. After several rounds of manual rebuilding and computational refinement, the model had R and R_{free} values of 18.9% and 23.8%, respectively. Representative electron density is shown in Figure 17. The region shown is part of the packing interface between the two copies of KsgA. In the final model, 92.2% of the residues had backbone torsion angles in the most favored regions of a Ramachandran plot, and 7.1% of the residues had torsions in additionally allowed regions; only 0.2% and 0.5% had torsions in generously allowed and disallowed regions, respectively. Data collection and refinement statistics are summarized in Table 3.

Structure analysis

The asymmetric unit is composed of two monomers, which share only a small contact surface (Figure 18). For each monomer, electron density was observed for residues 17-268 (out of a possible 293, which includes a poly histidine tag on the N-terminus). Least squares superposition of the two monomers indicated that they were not identical. The overall root mean squared deviation (RMSD) of the backbone atoms between the two monomers is 1.2 Å. This somewhat high RMSD between the two copies can be attributed to differing orientations of the C-terminal domain (residues 208-268) and of residues 140-160 with respect to the remainder of the N-terminal domain. When the two monomers are aligned using only residues 17-139 and 161-207, the RMSD drops to 0.5 Å; alignment of residues 140-160 and 208-268 yields an RMSD of 0.6 Å. While some of the structural differences between the monomers are likely due to constraints imposed by the crystal lattice, they may nonetheless have important consequences in the SAM binding site (see below).
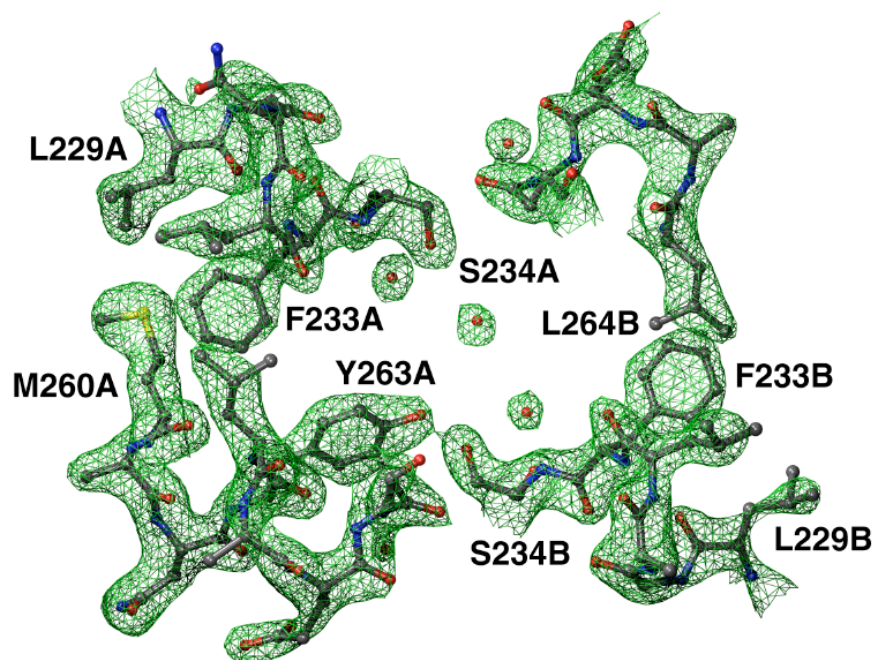
Figure 17. Representative region of KsgA, showing the interface between the two monomers of the asymmetric unit. The green mesh surface shows the $2F_o$-F_c map at 2.1 Å using phases calculated from the final refined model and contoured to 1 σ.

Table 3. X-ray data collection, phasing and refinement

Data Collection		
	Native	$K_3UO_2F_5$
Space group	C2	C2
Unit-cell parameters	a=173.9 Å, b=38.4 Å, c=83.0 Å, β=90.0°	a=174.5 Å, b=38.5 Å, c=83.0 Å, β=89.9°
Resolution (Overall / high resolution shell)	32.0–2.1 (2.14–2.10)	40–2.6 (2.64–2.60)
N_{obs}[a]	32474 (1261)	12478 (736)
Completeness (%)[a]	95.7 (76.6)	98.4 (68.9)
R_{symm}[b]	0.08 (0.33)	0.09 (0.31)
I/σ (I)	17.2 (2.7)	19.2 (3.2)
Multiplicity	4.0 (3.2)	6.9 (5.6)
Wilson Plot B ($Å^2$)	34.3	56.3
Optical Resolution (Å)[c]	1.65	
Phasing		
R_{deriv}[d]		0.31
R_{Cullis} (Isomorphous/Anomalous)[e]		0.734/0.642
Phasing Power (Isomorphous/Anomalous)		1.36/1.96
Figure of Merit (Centric/Acentric)	0.445/0.336	
Figure of Merit After Density Modification	0.871	
Refinement		
Resolution range (Overall / high resolution shell)	32.0–2.1 (2.16–2.10)	
Reflections (work/free)	28356/3159	
Protein Atoms	3876	
Water Atoms	237	
R_{work} (Overall / high resolution shell)	0.189 (0.218)	
R_{free} (Overall / high resolution shell)	0.238 (0.279)	
$R_{overall}$	0.194	
<B> ($Å^2$)	27.4	
Cruickshank's DPI (Å)[f]	0.291	
Estimated Maximal Coordinate Error	0.467	
RMSD Bond lengths (Å)	0.015	
RMSD Angles (º)	1.35	
Ramachandran Plot		
Most Favored (%)	93.6	
Additional Allowed (%)	5.7	
Generously Allowed (%)	0.2	
Forbidden (%)	0.5	

[a] For reflections with I_h>0 where I_h is the mean intensity of reflection h.

[b] $R_{symm} = \Sigma\Sigma_i |I_h-I_{hi}| / \Sigma\Sigma_i I_h$ where I_h is the mean intensity of reflection h. All data with I>-3σ are included.

[c] $W = (\sigma^2_{Patterson} + \sigma^2_{sph})^{1/2}$ the expected minimum distance between two resolved atom peaks[35].

[d] $R_{deriv} = \Sigma |F_{HP}-F_P| / \Sigma F_P$

[e] $R_{Cullis} = \Sigma ||F_{PH}-F_P| - |F_H (calc)|| / |F_{PH}-F_P|$

[f] $\sigma(x) = (N_{atoms}/N_{obs})^{1/2}c^{-1/3}d_{min}R_{free}$

Figure 18. Asymmetric unit of KsgA. The A monomer is shown at the top, and the B monomer is at the bottom.

The structure of the A monomer of KsgA is shown in Figure 19. The three-dimensional structure of KsgA is very similar to those of hDim1[109], ErmC'[107], and sc-mtTFB[106]. Figure 20 shows a structure-based sequence alignment of the four proteins. All four structures share a two-domain architecture with the larger N-terminal domain consisting of a mixed α/β structure and a smaller C-terminal domain consisting of 4-5 α-helices. The N-terminal domains of KsgA, hDim1 and ErmC' share the Rossman-like fold characteristic of SAM-utilizing methyltransferase, known as the Ado-Met dependent MTase fold[98]. This fold consists of alternating α-helices and β-strands, with a central seven-stranded β-sheet sandwiched between a variable number of α-helices. Strands 1-6 of the sheet are parallel, with strand 7 inserted antiparallel between strands 5 and 6. KsgA has three α-helices on one side of the sheet and four on the other, and an additional three 3_{10}-helices in the N-terminal domain. The C-terminal domain of KsgA consists of four α-helices and one 3_{10}-helix and, as in the other three structures, forms a cleft with the N-terminal domain. sc-mtTFB retains a similar topology to the other three enzymes but has an overall larger structure and an extra strand in its β-sheet.

The strong similarity between the KsgA/hDim1 and ErmC' structures is not surprising, since the two enzyme families perform very similar functions. The KsgA/Dim1 and Erm families are rRNA N^6-adenine methyltransferases, and both KsgA and ErmC' dimethylate their target adenosines. Both are members of the γ group of methyltransferases based on the order of conserved structural motifs[99]. KsgA clearly displays five of the nine structural motifs usually found in Ado-Met dependent methyltransferases. Residues in

(a)

(b)
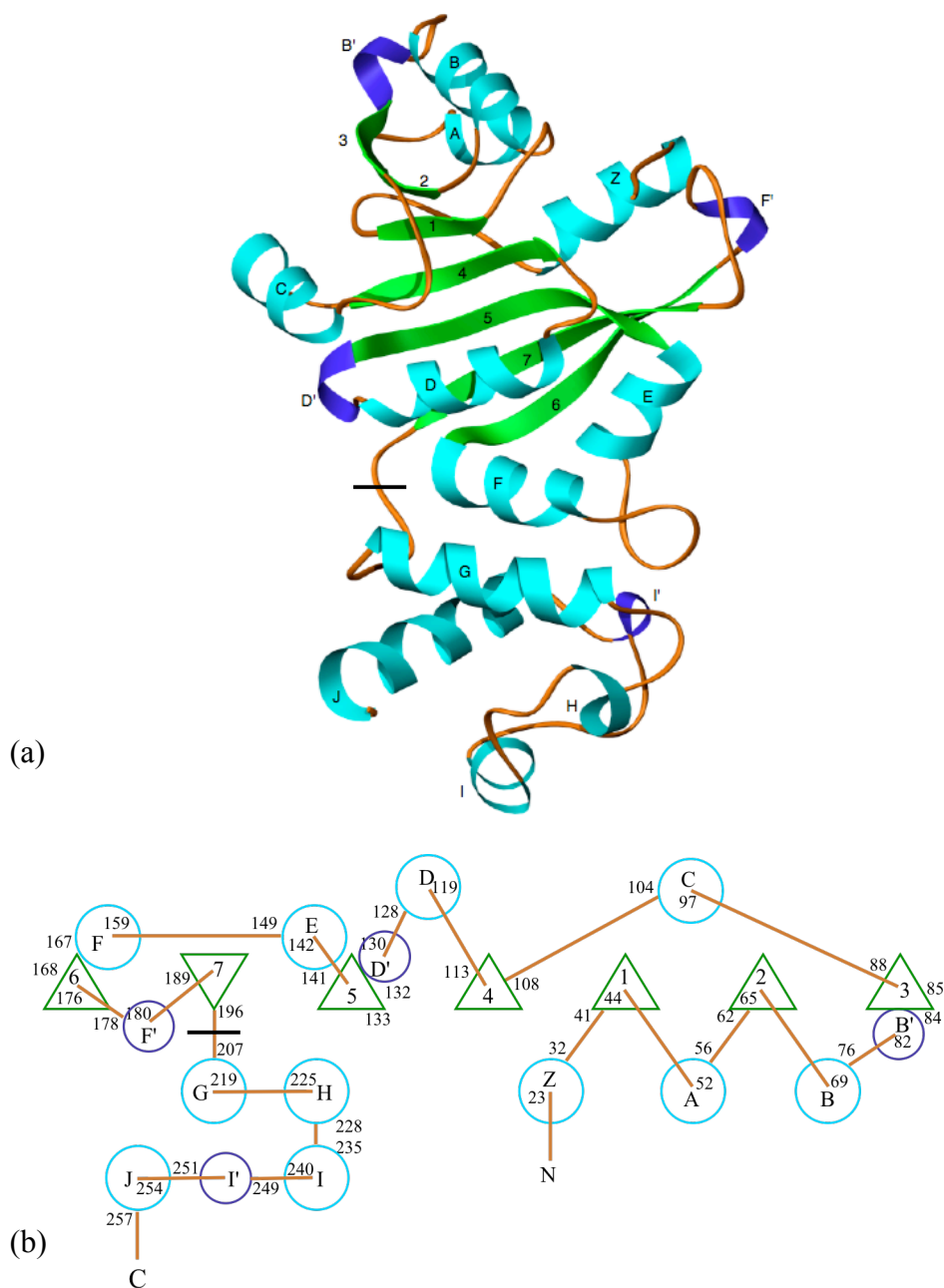
Figure 19. Diagram of the three dimensional structure of KsgA. β-strands are numbered and shown in green. Helices are denoted Z and A-J; α-helices are shown in cyan, 3_{10}-helices are magenta. The black line indicates the separation between the N- and C-terminal domains. (a) Ribbon diagram of the A monomer of KsgA. (b) Topology diagram of KsgA. β-strands are represented as triangles; α- and 3_{10}-helices are large and small circles, respectively. The residue numbers at the ends of secondary structures are noted.

```
                                                              αZ
                                                       ────────────
KsgA       ------------------MNNRVHQGHLARKRFGQNFLNDQFVIDSIVSAINPQK----
hDim1      MPKVKSGAIGRRRGRQEQRRELKSAGGLMFNTGIGQHILKNPLIINSIIDKAALRP----
sc-mtTFB   ---------------MSVPIPGIKDISKLKFFYGFKYLWNPTVYNKIFDKLDLTKTYKH
ErmC'      -----------------------MNEKNIKHSQNFITSKHNIDKIMTNIRLNE----


                β1              αA           β2           αB              β3
              ──────►       ─────────     ──────►     ─────────        ──────►
KsgA       --GQAMVEIGPGLAALTEPVGE--RLDQLTVIELDRDLAARLQT--HPFLGPKLTIYQQ-
hDim1      --TDVVLEVGPGTGNMTVKLLE--KAKKVVACELDPRLVAELHKRVQGTPVASKLQVLVG
sc-mtTFB   PEELKVLDLYPGVGIQSAIFYNKYCPRQYSLLEKRSSLYKFLNAKFEGS---PLQILKRD
ErmC'      --HDNIFEIGSGKGHFTLELVQR--CNFVTAIEIDHKLCKTTEN--KLVDHDNFQVLNK-
              ◄───────────►         ◄──►             ◄───────►
                    I                 II                  III


                αC                   β4            αD
             ─────────             ──────►      ─────────
KsgA       DAMTFNFGELAEKM-----G-Q-------PL-RVFGNLPYNISTPLMFHLFSYT------
hDim1      DVLKTDL--PF------------------FDTCVANLPYQISSPFVFKLLLHR------
sc-mtTFB   PYDWSTYSNLIDEERIFVPEVQSSDHINDKF-LTVANVTGEGSEGLIMQWLSCIGNKNWL
ErmC'      DILQFKF--PKN-------Q-S--------Y-KIFGNIPYNISTDIIRKIVFD-------
           ◄──►                         ◄──────►    ◄──────────►
            IV                            IV           V


              β5        αE                αF          β6
           ──────►   ─────────         ─────────   ──────►
KsgA       --DAIADMHFMLQKEVVNRLVAGPNSKAYGRLSVMAQYYCNVIPVLEV------------
hDim1      --PFFRCAILMFQREFALRLVAKPGDKLYCRLSINTQLLARVDHLMKV------------
sc-mtTFB   YRFGKVKMLLWMPSTTARKLLARPGMHSRSKCSVVREAFTDTKLIAISDANELKGFDSQC
ErmC'      --SIADEIYLIVEYGFAKRLLN-----TKRSLALFLMAEVDISILSMV------------
             ◄──►                            ◄──►
              VI                             VII


                        β7                           αG
                     ──────►                      ─────────        ──
KsgA       --------PPSAFTPPPKVDSAVVRLVPHATMPHPVK--DVRVLSRITTEAFNQRRKTI
hDim1      ---------GKNNFRPPPKVESSVVRIEPKNPPP-PI---NFQEWDGLVRITFVRKNKTL
sc-mtTFB   IEEWDPILFSAAEIWPTKGKPIALVEMDPID---FDF---DVDNWDYVTRHLMILKRTPL
ErmC'      ---------PREYFHPKPKVNSSLIRLNRK-KS--RISHKDKQKYNYFVMKWVN------
             ◄──────►
               VIII


            αH                                              αI
         ──────                                          ─────────    ──
KsgA       RNSLGNL----------------------------FSVEVLTGMGIDPAM-RAENISVA
hDim1      SAAFKSSAVQQLLEKNYRIHCSVHNIIIPEDFSIADKIQQILTSTGF-SDK-RARSMDID
sc-mtTFB   NTVMDSLGHGGQQ---------------------YFNSRIT-DKDLLKK-CPIDLTND
ErmC'      -KEYKKIFTK-------------------------NQFNNSLKHAGIDDLNNISFE


              αJ
           ─────────
KsgA       QYCQMANYLAENAPLQES
hDim1      DFIRLLHGFNAEGIHFS
sc-mtTFB   EFIYLTKLFMEWPFKPDILMDFVDMYQTEHSG
ErmC'      QFLSLFNSYKLFNK
```

Figure 20. Structure-based sequence alignment of KsgA (P06992), hDim1 (Q9UNQ2), sc-mtTFB (P14908), and ErmC' (P13596). Structures used for alignment were 1QYR[131] (KsgA), 1ZQ9[109] (hDim1), 1I4W[106] (sc-mtTFB) and 2ERC[107] (ErmC'). Blue lines and green arrows indicate KsgA's α-helices and β-strands, respectively. Double-headed arrows indicate motifs common to SAM-dependent methyltransferases. Bold residues in the ErmC' sequence are those which make direct contact with the SAM molecule.

these motifs have been shown to be involved in SAM binding in the ErmC' crystal structure[97], and in both SAM binding and interaction with the target adenosine in the DNA N^6-adenine methyltransferase M.TaqI[137, 138].

Motif I, often referred to as the G-loop, connects β1 with αA. It contains a canonical GXG sequence which in other methyltransferases forms part of the binding pocket for SAM. Motif II consists of an Asp or Glu residue (E66 in KsgA) followed by a hydrophobic residue (L67). In ErmC' the corresponding E59 forms hydrogen bonds with hydroxyl groups on the ribose of SAM; I60 might make van der Waals contact with the adenine of SAM. Motif III is also involved in SAM binding. The first residue of this motif is an Asp or Glu residue (D91 in KsgA) that, in DNA methyltransferases, interacts with the exocyclic N6 amine of the SAM adenine. This is followed in DNA methyltransferases by an aromatic residue and a hydrophobic residue; the NH group from the peptide backbone of the second residue forms a hydrogen bond with N1 of SAM. In Erm and KsgA enzymes the charged residue is followed by two hydrophobic residues.

Motif IV contains residues important in catalysis. The consensus sequence in DNA amino methyltransferases is (D/N)PPY; in Erm enzymes the consensus is (N/S)IP(Y/F) and the KsgA sequence is NLPY, residues 113-116. The first residue may form a hydrogen bond to the amino group in the target base and play a role in its deprotonation. Mutation of this residue is poorly tolerated in DNA methyltransferases[139-142]. In ErmC' the N101A mutant is inactive *in vivo* but retains some activity *in vitro*, although only about 10% of wild-type[143]. This reduction in activity is probably due to lack of proper SAM binding. The Y104A mutation in ErmC' was inactive both *in vivo* and *in vitro*, and similar results were

seen in M.TaqI[144]. These results highlight the importance of this residue, which is involved in stabilization of the target residue through stacking interactions.

Motif VI contains three adjacent hydrophobic residues at the C-terminal end of β5 (F138, M139 and L140 in KsgA) that are suggested to be involved in proper positioning of the target adenosine. Relative to ErmC', KsgA contains a five residue insert (151-155) which forms a loop between motif VI and motif VII, altering the trajectory of the backbone between αE and αF. The presence of this insert is conserved in the KsgA family, although the sequence of the five residues is not. Motif VIII contains a very highly conserved Phe residue (F181 in KsgA) that is part of the FXPXPXVXS sequence found in Erm and KsgA enzymes. Like Y116, this residue is involved in stabilization of the target base through stacking interactions. Mutation to Ala is very deleterious in the DNA methyltransferase M.TaqI (400-fold reduction of activity)[144], but activity of the Phe to Ala mutant is only reduced about 4-fold in ErmC'[143].

Motifs V, VII, and X are less well conserved in KsgA. Motif V in DNA methyltransferases is involved in SAM binding. The Leu residue in the DNA methyltransferase consensus sequence (N/D)LY makes van der Waals contact with the adenine of SAM[145]; this residue corresponds to I118 of KsgA, which may make a similar contact. Motif VII is weakly conserved among DNA methyltransferases and Erm enzymes, and we can find no equivalent of in KsgA. This motif may be involved in proper folding of DNA methyltransferases[146]. Motif X is also poorly conserved, if at all, but may comprise residues in αZ.

Structural comparisons

Figure 21 shows the A monomer of KsgA superimposed with the B monomer (Figure 21a), with hDim1 (Figure 21b), with ErmC' (Figure 21c), and with sc-mtTFB (Figure 21d). Table 4 lists regions of high structural correspondence between the proteins which were used to create the superpositions. Table 5 lists the RMSD values obtained from these superpositions. Not surprisingly, the RMSD between the two monomers of KsgA is slightly less when the two are superimposed using these regions, as opposed to using all backbone atoms. hDim1 has a slightly lower RMSD value with the B monomer (1.0 Å) than with the A monomer (1.2 Å); conversely, ErmC' has a lower RMSD value with the A monomer (2.3 Å) than with the B monomer (2.7 Å). sc-mtTFB deviates more strongly from hDim1 and ErmC' than from KsgA, with RMSD values of 2.4 Å (KsgA A and B monomers), 3.1 Å (hDim1), and 3.2 Å (ErmC').

Dim1 proteins have a large insert in the C-terminal domain compared to KsgA. This insert in hDim1 is shown with an arrow in Figure 21, and Figure 22 shows a slightly rotated, enlarged view. The insert falls between helices H and I, and results in an extra helix in hDim1 along with a section of random coil and a slightly longer helix I. Dim1 interacts with a variety of other proteins involved in pre-rRNA processing and maturation[24, 147-149]; it is tempting to speculate that some of these interactions are mediated by this "extra" region of Dim1.

mtTFB proteins have three inserts relative to KsgA proteins; all of these inserts lie in the N-terminal domain (Figure 23). Figure 23b shows the first of these inserts, which takes the form of a random coil leading into an extended beta strand. The second insert is
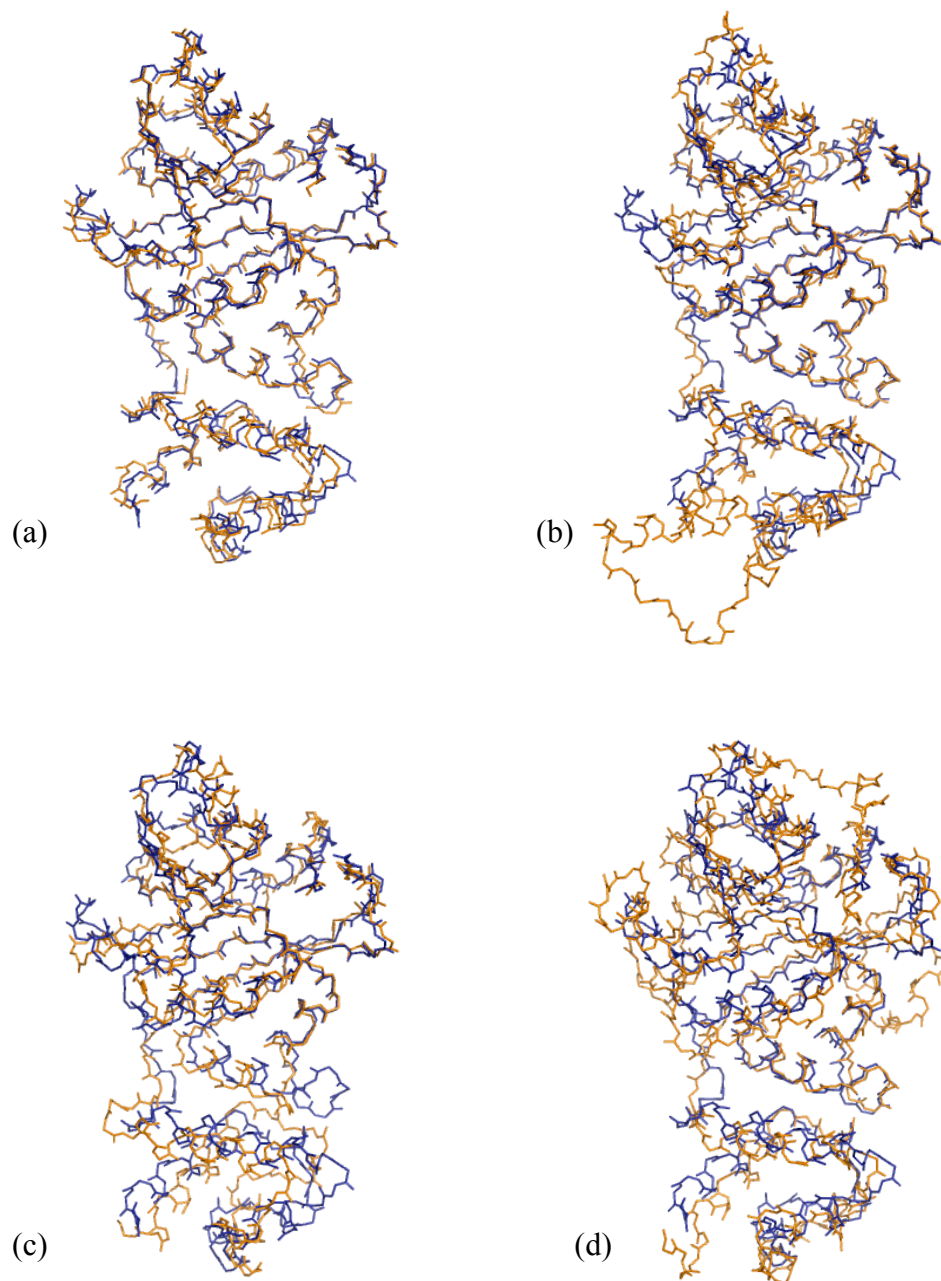
Figure 21. Superposition of the A monomer of KsgA with the B monomer (a), hDim1 (b), ErmC' (c), and sc-mtTFB (d). The KsgA A monomer is shown in blue, the other proteins are shown in orange.

Table 4. Regions of structural correspondence between KsgA, hDim1, ErmC', and sc-mtTFB.

Region	KsgA	hDim1	ErmC'	sc-mtTFB
1	17-37	36-56	10-30	20-40
2	38-57	57-76	31-40	47-66
3	58-77	77-96	51-70	69-88
4	84-96	106-118	77-89	94-106
5	108-129	123-144	96-117	131-152
6	131-150	146-165	118-137	162-181
7	156-176	171-191	138-158	187-207
8	177-197	192-212	159-179	229-249
9	206-219	219-232	187-200	254-267
10	248-266	289-307	223-241	301-319

Table 5. RMSD values of superpositions in Figure 21.

RMSD (Å)	KsgA (B)	hDim1	ErmC'	sc-mtTFB
KsgA (A)	0.9[a]	1.2	2.3	2.4
KsgA (B)		1.0	2.7	2.4
hDim1			1.9	3.1
ErmC'				3.2

[a]Superpositions were made based on the ranges listed in Table 4. When all backbone atoms are used KsgA (A) differs from KsgA (B) by 1.2 Å.
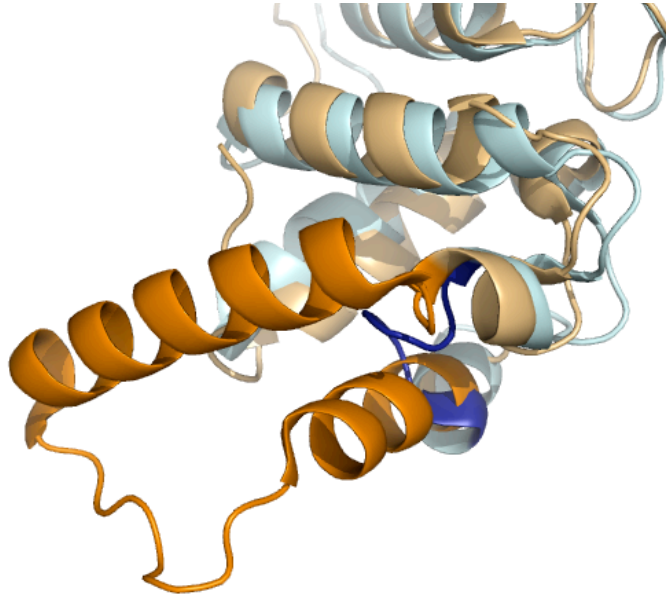
Figure 22. hDim1 insert. KsgA is shown in blue; hDim1 is shown in orange. The insert in hDim1 and the corresponding area in KsgA are shown in darker colors.
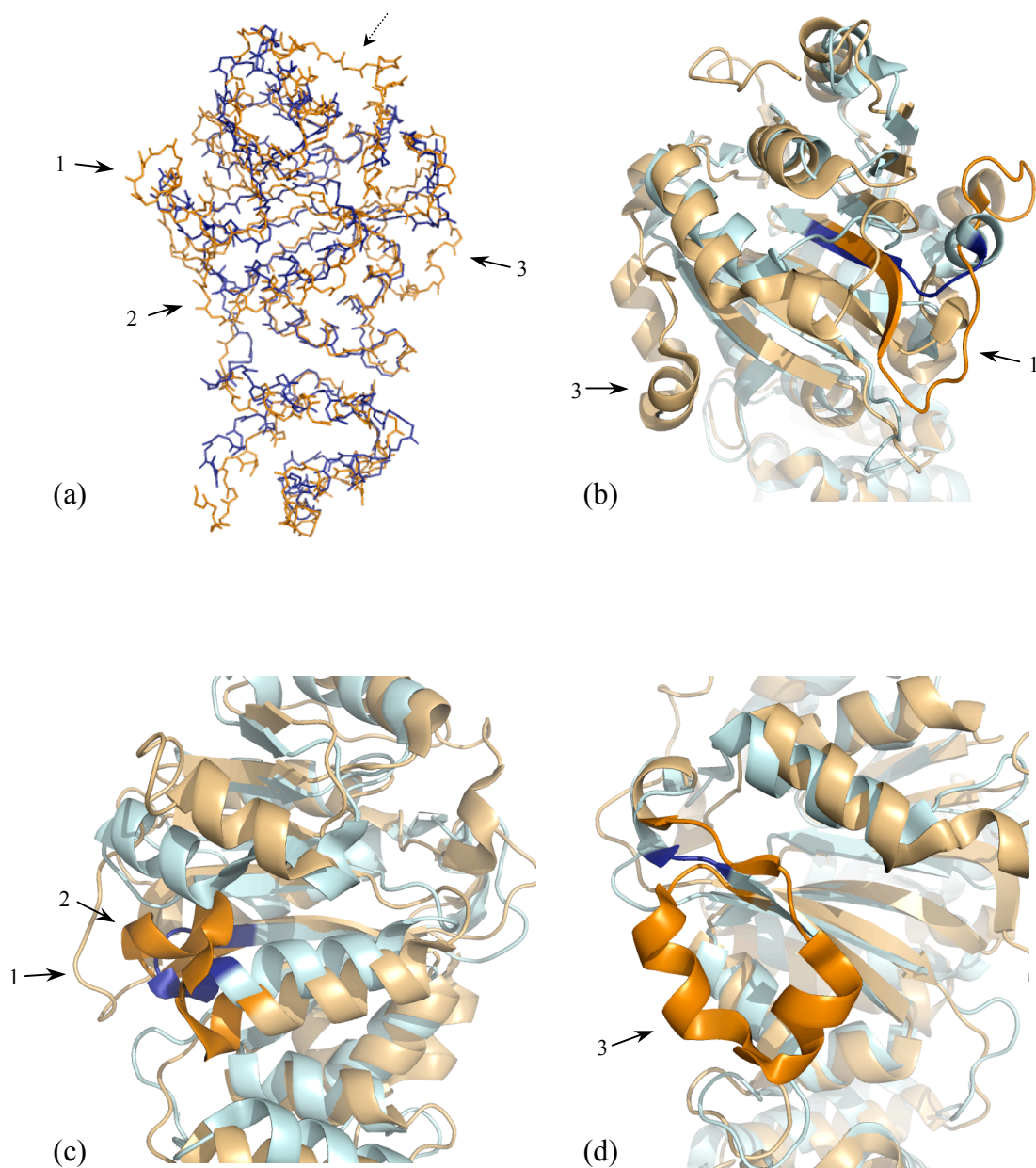
Figure 23. sc-mtTFB inserts. KsgA is shown in blue; sc-mtTFB is shown in orange. Inserts are indicated with solid arrows; inserts and their corresponding regions in KsgA are shown in darker colors in (b)-(d). The dashed arrow indicates the N-terminal end of sc-mtTFB. (a) Superposition from Figure 21. (b) Insert 1. (c) Insert 2. (d) Insert 3.

shown in Figure 23c and clusters together with insert 1. The third insert is shown in Figure 23d. This insert is on the other side of the protein from the first two, and consists of two alpha-helices which form a long loop on the protein surface. Inserts 1 and 2 are generally conserved in all mtTFB proteins; however, insert 1 is shorter in mtTFB1 than in mtTFB and mtTFB2, which have evolved as transcription factors perhaps at the expense of their full function as methyltransferases. Additionally, the presence of insert 3 is conserved in mtTFB and mtTFB2 proteins, but not in mtTFB1 proteins.

A Delphi surface map that reports modeled charge distribution reveals two regions of interest in the KsgA structure (Figure 24). In the N-terminal domain there is a negatively charged pocket which corresponds to residues in the canonical SAM-binding site; this pocket is also present in hDim1 and ErmC'. The pocket is approximately the correct size to accommodate a SAM molecule. Notably, a similar pocket is missing from the surface of sc-mtTFB. Figure 20 denotes residues in ErmC' that directly interact with SAM, all of which line the pocket. Of those twelve residues six are conserved in KsgA (G45, G47, E66, D91, N113 and P115), ErmC', sc-Dim1, and h-mtTFB, all proteins with demonstrated methyltransferase activity, while only three are conserved in sc-mtTFB. Three of the nonconserved residues in sc-mtTFB are likely to preclude sc-mtTFB from binding SAM. D91 of KsgA, which is invariant in the γ group of methyltransferases[99] is replaced by a Pro in sc-mtTFB. Based on the ErmC'/SAM co-structure[97] N19 of KsgA is predicted to form a hydrogen bond with the positively charged sulfur of SAM. In sc-mtTFB this Asn is replaced with a Lys that would likely produce an enormous charge repulsion between its side chain and SAM. The conserved Gly at position 45 in KsgA,
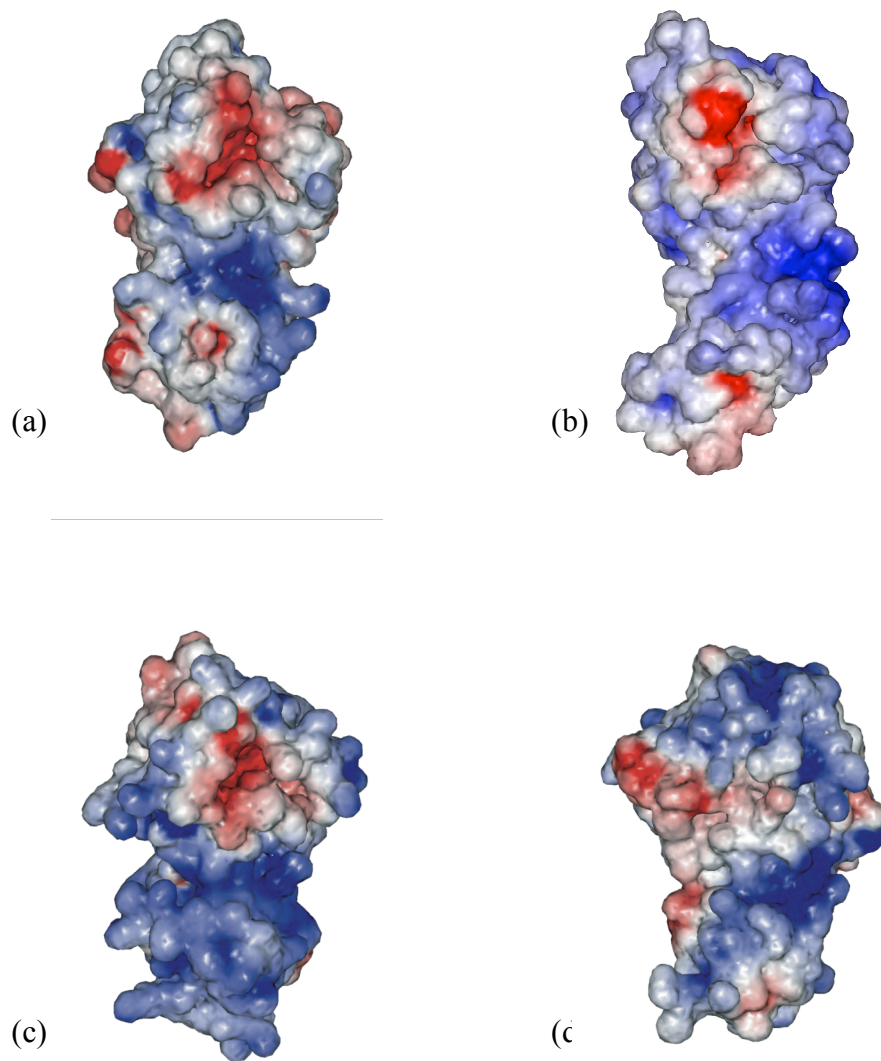
Figure 24. Delphi surface maps of (a) KsgA, (b) hDim1, (c) ErmC', and (d) sc-mtTFB. Red indicates areas of negative charge; areas of positive charge are shown in blue.

which in ErmC' makes a hydrogen bond to SAM through its carbonyl oxygen[97], is replaced by a Tyr in sc-mtTFB. The orientation of this Tyr[106] is such that its side chain would block SAM binding by steric hindrance. The lack of correspondence between several residues known to interact with SAM, along with the absence of methylated product in *S. cerevisiae* mitochondrial ribosomes, strongly suggests that sc-mtTFB is not able to act as a methyltransferase.

Also notable in the Delphi surface map is the large region of positive charge in the cleft between the N-terminal and C-terminal domains, present in all four enzymes. Bussiere *et al*. referred to the C-terminal domain of ErmC' as the RNA-recognition domain based on the observation that this cleft could accommodate an RNA helix[107], in much the same way as the DNA duplex is oriented in the M.TaqI methyltransferase ternary complex reported by Goedecke *et al*[150]. The importance of this region was recently tested by Maravic *et al*. by mutating charged and polar residues along the N-terminal and C-terminal faces of the cleft[151]. From this work they concluded that the C-terminal domain played no direct role in RNA binding, but on the basis of a series of deletion mutants probably helps to stabilize the larger N-terminal domain. Subsequent to that report Buriánková *et al*. reported the discovery of a *Mycobacterium tuberculosis* erm gene, *ermMT*, the product of which is an active Erm enzyme that lacks the C-terminal domain altogether[152]. Therefore, at least in some Erm methyltransferases the N-terminal domain alone is sufficient to bind target RNA.

The C-terminal domain of KsgA has fewer conserved residues than the N-terminal domain and has a structure substantially different from that of the same domain in ErmC'.

The first and last α-helices of the domain, αG and αJ in KsgA, are similar in composition and position, but the trajectories of the intervening segments follow very different paths. Five of the six C-terminal residues conserved in KsgA proteins (F218, R221, R222, K223, and R248) are located in the region between the two helices, and are spatially clustered (Figure 25). With the exception of F218 these conserved residues are outside of the domain interface, which might suggest a mode of RNA binding different from that proposed for ErmC'.

Substrate binding

As previously mentioned, the two copies of KsgA in the asymmetric unit differ considerably. To better understand the genesis of this difference we performed a double difference analysis on the two conformers of KsgA (data not shown). The plot indicates that the relatively large structural difference is the result of a large domain movement of the C-terminal domain relative to the N-terminal domain in the KsgA B monomer. This motion is propagated into the N-terminal domain through the residue 151-155 loop within the N-terminal domain and subsequently into the SAM binding site.

To further explore the effects of these domain movements on the putative active sites of the two KsgA monomers, the monomers were superimposed using Motifs I, II and III, which are involved in SAM binding. This superposition reveals a slightly compressed SAM binding pocket in the B monomer, along with a relatively large movement of the C-terminal domain (Figure 26). The active site cavities were also analyzed using the CASTp server[153] (cast.engr.uic.edu). For comparison, active site cavities for an ErmC' structure with SAM bound (PDB entry 1QAO) and for an ErmC' structure without SAM (PDB entry
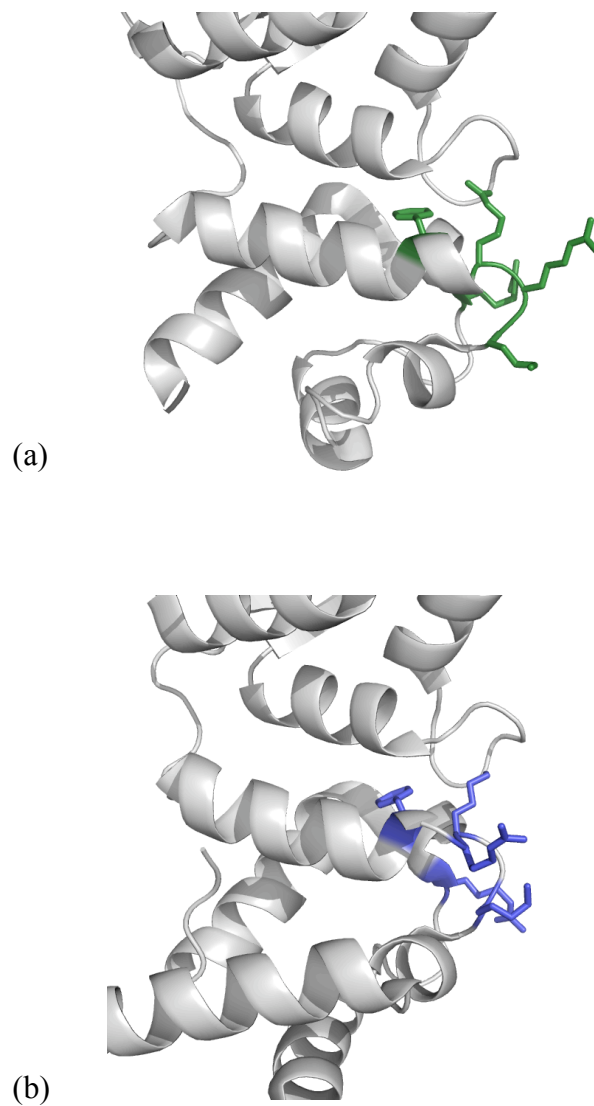
Figure 25. Conserved residues in the C-terminal domain of KsgA/Dim1. (a) KsgA. Residues F218, R221, R222, K223 and R248 are highlighted in green. (b) hDim1. Residues F231, R233, K234, K236 and R289 are highlighted in blue.

Figure 26. Superposition of the A (cyan) and B (orange) monomers, defined using SAM-binding motifs I, II, and III. (a) Backbone trace of the entire protein, highlighting an approximately 7 Å shift in the C-terminal domain. (b) Close-up of the SAM-binding pocket from (a), highlighting the movements of residues P115 and Q141.

1QAN) were also analyzed using CASTp. Active site cavity volumes and surface areas reported by CASTp are reported in Table 6.

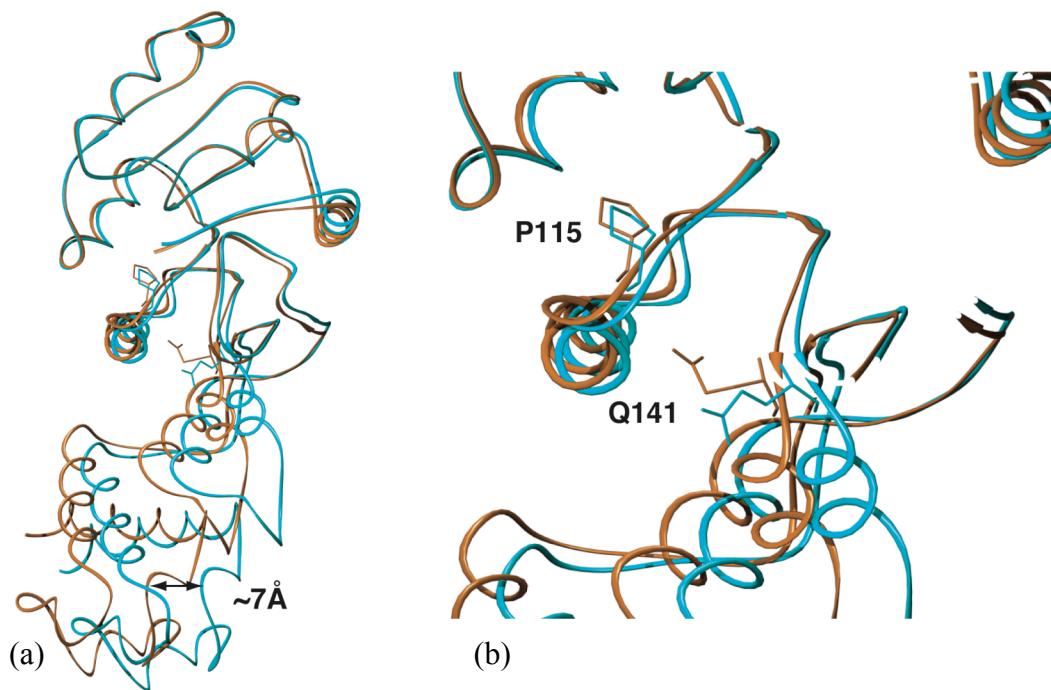Examination of the data in Table 6 indicates that there are no significant structural differences between the SAM binding pockets of ErmC' with or without a bound SAM molecule, suggesting that free ErmC' does not have to undergo any major conformational change in order to bind SAM. This is consistent with the observation that free ErmC' binds SAM with reasonably high affinity[154]. For KsgA, the volume of the putative SAM binding pocket in monomer A is similar to the volume of the binding pocket in ErmC'; however, for monomer B, the domain movements result in a 14% reduction in the cavity volume, which is largely the result of P115 being pushed into the cavity (Figure 26). In this conformation residue Q141 is oriented so that its side chain is pointing toward P115 from underneath, while in the A monomer conformation the Q141 side chain is oriented away from the binding pocket. The altered disposition of Q141 in the two copies of KsgA might affect the position of P115 within the SAM-binding pocket. The differences in the SAM binding cavities between the two KsgA monomers are particularly interesting in light of the observation that free KsgA is unable to bind SAM[47].

Our early expectation was that the structure of KsgA would present obvious clues as to why the free enzyme is unable to bind SAM tightly. Instead, the two copies provide a tantalizing hint to a possible allosteric mechanism. The two monomers in the asymmetric unit have very different packing interactions at a site involving highly conserved residues in the C-terminal domain (Figure 27). These interactions might approximate an important allosteric binding site whereby interaction of KsgA with a part of the 30S subunit separate

Table 6. Volume of active site cavities

Protein	Cavity volume[a]
KsgA A monomer	839 Å^3
KsgA B monomer	720 Å^3
ErmC' (with SAM)	812 Å^3
ErmC' (without SAM)	837 Å^3

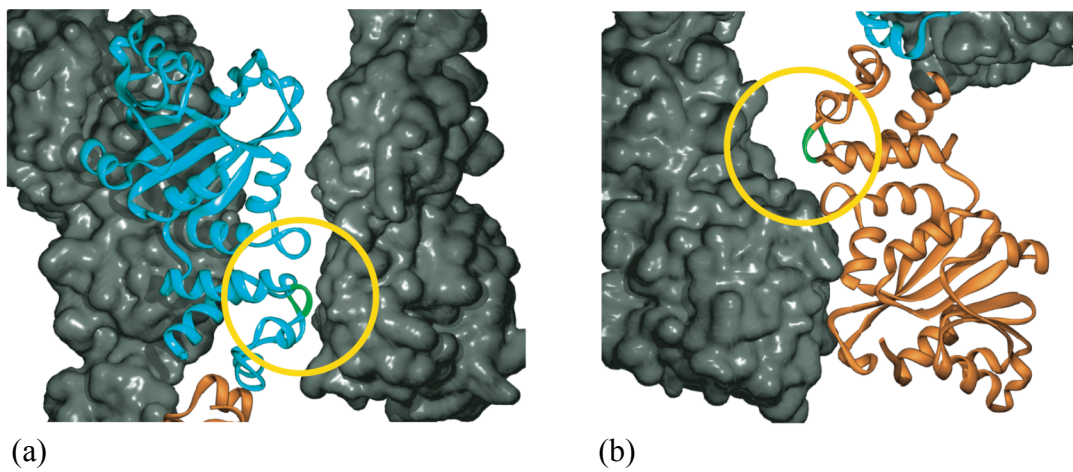[a]Calculations by CASTp: http://cast.engr.uic.edu/cast/

Figure 27. Packing interactions of (a) the A monomer and (b) the B monomer with their lattice neighbors, which are shown as solid surfaces. Residues R221, R222, K223, and R248 are shown in green, and the surrounding area is circled in yellow.

from the target adenosines in helix 45 could cause a conformational change, allowing SAM to bind in the binding pocket. Such a hypothesis is speculative and requires future experimental attention.

The Delphi map of KsgA shows a region of high positive charge in the cleft between the N-terminal and C-terminal domains (Figure 24). ErmC' and sc-mtTFB have similar clefts. Residues in this region in ErmC' have been implicated in recognition and binding of the substrate RNA[151], although only those on the N-terminal face of the cleft were shown to be important. Comparison of these three structures would suggest that sc-mtTFB retains its ability to bind to RNA, and there is some evidence to support this possibility.

The specifics of RNA binding to KsgA and all other RNA adenosine dimethyltransferases remain unknown. Through inspection of protein structures, assumed correspondence with DNA methyltransferases, and mutagenesis of ErmC', the body of evidence suggests that the likely orientation of helix 45 to KsgA is along the cleft formed between the two domains. Based on mutational analysis Maravic et al.[151] have constructed a model with the helical stem of domain V of 23S rRNA, the motif necessary and sufficient for substrate activity, positioned within interdomain cleft of ErmC'. The model satisfies many requirements: the target adenosine can be reasonably inserted into its active site pocket, while allowing for a large contact surface between the components in a way that maximizes the importance of those residues implicated in RNA binding as determined from a mutational analysis of ErmC'. In cartoon fashion the analogous complex between helix 45 of 16S rRNA and KsgA is shown in Figure 28. Given the close evolutionary ties
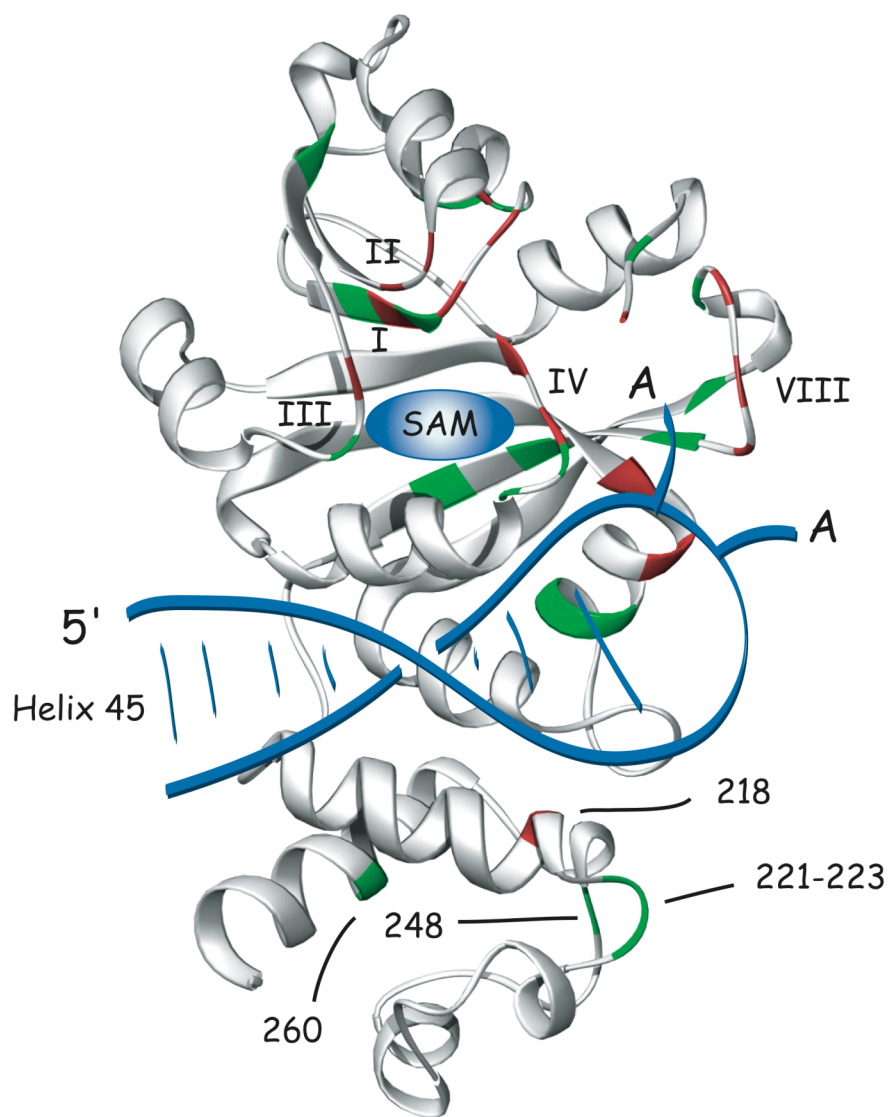
Figure 28. Cartoon model of Helix 45/KsgA binding interaction. The A monomer of KsgA is shown in white, with colored residues as in Figure 2. Helix 45 is shown in blue, with the target adenosines extending outward from the loop. Placement of the SAM molecule, also in blue, is based on the ErmC'/SAM co-crystal structure. Roman numerals indicate structural motifs, while numbers indicate the six residues that are conserved in the N-terminal domain (see text for details).

between ErmC' and KsgA it is tempting to believe that the enzymes bind their respective

substrates in similar ways and indeed this may be the case. However, while the overall

structures of KsgA and ErmC' are very similar, those similarities are concentrated within

the N-terminal domains, where the catalytic core resides, while the differences between the

two are concentrated in the C-terminal domains. In particular, the trajectories of the

peptide chain diverge considerably between αG and αJ, the only two structural units

conserved between KsgA and ErmC' in the C-terminal domains. When we align the

sequences of 30 Erm proteins, all from separate subfamilies, we find that there are no

residues in the C-terminal domain that are even moderately conserved according to T-

Coffee designation[130] (alignment not shown), which becomes less surprising with the

observation that at least one Erm, ErmMT, doesn't even have a C-terminal domain[152]. In

contrast, the KsgA/Dim1 family of methyltransferases has six absolutely conserved or

highly conserved residues in the C-terminal domain, four of which (R221, R222, K223,

and R248) are spatially clustered. This constellation is likely important for the enzyme's

function, and the nature of the four residues suggests that they may be involved in RNA

binding, either at helix 45 or elsewhere on the 16S RNA. If these residues constitute a

patch that interacts with the rRNA then KsgA probably does not bind helix 45 in the

manner that is thought to occur for ErmC' and its substrate RNA.

On the other hand, the conservation of positive charge in the C-terminal domain of

KsgA might be important for other reasons. Despite their similarities, KsgA and ErmC' are

fundamentally different enzymes. Functionally, KsgA must methylate two adjacent

adenosines, which means that at some point both A1518 and A1519 must flip into the

active site. Whether transfer of all four methyl groups requires one, two or four binding events between KsgA and the nascent ribosome is unknown. Additionally, KsgA is subject to more complex regulation than the Erm enzymes, which is possibly the result of an allosteric mechanism triggered by binding of KsgA to the partially formed 30S; such a mechanism would presumably involve RNA/enzyme contacts outside of the helix 45 region. Any or all of these extra functional constraints in KsgA might require altered or additional interactions with helix 45 and/or other regions of the 16S rRNA.

Mutational analysis

Using the structure of KsgA in combination with sequence alignments and mutational studies of related proteins, we chose eight residues in the catalytic core of KsgA for analysis (Figure 29). E66 is one of the most highly conserved residues in SAM-dependent methyltransferases; an acidic residue at this position is even found in the Rossman fold proteins, highlighting the importance of this residue in binding the adenosine moiety of SAM/NAD(P) in these enzymes[98, 99]. D91 is also predicted to be important for binding of SAM in the active site of KsgA. N113, P115, and Y116 are located in the highly conserved Motif IV. N113 has been proposed to be important for catalysis, and is located between the SAM and adenosine binding pockets. Y116 forms the floor of the adenosine-binding pocket and may help stabilize the adenosine into the pocket via stacking interactions. F181, P183 and P185, in Motif VIII, form the side wall of the adenosine-binding pocket; like Y116, F181 is proposed to interact with the target adenosine via stacking. Each of these residues was mutated to alanine, and Y116 and F181 were also mutated to tryptophan. In addition, the double mutant Y116A/F181A was constructed.
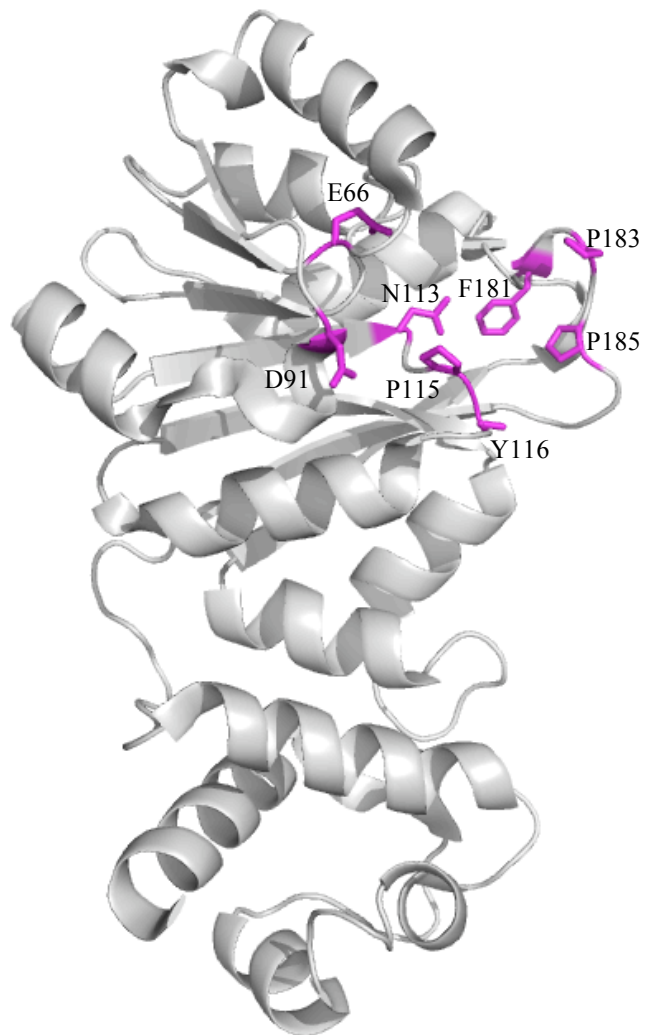
In vivo and *in vitro* analyses were performed as described in Chapter 2 (Figure 30). Most of the mutations studied have little to no effect on *in vivo* activity (Figure 30a), despite the generally high level of conservation of the mutated residues. Four notable exceptions are E66A and, to a lesser extent, D91A, Y116A, and Y116A/F181A. The E66A mutant has minimal activity *in vivo*, with an MIC close to that of the empty vector. D91A has impaired activity, but is able to restore some kasugamycin sensitivity to the bacteria. Both Y116A and Y116A/F181A have moderate activity *in vivo*. The N113A and F181A mutations are fully functional in our *in vivo* assay system. Mutation of P115, P183 or P185 to Ala has little to no measurable effect on *in vivo* activity. The conservative Y116W and F181W mutants had divergent effects. Y116W retained more activity than Y116A; this mutation resulted in near wild-type activity. F181W, conversely, has lower activity than F181A, possibly because of steric clashes involving the bulky Trp side chain; we did not study this mutant further.

Interpretation of the *in vivo* results alone is not quite straightforward. Although KsgA dimethylates two adenosines, at A1518 and A1519, the methylation status of A1519 seems to be the major determinant of resistance/sensitivity to kasugamycin. Mutation of A1518, and subsequent loss of dimethylation at this position, does not confer kasugamycin resistance[155] (Table 7). Mutation of A1519, on the other hand, leads to kasugamycin resistance comparable to that seen in *ksgA⁻* cells. Therefore, partial methylation by a KsgA mutant could still lead to kasugamycin sensitivity. In addition, there are other potential intermediate patterns of methylation which have not been correlated with resistance/sensitivity (Table 7). For this reason, the *in vivo* assay can be used qualitatively
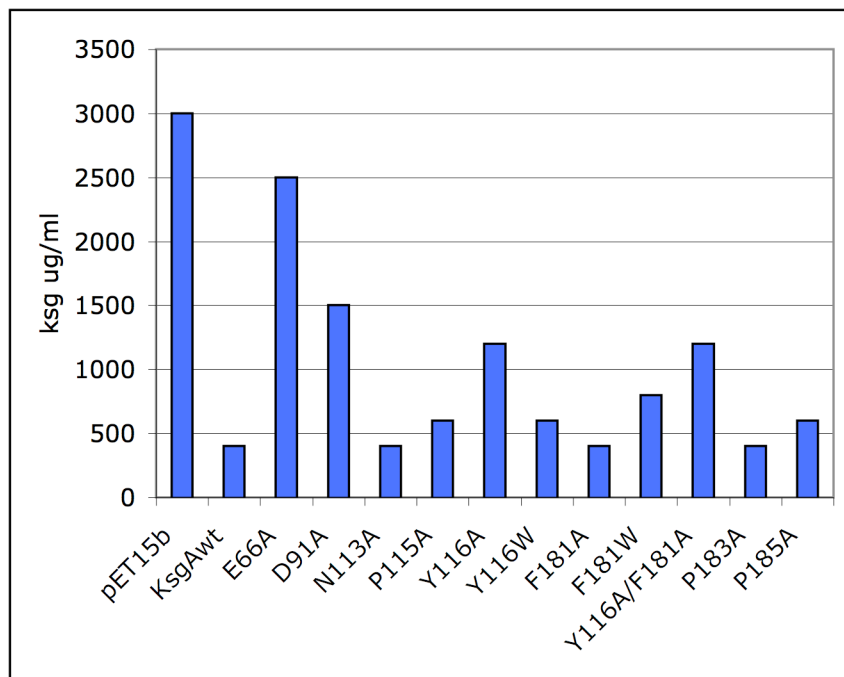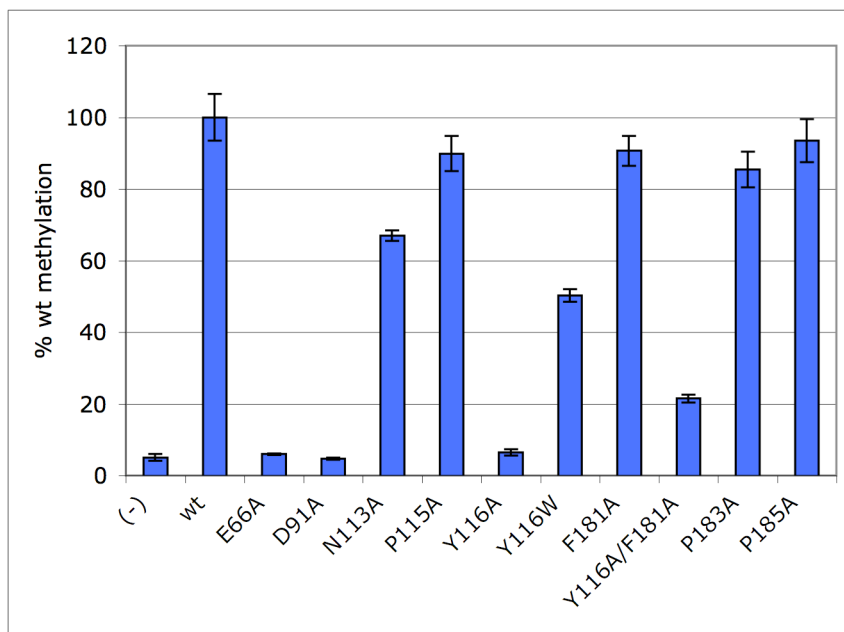
Figure 30. Activity of active site mutants. (a) *In vivo* activity as measured by kasugamycin resistance/sensitivity. (b) *In vitro* activity normalized to wild-type. Experiments were performed in triplicate; error bars represent standard deviation.

Table 7. Phenotypes produced by partial methylation of A1518 and A1519.

1518	1519	Phenotype[a]
**	**	S
-	**	S
-	-	R
**	-	R
*	-	nd
-	*	nd
*	*	nd
*	**	nd
**	*	nd

[a]S=ksg sensitive; R=ksg resistant; nd=not determined.

but is not a quantitative measure of the level of methylation produced by mutant KsgA

proteins. Therefore, *in vitro* analysis was used to assess methylating activity of the KsgA

mutants. Results were expressed as percent of wild-type methylation (Figure 30b). E66A is

severely impaired, showing almost no methylation after one hour. This is in good

agreement with the *in vivo* data showing almost complete loss of activity. D91A and

Y116A show a similar loss of function *in vitro* despite having some activity *in vivo*. The

Y116W mutant retains some function, with methyl transfer approximately 50% of wild-

type. Interestingly, the Y116A/F181A double mutant shows significantly higher activity

than the Y116A mutant, although still only approximately 25% of wild-type. N113A is

about 65% as active as wild-type KsgA, and P185A retains about 85% activity. P115A,

F181A and P185A show high activity *in vitro* which is not significantly different than

wild-type, and which agrees with their high activity *in vivo*.

These results are somewhat surprising in view of mutagenesis studies on related

enzymes such as ErmC' and M.TaqI. Most surprising are the results of the F181A

mutation. The equivalent mutation was highly deleterious in M.TaqI[144], and was only

poorly active in ErmC'[143]. Conversely, the F181A mutation in KsgA had activity

comparable to wild-type both *in vivo* and *in vitro*. This residue, along with Y116, is

postulated to be important for stabilization of the target base via stacking interactions;

clearly, any such interaction involving F181 does not contribute significantly to KsgA

activity. Also surprising was the *in vitro* activity of the Y116A/F181A double mutant.

Although this mutant was only 25% as active as wild-type, it is difficult to understand how

it could have any activity at all, especially given that the Y116A single mutant was

inactive *in vitro*. Finally, the activity of the N113A mutant was unexpected based upon comparisons with other enzymes. N113A showed wild-type activity *in vivo* and only a moderate loss of activity *in vitro*. The equivalent mutations in DNA methyltransferases generally results in significant loss of function[139-142]; in ErmC', the N101A mutation was inactive *in vivo* and barely active *in vitro*[143]. This residue is thought to be important for catalysis, but it seems to be mostly dispensable for KsgA activity.

Conclusions

KsgA is the only rRNA post-transcriptional modification enzyme that appears to be universally conserved in all three domains of life, a conservation that extends into mitochondria and chloroplasts[77-80, 86]. Here we report the 2.1 Å crystal structure of KsgA. This structure has helped direct mutational analysis and will aid in the probing of interactions between the enzyme and its substrate and target molecules. Extensive sequence alignment of KsgA orthologs from a wide range of organisms has identified residues that are highly conserved among species. Inclusion of eukaryotic orthologs such as Dim1 and h-mtTFB, which perform additional functions, emphasizes these conserved regions but also highlights variations. Alignments with related homologs, such as ErmC', identify both residues that are conserved between enzyme families and those that have diverged. The former group includes residues that are important for the common function of adenosine dimethylation, including the canonical SAM-binding site, while the latter may help identify residues that confer target specificity and diversity of function between enzymes.

Mutational analysis of KsgA reveals a surprising tolerance for mutation, even of highly conserved residues. The residues that were most important for enzyme activity are important for SAM binding (E66 and D91) and target base stabilization (Y116). Mutagenesis data also exist for the eukaryotic orthologs Dim1 and h-mtTFB[85, 86], and the KsgA structure will add to our understanding of these mutations and their consequences for enzyme function.

The three-dimensional structure of KsgA bears strong resemblance to those of ErmC', hDim1, and sc-mtTFB. All four proteins share a similar architecture and topology. All four have a region of high positive charge in the cleft between the N- and C-terminal domains; this region has been implicated in binding of ErmC' to its substrate RNA. This bolsters evidence that sc-mtTFB may bind RNA[106]. KsgA, hDim1 and ErmC' have well-formed, highly negatively charged pockets in the N-terminal domain which correspond to residues known to be involved in SAM binding. This pocket, and the associated negative charge, are absent in the sc-mtTFB three-dimensional structure, and this enzyme probably cannot bind SAM. Notably, *S. cerevisiae* mitochondrial ribosomes are not dimethylated on the analogous adenosines[124]. Sequence inserts characteristic of the Dim1 family and of the mtTFB family are seen in the hDim1 and sc-mtTFB structures relative to the KsgA structure. These inserts may prove fruitful for future studies of the rRNA processing and transcription factor functions that have been acquired by these protein lineages.

This similarity of the SAM-binding pockets of KsgA and ErmC' might have implications in the design of Erm inhibitors. Antibiotic resistance mediated by Erm enzymes has stimulated a search for compounds that can inhibit Erm activity and thus

restore sensitivity to resistant bacteria[156-158]. Many of these candidate inhibitors act by binding to the SAM binding site of the enzymes[157, 158]. Specificity has been tested by doing parallel inhibition experiments with other methyltransferases, such as catechol-O-methyltransferase (COMT), a eukaryotic enzyme, and EcoRI, a prokaryotic methyltransferase[156, 157]. However, despite sharing the same methyltransferase fold, these enzymes are not very closely related to the Erm enzymes. COMT is a small molecule methyltransferase and EcoRI is a DNA methyltransferase. KsgA and Dim1 are much more closely related to the Erms, and KsgA and ErmC' may share a nearly identical SAM binding pocket. This raises the possibility that Erm inhibitors that block the SAM-binding site could also interfere with KsgA function and perhaps with Dim1 function. The latter possibility could be a potential mode of toxicity for Erm inhibitors which act by blocking the SAM binding site.

Despite the high degree of structural similarity between KsgA and ErmC', KsgA has unique features which distinguish it from the Erm methyltransferases. While the Erm methyltransferases mono- or dimethylate a single adenosine in 23S rRNA, KsgA dimethylates two adenosines in 16S rRNA. It is clear from the structure that the active site pocket is not large enough to accommodate two adenosines simultaneously. This raises questions about KsgA's binding to its substrate RNA. Binding may occur only once, in such a way as to allow both nucleotides to flip into the active site successively. Alternatively, KsgA may bind in two different places, either in separate binding events or by moving along the helix, allowing each adenosine separate access to the active site. A related question involves the order of methylation, if any.

KsgA is subject to a complex regulation which is absent in the Erm enzymes. ErmC' can methylate RNA substrates as small as 32 nucleotides[97]. KsgA can bind to 16S rRNA[46], and to a small fragment of 16S which contains the target adenosines[159], but does not methylate naked 16S[44, 45]. KsgA requires 16S plus a defined subset of ribosomal proteins in order to be enzymatically active[45]. These observations suggest that the enzyme is allosterically regulated in some way, with at least partial assembly of the 30S required to trigger activity. It has been suggested that this trigger involves SAM binding. Poldermans *et al.* reported that SAM binds to KsgA so weakly that it is neither retained by the enzyme on a nitrocellulose filter nor does it move with the enzyme on a Sephadex G-25 column[47]. They suggested that SAM will only bind to KsgA upon interaction of the enzyme with fully or partially formed 30S subunits. The crystal structure of KsgA provides some support for this possibility. The asymmetric unit consists of two monomers in slightly different conformations. One of these, the B chain, seems unlikely to bind SAM due to steric hindrance by P115 and compression of the binding pocket. Both conformations contain a recognizable SAM-binding pocket, but the pocket is compressed in the B chain monomer in such a way that the remaining space is not large enough to accommodate a SAM molecule. If both conformations are biologically relevant then it is possible that the enzyme exists in the B form until binding of a competent substrate, at which point a conformational change to the A form is triggered that will allow SAM to enter its binding pocket.

Understanding of KsgA's regulation may also lend insight into the temporal disconnect between the two functions of Dim1. Dim1 is involved in early cleavage steps of

pre-rRNA, but does not dimethylate its target adenosines until late in the ribosome maturation process[83]. If Dim1 is regulated in a manner analogous to KsgA, this could explain how it performs its two functions at widely separated points in ribosome biogenesis, binding to the pre-rRNA to ensure proper cleavage but remaining enzymatically inactive until a subset of ribosomal proteins have assembled onto the maturing 40S subunit.

Experimental

Crystallization

Crystals were obtained by the hanging-drop vapor-diffusion method using a variety of conditions. Initial screening was done using Crystal Screens 1 and 2 from Hampton Research (Laguna Niguel, CA, USA). While none of these conditions produced crystals, we noted a strong solubility dependence on the presence of calcium acetate or magnesium acetate. This prompted us to investigate other divalent salts in a screen which closely resembled the Hampton Research PEG/Ion Screen. Needle showers were obtained with 0.2M calcium acetate and 20% w/v PEG 4000. Conditions were refined by optimizing the amount of PEG 4000 with a constant concentration of 0.2M calcium acetate. Generally, crystals took 3-4 four days to grow at room temperature, but on occasion crystal growth required as long as week. Some crystals grew singly, but most often they grew as clusters of two to three crystals, which could be separated to yield single crystals. A single heavy atom derivative was obtained by adding an equal volume of the precipitant solution saturated with potassium uranyl fluoride to a crystallization drop containing KsgA crystals. Crystals were soaked in this solution for three hours prior to data collection.

A single poor quality crystal was also obtained with 0.2M magnesium acetate and 20% w/v PEG 4000, and needle showers occurred with 30% or 40% ammonium sulfate and 0.1M NaPO$_4$, pH 8.0. These conditions were not optimized as we had already obtained several high quality crystals using PEG 4000 and calcium acetate.

Crystallographic analysis and structure determination

Crystals were cryoprotected by gradually increasing the amount of PEG 4000 to a final concentration of about 30% (w/v). Crystals were mounted in a nylon loop and cooled to 98 K in a stream of N$_2$ gas. Diffraction sets on the native and uranyl soaked crystals, covering 230° and 400° respectively, were collected as 2° oscillation frames on an R-AXIS IIC image plate detector using CuK$_\alpha$ radiation from a rotating anode generator operating at 50 kV and 100 mA with Osmic confocal optics. Data integration, merging and scaling were performed with the HKL suite[160].

Initial SIRAS phases were calculated from data obtained on the native and potassium uranyl fluoride crystals. Heavy atom positions were refined and initial phases were determined using SHARP[161] followed by density modification using Solomon[162] as implemented in the SHARP suite. An initial model consisting of approximately 80% of the backbone was automatically traced with the ARP/wARP program[163]. The initial model was refined using several cycles of manual rebuilding into SIGMAA weighted 2mF$_o$-dF$_c$ and/or composite omit maps[164] with Xfit[165] followed by simulated annealing in CNS[166]. In the final cycles, simulated annealing was followed by TLS refinement[167] and geometry restrained conjugate gradient maximum likelihood refinement in REFMAC version 5.1.24[168]. For each monomer, two TLS groups were defined consisting of residues 17-195

and 203-268, corresponding to the N-terminal and C-terminal domains. All refinements

were carried out without NCS restraints since attempts to introduce NCS restraints

between the two monomers resulted in significant increases in both R_{work} and R_{free}. Model

quality was monitored via the use of PROCHECK[169] and WHATCHECK[170]. Local

structural errors were identified with the aid of OOPS2[171] coupled with O[172] prior to

manual rebuilding with Xfit. No Ca^{2+} was found in the structure.

Charged surfaces for hDim1 (1ZQ9), ErmC' (1QAM), sc-mtTFB (1I4W) and KsgA

were calculated using the DELPHI module of the InsightII package (Accelrys, Inc., San

Diego, CA). In these calculations a protein formal charge set was used in which only Lys,

Arg, Asp and Glu residues were assumed to be charged. Volumes of active site cavities for

ErmC' and KsgA were calculated using the CASTp server (cast.engri.uic.edu). Active site

cavities identified by the CASTp server were visualized as molecular surfaces with the aid

of the Access program from the Ribbons package[173].

Mutagenesis

Primers were constructed which contained the mutations of interest flanked by 15-

20 bases of wild-type sequence (Table 8). Mutagenesis was performed using the

Stratagene QuikChange Kit according to the manufacturer's protocol. Briefly, wild-type

plasmid was denatured and the primers were annealed and extended by thermal cycling

with *PfuTurbo* DNA Polymerase (Stratagene). DpnI (Stratagene) was used to digest the

parental (non-mutant) DNA, which is naturally *dam* methylated. The mutated DNA is not

dam methylated, and thus is not susceptible to digestion by DpnI. The resulting plasmids

were transformed into *E. coli* XL1-Blue cells and mutants were confirmed by sequencing.

Table 8. Primers for active site mutants

Mutant	Primer
E66A	5'-CTG ACG GTC ATC GCA CTT GAC CGC GAT C-3'
D91A	5'- GAC GAT TTA TCA GCA GGC GGC GAT GAC CTT TAA CTT TGG-3'
N113A	5'- GCG TGT TTT CGG CGC CCT GCC TTA TAA CAT CTC C -3'
P115A	5'- TCG GCA ACC TGG CGT ATA ACA TCT CCA CGC CGT TG -3'
Y116A	5'- TTC GGC AAC CTG CCT GCG AAC ATC TCC ACG CCG -3'
Y116W	5'- TCG GCA ACC TGC CTT GGA ACA TCT CCA CGC C -3'
F181A	5'- GTA CCG CCG TCA GCC GCG ACA CCA CCA CCC -3'
F181W	5'- CCG CCG TCA GCC TGG ACA CCA CCA CCC -3'
P183A	5'- CGT CAG CCT TTA CAG CAC CAC CCA AAG TGG -3'
P185A	5'- CCT TTA CAC CAC CAG CCA AAG TGG ATT CCG C -3'

Protein expression and purification

Proteins were expressed and purified as in Chapter 2.

Activity assays

In vivo activity assays were performed as in Chapter 2. *In vitro* assays were performed as in Chapter 2, except that reactions were performed in a volume of 50μl and the entire reaction was stopped at sixty minutes. Reactions contained 10 pmol protein.

Generation of molecular figures

The modeling software package Ribbons[173] was used in the construction of Figures 17, 18a, and 25-27, which were rendered using the software package Povray (www.povray.org). The molecular visualization program PyMOL (www.pymol.org) was used in the construction of Figures 19-21, 24, and 28.

KsgA/30S Interaction

KsgA has a complex interaction with 30S subunits. The protein can bind to free 16S, but is not enzymatically active unless the 16S has bound at least a subset of ribosomal proteins; however, the basis for this regulation remains unknown. To gain a better understanding of KsgA's activity, we needed to study its binding to the 30S substrate. Since KsgA can bind to 16S rRNA, and since there are no ribosomal proteins in the immediate area of helix 45, we can presume that KsgA's binding to 30S is determined solely by the RNA. van Gemen *et al.* showed that KsgA binds to the colicin fragment of 16S rRNA[159], which consists of the 49 3'-terminal nucleotides; this fragment includes helix 45 and flanking sequences. Based on this data, the common belief was that KsgA binding is localized to the stem of helix 45. This assumption was strengthened by experiments with the closely related ErmC' homolog (see Figure 27). However, upon further consideration, this binding model is inadequate. First, there is the problem of helix 45 accessibility. In the 30S structure, the loop of helix 45 is tucked into the base of helix 44, and the stem of helix 45 is not exposed (Figure 31). Although KsgA can methylate a pre-30S particle, and thus may normally act at a point during subunit maturation at which helix 45 is not buried, KsgA is able to methylate 30S subunits, and must have access to its binding site in this context. Also, if KsgA is bound to helix 45 as shown in Figure 27, it is difficult to understand its ability to dimethylate both adenosines in the loop of the helix. If
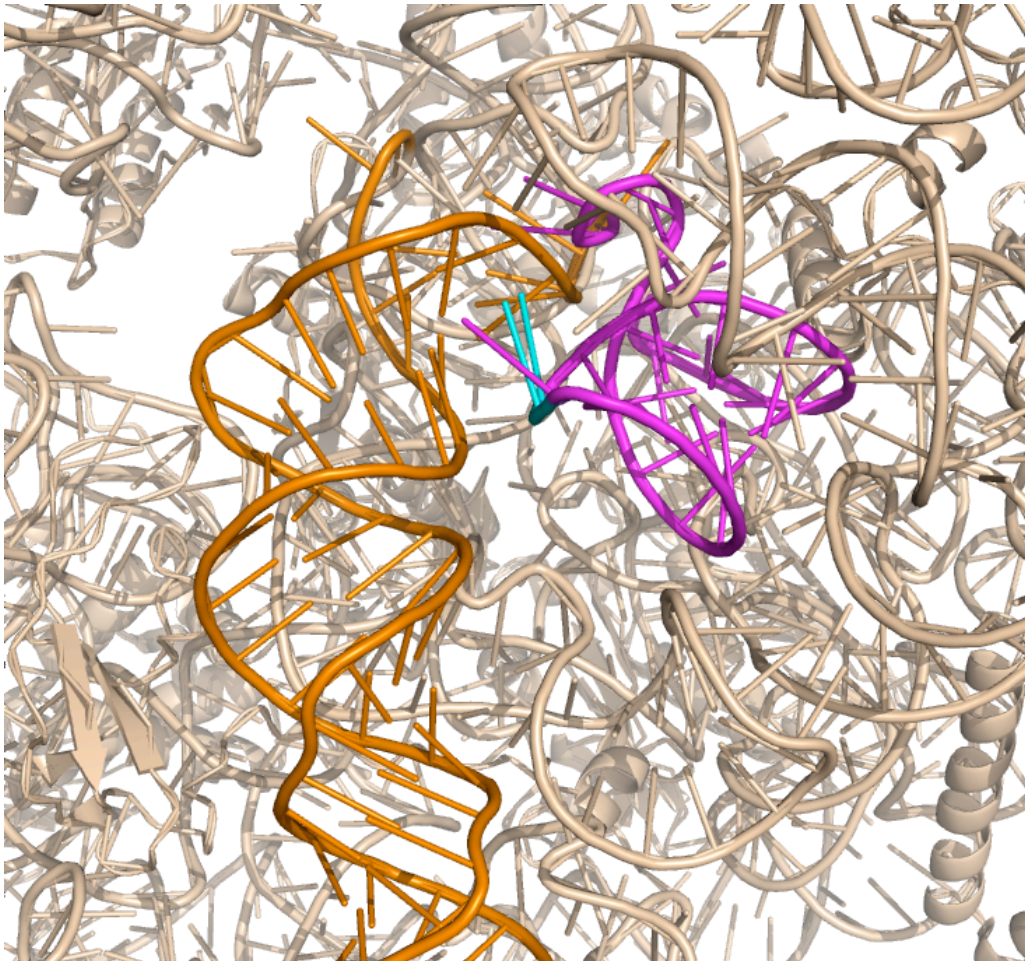
Figure 31. Helix 45 inaccessibility. Helix 44 is shown in orange, helix 45 in magenta, and A1518 and A1519 are cyan.

binding were optimized for access of one adenosine into the binding pocket, then the other adenosine would seem less likely to have access to the pocket. Two equally unsatisfactory alternatives are that binding is such that both adenosines can reach the pocket, although probably not in an optimal way, or that there are two separate binding interactions on helix 45, so that each adenosine can be methylated.

In order to get a clearer picture of KsgA's interaction with 30S, we used a powerful technique called directed hydroxyl radical probing in combination with primer extension[174, 175]. This approach has been used successfully to model interaction of the 30S subunit with ribosomal proteins[174, 176-181], with elongation factor[182], with release factor[183], with initiation factor[184], and with ribosome recycling factor[185]. Directed hydroxyl radical probing of 16S rRNA has allowed us to model the interaction of KsgA with the 30S subunit. This model sheds light on KsgA's substrate requirements, and may also help explain the temporal separation of Dim1's two functions in eukaryotic ribosome biogenesis.

Hydroxyl radical probing

In order to probe the RNA with hydroxyl radicals, an Fe(II) is first attached to a single cysteine residue in the protein. KsgA naturally contains two cysteine residues, at positions 168 and 258. We mutated C258 to alanine, leaving the protein with a single cysteine at residue 168 (C168). We then mutated C168 to alanine to produce a mutant with no cysteine residues (Cys-less), which was used as a control; we used the double mutant to construct KsgA proteins with a single cysteine residue at positions 154, 182, 219 and 231 (C154, C182, C219, C231). These residues were chosen based on three criteria. First, they

are all surface-exposed and accessible to solvent. Second, none of these residues are highly

conserved, so we didn't expect mutations at these positions to affect protein activity. Third,

we wanted residues that were widely spaced in the protein's structure, so that we could

maximize the amount of non-redundant information gained. All of these mutants except

C219 were fully active in our *in vivo* activity assay (Figure 32); we discontinued work on

the N219C mutant. Figure 33 shows the four single-cysteine KsgA mutants C154, C168,

C182, and C231. Mutant proteins were shown to be soluble and purified to homogeneity as

in Chapter 2.

Fe(II) derivatization, hydroxyl radical probing, and primer extension were

performed by Zhili Xu, under the direction of Dr. Gloria Culver. The Fe(II) moiety was

tethered to each cysteine residue via a flexible linker, 1-(*p*-bromoacetamidobenzyl)-EDTA

(BABE) and the derivatized proteins were incubated with 30S subunits to allow complex

formation (Figure 34). Upon addition of H_2O_2, and in the presence of ascorbic acid,

hydroxyl radicals were generated in the area of the Fe(II) by Fenton chemistry. These

radicals cleave the surrounding RNA backbone; since the radicals are short-lived, the

cleavage is localized to areas of the RNA in close proximity to the Fe(II), and therefore to

the cysteine residue. After cleavage the RNA was extracted from the 30S subunits,

annealed with labeled primers, and subjected to primer extension reactions (Figure 35).

The extension products were separated on a gel, and cleavage sites calculated based on the

length of the extension products. Figure 36 depicts representative cleavage patterns

produced by each cysteine mutant. Bands on the gel correspond to extension stops;

comparison with uncleaved RNA (Cys-less) differentiates stops due to cleavage events
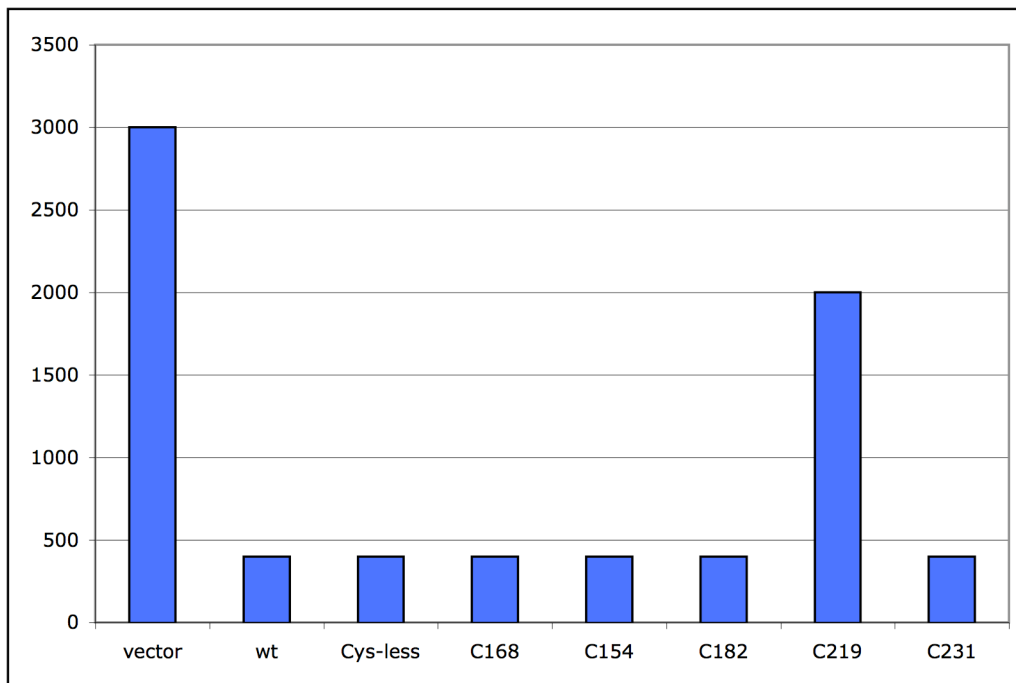
Figure 32. *In vivo* activity assay of cysteine mutants.
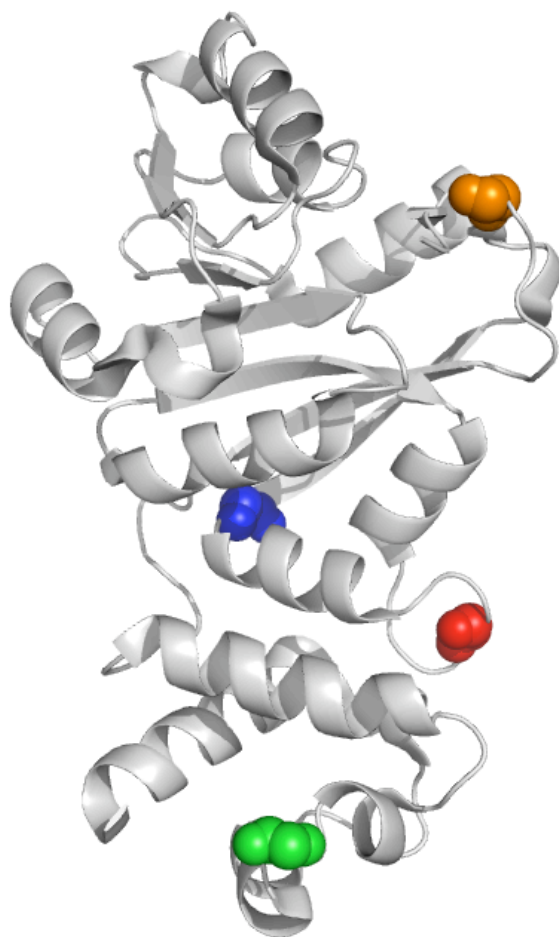
Figure 33. Cysteine mutants. Backbone atoms of residues which were mutated to cysteine are shown as spheres. C154 is red, C168 is blue, C182 is orange, and C231 is green.
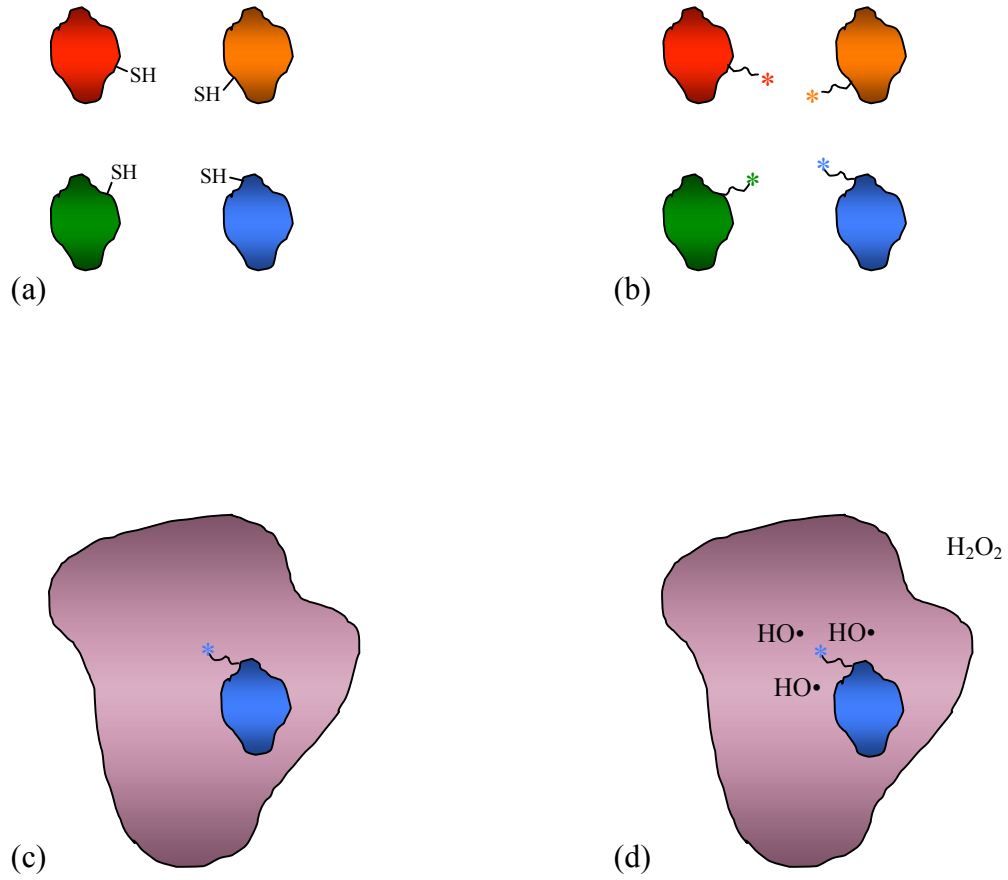
Figure 34. Hydroxyl radical probing. Proteins containing a singles cysteine residue are constructed (a), and Fe(II) is attached to these proteins via a flexible linker (b). The proteins are incubated with 30S (c) to allow complex formation. Hydrogen peroxide is added and generates hydroxyl radicals, which cleave the rRNA backbone.

(a)

(b)
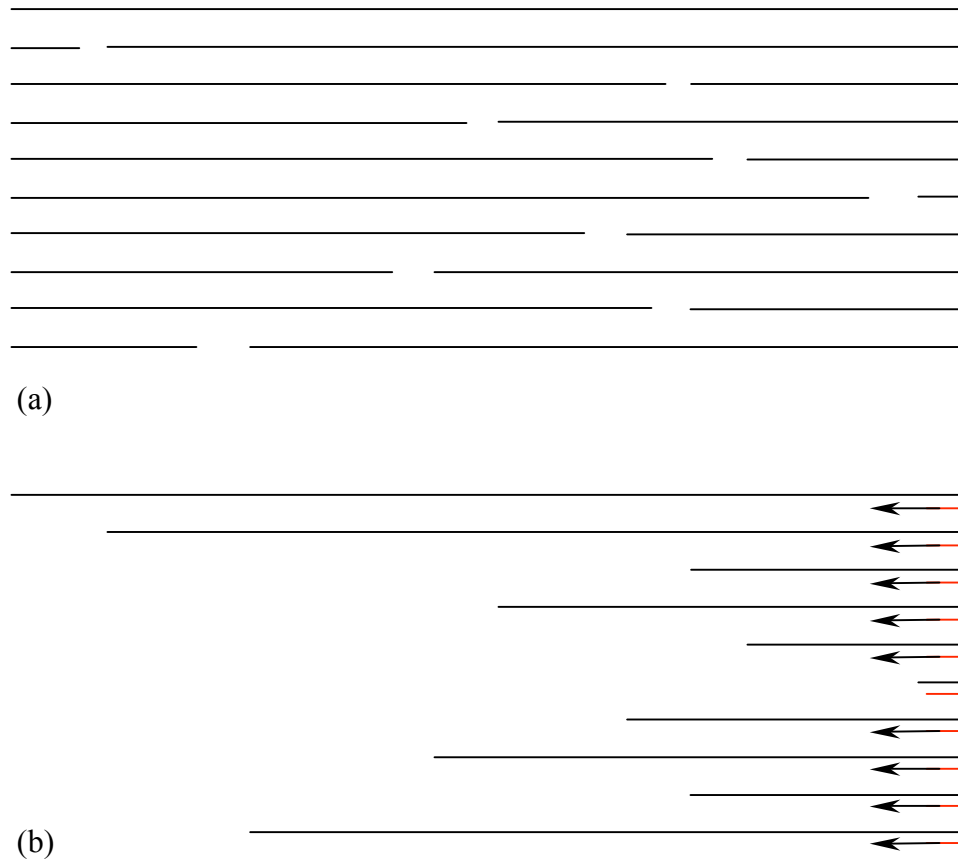
Figure 35. Primer extension. Cleaved RNA (a) is annealed to labeled primers (b). The primers are extended using free nucleotides and RNA polymerase. Extension products are separated by gel electrophoresis as in Figure 36.
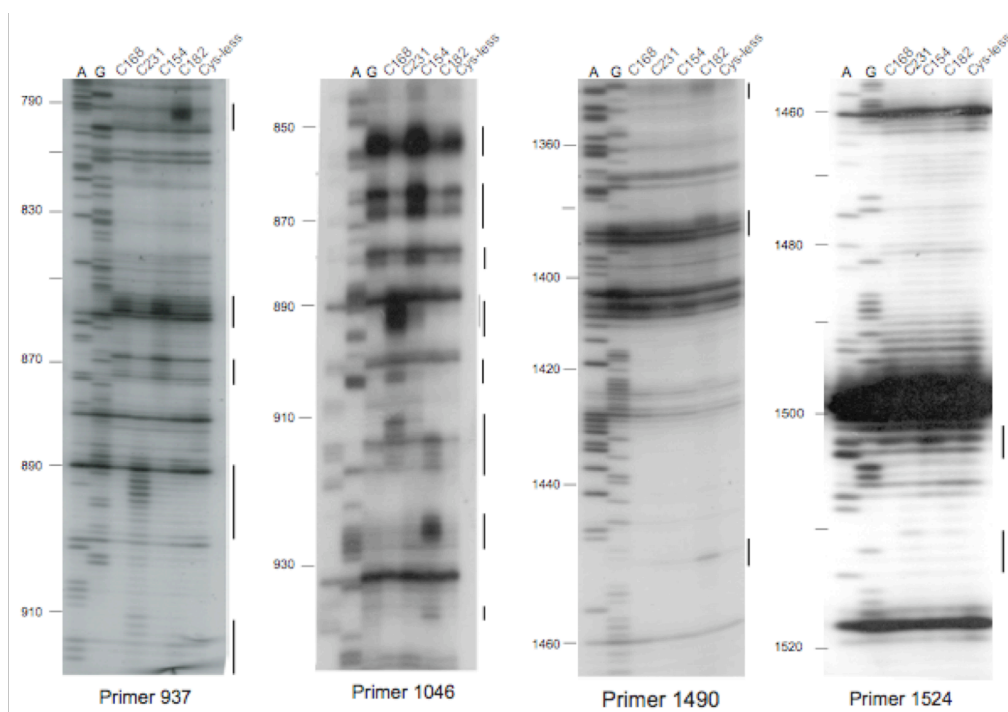
Figure 36. Representative primer extension. Each of the four gels represents an extension reaction using a different primer.

from those resulting from secondary structure. Portions of the RNA backbone which are in close proximity to the Fe(II) will be cleaved more often than positions which are further away; more cleavage events occurring at a particular site correspond to a heavier band on the gel.

Binding model

The first step in building our model was to visualize the cleavage sites within the context of the 30S subunit. Figure 37 graphically depicts the cleavage patterns produced by the four Fe(II) positions (attached to residues 154, 168, 182 and 231). Larger spheres represent stronger primer extension stops, indicating closer proximity to the Fe(II). The colors correspond to the positions of the four single cysteines (Figure 33). Using these cleavage patterns, we manually docked KsgA onto the 30S subunit. Docking was done in such a way as to position each cysteine in proximity to its resultant cleavages. The model was built to maximize favorable intermolecular contacts between KsgA and the 30S. Consideration was also given to placing the target adenosine binding pocket in the closest proximity to A1518 and A1519 possible while maintaining a minimum of steric clash. In an effort to achieve the best possible interface between KsgA and 30S, the two domains of KsgA were moved independently: the entire KsgA protein was docked optimizing the interface of the N-terminal domain, then the C-terminal domain was moved to refine its interface with 30S. Overall, this dislocation of the C-terminal domain relative to the N-terminal domain was modest and well within the range of motion known to be possible for KsgA[131]. The final binding model is shown in Figure 38. The model demonstrates a mode of binding which does not rely on helix 45; rather, KsgA binding to the 30S subunit is
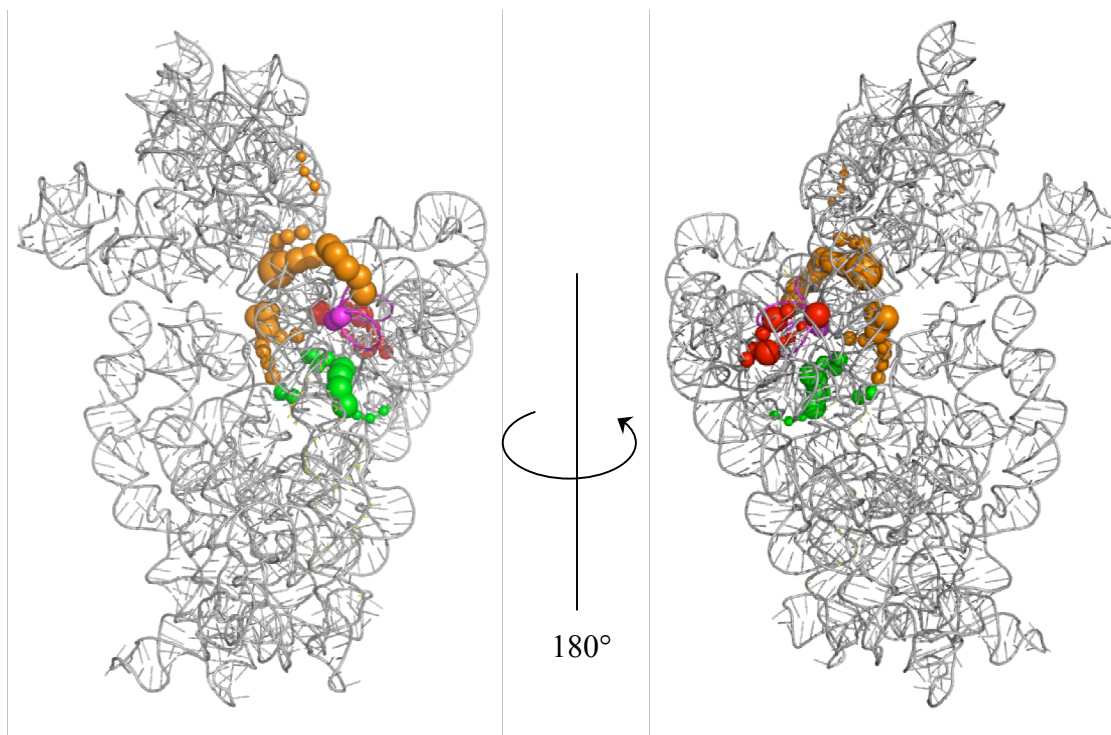
Figure 37. Cleavage patterns produced by the four cysteine mutants. Colors correspond to Figure 33. Helix 45 is represented in magenta, with the target adenosines shown as magenta spheres. Blue spheres corresponding to cleavage by C68 are hidden by the orange spheres.
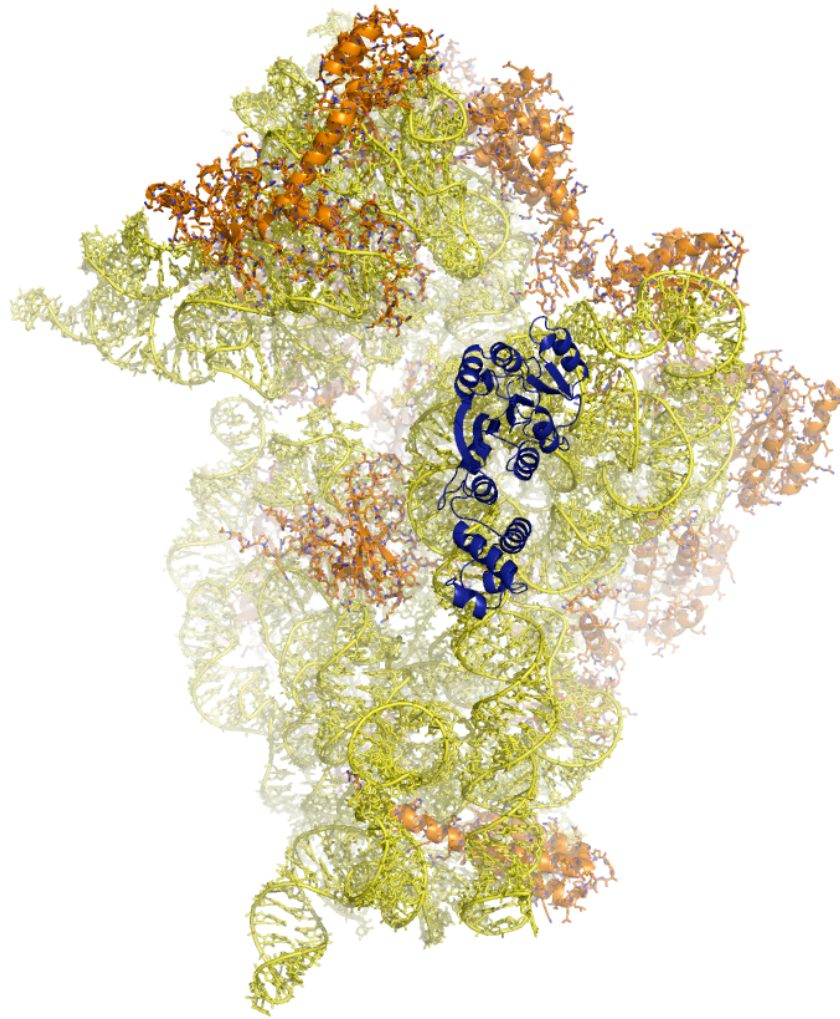
Figure 38. Model of KsgA binding to 30S. 16S rRNA is shown in yellow; ribosomal proteins are orange; KsgA is blue.

chiefly determined by interaction with helix 44. The cleavage data virtually exclude helix 45 as a binding site for KsgA.

A rotated, close-up view of KsgA, helix 45 and the proximal portion of helix 44 are shown in Figure 39. In our model, a long edge of KsgA is aligned with the long axis of helix 44 of 16S rRNA and encompasses a footprint extending from the G1497-C1404 pairing to the G1482/A1418 pairing. There is good shape complementarity between ksgA and helix 44 with alternating minor, major, and minor groove interactions.

The model also shows interactions with the loop of helix 45 and the loop of helix 24a (the so-called 790 loop) of the central domain (Figure 40). In neither case is the observed interaction satisfying. In the first case, the helix 45 loop is buried into the minor groove of helix 44 in such a way that will undoubtedly require a significant rearrangement before A1518 and A1519 are brought close enough to the active site of KsgA to support methylation. In the second case, the backbone region of A790 and G791 sterically clashes with amino acids 182, 183, and 184 of KsgA.

The two above problems highlight the most significant shortcoming of our model. Namely, the 30S structure used in our docking is from the high Mg^{2+}, translationally active conformation, whereas KsgA only operates on the low Mg^{2+}, translationally inactive conformation[48]. At present the latter conformational state has not been determined crystallographically. However, chemical probing of the two states was done. In that study, the authors pinpointed the residues that undergo conformation change in going from the high Mg^{2+} state to the low Mg^{2+} state. Those changes are predominately located within the decoding region of 30S, mapping to helix 28 and the P-site[52]. Other nearby locations

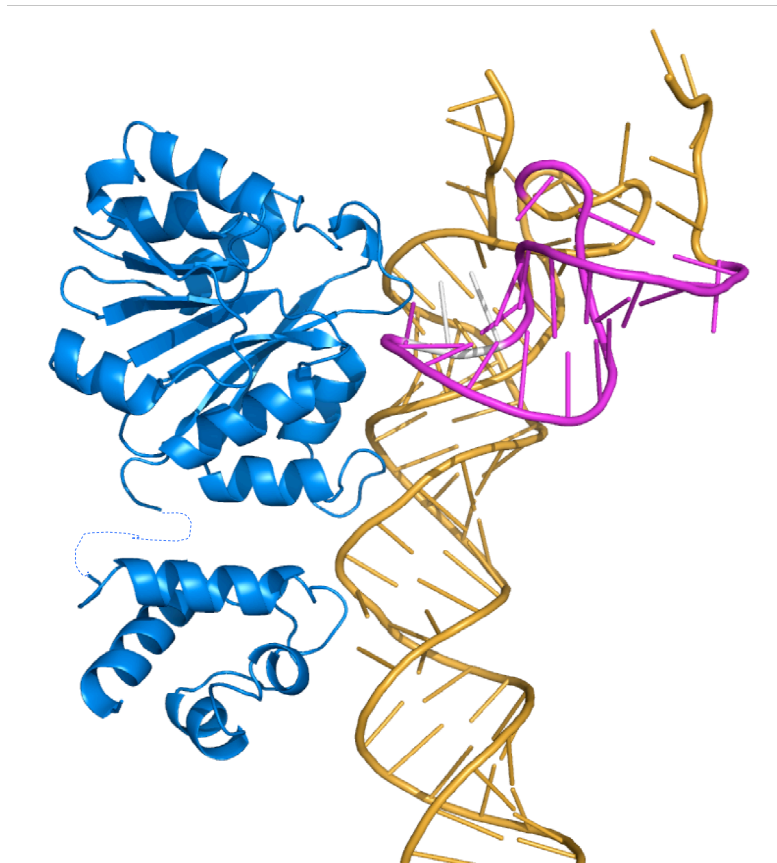Figure 39. Interaction of KsgA (blue) with helix 44 (orange). Helix 45 is shown in magenta; the target adenosines are white. The dashed line in KsgA indicates the break between the two domains required for model building (see text).
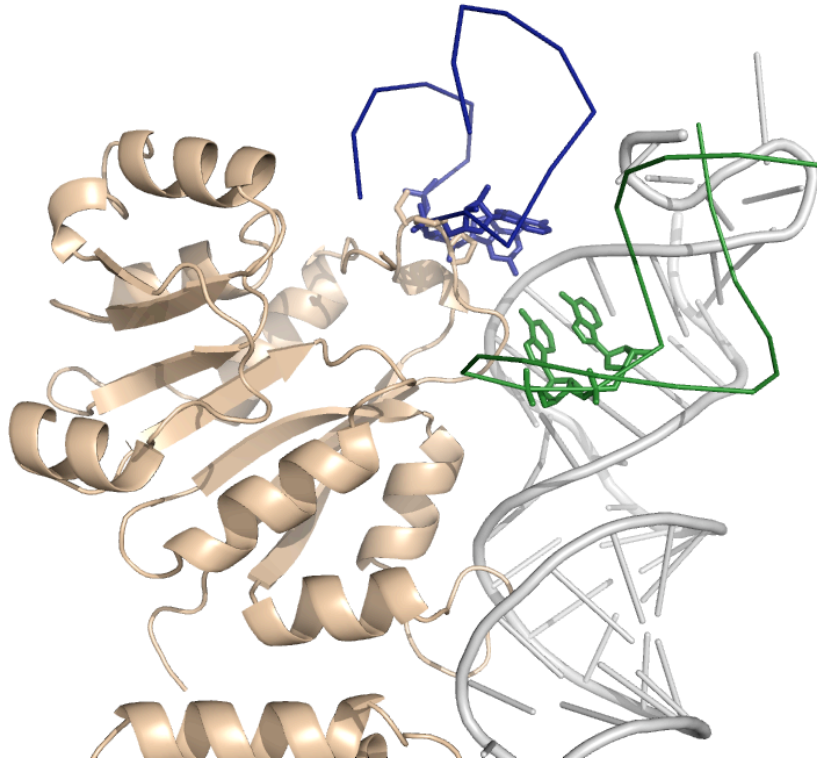
Figure 40. Interaction of KsgA with helix 24a (blue) and helix 45 (green). Nucleotides 792-793 and 1518-1519 are highlighted. Helix 45 is shown in gray.

include the interface of the 790 loop and helix 45 loop with the A-site region of helix 44, and the interface between helix 27 and helix 44. Movement of the 790 loop and the helix 45 loop could relieve the clash present between KsgA and A790 and G791 and allow for the required disposition of A1518 and A1519 close to the active site of KsgA. Indeed, A1518 and A1519 are known to be solvent exposed in the low Mg^{2+} state, but buried in the high Mg^{2+} state[53], further supporting the idea that the loop of helix 45 is capable of significant conformational flexibility.

Figure 41 highlights regions in KsgA that fit into the major and minor grooves of the helix. Residues 186 and 187 interact with the minor groove in close proximity to helix 45. These residues are part of the very highly conserved Motif VIII, which helps form the adenosine binding pocket in other methyltransferases. In KsgA these residues are lysine and valine. The valine is conserved in all KsgA and Dim1 enzymes, and is also conserved in mtTFB1 enzymes. The lysine is conserved as a lysine or arginine in KsgA/Dim1, although not absolutely. K142 and E143 interact in the same part of the minor groove, just below K186 and V187. E143 is conserved in KsgA/Dim1/mtTFB1, while K142 is loosely conserved as a positively charged residue. None of these residues are conserved in mtTFB2 or mtTFB proteins. The loop consisting of residues 153-156 interacts with the major groove below helix 45. The presence of this loop is conserved in KsgA, Dim1 and mtTFB enzymes, although there is little sequence conservation in this region. This loop has been lost entirely in the Erm family of enzymes, which have evolved to bind to 23S rRNA instead of 16S. A third interaction is provided by a group of positive residues in the C-terminal domain of KsgA, which interact with the minor groove below the 153-156 loop.
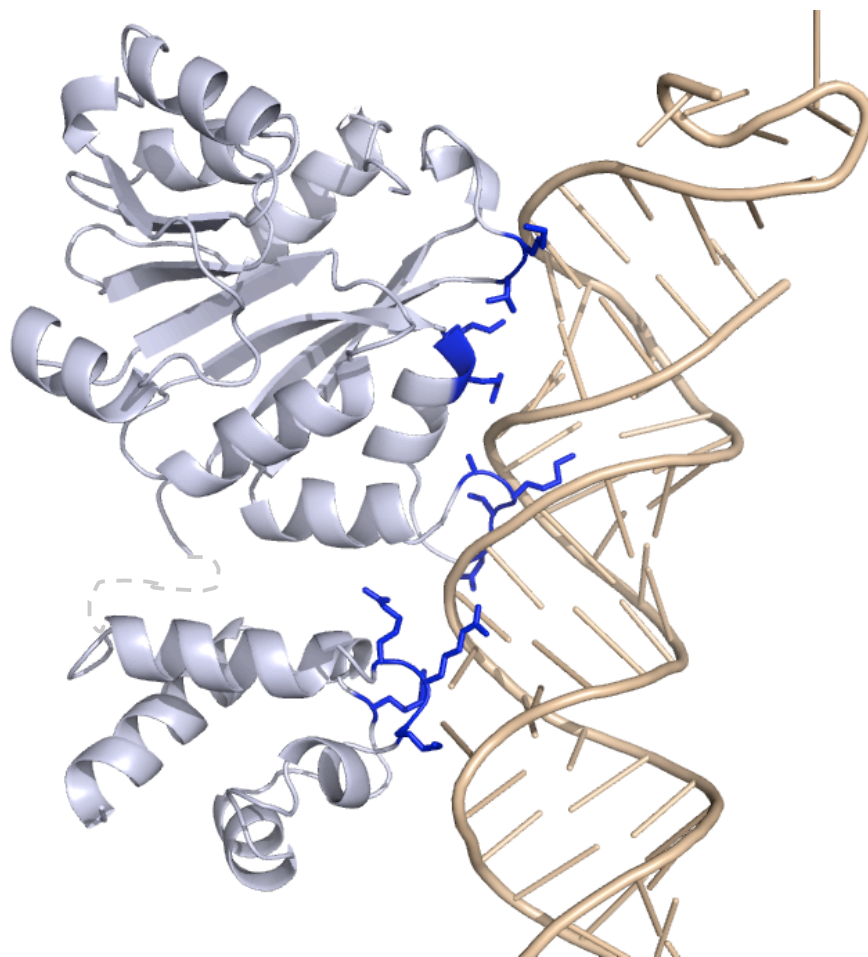
Figure 41. Interactions between KsgA and the major and minor grooves of helix 44. Labels refer to residues discussed in the text.

This group consists of R221, R222, K223 and R248. The presence of this cluster of lysine and arginine residues is preserved in KsgA/Dim1, even if the exact sequence of the residues is not conserved.

Since the KsgA/30S binding interaction is conserved among disparate organisms, we would expect high conservation of the amino acid and nucleotide residues involved in binding, and any model of the interaction should display this conservation. Figure 42 shows that the patterns of conservation in KsgA and helix 44 agree with our model. The most conserved region of helix 44 is the base of the helix, where it interacts with KsgA. The more distal portion of the helix is not very well conserved. Likewise, the most highly conserved residues in KsgA are found in the active site and along the face of the protein which interacts with 16S, and the least well conserved residues are on the side opposite RNA binding. Further validation of the model was given by the results of a gel-shift experiment. We constructed an RNA oligo corresponding to the region of helix 44 which should interact with KsgA (Figure 43). As shown in Figure 44, KsgA was clearly able to shift this oligo. A final piece of evidence comes from measurements of KsgA binding to 30S subunits. Poldermans *et al.* showed that KsgA binds with similar affinity to both methylated and unmethylated subunits[47]. This demonstrates that the target adenosines are not determinants for KsgA binding. We have begun using Biacore experiments to corroborate this observation, and preliminary data show the two binding affinities to be comparable, with K_D values of 7 nM and 4 nM for methylated and unmethylated subunits, respectively.
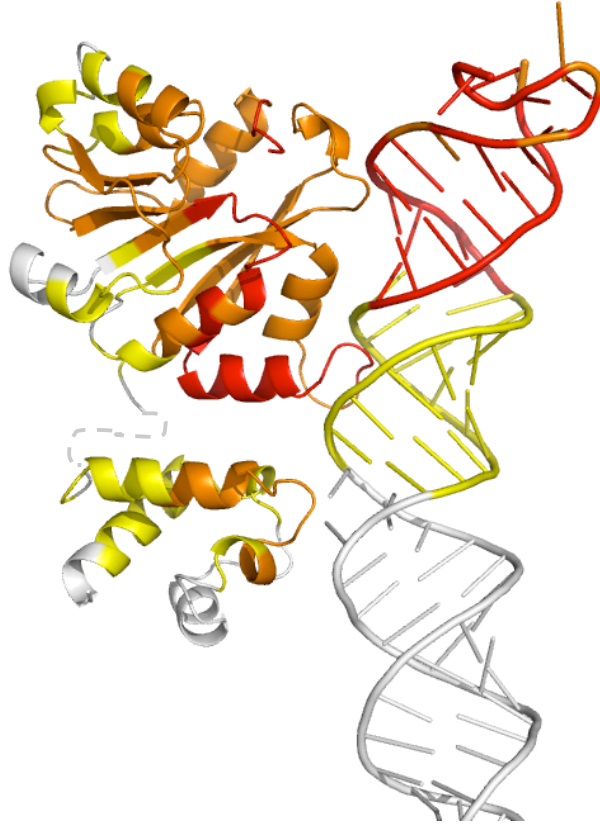
Figure 42. Conservation of residues in KsgA and helix 44. RNA conservation was adapted from the minimal ribosome modeled by Mears *et al*[110]; red denotes sequences which are greater than 98% conserved; orange denotes sequences which are 90-98% conserved; yellow denotes sequences which are less than 90% conserved. White denotes sequences which are present in less than 95% of organisms. KsgA conservation was measured by the Tcoffee program[130] as strength of alignment of thirty-four KsgA, Dim1 and mtTFB1 sequences. Red denotes areas of strong alignment, and therefore high sequence conservation; orange sequences are moderately conserved; yellow sequences are poorly conserved. White denotes areas which are not conserved and thus do not align well.

```
                                \  /
                                 \/
                                G-C
        G-C                     *C•A
        G-C                     C     A
        C-G                     C   U*
        G-C                      C-G
        U   U                    G-C
        C-G                      U   U
            A                    *C-G
        A                            A
            A                    A
        C-G  A                       A
        A-U                      C-G  A
        C-G                      A-U
        C-G                      C-G
        A   G                    C-G
        U•G                      A   G
        G•U                      U•G
        G-C                      G•U
        G   A                    G-C
        A   G                    G   A
        G•U                      A   G
        U-A                      G•U
        A-U                      U-A
        G-C                      G-C
(a)     G   A      (b)           G•U
        A A                      G•U
                                 U-A
                                 U•G

                                 | |
```
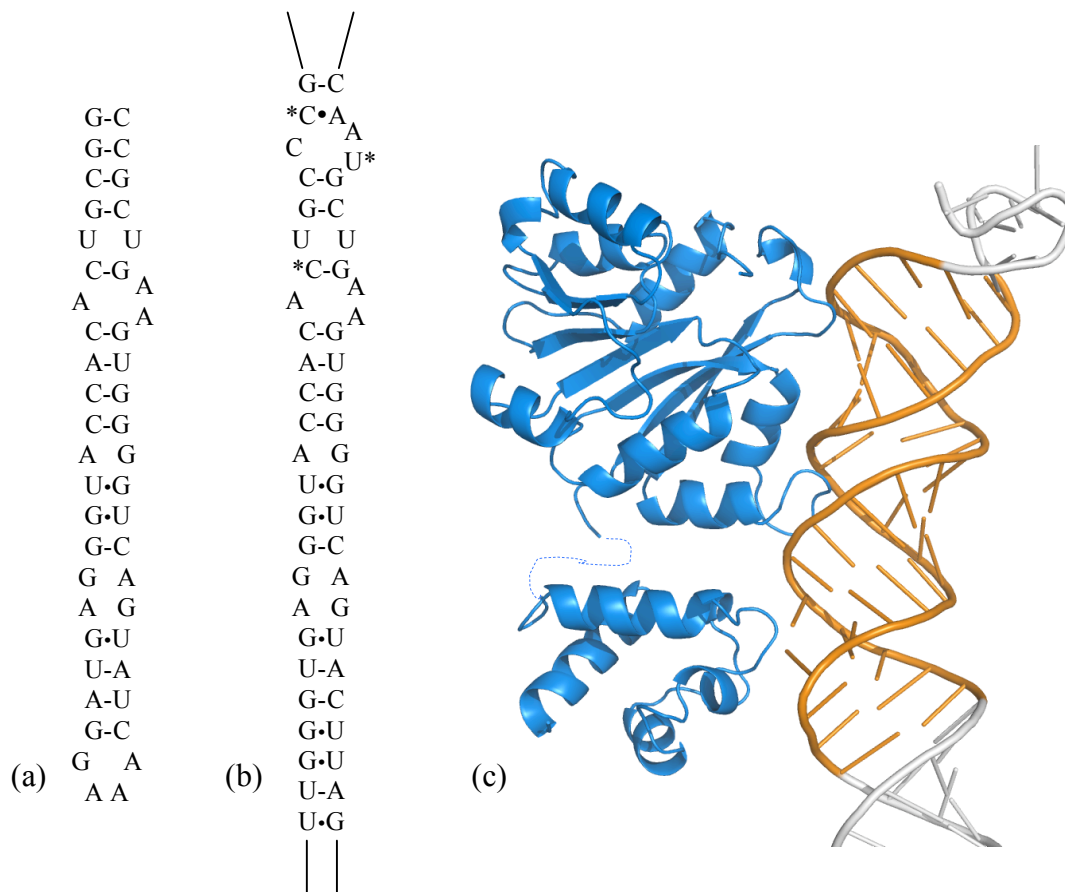
Figure 43. Helix 44 oligo design. (a). h44 oligo. (b) Base of helix 44. Slanted lines indicate where the base of the helix joins the rest of the rRNA; straight indicate the remaining portion of helix 44, which has been left out for clarity. Stars indicates modified nucleotides. (c) Binding model. The portion of helix 44 corresponding to the h44 oligo is shown in orange.
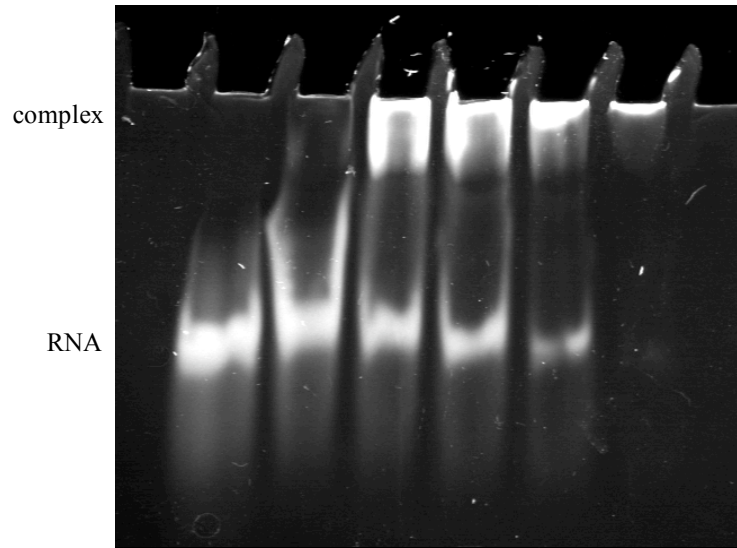
Figure 44. Gel-shift assay with KsgA and h44 oligo. Lane 1 contains h44 with no KsgA; lanes 2-6 contain increasing amounts of KsgA.

Conclusions

Based on gel-shift analysis and directed hydroxyl radical probing, we have

constructed a model of KsgA's binding interaction with 30S, specifically with the base of

helix 44. This interaction is mediated by numerous contacts between KsgA and the major

and minor grooves of the helix. Notably, helix 45 does not contribute significantly to KsgA

binding, and in fact would require a conformational rearrangement in order to bring the

target adenosines into KsgA's active site. We believe that the source of this rearrangement

is the difference between translationally inactive and translationally active subunits. The

model does not explain earlier observations that KsgA binds to the colicin fragment of 16S

rRNA[159], which contains helix 45 but not helix 44. However, it should be noted that this

study found a K_D value of 200nM for KsgA binding to the colicin fragment, which could

be explained by a nonspecific binding interaction.

Our binding model may explain KsgA's requirement for some assembly of the

subunit to occur before methylation, despite the fact that KsgA can bind to free 16S

rRNA[46]. In this scenario, KsgA binds to pre-16S rRNA early in the ribosome assembly

process. However, further maturation, including assembly of proteins S4, S6, S8, S11, and

S15-S18, is required to bring helix 45 into a position where it is accessible from helix 44.

A similar explanation may lie behind the temporal separation of Dim1's two functions in

eukaryotic ribosome biogenesis. Processing of the A_1 and A_2 cleavage sites requires early

binding of Dim1 to the pre-rRNA, while methylation is a very late step in assembly,

probably because of the delay in availability of the target adenosines to Dim1's binding

site.

Experimental

Mutagenesis

Mutagenesis was performed as in Chapter 3. Table 9 lists the primers used.

Protein expression and purification

Proteins were expressed and purified as in Chapter 2.

Activity assays

In vivo activity assays were performed as in Chapter 2.

Hydroxyl radical probing

Hydroxyl radical probing and primer extension were performed by Zhili Xu, under the direction of Dr. Gloria Culver, as described[175].

Model building

The KsgA binding model was built manually using the program PyMOL (www.pymol.org). Pymol was also used in the generation of Figures 31, 33, 37-42, and 43c. Structures used were 1QYR[131] (KsgA) and 1J5E[7] (30S).

Table 9. Primers for cysteine mutants.

Mutant	Primer
C168A	5'- GTC ATG GCG CAA TAC TAT GCC AAT GTG ATC CCG GTA CTG G -3'
C258A	5'- CTC TGT CGC GCA ATA TGC CCA GAT GGC GAA CTA TCT GG -3'
S154C	5'- GCA GGA CCG AAC TGC AAA GCG TAT GGT CG -3'
T182C	5'- CCG CCG TCA GCC TTT TGC CCA CCA CCC AAA GTG -3'
N219C	5'- CCA CCG AAG CCT TTT GCC AGC GTC GTA AAA CAT TCG -3
N231C	5'- CGT AAC AGC CTC GGC TGC CTG TTT AGC GTC G -3'

Ongoing and future work

The KsgA/Dim1 proteins are intimately involved in ribosome biogenesis, a process that is fundamental to cellular life. These proteins play a role important enough to have been carried along throughout all of evolution. We have characterized this family of proteins structurally and biochemically, and have explored their conservation. KsgA proteins from distinct evolutionary kingdoms can function in a bacterial system, demonstrating a common conservation of both the protein and the key structures of the 30S subunit. The next question to ask is whether diverse orthologs can function in a eukaryotic system. Eukaryotic Dim1 enzymes are vital for pre-rRNA processing in addition to their methyltransferase function. The genesis of this second function is unknown; it is possible that bacterial KsgA enzymes have the potential to mediate these pre-RNA cleavage steps, and that this is not a "new" function but is rather an undiscovered role played by all KsgA/Dim1 enzymes in ribosome biogenesis. Our lab is currently exploring this question by two separate approaches. First, we are investigating the ability of the bacterial KsgA and the archaeal Dim1 to complement for ScDim1 knockout in yeast. These experiments will tell us if evolutionarily distinct orthologs are competent for either or both of Dim1's functions in a eukaryotic system. If KsgA can methylate yeast 40S, that will be further evidence of conservation of structural cues recognized by these enzymes. If KsgA can

complement for Dim1's processing function, that would suggest similar processing steps in bacterial rRNA maturation. Bacterial and eukaryotic ribosome biogenesis are thought to follow very different pathways; evidence of similarities in major assembly steps would be very exciting.

Another ongoing project, which we are performing in collaboration with Dr. Gloria Culver's lab, is an in-depth examination of KsgA in bacterial ribosome maturation. Although Leveque *et al*. showed that KsgA is not strictly essential[57], we have hypothesized that KsgA may be required under non-optimal conditions such as low temperature. Dr. Culver is currently performing experiments with a *ksgA⁻* knockout strain, looking for effects of knockout aside from lack of methylation at A1518 and A1519.

Mutation of active site residues had surprisingly little effect on KsgA activity. We are currently performing kinetic experiments on these mutant proteins to uncover subtle effects on activity, and to help explain the defects in activity that we did see with a few residues. The most deleterious mutations were of residues important for SAM binding (E66A and D91A) and for target adenosine binding (Y116A). We predicted that mutation of N113 would have a large effect, based on analysis of related methyltransferases. This residue has been proposed to be important for catalysis, possibly by accepting a proton from the target amine, thereby increasing its nucleophilicity. However, the N113A mutation had only a moderate effect on activity. Further work is needed to elucidate the role of this residue in the methyl transfer, and to understand details of the enzyme's mechanism. The F181A mutation also showed unexpectedly high activity. In other methyltransferases, this residue, along with Y116, is important for stabilization of the

target adenosine. However, in KsgA, only Y116 seemed to be important for activity. This might suggest a slightly different mode of adenosine binding in KsgA, a possibility which we would like to explore crystallographically. Planned experiments include crystallization of KsgA with adenosine and adenosine analogs in an attempt to solve the structure of the target adenosine bound into the active site.

Another outstanding question we would like to address involves the order (if any) of methyl transfer. Kinetic studies of wild-type and mutant KsgA proteins may help us probe KsgA's mechanism. A combination of techniques will be used to study the methyl transfer in detail. By stopping methylation reactions at early time-points, we can obtain snapshots of partially methylated intermediates. HPLC analysis will allow us to determine amounts of mono- and dimethylated adenosine; and primer extension will help pinpoint which of the two target residues is methylated, since primer extension reactions will not read through the modified bases. These three methods used in concert should reveal whether the two adenosines are methylated in a specific order, or if the four methyl groups are transferred randomly.

Our model of the binding interaction between KsgA and 30S illuminates aspects of KsgA's activity. This model helps explain KsgA's complex substrate requirements, and may account for the temporal disconnect between Dim1's rRNA processing and methylation functions in eukaryotes. We are currently using this model to direct mutational studies, in order to test potential RNA-binding residues in KsgA. Given the large contact area between KsgA and helix 44, we expect that drastic mutations will have to be made in order to have a significant impact on binding. We would also like to characterize the

interaction by NMR and/or X-ray crystallography. We are performing preliminary NMR experiments, with the ultimate goal of solving the structure of a KsgA/helix 44 complex, using the h44 oligo described in Chapter 4. Future work will consist of crystallographic experiments using KsgA bound to the 30S subunit. Results from these structural studies will help define binding on a molecular basis, identifying specific amino acid and nucleotide residues which mediate the binding interaction.

The work presented here represents a substantial contribution to the KsgA literature, and to the ribosome field in general. We believe that the KsgA/Dim1 family is an important factor in the ribosome assembly process in all kingdoms of life, and that a better understanding of these enzymes will lead to a more thorough understanding of ribosome biogenesis in molecular detail.

<u>Literature Cited</u>

Literature Cited

1. Traub P., Nomura M. (1968) Structure and function of E. coli ribosomes. V. Reconstitution of functionally active 30S ribosomal particles from RNA and proteins. *Proc. Natl. Acad. Sci. U. S. A.* **59,** 777-784.

2. Held W. A., Mizushima S. & Nomura M. (1973) Reconstitution of Escherichia coli 30 S ribosomal subunits from purified molecular components. *J. Biol. Chem.* **248,** 5720-5730.

3. Nierhaus K. H., Dohme F. (1974) Total reconstitution of functionally active 50S ribosomal subunits from Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* **71,** 4713-4717.

4. Krzyzosiak W., Denman R., Nurse K., Hellmann W., Boublik M., Gehrke C. W., Agris P. F. & Ofengand J. (1987) In vitro synthesis of 16S ribosomal RNA containing single base changes and assembly into a functional 30S ribosome. *Biochemistry* **26,** 2353-2364.

5. Culver G. M., Noller H. F. (1999) Efficient reconstitution of functional Escherichia coli 30S ribosomal subunits from a complete set of recombinant small subunit ribosomal proteins. *RNA* **5,** 832-843.

6. Ban N., Nissen P., Hansen J., Moore P. B. & Steitz T. A. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. *Science* **289,** 905-920. 8744 [pii].

7. Wimberly B. T., Brodersen D. E., Clemons W. M.,Jr, Morgan-Warren R. J., Carter A. P., Vonrhein C., Hartsch T. & Ramakrishnan V. (2000) Structure of the 30S ribosomal subunit. *Nature* **407,** 327-339. 10.1038/35030006 [doi].

8. Harms J., Schluenzen F., Zarivach R., Bashan A., Gat S., Agmon I., Bartels H., Franceschi F. & Yonath A. (2001) High resolution structure of the large ribosomal subunit from a mesophilic eubacterium. *Cell* **107,** 679-688. S0092-8674(01)00546-3 [pii].

9. Yusupov M. M., Yusupova G. Z., Baucom A., Lieberman K., Earnest T. N., Cate J. H. & Noller H. F. (2001) Crystal structure of the ribosome at 5.5 A resolution. *Science* **292,** 883-896. 10.1126/science.1060089 [doi]; 1060089 [pii].

10. Schuwirth B. S., Borovinskaya M. A., Hau C. W., Zhang W., Vila-Sanjurjo A., Holton J. M. & Cate J. H. (2005) Structures of the bacterial ribosome at 3.5 A resolution. *Science* **310,** 827-834. 310/5749/827 [pii]; 10.1126/science.1117230 [doi].

11. Selmer M., Dunham C. M., Murphy F. V.,4th, Weixlbaumer A., Petry S., Kelley A. C., Weir J. R. & Ramakrishnan V. (2006) Structure of the 70S ribosome complexed with mRNA and tRNA. *Science* **313,** 1935-1942. 1131127 [pii]; 10.1126/science.1131127 [doi].

12. Lovgren J. M., Wikstrom P. M. (2001) The rlmB gene is essential for formation of Gm2251 in 23S rRNA but not for ribosome maturation in Escherichia coli. *J. Bacteriol.* **183,** 6957-6960. 10.1128/JB.183.23.6957-6960.2001 [doi].

13. Inoue K., Alsina J., Chen J. & Inouye M. (2003) Suppression of defective ribosome assembly in a rbfA deletion mutant by overexpression of Era, an essential GTPase in Escherichia coli. *Mol. Microbiol.* **48,** 1005-1016. 3475 [pii].

14. Himeno H., Hanawa-Suetsugu K., Kimura T., Takagi K., Sugiyama W., Shirata S., Mikami T., Odagiri F., Osanai Y., Watanabe D., Goto S., Kalachnyuk L., Ushida C. & Muto A. (2004) A novel GTPase activated by the small subunit of ribosome. *Nucleic Acids Res.* **32,** 5303-5309. 32/17/5303 [pii]; 10.1093/nar/gkh861 [doi].

15. Bylund G. O., Wipemo L. C., Lundberg L. A. & Wikstrom P. M. (1998) RimM and RbfA are essential for efficient processing of 16S rRNA in Escherichia coli. *J. Bacteriol.* **180,** 73-82.

16. Kaczanowska M., Ryden-Aulin M. (2005) The YrdC protein--a putative ribosome maturation factor. *Biochim. Biophys. Acta* **1727,** 87-96. S0167-4781(04)00263-5 [pii]; 10.1016/j.bbaexp.2004.11.010 [doi].

17. Srivastava A. K., Schlessinger D. (1989) Processing pathway of Escherichia coli 16S precursor rRNA. *Nucleic Acids Res.* **17,** 1649-1663.

18. Mizushima S., Nomura M. (1970) Assembly mapping of 30S ribosomal proteins from E. coli. *Nature* **226,** 1214.

19. Traub P., Nomura M. (1969) Structure and function of Escherichia coli ribosomes. VI. Mechanism of assembly of 30 s ribosomes studied in vitro. *J. Mol. Biol.* **40,** 391-413.

20. Maki J. A., Schnobrich D. J. & Culver G. M. (2002) The DnaK chaperone system facilitates 30S ribosomal subunit assembly. *Mol. Cell* **10,** 129-138. S1097276502005622 [pii].

21. Nashimoto H., Held W., Kaltschmidt E. & Nomura M. (1971) Structure and function of bacterial ribosomes. XII. Accumulation of 21 s particles by some cold-sensitive mutants of Escherichia coli. *J. Mol. Biol.* **62,** 121-138.

22. Hage A. E., Alix J. H. (2004) Authentic precursors to ribosomal subunits accumulate in Escherichia coli in the absence of functional DnaK chaperone. *Mol. Microbiol.* **51,** 189-201. 3813 [pii].

23. Venema J., Tollervey D. (1999) Ribosome synthesis in Saccharomyces cerevisiae. *Annu. Rev. Genet.* **33,** 261-311. 10.1146/annurev.genet.33.1.261 [doi].

24. Schafer T., Strauss D., Petfalski E., Tollervey D. & Hurt E. (2003) The path from nucleolar 90S to cytoplasmic 40S pre-ribosomes. *EMBO J.* **22,** 1370-1380. 10.1093/emboj/cdg121 [doi].

25. Osheim Y. N., French S. L., Keck K. M., Champion E. A., Spasov K., Dragon F., Baserga S. J. & Beyer A. L. (2004) Pre-18S ribosomal RNA is structurally compacted into the SSU processome prior to being cleaved from nascent transcripts in Saccharomyces cerevisiae. *Mol. Cell* **16,** 943-954. S1097276504007233 [pii]; 10.1016/j.molcel.2004.11.031 [doi].

26. Mangiarotti G., Chiaberge S. (1997) Reconstitution of functional eukaryotic ribosomes from Dictyostelium discoideum ribosomal proteins and RNA. *J. Biol. Chem.* **272,** 19682-19687.

27. Pace N. R. (1997) A molecular view of microbial diversity and the biosphere. *Science* **276,** 734-740.

28. Bult C. J., White O., Olsen G. J., Zhou L., Fleischmann R. D., Sutton G. G., Blake J. A., FitzGerald L. M., Clayton R. A., Gocayne J. D., Kerlavage A. R., Dougherty B. A., Tomb J. F., Adams M. D., Reich C. I., Overbeek R., Kirkness E. F., Weinstock K. G., Merrick J. M., Glodek A., Scott J. L., Geoghagen N. S. & Venter J. C. (1996) Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii. *Science* **273,** 1058-1073.

29. Sanchez M. E., Urena D., Amils R. & Londei P. (1990) In vitro reassembly of active large ribosomal subunits of the halophilic archaebacterium Haloferax mediterranei. *Biochemistry* **29,** 9256-9261.

30. Sanchez E., Amils R. (1995) Absolute requirement of ammonium sulfate for reconstitution of active 70S ribosomes from the extreme halophilic archaeon Haloferax mediterranei. *Eur. J. Biochem.* **233,** 809-814.

31. Sanchez M. E., Londei P. & Amils R. (1996) Total reconstitution of active small ribosomal subunits of the extreme halophilic archaeon Haloferax mediterranei. *Biochim. Biophys. Acta* **1292,** 140-144.

32. Limbach P. A., Crain P. F. & McCloskey J. A. (1994) Summary: the modified nucleosides of RNA. *Nucleic Acids Res.* **22,** 2183-2196.

33. Kiss-Laszlo Z., Henry Y., Bachellerie J. P., Caizergues-Ferrer M. & Kiss T. (1996) Site-specific ribose methylation of preribosomal RNA: a novel function for small nucleolar RNAs. *Cell* **85,** 1077-1088. S0092-8674(00)81308-2 [pii].

34. Ganot P., Bortolin M. L. & Kiss T. (1997) Site-specific pseudouridine formation in preribosomal RNA is guided by small nucleolar RNAs. *Cell* **89,** 799-809. S0092-8674(00)80263-9 [pii].

35. Fromont-Racine M., Senger B., Saveanu C. & Fasiolo F. (2003) Ribosome assembly in eukaryotes. *Gene* **313,** 17-42. S0378111903006292 [pii].

36. Omer A. D., Ziesche S., Decatur W. A., Fournier M. J. & Dennis P. P. (2003) RNA-modifying machines in archaea. *Mol. Microbiol.* **48,** 617-629. 3483 [pii].

37. Omer A. D., Lowe T. M., Russell A. G., Ebhardt H., Eddy S. R. & Dennis P. P. (2000) Homologs of small nucleolar RNAs in Archaea. *Science* **288,** 517-522. 8415 [pii].

38. Tang T. H., Bachellerie J. P., Rozhdestvensky T., Bortolin M. L., Huber H., Drungowski M., Elge T., Brosius J. & Huttenhofer A. (2002) Identification of 86 candidates for small non-messenger RNAs from the archaeon Archaeoglobus fulgidus. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 7536-7541. 10.1073/pnas.112047299 [doi].

39. Watanabe Y., Gray M. W. (2000) Evolutionary appearance of genes encoding proteins associated with box H/ACA snoRNAs: cbf5p in Euglena gracilis, an early diverging eukaryote, and candidate Gar1p and Nop10p homologs in archaebacteria. *Nucleic Acids Res.* **28,** 2342-2352.

40. Speckmann W. A., Li Z. H., Lowe T. M., Eddy S. R., Terns R. M. & Terns M. P. (2002) Archaeal guide RNAs function in rRNA modification in the eukaryotic nucleus. *Curr. Biol.* **12,** 199-203. S0960982202006553 [pii].

41. Van Knippenberg P. H., Van Kimmenade J. M. & Heus H. A. (1984) Phylogeny of the conserved 3' terminal structure of the RNA of small ribosomal subunits. *Nucleic Acids Res.* **12,** 2595-2604.

42. Cannone J. J., Subramanian S., Schnare M. N., Collett J. R., D'Souza L. M., Du Y., Feng B., Lin N., Madabusi L. V., Muller K. M., Pande N., Shang Z., Yu N. & Gutell R. R. (2002) The comparative RNA web (CRW) site: an online database of

comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics* **3,** 2.

43. Helser T. L., Davies J. E. & Dahlberg J. E. (1971) Change in methylation of 16S ribosomal RNA associated with mutation to kasugamycin resistance in Escherichia coli. *Nat. New Biol.* **233,** 12-14.

44. Helser T. L., Davies J. E. & Dahlberg J. E. (1972) Mechanism of kasugamycin resistance in Escherichia coli. *Nat. New Biol.* **235,** 6-9.

45. Thammana P., Held W. A. (1974) Methylation of 16S RNA during ribosome assembly in vitro. *Nature* **251,** 682-686.

46. Desai P. M., Rife J. P.

47. Poldermans B., Roza L. & Van Knippenberg P. H. (1979) Studies on the function of two adjacent N6,N6-dimethyladenosines near the 3' end of 16 S ribosomal RNA of Escherichia coli. III. Purification and properties of the methylating enzyme and methylase-30 S interactions. *J. Biol. Chem.* **254,** 9094-9100.

48. Desai P. M., Rife J. P. (2006) The adenosine dimethyltransferase KsgA recognizes a specific conformational state of the 30S ribosomal subunit. *Arch. Biochem. Biophys.* **449,** 57-63. S0003-9861(06)00072-5 [pii]; 10.1016/j.abb.2006.02.028 [doi].

49. Zamir A., Miskin R. & Elson D. (1969) Interconversions between inactive and active forms of ribosomal subunits. *FEBS Lett.* **3,** 85-88. 0014579369801031 [pii].

50. Zamir A., Miskin R. & Elson D. (1971) Inactivation and reactivation of ribosomal subunits: amino acyl-transfer RNA binding activity of the 30 s subunit of Escherichia coli. *J. Mol. Biol.* **60,** 347-364.

51. Guerin M. F., Hayes D. H. (1987) Comparison of active and inactive forms of the E. coli 30S ribosomal subunits. *Biochimie* **69,** 965-974. 0300-9084(87)90230-6 [pii].

52. Moazed D., Van Stolk B. J., Douthwaite S. & Noller H. F. (1986) Interconversion of active and inactive 30 S ribosomal subunits is accompanied by a conformational change in the decoding region of 16 S rRNA. *J. Mol. Biol.* **191,** 483-493.

53. Thammana P., Cantor C. R. (1978) Studies on ribosome structure and interactions near the m62Am62A sequence. *Nucleic Acids Res.* **5,** 805-823.

54. Backendorf C., Ravensbergen C. J., Van der Plas J., van Boom J. H., Veeneman G. & Van Duin J. (1981) Basepairing potential of the 3' terminus of 16S RNA: dependence on the functional state of the 30S subunit and the presence of protein S21. *Nucleic Acids Res.* **9,** 1425-1444.

55. Basturea G. N., Rudd K. E. & Deutscher M. P. (2006) Identification and characterization of RsmE, the founding member of a new RNA base methyltransferase family. *RNA* **12,** 426-434. rna.2283106 [pii]; 10.1261/rna.2283106 [doi].

56. Andersen N. M., Douthwaite S. (2006) YebU is a m5C methyltransferase specific for 16 S rRNA nucleotide 1407. *J. Mol. Biol.* **359,** 777-786. S0022-2836(06)00456-6 [pii]; 10.1016/j.jmb.2006.04.007 [doi].

57. Leveque F., Blanchin-Roland S., Fayat G., Plateau P. & Blanquet S. (1990) Design and characterization of Escherichia coli mutants devoid of Ap4N-hydrolase activity. *J. Mol. Biol.* **212,** 319-329.

58. Poldermans B., Van Buul C. P. & Van Knippenberg P. H. (1979) Studies on the function of two adjacent N6,N6-dimethyladenosines near the 3' end of 16 S ribosomal RNA of Escherichia coli. II. The effect of the absence of the methyl groups on initiation of protein biosynthesis. *J. Biol. Chem.* **254,** 9090-9093.

59. Van Charldorp R., Heus H. A. & Van Knippenberg P. H. (1981) Adenosine dimethylation of 16S ribosomal RNA: effect of the methylgroups on local conformational stability as deduced from electrophoretic mobility of RNA fragments in denaturing polyacrylamide gels. *Nucleic Acids Res.* **9,** 267-275.

60. Van Charldorp R., Heus H. A., Van Knippenberg P. H., Joordens J., De Bruin S. H. & Hilbers C. W. (1981) Destabilization of secondary structure in 16S ribosomal RNA by dimethylation of two adjacent adenosines. *Nucleic Acids Res.* **9,** 4413-4422.

61. van Knippenberg P. H., Heus H. A. (1983) The conformation of a conserved stem-loop structure in ribosomal RNA. *J. Biomol. Struct. Dyn.* **1,** 371-381.

62. Rife J. P., Cheng C. S., Moore P. B. & Strobel S. A. (1998) N 2-methylguanosine is iso-energetic with guanosine in RNA duplexes and GNRA tetraloops. *Nucleic Acids Res.* **26,** 3640-3644. gkb615 [pii].

63. Wickstrom E., Heus H. A., Haasnoot C. A. & van Knippenberg P. H. (1986) Circular dichroism and 500-MHz proton magnetic resonance studies of the interaction of Escherichia coli translational initiation factor 3 protein with the 16S ribosomal RNA 3' cloacin fragment. *Biochemistry* **25,** 2770-2777.

64. Igarashi K., Kishida K., Kashiwagi K., Tatokoro I., Kakegawa T. & Hirose S. (1981) Relationship between methylation of adenine near the 3' end of 16-S ribosomal RNA and the activity of 30-S ribosomal subunits. *Eur. J. Biochem.* **113,** 587-593.

65. Sorensen M. A., Fricke J. & Pedersen S. (1998) Ribosomal protein S1 is required for translation of most, if not all, natural mRNAs in Escherichia coli in vivo. *J. Mol. Biol.* **280,** 561-569. S0022-2836(98)91909-X [pii]; 10.1006/jmbi.1998.1909 [doi].

66. Tedin K., Resch A. & Blasi U. (1997) Requirements for ribosomal protein S1 for translation initiation of mRNAs with and without a 5' leader sequence. *Mol. Microbiol.* **25,** 189-199.

67. Chin K., Shean C. S. & Gottesman M. E. (1993) Resistance of lambda cI translation to antibiotics that inhibit translation initiation. *J. Bacteriol.* **175,** 7471-7473.

68. Poldermans B., Goosen N. & Van Knippenberg P. H. (1979) Studies on the function of two adjacent N6,N6-dimethyladenosines near the 3' end of 16 S ribosomal RNA of Escherichia coli. I. The effect of kasugamycin on initiation of protein synthesis. *J. Biol. Chem.* **254,** 9085-9089.

69. Moll I., Hirokawa G., Kiel M. C., Kaji A. & Blasi U. (2004) Translation initiation with 70S ribosomes: an alternative pathway for leaderless mRNAs. *Nucleic Acids Res.* **32,** 3354-3363. 10.1093/nar/gkh663 [doi]; 32/11/3354 [pii].

70. van Buul C. P., Visser W. & van Knippenberg P. H. (1984) Increased translational fidelity caused by the antibiotic kasugamycin and ribosomal ambiguity in mutants harbouring the ksgA gene. *FEBS Lett.* **177,** 119-124. 0014-5793(84)80994-1 [pii].

71. Ryden-Aulin M., Shaoping Z., Kylsten P. & Isaksson L. A. (1993) Ribosome activity and modification of 16S RNA are influenced by deletion of ribosomal protein S20. *Mol. Microbiol.* **7,** 983-992.

72. Melancon P., Gravel M., Boileau G. & Brakier-Gingras L. (1987) Reassembly of active 30S ribosomal subunits with an unmethylated in vitro transcribed 16S rRNA. *Biochem. Cell Biol.* **65,** 1022-1030.

73. Blanchin-Roland S., Blanquet S., Schmitter J. M. & Fayat G. (1986) The gene for Escherichia coli diadenosine tetraphosphatase is located immediately clockwise to folA and forms an operon with ksgA. *Mol. Gen. Genet.* **205,** 515-522.

74. Mechulam Y., Fromant M., Mellot P., Plateau P., Blanchin-Roland S., Fayat G. & Blanquet S. (1985) Molecular cloning of the Escherichia coli gene for diadenosine 5',5'''-P1,P4-tetraphosphate pyrophosphohydrolase. *J. Bacteriol.* **164,** 63-69.

75. Farr S. B., Arnosti D. N., Chamberlin M. J. & Ames B. N. (1989) An apaH mutation causes ApppppA to accumulate and affects motility and catabolite repression in Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* **86,** 5010-5014.

76. Johnstone D. B., Farr S. B. (1991) ApppppA binds to several proteins in Escherichia coli, including the heat shock and oxidative stress proteins DnaK, GroEL, E89, C45 and C40. *EMBO J.* **10,** 3897-3904.

77. Van Buul C. P., Damm J. B. & Van Knippenberg P. H. (1983) Kasugamycin resistant mutants of Bacillus stearothermophilus lacking the enzyme for the methylation of two adjacent adenosines in 16S ribosomal RNA. *Mol. Gen. Genet.* **189,** 475-478.

78. Lafontaine D., Delcour J., Glasser A. L., Desgres J. & Vandenhaute J. (1994) The DIM1 gene responsible for the conserved m6(2)Am6(2)A dimethylation in the 3'-terminal loop of 18 S rRNA is essential in yeast. *J. Mol. Biol.* **241,** 492-497. S0022283684715257 [pii].

79. Housen I., Demonte D., Lafontaine D. & Vandenhaute J. (1997) Cloning and characterization of the KlDIM1 gene from Kluyveromyces lactis encoding the m2(6)A dimethylase of the 18S rRNA. *Yeast* **13,** 777-781. 10.1002/(SICI)1097-0061(19970630)13:8<777::AID-YEA140>3.0.CO;2-1 [pii].

80. Tokuhisa J. G., Vijayan P., Feldmann K. A. & Browse J. A. (1998) Chloroplast development at low temperatures requires a homolog of DIM1, a yeast gene encoding the 18S rRNA dimethylase. *Plant Cell* **10,** 699-711.

81. McCulloch V., Seidel-Rogol B. L. & Shadel G. S. (2002) A human mitochondrial transcription factor is related to RNA adenine methyltransferases and binds S-adenosylmethionine. *Mol. Cell. Biol.* **22,** 1116-1125.

82. O'Farrell H. C., Pulicherla N., Desai P. M. & Rife J. P. (2006) Recognition of a complex substrate by the KsgA/Dim1 family of enzymes has been conserved throughout evolution. *RNA* **12,** 725-733. rna.2310406 [pii]; 10.1261/rna.2310406 [doi].

83. Lafontaine D., Vandenhaute J. & Tollervey D. (1995) The 18S rRNA dimethylase Dim1p is required for pre-ribosomal RNA processing in yeast. *Genes Dev.* **9,** 2470-2481.

84. Tokuhisa J. G.

85. Lafontaine D. L., Preiss T. & Tollervey D. (1998) Yeast 18S rRNA dimethylase Dim1p: a quality control mechanism in ribosome synthesis? *Mol. Cell. Biol.* **18,** 2360-2370.

86. Seidel-Rogol B. L., McCulloch V. & Shadel G. S. (2003) Human mitochondrial transcription factor B1 methylates ribosomal RNA at a conserved stem-loop. *Nat. Genet.* **33,** 23-24. 10.1038/ng1064 [doi]; ng1064 [pii].

87. Rantanen A., Gaspari M., Falkenberg M., Gustafsson C. M. & Larsson N. G. (2003) Characterization of the mouse genes for mitochondrial transcription factors B1 and B2. *Mamm. Genome* **14,** 1-6. 10.1007/s00335-002-2218-z [doi].

88. Cotney J., Shadel G. S. (2006) Evidence for an Early Gene Duplication Event in the Evolution of the Mitochondrial Transcription Factor B Family and Maintenance of rRNA Methyltransferase Activity in Human mtTFB1 and mtTFB2. *J. Mol. Evol.* **63,** 707-717. 10.1007/s00239-006-0075-1 [doi].

89. Falkenberg M., Gaspari M., Rantanen A., Trifunovic A., Larsson N. G. & Gustafsson C. M. (2002) Mitochondrial transcription factors B1 and B2 activate transcription of human mtDNA. *Nat. Genet.* **31,** 289-294. 10.1038/ng909 [doi]; ng909 [pii].

90. Matsushima Y., Garesse R. & Kaguni L. S. (2004) Drosophila mitochondrial transcription factor B2 regulates mitochondrial DNA copy number and transcription in schneider cells. *J. Biol. Chem.* **279,** 26900-26905. 10.1074/jbc.M401643200 [doi]; M401643200 [pii].

91. Matsushima Y., Adan C., Garesse R. & Kaguni L. S. (2005) Drosophila mitochondrial transcription factor B1 modulates mitochondrial translation but not transcription or DNA copy number in Schneider cells. *J. Biol. Chem.* **280,** 16815-16820. M500569200 [pii]; 10.1074/jbc.M500569200 [doi].

92. Carrodeguas J. A., Yun S., Shadel G. S., Clayton D. A. & Bogenhagen D. F. (1996) Functional conservation of yeast mtTFB despite extensive sequence divergence. *Gene Expr.* **6,** 219-230.

93. Shutt T. E., Gray M. W. (2006) Homologs of mitochondrial transcription factor B, sparsely distributed within the eukaryotic radiation, are likely derived from the dimethyladenosine methyltransferase of the mitochondrial endosymbiont. *Mol. Biol. Evol.* **23,** 1169-1179. msk001 [pii]; 10.1093/molbev/msk001 [doi].

94. Skinner R., Cundliffe E. & Schmidt F. J. (1983) Site of action of a ribosomal RNA methylase responsible for resistance to erythromycin and other antibiotics. *J. Biol. Chem.* **258,** 12702-12706.

95. van Buul C. P., van Knippenberg P. H. (1985) Nucleotide sequence of the ksgA gene of Escherichia coli: comparison of methyltransferases effecting dimethylation of adenosine in ribosomal RNA. *Gene* **38,** 65-72.

96. Weisblum B. (1995) Erythromycin resistance by ribosome modification. *Antimicrob. Agents Chemother.* **39,** 577-585.

97. Schluckebier G., Zhong P., Stewart K. D., Kavanaugh T. J. & Abad-Zapatero C. (1999) The 2.2 A structure of the rRNA methyltransferase ErmC' and its complexes with cofactor and cofactor analogs: implications for the reaction mechanism. *J. Mol. Biol.* **289,** 277-291. 10.1006/jmbi.1999.2788 [doi]; S0022-2836(99)92788-2 [pii].

98. Fauman, E. B., Blumenthal, R. M. & Cheng, X. (1998). Structure and evolution of AdoMet-dependent methyltransferases. In *S-adenosylmethionine-dependent methyltransferases: Structures and functions* (Cheng, S., Blumenthal, R. M., eds.), pp. 1-38, World Scientific Publishing Co. Pte. Ltd., Singapore.

99. Malone T., Blumenthal R. M. & Cheng X. (1995) Structure-guided analysis reveals nine sequence motifs conserved among DNA amino-methyltransferases, and suggests a catalytic mechanism for these enzymes. *J. Mol. Biol.* **253,** 618-632. S0022-2836(85)70577-3 [pii]; 10.1006/jmbi.1995.0577 [doi].

100. Denoya C., Dubnau D. (1989) Mono- and dimethylating activities and kinetic studies of the ermC 23 S rRNA methyltransferase. *J. Biol. Chem.* **264,** 2615-2624.

101. Van Buul C. P., Hamersma M., Visser W. & Van Knippenberg P. H. (1984) Partial methylation of two adjacent adenosines in ribosomes from Euglena gracilis chloroplasts suggests evolutionary loss of an intermediate stage in the methyl-transfer reaction. *Nucleic Acids Res.* **12,** 9205-9208.

102. Cunningham P. R., Weitzmann C. J., Nurse K., Masurel R., Van Knippenberg P. H. & Ofengand J. (1990) Site-specific mutation of the conserved m6(2)A m6(2)A residues of E. coli 16S ribosomal RNA. Effects on ribosome function and activity of the ksgA methyltransferase. *Biochim. Biophys. Acta* **1050,** 18-26.

103. Wu J. C., Santi D. V. (1987) Kinetic and catalytic mechanism of HhaI methyltransferase. *J. Biol. Chem.* **262,** 4778-4786.

104. Lerner C. G., Gulati P. S. & Inouye M. (1995) Cold-sensitive conditional mutations in Era, an essential Escherichia coli GTPase, isolated by localized random polymerase chain reaction mutagenesis. *FEMS Microbiol. Lett.* **126,** 291-298. 037810979500025Z [pii].

105. Lu Q., Inouye M. (1998) The gene for 16S rRNA methyltransferase (ksgA) functions as a multicopy suppressor for a cold-sensitive mutant of era, an essential RAS-like GTP-binding protein in Escherichia coli. *J. Bacteriol.* **180,** 5243-5246.

106. Schubot F. D., Chen C. J., Rose J. P., Dailey T. A., Dailey H. A. & Wang B. C. (2001) Crystal structure of the transcription factor sc-mtTFB offers insights into mitochondrial transcription. *Protein Sci.* **10,** 1980-1988.

107. Bussiere D. E., Muchmore S. W., Dealwis C. G., Schluckebier G., Nienaber V. L., Edalji R. P., Walter K. A., Ladror U. S., Holzman T. F. & Abad-Zapatero C. (1998) Crystal structure of ErmC', an rRNA methyltransferase which mediates antibiotic resistance in bacteria. *Biochemistry* **37,** 7103-7112. 10.1021/bi973113c [doi]; bi973113c [pii].

108. Yu L., Petros A. M., Schnuchel A., Zhong P., Severin J. M., Walter K., Holzman T. F. & Fesik S. W. (1997) Solution structure of an rRNA methyltransferase (ErmAM) that confers macrolide-lincosamide-streptogramin antibiotic resistance. *Nat. Struct. Biol.* **4,** 483-489.

109. Dong A., Wu H., Zeng H., Loppnau P., Sundstrom M., Arrowsmith C., Edwards A., Bochkarev A. & Plotnikov A. Crystal structure of human Dimethyladenosine transferase with SAM.

110. Mears J. A., Cannone J. J., Stagg S. M., Gutell R. R., Agrawal R. K. & Harvey S. C. (2002) Modeling a minimal ribosome based on comparative sequence analysis. *J. Mol. Biol.* **321,** 215-234. S0022283602005685 [pii].

111. Lee S. G., Evans W. R. (1971) Hybrid ribosome formation from Escherichia coli and chloroplast ribosome subunits. *Science* **173,** 241-242.

112. Klein H. A., Ochoa S. (1972) Peptide synthesis by prokaryotic-eukaryotic hybrid ribosomes. *J. Biol. Chem.* **247,** 8122-8128.

113. Boublik M., Wydro R. M., Hellmann W. & Jenkins F. (1979) Structure of functional A salina -- E coli hybrid ribosome by electron microscopy. *J. Supramol. Struct.* **10,** 397-404. 10.1002/jss.400100403 [doi].

114. Altamura S., Cammarano P. & Londei P. (1986) Archaebacterial and eukaryotic ribosomal subunits can form active hybrid ribosomes. *FEBS Lett.* **204,** 129-133. 0014-5793(86)81400-4 [pii].

115. Londei P., Altamura S., Caprini E. & Martayan A. (1991) Translation and ribosome assembly in extremely thermophilic archaebacteria. *Biochimie* **73,** 1465-1472.

116. Thompson J., Musters W., Cundliffe E. & Dahlberg A. E. (1993) Replacement of the L11 binding region within E.coli 23S ribosomal RNA with its homologue from yeast: in vivo and in vitro analysis of hybrid ribosomes altered in the GTPase centre. *EMBO J.* **12,** 1499-1504.

117. Weglohner W., Junemann R., von Knoblauch K. & Subramanian A. R. (1997) Different consequences of incorporating chloroplast ribosomal proteins L12 and S18 into the bacterial ribosomes of Escherichia coli. *Eur. J. Biochem.* **249,** 383-392.

118. Uhlein M., Weglohner W., Urlaub H. & Wittmann-Liebold B. (1998) Functional implications of ribosomal protein L2 in protein biosynthesis as shown by in vivo replacement studies. *Biochem. J.* **331 (Pt 2),** 423-430.

119. Uchiumi T., Hori K., Nomura T. & Hachimori A. (1999) Replacement of L7/L12.L10 protein complex in Escherichia coli ribosomes with the eukaryotic counterpart changes the specificity of elongation factor binding. *J. Biol. Chem.* **274,** 27578-27582.

120. Terasaki M., Suzuki T., Hanada T. & Watanabe K. (2004) Functional compatibility of elongation factors between mammalian mitochondrial and bacterial ribosomes: characterization of GTPase activity and translation elongation by hybrid ribosomes bearing heterologous L7/12 proteins. *J. Mol. Biol.* **336,** 331-342. S0022283603015213 [pii].

121. Nomura T., Nakano K., Maki Y., Naganuma T., Nakashima T., Tanaka I., Kimura M., Hachimori A. & Uchiumi T. (2006) In vitro reconstitution of the GTPase-associated centre of the archaebacterial ribosome: the functional features observed in a hybrid form with Escherichia coli 50S subunits. *Biochem. J.* **396,** 565-571. BJ20060038 [pii]; 10.1042/BJ20060038 [doi].

122. Culver G. M. (2003) Assembly of the 30S ribosomal subunit. *Biopolymers* **68,** 234-249. 10.1002/bip.10221 [doi].

123. Kiss T. (2001) Small nucleolar RNA-guided post-transcriptional modification of cellular RNAs. *EMBO J.* **20,** 3617-3622. 10.1093/emboj/20.14.3617 [doi].

124. Klootwijk J., Klein I. & Grivell L. A. (1975) Minimal post-transcriptional modification of yeast mitochondrial ribosomal RNA. *J. Mol. Biol.* **97,** 337-350.

125. Steege D. A., Graves M. C. & Spremulli L. L. (1982) Euglena gracilis chloroplast small subunit rRNA. Sequence and base pairing potential of the 3' terminus, cleavage by colicin E3. *J. Biol. Chem.* **257,** 10430-10439.

126. Noon K. R., Bruenger E. & McCloskey J. A. (1998) Posttranscriptional modifications in 16S and 23S rRNAs of the archaeal hyperthermophile Sulfolobus solfataricus. *J. Bacteriol.* **180,** 2883-2888.

127. Fan J., Schnare M. N. & Lee R. W. (2003) Characterization of fragmented mitochondrial ribosomal RNAs of the colorless green alga Polytomella parva. *Nucleic Acids Res.* **31,** 769-778.

128. Schnare M. N., Heinonen T. Y., Young P. G. & Gray M. W. (1986) A discontinuous small subunit ribosomal RNA in Tetrahymena pyriformis mitochondria. *J. Biol. Chem.* **261,** 5187-5193.

129. Lambowitz A. M., Luck D. J. (1976) Studies on the poky mutant of eurospora crassa. Fingerprint analysis of mitochondrial ribosomal RNA. *J. Biol. Chem.* **251,** 3081-3095.

130. Notredame C., Higgins D. G. & Heringa J. (2000) T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* **302,** 205-217. 10.1006/jmbi.2000.4042 [doi]; S0022-2836(00)94042-7 [pii].

131. O'Farrell H. C., Scarsdale J. N. & Rife J. P. (2004) Crystal structure of KsgA, a universally conserved rRNA adenine dimethyltransfase in Escherichia coli. *J. Mol. Biol.* **339,** 337-353. 10.1016/j.jmb.2004.02.068 [doi]; S0022283604002487 [pii].

132. Blaha G., Stelzl U., Spahn C. M., Agrawal R. K., Frank J. & Nierhaus K. H. (2000) Preparation of functional ribosomal complexes and effect of buffer conditions on tRNA positions observed by cryoelectron microscopy. *Methods Enzymol.* **317,** 292-309.

133. Gehrke C. W., Kuo K. C. (1989) Ribonucleoside analysis by reversed-phase high-performance liquid chromatography. *J. Chromatogr.* **471,** 3-36.

134. Rife J. P., Chang C. S., Moore P. B. & Strobel S. A. (1998) The Synthesis of RNA Containing the Modified Nucleotides N2-methylguanosine and N6,N6-dimethyladenosine. *Nucleosides & Nucleotides* **17,** 2281-2288.

135. Martin J. L., McMillan F. M. (2002) SAM (dependent) I AM: the S-adenosylmethionine-dependent methyltransferase fold. *Curr. Opin. Struct. Biol.* **12,** 783-793. S0959440X02003913 [pii].

136. Schubert H. L., Blumenthal R. M. & Cheng X. (2003) Many paths to methyltransfer: a chronicle of convergence. *Trends Biochem. Sci.* **28,** 329-335. S0968000403000902 [pii].

137. Labahn J., Granzin J., Schluckebier G., Robinson D. P., Jack W. E., Schildkraut I. & Saenger W. (1994) Three-dimensional structure of the adenine-specific DNA methyltransferase M.Taq I in complex with the cofactor S-adenosylmethionine. *Proc. Natl. Acad. Sci. U. S. A.* **91,** 10957-10961.

138. Schluckebier G., Labahn J., Granzin J. & Saenger W. (1998) M.TaqI: possible catalysis via cation-pi interactions in N-specific DNA methyltransferases. *Biol. Chem.* **379,** 389-400.

139. Guyot J. B., Grassi J., Hahn U. & Guschlbauer W. (1993) The role of the preserved sequences of Dam methylase. *Nucleic Acids Res.* **21,** 3183-3190.

140. Willcock D. F., Dryden D. T. & Murray N. E. (1994) A mutational analysis of the two motifs common to adenine methyltransferases. *EMBO J.* **13,** 3902-3908.

141. Kong H., Smith C. L. (1997) Substrate DNA and cofactor regulate the activities of a multi-functional restriction-modification enzyme, BcgI. *Nucleic Acids Res.* **25,** 3687-3692. gka592 [pii].

142. Roth M., Helm-Kruse S., Friedrich T. & Jeltsch A. (1998) Functional roles of conserved amino acid residues in DNA methyltransferases investigated by site-directed mutagenesis of the EcoRV adenine-N6-methyltransferase. *J. Biol. Chem.* **273,** 17333-17342.

143. Maravic G., Feder M., Pongor S., Flogel M. & Bujnicki J. M. (2003) Mutational analysis defines the roles of conserved amino acid residues in the predicted catalytic pocket of the rRNA:m6A methyltransferase ErmC'. *J. Mol. Biol.* **332,** 99-109. S0022283603008635 [pii].

144. Pues H., Bleimling N., Holz B., Wolcke J. & Weinhold E. (1999) Functional roles of the conserved aromatic amino acid residues at position 108 (motif IV) and position 196 (motif VIII) in base flipping and catalysis by the N6-adenine DNA methyltransferase from Thermus aquaticus. *Biochemistry* **38,** 1426-1434. 10.1021/bi9818016 [doi]; bi9818016 [pii].

145. Schluckebier G., O'Gara M., Saenger W. & Cheng X. (1995) Universal catalytic domain structure of AdoMet-dependent methyltransferases. *J. Mol. Biol.* **247,** 16-20. S0022-2836(84)70117-3 [pii]; 10.1006/jmbi.1994.0117 [doi].

146. Cheng X. (1995) Structure and function of DNA methyltransferases. *Annu. Rev. Biophys. Biomol. Struct.* **24,** 293-318.

147. Gavin A. C., Bosche M., Krause R., Grandi P., Marzioch M., Bauer A., Schultz J., Rick J. M., Michon A. M., Cruciat C. M., Remor M., Hofert C., Schelder M., Brajenovic M., Ruffner H., Merino A., Klein K., Hudak M., Dickson D., Rudi T., Gnau V., Bauch A., Bastuck S., Huhse B., Leutwein C., Heurtier M. A., Copley R. R., Edelmann A., Querfurth E., Rybin V., Drewes G., Raida M., Bouwmeester T., Bork P., Seraphin B., Kuster B., Neubauer G. & Superti-Furga G. (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415,** 141-147. 10.1038/415141a [doi]; 415141a [pii].

148. Grandi P., Rybin V., Bassler J., Petfalski E., Strauss D., Marzioch M., Schafer T., Kuster B., Tschochner H., Tollervey D., Gavin A. C. & Hurt E. (2002) 90S pre-ribosomes include the 35S pre-rRNA, the U3 snoRNP, and 40S subunit processing factors but predominantly lack 60S synthesis factors. *Mol. Cell* **10,** 105-115. S1097276502005798 [pii].

149. Vanrobays E., Gelugne J. P., Caizergues-Ferrer M. & Lafontaine D. L. (2004) Dim2p, a KH-domain protein required for small ribosomal subunit synthesis. *RNA* **10,** 645-656.

150. Goedecke K., Pignot M., Goody R. S., Scheidig A. J. & Weinhold E. (2001) Structure of the N6-adenine DNA methyltransferase M.TaqI in complex with DNA and a cofactor analog. *Nat. Struct. Biol.* **8,** 121-125. 10.1038/84104 [doi].

151. Maravic G., Bujnicki J. M., Feder M., Pongor S. & Flogel M. (2003) Alanine-scanning mutagenesis of the predicted rRNA-binding domain of ErmC' redefines the substrate-binding site and suggests a model for protein-RNA interactions. *Nucleic Acids Res.* **31,** 4941-4949.

152. Buriankova K., Doucet-Populaire F., Dorson O., Gondran A., Ghnassia J. C., Weiser J. & Pernodet J. L. (2004) Molecular basis of intrinsic macrolide resistance in the Mycobacterium tuberculosis complex. *Antimicrob. Agents Chemother.* **48,** 143-150.

153. Liang J., Edelsbrunner H. & Woodward C. (1998) Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design. *Protein Sci.* **7,** 1884-1897.

154. Denoya C. D., Dubnau D. (1987) Site and substrate specificity of the ermC 23S rRNA methyltransferase. *J. Bacteriol.* **169,** 3857-3860.

155. Vila-Sanjurjo A., Squires C. L. & Dahlberg A. E. (1999) Isolation of kasugamycin resistant mutants in the 16 S ribosomal RNA of Escherichia coli. *J. Mol. Biol.* **293,** 1-8. 10.1006/jmbi.1999.3160 [doi]; S0022-2836(99)93160-1 [pii].

156. Clancy J., Schmieder B. J., Petitpas J. W., Manousos M., Williams J. A., Faiella J. A., Girard A. E. & McGuirk P. R. (1995) Assays to detect and characterize synthetic agents that inhibit the ErmC methyltransferase. *J. Antibiot. (Tokyo)* **48,** 1273-1279.

157. Hajduk P. J., Dinges J., Schkeryantz J. M., Janowick D., Kaminski M., Tufano M., Augeri D. J., Petros A., Nienaber V., Zhong P., Hammond R., Coen M., Beutel B., Katz L. & Fesik S. W. (1999) Novel inhibitors of Erm methyltransferases from NMR and parallel synthesis. *J. Med. Chem.* **42,** 3852-3859. jm990293a [pii].

158. Hanessian S., Sgarbi P. W. (2000) Design and synthesis of mimics of S-adenosyl-L-homocysteine as potential inhibitors of erythromycin methyltransferases. *Bioorg. Med. Chem. Lett.* **10,** 433-437. S0960-894X(00)00021-4 [pii].

159. van Gemen B., Twisk J. & van Knippenberg P. H. (1989) Autogenous regulation of the Escherichia coli ksgA gene at the level of translation. *J. Bacteriol.* **171,** 4002-4008.

160. Otwinowski Z., Minor W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzmol.* **276,** 307-326.

161. de La Fortelle E., Bricogne G. (1997) Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* **276,** 472-494.

162. Abrahams J. P. (1997) Bias reduction in phase refinement by modified interference functions: introducing the gamma correction. *Acta Crystallogr. D Biol. Crystallogr.* **53,** 371-376. 10.1107/S0907444996015272 [doi]; S0907444996015272 [pii].

163. Perrakis A., Sixma T. K., Wilson K. S. & Lamzin V. S. (1997) wARP: improvement and extension of crystallographic phases by weighted averaging of multiple-refined dummy atomic models. *Acta Crystallogr. D Biol. Crystallogr.* **53,** 448-455. 10.1107/S0907444997005696 [doi]; S0907444997005696 [pii].

164. Read R. J. (1986) Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallographica Section A* **42,** 140-149.

165. McRee D. E. (1999) XtalView/Xfit--A versatile program for manipulating atomic coordinates and electron density. *J. Struct. Biol.* **125,** 156-165. S1047-8477(99)94094-7 [pii]; 10.1006/jsbi.1999.4094 [doi].

166. 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 Crystallogr. D Biol. Crystallogr.* **54 (Pt 5),** 905-921.

167. Schomaker V., Trueblood K. N. (1968) On the rigid-body motion of molecules in crystals. *Acta Crystallographica Section B* **24,** 63-76.

168. Murshudov G. N., Vagin A. A. & Dodson E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53,** 240-255. 10.1107/S0907444996012255 [doi]; S0907444996012255 [pii].

169. R. A. Laskowski, M. W. MacArthur, D. S. Moss and J. M.,Thornton. PROCHECK: a program to check the stereochemical quality of protein structures.

170. Hooft R. W., Vriend G., Sander C. & Abola E. E. (1996) Errors in protein structures. *Nature* **381,** 272. 10.1038/381272a0 [doi].

171. Kleywegt G. J., Jones T. A. (1996) Efficient rebuilding of protein structures. *Acta Crystallogr. D Biol. Crystallogr.* **52,** 829-832. 10.1107/S0907444996001783 [doi]; S0907444996001783 [pii].

172. Jones T. A., Zou J. Y., Cowan S. W. & Kjeldgaard M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47 (Pt 2),** 110-119.

173. Carson M. (1991) \it RIBBONS 2.0. *Journal of Applied Crystallography* **24,** 958-961.

174. Heilek G. M., Marusak R., Meares C. F. & Noller H. F. (1995) Directed hydroxyl radical probing of 16S rRNA using Fe(II) tethered to ribosomal protein S4. *Proc. Natl. Acad. Sci. U. S. A.* **92,** 1113-1116.

175. Culver G. M., Noller H. F. (2000) Directed hydroxyl radical probing of RNA from iron(II) tethered to proteins in ribonucleoprotein complexes. *Methods Enzymol.* **318,** 461-475.

176. Powers T., Noller H. F. (1995) Hydroxyl radical footprinting of ribosomal proteins on 16S rRNA. *RNA* **1,** 194-209.

177. Benard L., Philippe C., Ehresmann B., Ehresmann C. & Portier C. (1996) Pseudoknot and translational control in the expression of the S15 ribosomal protein. *Biochimie* **78,** 568-576. S0300-9084(96)80003-4 [pii].

178. Heilek G. M., Noller H. F. (1996) Directed hydroxyl radical probing of the rRNA neighborhood of ribosomal protein S13 using tethered Fe(II). *RNA* **2,** 597-602.

179. Heilek G. M., Noller H. F. (1996) Site-directed hydroxyl radical probing of the rRNA neighborhood of ribosomal protein S5. *Science* **272,** 1659-1662.

180. Culver G. M., Noller H. F. (1998) Directed hydroxyl radical probing of 16S ribosomal RNA in ribosomes containing Fe(II) tethered to ribosomal protein S20. *RNA* **4,** 1471-1480.

181. Miyamoto A., Usui M., Yamasaki N., Yamada N., Kuwano E., Tanaka I. & Kimura M. (1999) Role of the N-terminal region of ribosomal protein S7 in its interaction with 16S rRNA which binds to the concavity formed by the beta-ribbon arm and the alpha-helix. *Eur. J. Biochem.* **266,** 591-598. ejb901 [pii].

182. Wilson K. S., Noller H. F. (1998) Mapping the position of translational elongation factor EF-G in the ribosome by directed hydroxyl radical probing. *Cell* **92,** 131-139. S0092-8674(00)80905-8 [pii].

183. Wilson K. S., Ito K., Noller H. F. & Nakamura Y. (2000) Functional sites of interaction between release factor RF1 and the ribosome. *Nat. Struct. Biol.* **7,** 866-870. 10.1038/82818 [doi].

184. Dallas A., Noller H. F. (2001) Interaction of translation initiation factor 3 with the 30S ribosomal subunit. *Mol. Cell* **8,** 855-864. S1097-2765(01)00356-2 [pii].

185. Lancaster L., Kiel M. C., Kaji A. & Noller H. F. (2002) Orientation of ribosome recycling factor in the ribosome from directed hydroxyl radical probing. *Cell* **111,** 129-140. S0092867402009388 [pii].

<div align="center"><u>VITA</u></div>

Heather Colleen O'Farrell
Born: March 22, 1974; San Francisco, CA.
Citizenship: US

Education

Virginia Commonwealth University -Medical Campus	8/2002-present
Ph.D. candidate, Biochemistry	
University of Virginia	9/1992-5/1996
B. A., Biology	

Work experience

Virginia Commonwealth University	11/2001-8/2002
Department of Medicinal Chemistry	
Laboratory Technician	
BioCache Pharmaceuticals	1/1999-11/2001
Research Associate	
National Institutes of Health	10/1996-5/1997
NICHD, ERRB	
Pre-IRTA Fellow	

Publications

O'Farrell, HC, Pulicherla, N, Desai, PM, Rife, JP. (2006). Recognition of a complex substrate by the KsgA/Dim1 family of enzymes has been conserved throughout evolution. *RNA* **12**, 725-733.

O'Farrell, HC, Scarsdale, JN, Rife, JP. (2004). Crystal structure of KsgA, a universally conserved rRNA adenine dimethyltransferase in Escherichia coli. *J. Mol. Biol.* **339**, 337-353.

O'Farrell, HC, Musayev, FN, Scarsdale, JN, Wright, HT, Rife, JP. (2003). Crystallization and preliminary X-ray diffraction analysis of KsgA, a universally conserved RNA adenine dimethyltransferase in Escherichia coli. *Acta Crystallog. sect. D* **59**, 1490-1492.

Arora KK, Krsmanovic LZ, Mores N, **O'Farrell H**, Catt KJ. (1998). Mediation of cyclic AMP signaling by the first intracellular loop of the gonadotropin-releasing hormone receptor. *J. Biol. Chem.* **273**, 25581-25586.

Presentations

O'Farrell HC. 2006. Conserved methylation in ribosome biogenesis: the KsgA/Dim1 family of enzymes. Oral presentation at the 7th International Conference on Ribosome Synthesis.

Plicherla N, **O'Farrel HC**, Desai PM, Rife JP. 2006. Yeast 18S rRNA dimethylase, Dim1, complements the function of its orthologs. Poster presentation at the 7th International Conference on Ribosome Synthesis.

Xu Z, **O'Farrell HC**, Rife JP, Culver GM. 2006. Mapping the rRNA environment of KsgA using directed hydroxyl radical probing. Poster presentation at the 11th Annual Meeting of the RNA Society.

O'Farrell HC, Rife JP. 2006. Mutational analysis of conserved residues in the active site of the universally conserved methyltransferase KsgA. Oral presentation at the John C. Forbes Graduate Student Honors Colloquium.

O'Farrell HC, Rife JP. 2004. Mutational Analysis of KsgA, an evolutionarily conserved adenosine dimethyltransferase. Poster presentation at the Daniel T. Watts Research Poster Symposium.

O'Farrell HC. 2003. Crystallization and structure solution of KsgA, a universally conserved adenosine dimethyltransferase. Oral presentation at the annual meeting of the American Society for Microbiology, Virginia branch.

O'Farrell HC, Scarsdale JN, Wright HT, Rife JP. 2003. Crystallization and structure solution of KsgA, a universally conserved adenosine dimethyltransferase. Poster presentation at the Daniel T. Watts Resarch Poster Symposium. Poster presentation at the Symposium on RNA Biology V: RNA, Tool and Target.

Arora KK, Chen HC, **O'Farrell H**. 1997. Mutational analysis of the potential N-linked glycosylation sites of the gonadotropin-releasing hormone receptor. Poster presentation at the 79th Annual Meeting of The Endocrine Society.

Awards

Phi Kappa Phi Scholarship.	2005
C. C. Clayton Award.	2004
First place, Graduate Student Presentation Competition. American Society for Microbiology, Virginia branch.	2003