Published online 6 December 2007

Nucleic Acids Research, 2008, Vol. 36, No. 2 ***629–639***
doi:10.1093/nar/gkm1074

The yeast ribosome synthesis factor Emg1 is a novel member of the superfamily of alpha/beta knot fold methyltransferases

Nicolas Leulliot[1,*], Markus T. Bohnsack[2], Marc Graille[1], David Tollervey[2] and Herman Van Tilbeurgh[1]

[1]Institut de Biochimie et de Biophysique Moléculaire et Cellulaire, UMR8619, Bât 430, Université de Paris-Sud, 91405 Orsay Cedex, France and [2]Wellcome Trust Centre for Cell Biology, University of Edinburgh, EH9 3JR, UK

Received October 10, 2007; Revised November 12, 2007; Accepted November 14, 2007

ABSTRACT

Emg1 was previously shown to be required for maturation of the 18S rRNA and biogenesis of the 40S ribosomal subunit. Here we report the determination of the crystal structure of Emg1 at 2 Å resolution in complex with the methyl donor, *S*-adenosyl-methionine (SAM). This structure identifies Emg1 as a novel member of the alpha/beta knot fold methyltransferase (SPOUT) superfamily. In addition to the conserved SPOUT core, Emg1 has two unique domains that form an extended surface, which we predict to be involved in binding of RNA substrates. A point mutation within a basic patch on this surface almost completely abolished RNA binding *in vitro*. Three point mutations designed to disrupt the interaction of Emg1 with SAM each caused > 100-fold reduction in SAM binding *in vitro*. Expression of only Emg1 with these mutations could support growth and apparently normal ribosome biogenesis in strains genetically depleted of Emg1. We conclude that the catalytic activity of Emg1 is not essential and that the presence of the protein is both necessary and sufficient for ribosome biogenesis.

INTRODUCTION

Ribosome biogenesis in the yeast *Saccharomyces cerevisiae* is a highly complex process involving the coordinated action of more than 200 cofactors, including non-ribosomal proteins and small nucleolar RNPs (snoRNPs) (1–4). The ribosome itself consists of 79 ribosomal proteins and four ribosomal RNAs (rRNAs), of which 18S, 5.8S and 25S are processed in multiple steps from a single precursor transcribed by RNA polymerase I (Pol I), whereas the 5S rRNA is independently transcribed by Pol III. Assembly of the ribosomal proteins onto the rRNA is coordinated with the pathway of maturation of the rRNAs, which involves multiple processing steps and numerous covalent nucleotide modifications. Most nucleotide modifications are the conversion of uridine to pseudouridine, mediated by box H/ACA snoRNPs and methylation of the 2′-OH of ribose, directed by box C/D snoRNPs (5–7).

In addition to the modifications directed by snoRNPs, a small number of base modifications are generated by specific RNA methyltransferases (MTases). Dim1 is the best-studied rRNA MTase in yeast and is required for the dimethylation that forms a $m^6_2Am^6_2A$ doublet at the 3′-end of the 18S rRNA (8,9). Dim1 is essential for viability and for 18S rRNA synthesis in yeast. However, analyses of catalytically inactive mutants of Dim1 and mutation of the target site in the rRNA showed that dimethylation activity was not required for 18S rRNA synthesis. Lack of the modification did, however, inhibit 40S subunit function in translation *in vitro* (9). The requirement for Dim1 in 18S synthesis was suggested to reflect the activity of a quality control mechanism that verifies the presence of the MTase to ensure the synthesis of fully functional ribosomes.

Emg1 (also called Nep1) is a 28-kDa protein conserved from Eukaryotes to Archaea, and is essential for cell growth in yeast. Expression of *EMG1* shows the environmental stress responsive pattern that is characteristic of many ribosomal proteins (10) and Emg1 is associated with the SSU processome, a large complex required for the production of mature 18S RNA and biogenesis of the 40S subunit (11). Consistent with these observations, depletion of Emg1 inhibits pre-rRNA processing in the pathway of 18S rRNA synthesis. Two-hybrid screening and co-immunoprecipitation revealed that Emg1 interacts

*To whom correspondence should be addressed. Tel: 33 169155047; Fax: 33 169853715; Email: nicolas.leulliot@u-psud.fr

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

with Nop14, which is required for nuclear localization of Emg1. The nucleolar localization of Nop14 is dependent on Noc4 and the Nop14–Noc4 complex is also required for biogenesis of the small ribosomal subunit (10,11).

Screens for suppressors of defects in *EMG1* have identified mutations in several genes encoding ribosome synthesis factors (12,13). In particular, suppression was conferred by deletion of the gene encoding the box C/D snoRNA snR57, which is responsible for methylation of G_{1570} in the 18S RNA (12). This suggested that either the loss of methylation at G_{1570} or the absence of snR57 binding at this site could compensate for the absence of Emg1. Over-expression of the ribosomal protein Rps19 also partially compensated for the loss of function of Emg1 (12), indicating that Emg1 may be involved in the recruitment and/or assembly of Rps19 on the pre-40S ribosome subunit. These observations correlate with evidence from a three-hybrid screen, showing that Emg1 interacts with an 18S rRNA sequence in a loop between helices 33 and 47 (12). This region contains G_{1570}, the target of snR57 methylation, and is in close proximity with the proposed Rps19-binding site. Finally, a multi-copy suppressor screen identified *SAM2* (14), which encodes *S*-adenosyl-methionine (SAM) synthase. This suggested that the Emg1 mutation was suppressed by a high SAM concentration, and that Emg1 is involved in a methylation reaction. An extensive bioinformatical analysis of the alpha/beta knot fold MTase proteins, predicted Emg1 as a potential member of the SPOUT superfamily (15).

Here we present the crystal structure of Emg1, and report that the protein belongs to the alpha/beta knot fold MTase superfamily. Emg1 defines a new structural class in this superfamily as it contains the core structure characteristic for members of the superfamily, but also carries two extra domains that are conserved in Emg1 orthologs, but absent from other members of the superfamily. These contribute to an extended surface that we predict to be involved in binding of the RNA substrate. We further show that Emg1 displays robust SAM-binding activity *in vitro*, but this activity is not required for ribosome biogenesis or growth *in vivo*.

MATERIALS AND METHODS

Cloning, expression and purification of recombinant Emg1

The *EMG1* ORF was amplified by PCR using genomic DNA of *S. cerevisiae* strain S288C as a template. An additional sequence coding for a 6-histidine tag was introduced at the 3′-end of the ORF during amplification. The PCR product was then cloned into a derivative of pET9 vector. Expression was done at 37°C using the *Escherichia coli* Rosetta strain. The His-tagged protein was purified on a Ni–NTA column (Qiagen Inc.) followed by gel filtration. Selenomethionine-substituted Emg1 was produced and purified as the native protein. For binding assays the *EMG1* ORF was cloned into a pQE80 (Qiagen) derivative for expression with an N-terminal His_{10}-zz- or His_{10}-tag. The protein was purified as described above.

Table 1. Data collection and refinement statistics

	Free Emg1	SAM bound Emg1
Wavelength (Å)	0.97953 (Se peak)	0.93100
Space group	$P2_12_12_1$	$P2_12_12_1$
Unit-cell parameters *a*, *b*, *c* (Å)	42.5 76.6 87.6	42.0 78.7 89.6
Resolution (Å)	19.0–1.94	34.2–2.0
Total number of reflections	69 469	91 942
Total number of unique reflections	20 207	20 687
Multiplicity	3.4 (1.7)	4.4 (4.6)
R_{merge}[a]	4.5 (14.2)	6.7 (44.2)
$I/\sigma(I)$	18.2 (4.0)	15.1 (3.0)
Overall completeness (%)	93.2 (61.1)	96.6 (100.0)
Resolution refinement (Å)	19.0–2.0	20.0–2.0
Reflections (working/test)	18 466/990	19 570/1055
R_{cryst}/R_{free}[b]	21.6/27.0	19.7/24.7
Non-hydrogen atoms	1941	1890
Water molecules	225	141
Bonds (Å)	0.007	0.016
Angles (°)	0.943	1.569
Mean B factor (Å^2) Protein/SAM	17.4	26.6/21.5
Ramachandran analysis Most-favored/allowed(%)	94.2/5.8	91.3/8.7

[a] $R_{merge} = \sum_h \sum_i |I_{hi} - \langle I_h \rangle| / \sum_h \sum_i I_{hi}$, were I_{hi} is the *i*th observation of the reflection *h*, while $\langle I_h \rangle$ is the mean intensity of reflection *h*.
[b] $R_{factor} = \sum ||F_o| - |F_c|| / |F_o|$. R_{free} was calculated with a set of randomly selected reflections (5%).

Crystallization of Emg1

Crystals of Emg1 were obtained from a 1:1 mixture of protein (10 mg/ml) with 25% PEG4K, 0.2 M Li2SO4, 0.1 M Tris pH7.5, 20% glycerol. Crystals were transferred to the precipitant solution with the glycerol concentration raised to 30% prior to flash freezing in liquid nitrogen. The selenomethionine-substituted protein crystals belong to the $P2_12_12_1$ space group with one molecule per asymmetric unit and diffracted to 2 Å on the BM30A beamline at the ESRF, and to 2 Å on the ID14-4 beamline for the SAM soaks. The phases were obtained by single anomalous dispersion collected on the Selenium peak. The Selenium sites were found using ShelX (16) and the phases were determined using Solve (17). Eighty percent of the protein was automatically rebuilt using Resolve (18) and manually completed using TURBO and O. A few regions of the protein were not visible in the electron density, probably due to the structural disorder. Analysis of the B-factors shows that the two Emg1-specific subdomains are less well ordered than the rest of the protein. The structural disorder is visible in the poorly defined βc′–βd′ loop and the missing eight residues in the βa′, βb′ loop. Two sulfate ions could also be modeled in the structure. Cell parameters and data collection statistics are reported in Table 1. The coordinates have been deposited at the PDB under the codes 2v3j and 2v3k.

Dimerization analysis, RNA and SAM-binding experiments

For gel filtration, marker proteins were dissolved and the Superose 12 column (GE Healthcare) was run in a buffer containing 50 mM Tris–HCl pH 7.5, 50 mM NaCl, 5 mM $MgCl_2$.

Binding assays with RNA and for dimerization were performed in the same buffer as above. For dimerization, HisZZ- and His-tagged WT and R88D Emg1 were incubated in the combinations shown in Figure 4B. After incubation, proteins were bound to IgG Sepharose (GE Healthcare), beads were washed, proteins eluted and separated by SDS–PAGE followed by Coomassie staining. For RNA binding, radiolabeled [^{32}P-adenosine] 35 nucleotide *in vitro* transcripts either containing nucleotides 1553–1576 of the 18S rRNA or an unrelated sequence (19) were generated. HisZZ-tagged WT or R88D Emg1 was incubated with the transcripts and proteins as well as bound RNA were retrieved on IgG Sepharose. After elution, RNA was extracted, separated on a 6% polyacrylamide/urea gel and analyzed by autoradiography and phosphoimaging.

For SAM binding, recombinant HisZZ-tagged WT and mutant Emg1 were immobilized on IgG Sepharose and incubated with [^{3}H-methyl] SAM in buffer containing 50 mM Tris–HCl pH 7.5, 100 mM NaCl, 5 mM $MgCl_2$ 5 mM β-mercaptoethanol. After extensive washing in binding buffer, elution was performed in buffer containing 80 mM Tris–HCl pH 6.8, 0.5 M sucrose, 4% SDS, 50 mM DTT. The eluates were directly used for scintillation counting in Ultima Gold XR (Perkin Elmer). The data obtained were normalized to recovery from the control (matrix without Emg1), which was set to 1.

Complementation and analysis of ribosome biogenesis

Wild-type *EMG1* (including 457 nt upstream and 301 nt downstream) was cloned BamHI–EcoRI into pRS413 for expression from its own promoter and the indicated mutations were made in the coding sequence by site-directed mutagenesis. Plasmids were transformed into the strain BY4741 (MATa; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0) with genomic *EMG1* under control of a *GAL1* promoter. Growth was measured in YP medium (1% yeast extract, 2% peptone) containing either 2% glucose for depletion of the endogenous copy or 2% galactose at 30°C (Figure 6A–C) or 25°C (Figure 6D–F). For analyses of ribosome biogenesis, total RNA was isolated, separated by gel electrophoresis, transferred by northern blot and probed for intermediates and mature rRNA essentially as described (20). Hybridization probes used are listed in Table 2.

Table 2. Probes used for northern hybridization

Name	Sequence
004	CGGTTTTAATTGTCCTA
005	ATGAAAACTCCACAGTG
007	CTCCGCTTATTGATATGC
008	CATGGCTTAATCTTTGAGAC
017	GCGTTCTTCATCGATGC
020	TGAGAAGGAAATGACGCT
041	CTACTCGGTCAGGCTC
306	GCATCTTACGATACCTG

RESULTS

Structure determination

The structure of *S. cerevisiae* Emg1 was determined by single anomalous dispersion on seleno-methionine substituted protein. The seleno-methionine crystals diffracted to 2Å resolution with one molecule in the asymmetric unit. The seleno-methionine crystals were also used for soaking experiments with SAM. Some regions of Emg1 were not visible in the electron density and are probably unstructured or flexible. The first region contains the 23 N-terminal residues, which are poorly conserved in eukaryotes and absent in archaeal Emg1 homologs. This region is evidently dispensable for the core structure of Emg1 but might be important for *in vivo* function. Another unresolved region comprises residues 56–64, which are present in a loop that is poorly conserved in length and sequence in higher Eukaryotes and Archaea.

The topology of the protein is shown in Figure 1A. The structure of Emg1 is built around a central eight-stranded β-sheet with α helical insertions (Figure 1B). From the topology of the protein and the presence of a deep trefoil knot (Figure 2), the protein can unambiguously be assigned to the α/β knot MTase (SPOUT) superfamily. Secondary structure elements that belong to the core α/β knot fold are colored in red, whereas Emg1-specific additions are shown in blue. Three α helices (αA, αE, αd′) pack on one face of the β sheet (from β1 to β6), forming the conserved structural core, while the Emg1-specific insertions (dark blue in Figure 1B) cluster on the side of the β sheet.

The core α/β knot fold is centered on a five-stranded parallel β-sheet with a strand order β2β1β5β4β6. The β3 strand, which extends the β-sheet after β2, is found in most members of the superfamily but Emg1 is the only case where this strand is antiparallel to β2. This can be attributed to the fact that the αB helix inserted between β2 and β3 is absent in Emg1, and replaced by a short hairpin loop in order to make room for the Emg1-specific insertions. Several α helical insertions are found in other α/β knot MTase superfamily members (21) but only αA (inserted between β1 and β2) and αE (at the C-terminus) are present in Emg1. The αC helix is replaced by a short 3-10 helix (HC).

Emg1 contains an additional prominent subdomain composed of two short β-strands (βa′, βb′), helix αa′ and a short 3-10 helix (Hb′), which is inserted between β1 and αA. A second insertion between β3 and HC is composed of αc′, βc′, βd′. The βc′, βd′ strands are antiparallel and extend the β-sheet after β3 to form a continuous eight-stranded β-sheet. Helix αc′ stacks against the βc′, βd′, β3 and β2 strands and αA helix forming extensive hydrophobic interactions. The antiparallel βa′ and βb′ strands are sandwiched between αa′ and αc′, so that the two insertions create a continuous surface on one side of the protein.

Emg1 is a novel member of the alpha/beta knot fold MTase family

Consistent with the topology and trefoil knot present in Emg1, a search for structural neighbors on the EBI SSM server (22) found hits with rRNA and tRNA MTase

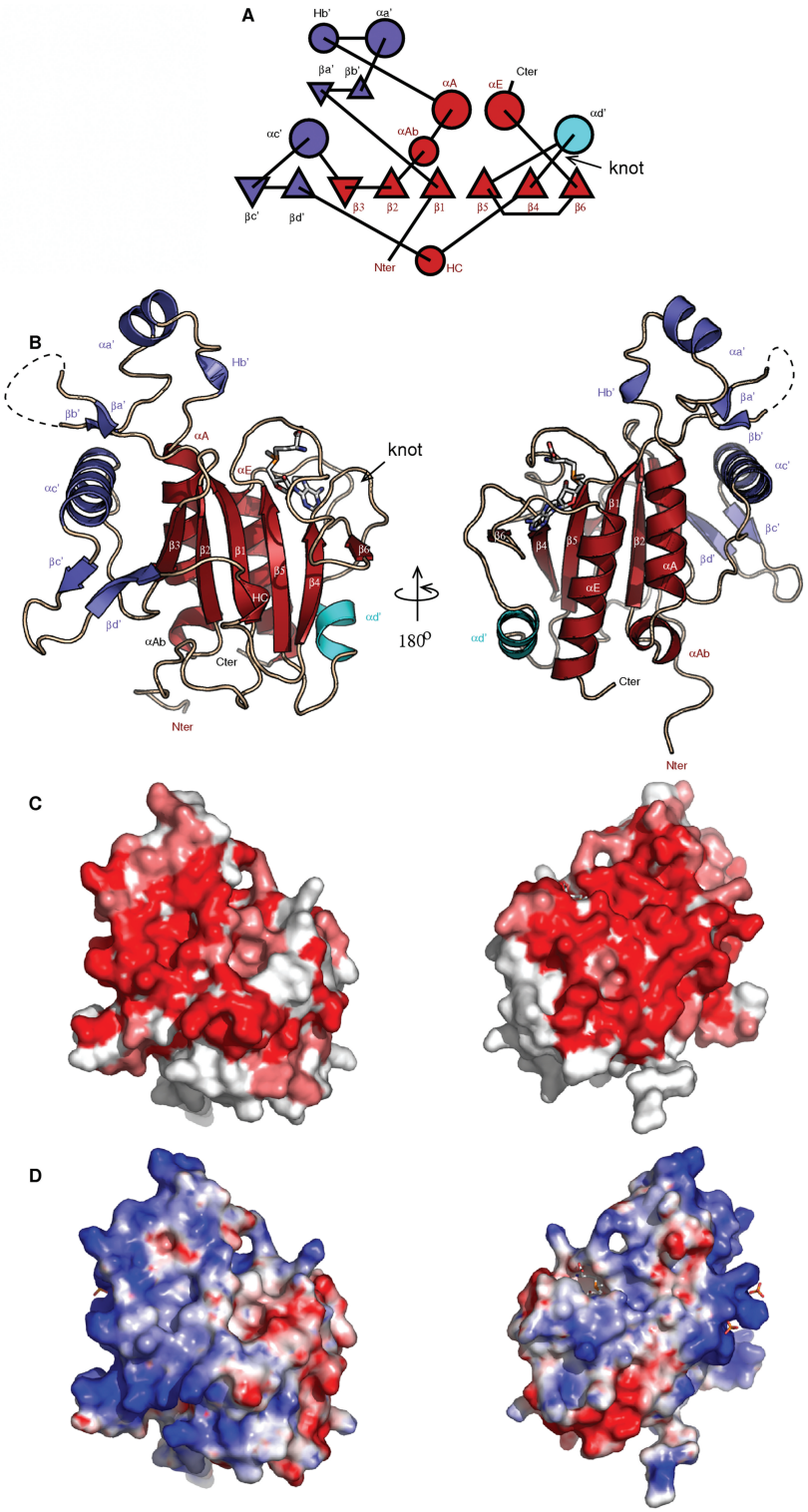

Figure 1. The structure of Emg1. (**A**) Topology diagram of the Emg1 protein. The secondary structure of the core alpha/beta knot fold of the methyltransferase is colored in red and the Emg1-specific insertions are in blue. The 3–10 helices are indicated by the capital letter H, while canonical helices and strands are indicated by Greek letters. (**B**) Ribbon representation of the Emg1 protein with the same color code as panel A. The SAM is shown in stick representation. The right view is turned by 180° in respect to the left view. (**C**) Surface representation of the Emg1 protein. The surface is colored in shades of red corresponding to sequence conservation, from white (less conserved) to dark red (most conserved). Only eukaryote sequences were included in the determination of the degree of conservation [calculated with rate4site (31)]. The two orientations are identical to those shown in panel B and the SAM is represented as sticks. (**D**) Electrostatic potential surface representation of the Emg1 protein. The two orientations are identical to panel B and C and the SAM is represented as sticks. The Emg1-specific insertion (blue in panel B) is highly conserved (panel C) and displays a strong positive surface, ideal for RNA binding. Two sulfates bound to this surface are shown in stick representation. The potential was calculated using MEAD (32).

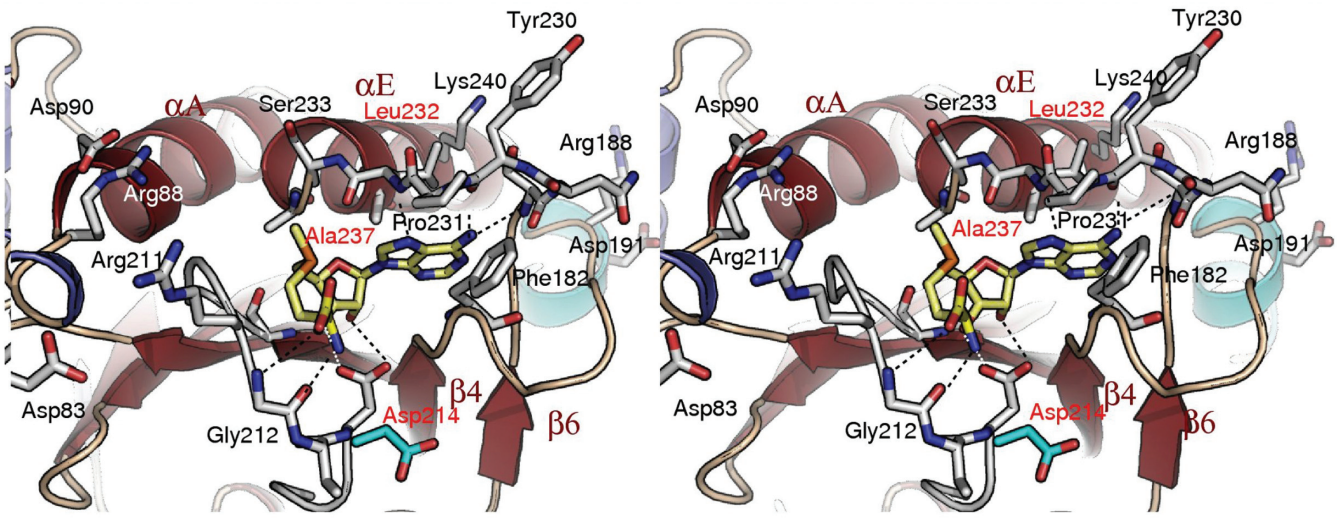

Figure 2. Emg1 SAM-binding site. Stereo representation of the SAM-binding site. SAM is shown in yellow. The most relevant residues of Emg1 are shown in stick representation. Hydrogen bonds are shown in dashed lines. The numbering of residues mutated below is colored in red. The conformation of Asp214 in the apo protein is shown in blue sticks.

proteins, notably the Rlmb 23S rRNA G_{2251} 2′*O*-methyltransferase (pdb code 1gz0, *Z*-score 5.9 and r.m.s.d. 2.64 Å for 119 aligned residues), TrmH tRNA (guanosine-2′*O*-) methyltransferase (pdb code 1v2x, *Z*-score 5.3 and r.m.s.d. 2.64 Å for 109 aligned residues), TrmD tRNA (guanosine-1) MTase (pdb code 1p9p, *Z*-score 4.1 and r.m.s.d. 2.55 Å for 101 aligned residues).

The core of Emg1 superimposes well with the same region of all members of the alpha/beta knot fold superfamily, whose core fold also comprises β1 to β6 and αA, αE and HC (regions colored in red in Figure 1A and B). However, Emg1 has a unique topology outside the core fold (15), and comparisons with all other alpha/beta knot fold proteins failed to identify proteins containing insertions structurally similar to Emg1. We therefore propose that Emg1 represents the founding member of a new subfamily of alpha/beta knot fold MTases.

Emg1 directly interacts with RNA

The Emg1 protein is conserved in Eukaryotes and Archaea. The sequence alignment is shown in Figure 3. In Figure 1C, the surface of Emg1 is colored in increasing shades of red reflecting residue conservation in Eukaryotes. Emg1 has an extended conserved surface patch, which includes surface residues on the entire circumference of the protein: the putative catalytic site, the αA and αE helical surface, the HC helical surface and the Emg1 specific insertions. The electrostatic potential mapped on the surface of the protein (shown in Figure 1D), shows that this conserved surface is mostly basic. The most basic region comprises the Emg1-specific insertions and extends to the putative catalytic site. This surface is therefore ideally suited to bind nucleic acids.

To confirm this hypothesis, the RNA-binding activity of Emg1 was analyzed *in vitro*. HisZZ-tagged Emg1 was incubated together with the *in vitro* transcribed 18S rRNA fragment (labeled 18S in Figure 4A) with which the protein was proposed to interact (12), or a transcript from pBluescript (labeled pBS in Figure 4A) (19). WT Emg1 retrieved both transcripts, but the RNA with the sequence of 18S was more enriched than the pBS transcript (Figure 4A).

Based on the crystal structure, Arg 88 is a residue of the Emg1-specific insertion, conserved in Eukaryotes and Archaea, which lies in the putative RNA-binding site (Figure 1D) near the SAM-binding site (Figure 2). Emg1 with an R88D mutation was strongly compromised for RNA binding on both substrates, with recovery reduced almost to background levels (see quantification in Figure 4A). These results confirmed the prediction that Arg 88 is involved in RNA substrate binding.

We also tested for the putative MTase activity of Emg1 *in vitro*, using the conditions under which RNA binding was observed. The wild-type protein and mutants lacking SAM-binding activity were incubated with the 18S and pBS RNA substrates in the presence of [^{3}H]-labeled SAM. Following incubation, the RNA was re-extracted and methylation was either directly analyzed by scintillation counting or the RNA separated on a polyacrylamide gel followed by northern transfer and PhosphorImager analysis. No [^{3}H]-methyl incorporation into the RNA substrates by wild-type or mutant Emg1 was detected under any condition tested (data not shown), which could indicate a lack of the correct substrate in the assay or that the protein is active only in the context of the pre-ribosome.

Dimerization of Emg1

Other members of the alpha/beta knot fold MTase superfamily have been shown to dimerize, and two different homodimerization modes were found in the structures of TrmD and TrmH (15). In these cases, dimerization always involves the αA and αE helical side of

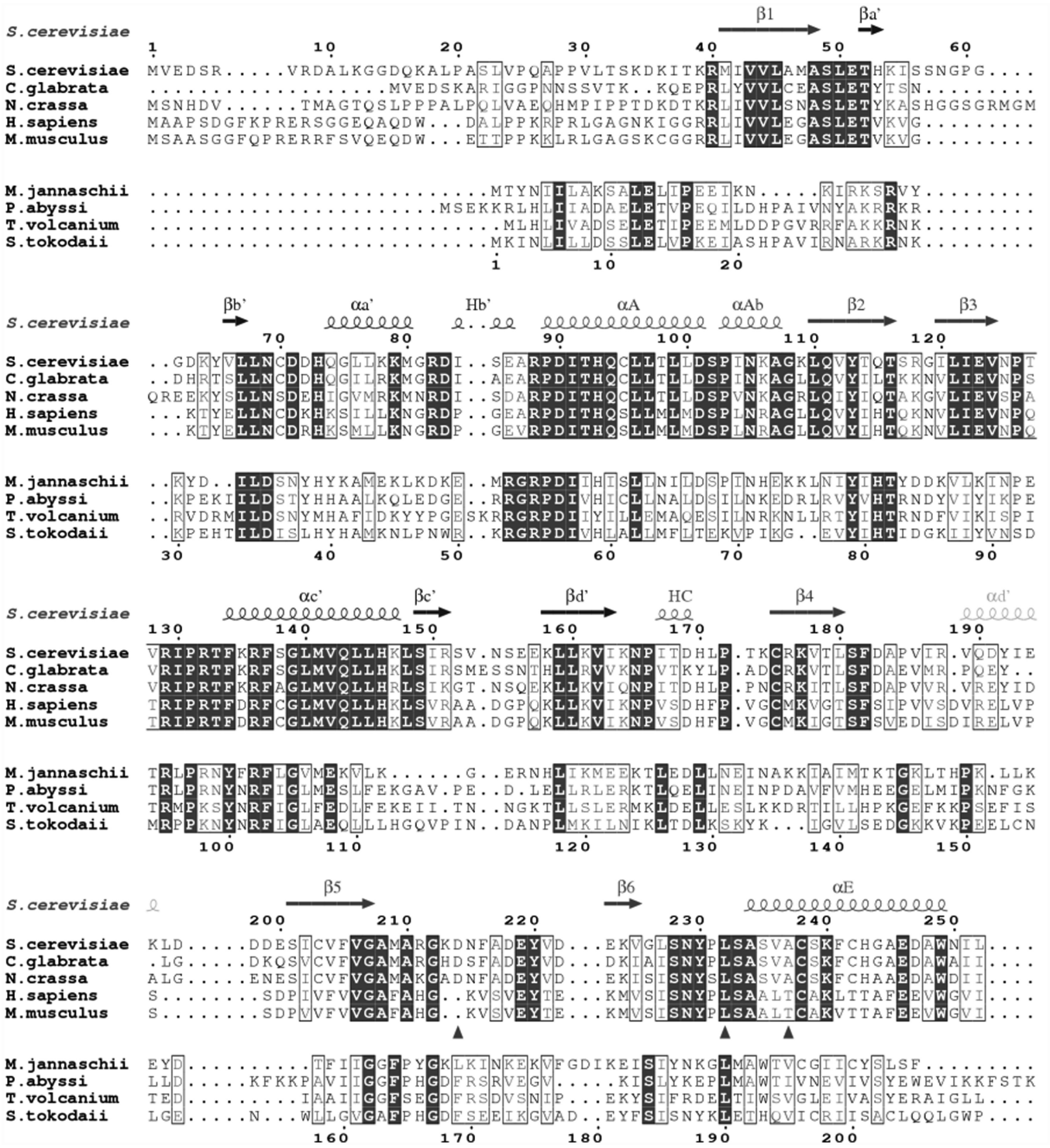

Figure 3. Sequence alignment of Emg1 proteins of two different phylogenetic groups. (1) Eukaryotes and (2) Archaea. The secondary structure of Emg1 is shown above the alignment. Residues mutated in the SAM-binding site are indicated with a triangle. The figure was generated with a special version of ESPript (33).

the protein, but the orientation of the two monomers varies. Dimerization provides an extended surface for RNA binding and completes the catalytic site, with the catalytic residue being provided by the opposing monomer (23,24). As the crystal structure did not provide evidence for a similar dimeric form for Emg1, dimerization was analyzed by binding experiments and gel filtration (Figure 4B and C). Two Emg1 fusion constructs were expressed separately; a fusion of Emg1 with two copies of the Z-domain of protein A and His_{10} (Hzz-Emg1) and

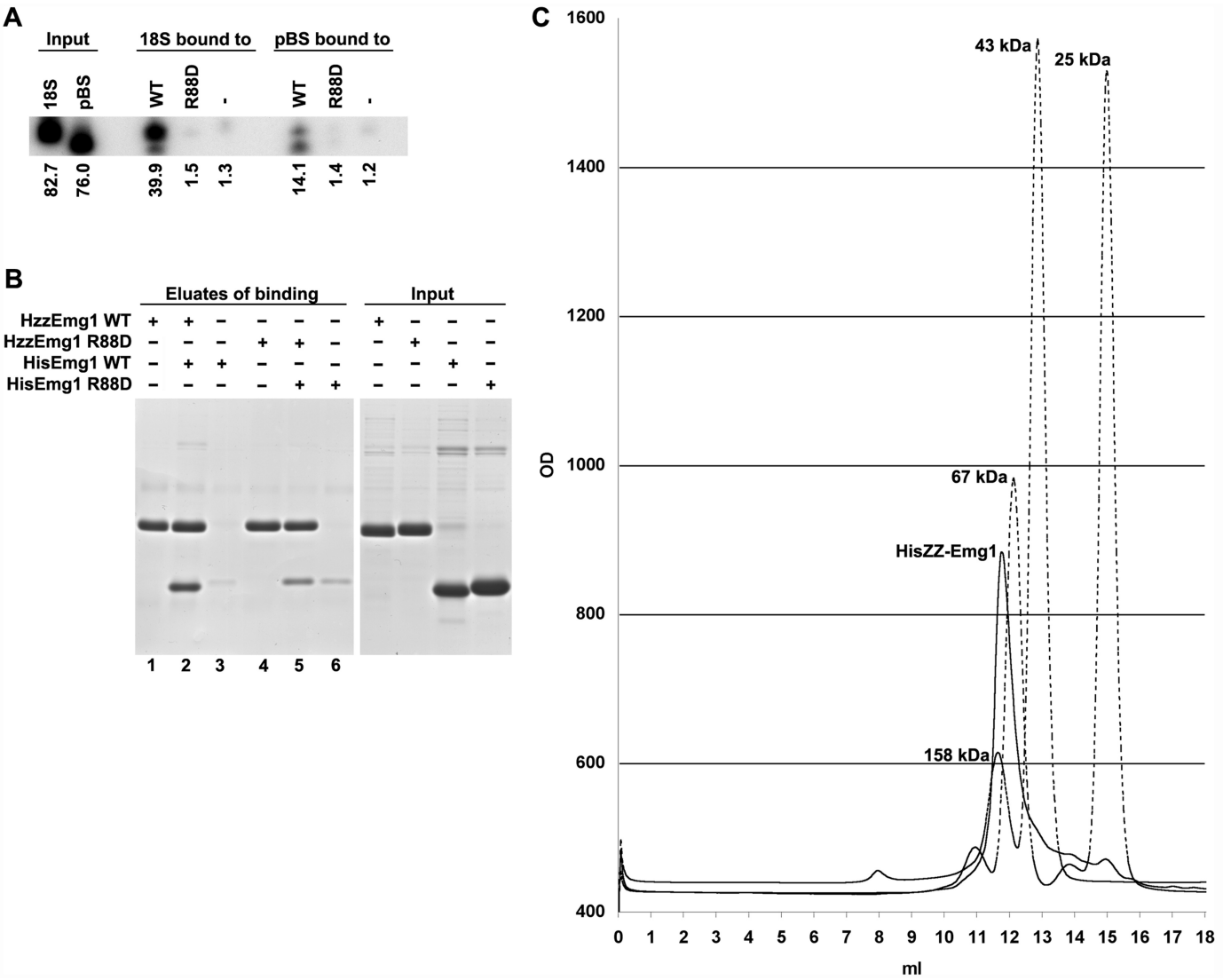

Figure 4. Emg1 dimerizes and directly binds RNA. (**A**) Emg1 binds RNA directly. *In vitro* transcribed, [^{32}P]-labeled RNA was incubated with HisZZ-tagged wild-type (WT) or R88D Emg1 and retrieved on IgG Sepharose. RNA was extracted from eluates, separated on a 6% polyacrylamide/urea gel and analyzed by autoradiography and phosphoimaging. An autoradiograph is shown with relative intensities measured by phosphoimager indicated below. While the WT protein binds RNA efficiently, binding is strongly compromised for the R88D mutant. (**B**) Dimerization of Emg1 analyzed by binding experiment. Recombinant HisZZ-tagged WT (lanes 1–3) or R88D Emg1 (lanes 4–6) was incubated with His-tagged Emg1 and retrieved on IgG Sepharose. Eluted samples were analyzed after SDS–PAGE and Coomassie staining, with inputs shown on the right. The HisZZ-tagged WT Emg1 pulls down the His-tagged fusion, showing dimerization of Emg1. The R88D mutation reduces dimerization. (**C**) Gel filtration to analyze the dimerization status of Emg1. Gel filtration was performed on a Superose 12 column with marker proteins of 25 (Chymotrypsinogen), 43 (Ovalbumin), 67 (Albumin), 158 kDa (Aldolase) and HisZZ-Emg1. An overlay of the elution profiles is shown. The fusion protein has a calculated molecular weight of ~46 kDa for the monomer, but the large majority of it elutes between the 158 and the 67 kDa markers, showing that it forms a homodimer in solution.

a fusion with Emg1 only a His_{10}-tag (His-Emg1). The two constructs were mixed and incubated together, and the Hzz-Emg1 was then specifically retrieved on IgG Sepharose. Co-precipitation of His-Emg1 with Hzz-Emg1 was clearly seen by SDS–PAGE (Figure 4B, lanes 1–3), showing that the differentially tagged variants formed heterodimers. We note that His-Emg1 carrying the R88D mutation resulted also in reduced dimerization (Figure 4B, lanes 4–6). The reduction in dimerization efficiency was, however, clearly less than the reduction in RNA binding, suggesting that it is not causal.

Dimerization of Emg1 was independently confirmed by gel filtration. Hzz-Emg1 was used in these analyses, as Emg1 alone lacks sufficient UV absorbing residues to give a strong signal. As shown in Figure 4C, Hzz-Emg1 elutes at an appropriate position for the dimer, between the 67 kDa (Albumin) and the 158 kDa (Aldolase) marker proteins. Only a small shoulder can be detected at the expected position for the monomer, slightly before the 43-kDa marker (Ovalbumin). Together, these data demonstrate that Emg1 forms a homodimer in solution, which is probably the active form of the protein.

Location of the catalytic site

All the members of the alpha/beta knot fold family are predicted or known to be SAM-dependent MTases. For several proteins, the location of the SAM-binding site was determined by co-crystallization with either SAM or SAH. In all solved complexes, the SAM molecule binds in the knotted region [for example, YibK (25) and TrmD (23)]. The particular structure of the knot is at the center of the catalytic site of the enzyme and therefore necessary for its function.

Superposition of Emg1 on the structure of knotted MTase proteins in complex with SAH or SAM revealed that Emg1 contains an appropriate cavity to accommodate the methyl-donor. We confirmed that Emg1 binds SAM at this site by binding experiments (see below) and by soaking Emg1 crystals in mother liquor containing 2 mM SAM and solving the structure of the Emg1/SAM binary complex (Figure 2). The residual electron density unambiguously showed the presence of SAM in the expected pocket. The SAM-binding site in Emg1 is located in the knotted region and is composed of residues from the β4-αd′, β5-β6 and β6-αE loops (Figure 2). The adenine base is deeply buried in a cavity lined with hydrophobic residues. The ceiling of the cavity is closed by stacking interactions between Phe182 from the β4-αd′ loop and Pro231 from the β6-αE loop. The transferred methyl group of SAM is oriented towards the αA and αE face. Comparison of Apo Emg1 and SAM-bound Emg1 showed only limited structural rearrangements; a slight movement of the β5–β6 loop and a 180° reorientation of the Asp214 side chain which swings around to form a hydrogen bond with the 3′O of the SAM ribose moiety and a salt bridge with the nitrogen of the SAM methionine moiety (Figure 2). Although Asp214 is conserved in fungal Emg1 homologs, it is absent in mammals and Archaea. Six other hydrogen bonds are formed with the SAM moiety, all involving backbone atoms of Gly207, Gly212, Leu232, Ser228 and Tyr230. All of these residues are conserved in Eukaryotes and the glycines are also conserved in Archaea.

Mutations in the catalytic site of Emg1 abolish SAM binding

Based on the crystal structure (Figures 1 and 2), residues around the predicted catalytic site were selected to generate Emg1 mutants that are defective in the binding of SAM. As the SAM-binding pocket is largely lined by hydrophobic residues, single mutations were chosen that introduce longer side chains which protrude into the cavity or bring in charges. All hydrogen bonds to SAM are formed with the peptide backbone, with the exception of Asp214 (Figure 2), which was mutated to arginine for charge inversion. The mutations analyzed were D214R, L232D, L232S, A237D and A237R. All recombinant proteins except L232D were soluble and remained in the supernatant during ultracentrifugation (data not shown), suggesting that general folding was not affected by the other mutations. The mutations D214R, L232S and A237D (labeled in red in Figure 2) were chosen for further analysis.

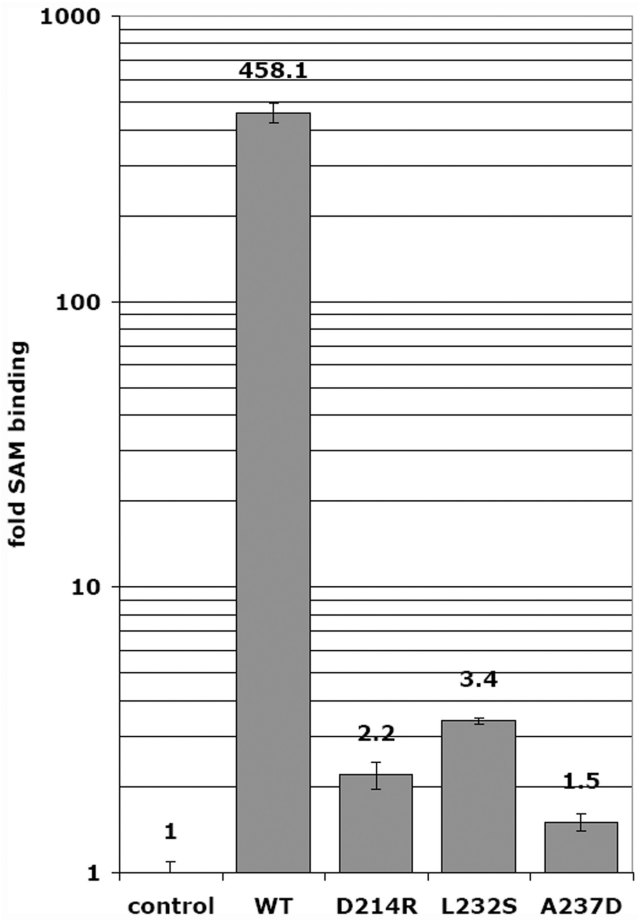

Figure 5. Mutations in residues flanking the hydrophobic cavity in Emg1 abolish SAM binding. Recombinant wild-type Emg1 and mutant proteins were immobilized and incubated with [^{3}H] SAM. After washing, bound fractions were eluted and the recovery of [^{3}H] SAM was determined by scintillation counting. The results were normalized to the control (matrix without Emg1). All Emg1 mutants analyzed had almost entirely lost the ability to bind SAM.

The ability of the WT and mutant forms Emg1 to bind [^{3}H]-labeled SAM was analyzed *in vitro* (Figure 5). Mutation of residues lining the SAM-binding site resulted in almost complete loss of SAM binding. This strongly indicates that these mutants will be catalytically inactive, since SAM is required as the methyl-donor for the MTase reaction.

Emg1 mutants that fail to bind SAM can support growth

A yeast strain was generated in which the genomic *EMG1* was under the control of the repressible *GAL1* promoter (see Materials and Methods section). Plasmids that expressed either wild-type Emg1 or one of the mutants defective in binding of SAM (Figure 6A–C) or RNA (Figure 6D–F) were introduced into this strain. The ability of these constructs or the empty plasmid (vector) to support growth was tested following depletion of Emg1 by transfer to glucose medium. Serial dilutions of the yeast cultures were spotted on solid medium containing

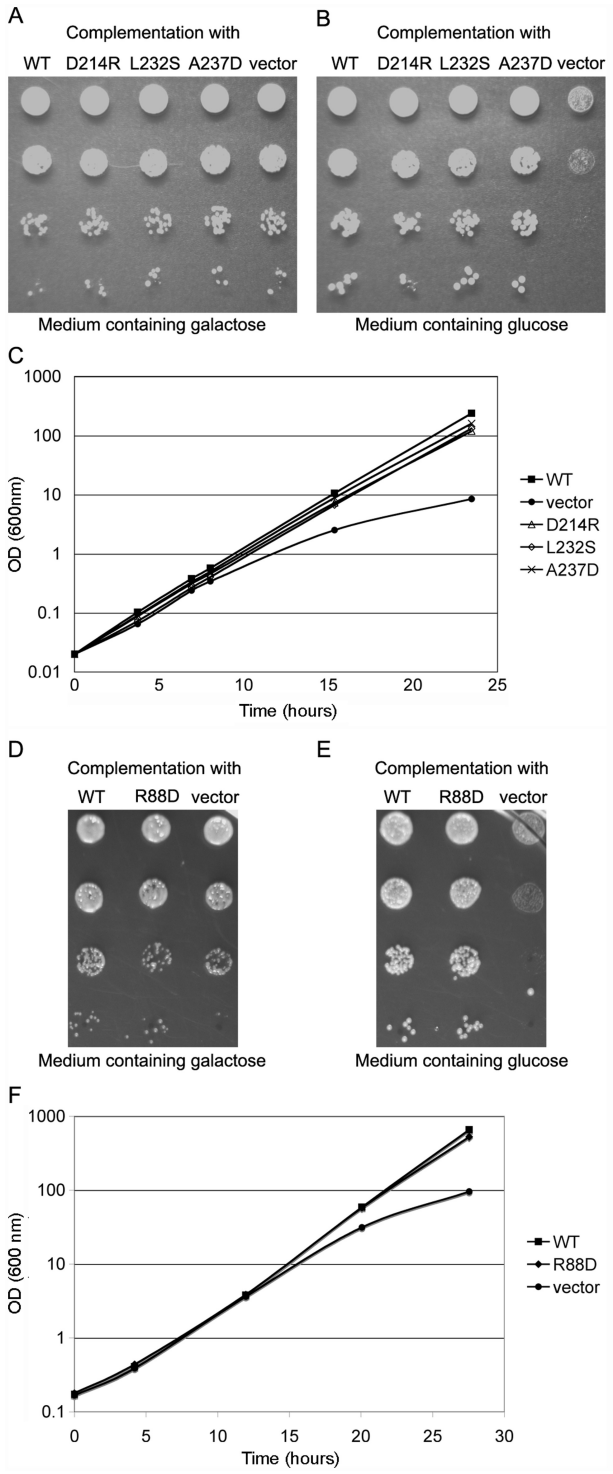

Figure 6. Emg1 mutants can complement for growth upon depletion of endogenous Emg1. Growth complementation in a background where the genomic *EMG1* gene is under control of the *GAL1* promoter was tested with plasmid-derived wild-type or mutant Emg1, or empty plasmid. Three SAM-binding mutants (**A**–**C**) and the RNA-binding mutant R88D (**D**–**F**) were tested for growth complementation. Dilutions of cultures were spotted on plates containing galactose (A and D) or under depletion conditions on glucose (B and E). Growth of the strains was also analyzed in liquid culture (C and F). Both wild-type Emg1 as well as the RNA- and SAM-binding defective mutants complement the growth phenotype of Emg1 depletion, whereas cells containing only the empty plasmid show strong growth impairment.

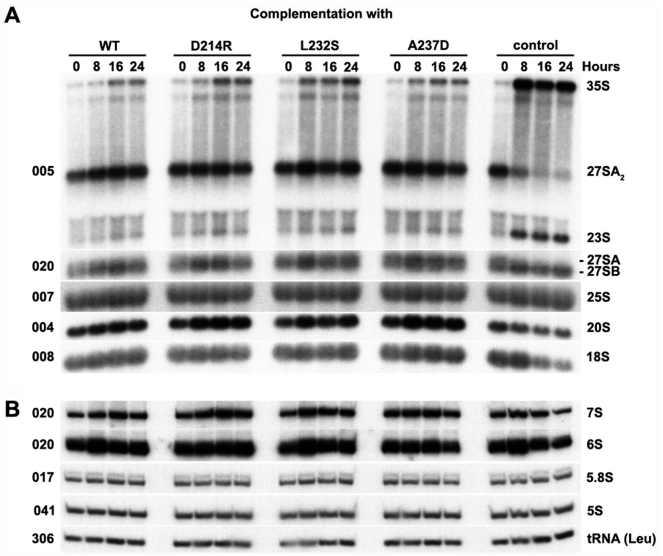

Figure 7. Mutations in the SAM-binding site of Emg1 do not affect ribosome biogenesis. As described in Figure 6, strains depleted of Emg1 were complemented with plasmid-derived wild-type or mutant Emg1, or with the empty vector (control). Total RNA was isolated before depletion (0 h) and after Emg1 depletion for 8, 16 and 24 h. RNA was separated on 1.2% agarose glyoxal gels (**A**) or 8% polyacrylamide/urea gels (**B**), and analyzed by northern hybridization. Probe numbers are shown on the left of the corresponding panels. rRNA intermediates and mature rRNAs are labeled on the right. In the control strain, depletion of Emg1 leads to strong ribosome biogenesis defects, whereas the strains complemented with wild-type Emg1 or the mutants do not show such effects.

galactose (Figure 6A and D) or glucose (Figure 6B and E). Expression of plasmid-borne, wild-type Emg1 or any of the mutants tested fully complement for growth, whereas cells containing the empty vector showed strong growth inhibition. These results were confirmed in liquid culture in the presence of glucose (Figure 6C and F). Control cells carrying only the vector showed a clear growth phenotype upon depletion of endogenous Emg1 in glucose-containing medium, whereas cells expressing wild-type or mutant Emg1 showed no growth defects.

The catalytic activity of Emg1 is not required for ribosome biogenesis in yeast

Emg1 is required for ribosome synthesis in yeast (10,11,14), and the mutants were also tested for complementation of the ribosome biogenesis defect. Cells were grown in galactose medium and transferred to glucose medium to allow depletion of Emg1. Samples were taken for analysis during growth on galactose (0 h sample) and 8, 16 and 24 h after transfer to glucose. Total RNA was isolated and analyzed for the levels of rRNA and pre-rRNA by northern hybridization (Figure 7). In the control cells transformed with the empty vector clear accumulation of 35S and 23S was observed, together with depletion of $27SA_2$ and 20S pre-rRNA and mature 18S rRNA. Notably, all mutants defective in SAM binding fully complemented the pre-rRNA and rRNA defects seen on Emg1 depletion (Figure 7). This strongly suggests that the

proposed SAM-dependent MTase activity of Emg1 is not required for ribosome biogenesis, even though the presence of the protein is essential.

DISCUSSION

The crystal structure of Emg1 revealed that it is a novel member of the alpha/beta knot fold MTase superfamily (Figure 1). The alpha/beta knot fold in SCOP (26) currently encompasses five families: (i) The YbeA-like family (pdb code 1ns5), with four uncharacterized members, which do not contain any specific additions to the core fold. (ii) The TrmD tRNA (m1G37)-MTases methylate the N1 position of guanine in tRNA. Their fold is similar to the YbeA-like family but contains an all-alpha C-terminal subdomain. (iii) The SpoU-like RNA 2′-*O* ribose MTase family contains both tRNA (TrmH) and rRNA (RlmB) MTases, which methylate the ribose 2′-OH of guanosine at specific positions of tRNA or rRNA, respectively. RlmB contains an N-terminal domain with analogy to the L30 ribosomal protein linked to the catalytic subunit by a flexible linker. (iv) The archaeal *Methanobacterium thermoautotrophicum* MT1 protein family, which is unique due to the insertion of an OB fold subdomain between β2 and β3. (v) The YggJ-like family that contains a N-terminal L5 ribosomal like domain. The core of Emg1 could be superimposed well with the core region of all members of the alpha/beta knot fold MTase superfamily. However, the Emg1 structure shows insertions that make it clearly distinct from each of these families. We therefore confirmed the prediction that Emg1 is the founding member of an additional alpha/beta knot fold MTase subfamily (15).

The Emg1-specific insertions consist of two subdomains that form an extended surface on one side of the protein. This contributes to a basic patch around the active site, which is very likely involved in substrate rRNA binding (Figure 1D), and Emg1 was indeed shown to bind RNA directly (Figure 4C). Mutation of Arginine 88 (R88D), which is located within the basic patch, almost completely abolishes the interaction with the RNA, supporting the proposed function of this region in substrate binding. This conclusion is further supported by the finding of two sulfate ions in the structure that could mimic phosphates of the RNA backbone. The two sulfates are bound to the protein by salt bridges to Lys135/Arg136 and Arg136/Arg129. These three residues are located on the Emg1-specific region, αc′ helix and αc′-β3 loop and Arg136/Arg129 are absolutely conserved in both Eukaryotes and Archaea. This strongly suggests that this extended Emg1-specific region is also involved in RNA binding.

The results from the RNA-binding experiments are consistent with the findings of Buchhaupt *et al.* (2006), where a modified yeast three-hybrid approach was performed to identify a sequence derived from the 18S rRNA that was bound by Emg1 with high affinity. Our *in vitro* analyses showed that Emg1 binds to this 18S sequence more strongly than an RNA of unrelated sequence. However, *in vitro* binding was clearly not highly specific.

There is a striking difference between the sequence conservation of Emg1 in Eukaryotes (on average 50% identical) versus that in Archaea (on average 30% identical) (Figure 3). The alignment suggests that the archaeal and eukaryotic Emg1 orthologs have the same characteristic fold but constitute two separate families, which might indicate that Emg1 has overlapping but not identical functions in both kingdoms of life. Surface residues conserved between Eukaryotes and Archaea cluster on one side of the protein: the SAM-binding site, the region immediately proximate to the methyl donor, the αA and αE face of the protein and the Emg1-specific extensions (Figure 1C, left panel). In contrast, residues on the opposite side of the protein (Helix HC side, Figure 1C right panel) are much less conserved between Eukaryotes and Archaea. While the αA and αE conserved face of Emg1 is likely involved in dimerization and/or substrate binding, the opposite, less conserved face, may interact with additional factors that are not conserved between Eukaryotes and Archaea. A candidate for such a factor could be Nop14, which binds to Emg1 in yeast (10) but has no clear archaeal homolog.

A feature shared between all previously characterized members of the alpha/beta knot MTase families is the presence of dimers in the crystal forms, and some members were confirmed to form dimers in solution (27). The detailed dimerization mode varies, but always involves the αA and αE helical surface of the protein. Moreover, mutants of the SpoU family protein Yibk, engineered to be monomeric rather than dimeric, were unable to bind SAH, suggesting that dimerization is required for structure and function of alpha-beta knot MTase (28). In contrast, Emg1 crystallized as a monomer, possibly due to differences in the dimerization interface, the composition of the active site or disruption of the dimer by the crystallization liquor. Stable dimerization was, however, confirmed for Emg1 in solution, both by gel filtration and binding experiments (Figure 4).

Based on the structure of the SAM-binding region of Emg1, site-directed mutagenesis was performed to obtain mutants defective in SAM binding. All mutants analyzed were found to be defective in binding SAM *in vitro* (Figure 5), confirming predictions based on the structural data. The almost complete loss of affinity for SAM, the MTase cofactor, strongly indicates that these mutants have lost the capacity to perform a MTase reaction.

Emg1 is essential for viability in yeast, and we assessed whether the essential function requires its MTase activity or RNA binding. Three *emg1* mutants that had lost SAM binding as well as the R88D mutant that lacks RNA-binding activity were able to fully complement the growth defect seen upon depletion of endogenous Emg1, suggesting that the presence of the catalytically inactive protein was sufficient to support growth (Figure 6). Analysis of ribosome biogenesis in the SAM-binding mutants revealed that expression of only the inactive forms of Emg1 also supported normal pre-rRNA processing and ribosome synthesis (Figure 7). We conclude that the putative SAM-dependant MTase activity of Emg1 is not required for ribosome biogenesis or growth in yeast.

Similar results were previously obtained for the MTase Dim1 and for methylation guide snoRNPs. Dim1 dimethylates two adjacent adenosine residues in the 18S rRNA (9). Analyses of Dim1 showed that, as for Emg1, the presence of catalytically inactive protein was sufficient for ribosome biogenesis (9). This leads to two possible interpretations. Both MTases may play structural roles in the pre-ribosome that are distinct from their enzymatic activities. Alternatively, or in addition, quality control mechanisms may have evolved to check for the presence of the protein rather than for the successful modification of the rRNA. Methylation guide snoRNPs carry out 2′-OH methylation at several sites on the rRNA. Strains lacking the activity of any individual methylation guide snoRNA were also viable (29), whereas loss of the activity of the snoRNA-associated MTase Nop1, was lethal (30). These data reveal that loss of rRNA methylation at any single site tested has little effect on ribosome biogenesis or function in yeast, whereas loss of methylation at multiple sites impairs ribosome synthesis and/or translation.

ACKNOWLEDGEMENTS

We would like to thank P. Gouet for a special version of ESPript used in Figure 3. Work at IBBMC-Université Paris-Sud was supported by the European 3D-repertoire program (LSHG-CT-2005-512028). M.T.B. was funded by a FEBS Long term fellowship. D.T. is funded by the Wellcome Trust. Funding to pay the Open Access publication charges for this article was provided by BIORIB grant (BLAN07-1_194553) from the Agence Nationale de la Recherche (ANR Blanche).

Conflict of interest statement. None declared.

REFERENCES

1. Venema,J. and Tollervey,D. (1999) Ribosome synthesis in Saccharomyces cerevisiae. *Annu. Rev. Genet.*, **33**, 261–311.
2. Fromont-Racine,M., Senger,B., Saveanu,C. and Fasiolo,F. (2003) Ribosome assembly in eukaryotes. *Gene*, **313**, 17–42.
3. Wade,C.H., Umbarger,M.A. and McAlear,M.A. (2006) The budding yeast rRNA and ribosome biosynthesis (RRB) regulon contains over 200 genes. *Yeast*, **23**, 293–306.
4. Granneman,S. and Baserga,S.J. (2004) Ribosome biogenesis: of knobs and RNA processing. *Exp. Cell Res.*, **296**, 43–50.
5. Bachellerie,J.P., Cavaille,J. and Huttenhofer,A. (2002) The expanding snoRNA world. *Biochimie*, **84**, 775–790.
6. Ganot,P., Bortolin,M.L. and Kiss,T. (1997) Site-specific pseudouridine formation in preribosomal RNA is guided by small nucleolar RNAs. *Cell*, **89**, 799–809.
7. Kiss-Laszlo,Z., Henry,Y., Bachellerie,J.P., Caizergues-Ferrer,M. and Kiss,T. (1996) Site-specific ribose methylation of preribosomal RNA: a novel function for small nucleolar RNAs. *Cell*, **85**, 1077–1088.
8. Lafontaine,D., Delcour,J., Glasser,A.L., Desgres,J. and 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.
9. Lafontaine,D.L., Preiss,T. and Tollervey,D. (1998) Yeast 18S rRNA dimethylase Dim1p: a quality control mechanism in ribosome synthesis? *Mol. Cell. Biol.*, **18**, 2360–2370.
10. Liu,P.C. and Thiele,D.J. (2001) Novel stress-responsive genes EMG1 and NOP14 encode conserved, interacting proteins required for 40S ribosome biogenesis. *Mol. Biol. Cell*, **12**, 3644–3657.
11. Bernstein,K.A., Gallagher,J.E., Mitchell,B.M., Granneman,S. and Baserga,S.J. (2004) The small-subunit processome is a ribosome assembly intermediate. *Eukaryot. Cell*, **3**, 1619–1626.
12. Buchhaupt,M., Meyer,B., Kotter,P. and Entian,K.D. (2006) Genetic evidence for 18S rRNA binding and an Rps19p assembly function of yeast nucleolar protein Nep1p. *Mol. Genet. Genomics*, **276**, 273–284.
13. Buchhaupt,M., Kotter,P. and Entian,K.D. (2007) Mutations in the nucleolar proteins Tma23 and Nop6 suppress the malfunction of the Nep1 protein. *FEMS Yeast Res*, **7**, 771–781.
14. Eschrich,D., Buchhaupt,M., Kotter,P. and Entian,K.D. (2002) Nep1p (Emg1p), a novel protein conserved in eukaryotes and archaea, is involved in ribosome biogenesis. *Curr. Genet.*, **40**, 326–338.
15. Tkaczuk,K.L., Dunin-Horkawicz,S., Purta,E. and Bujnicki,J.M. (2007) Structural and evolutionary bioinformatics of the SPOUT superfamily of methyltransferases. *BMC Bioinformatics*, **8**, 73.
16. Uson,I. and Sheldrick,G.M. (1999) Advances in direct methods for protein crystallography. *Curr. Opin. Struct. Biol.*, **9**, 643–648.
17. Terwilliger,T.C. and Berendzen,J. (1999) Automated MAD and MIR structure solution. *Acta Cryst. D*, **55**, 849–861.
18. Terwilliger,T.C. (1999) Reciprocal-space solvent flattening. *Acta Crystallogr. D Biol. Crystallogr.*, **55**, 1863–1871.
19. Mitchell,P., Petfalski,E., Shevchenko,A., Mann,M. and Tollervey,D. (1997) The exosome: a conserved eukaryotic RNA processing complex containing multiple 3′–>5′ exoribonucleases. *Cell*, **91**, 457–466.
20. Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: a Laboratory Manual*. Cold Spring Harbor, New York.
21. Ahn,H.J., Kim,H.W., Yoon,H.J., Lee,B.I., Suh,S.W. and Yang,J.K. (2003) Crystal structure of tRNA(m1G37)methyltransferase: insights into tRNA recognition. *EMBO J.*, **22**, 2593–2603.
22. Krissinel,E. and Henrick,K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 2256–2268.
23. Elkins,P.A., Watts,J.M., Zalacain,M., van Thiel,A., Vitazka,P.R., Redlak,M., Andraos-Selim,C., Rastinejad,F. and Holmes,W.M. (2003) Insights into catalysis by a knotted TrmD tRNA methyltransferase. *J. Mol. Biol.*, **333**, 931–949.
24. Watanabe,K., Nureki,O., Fukai,S., Ishii,R., Okamoto,H., Yokoyama,S., Endo,Y. and Hori,H. (2005) Roles of conserved amino acid sequence motifs in the SpoU (TrmH) RNA methyltransferase family. *J. Biol. Chem.*, **280**, 10368–10377.
25. Lim,K., Zhang,H., Tempczyk,A., Krajewski,W., Bonander,N., Toedt,J., Howard,A., Eisenstein,E. and Herzberg,O. (2003) Structure of the YibK methyltransferase from Haemophilus influenzae (HI0766): a cofactor bound at a site formed by a knot. *Proteins*, **51**, 56–67.
26. Murzin,A.G., Brenner,S.E., Hubbard,T. and Chothia,C. (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.*, **247**, 536–540.
27. Nureki,O., Watanabe,K., Fukai,S., Ishii,R., Endo,Y., Hori,H. and Yokoyama,S. (2004) Deep knot structure for construction of active site and cofactor binding site of tRNA modification enzyme. *Structure*, **12**, 593–602.
28. Mallam,A.L. and Jackson,S.E. (2007) The dimerization of an alpha/beta-knotted protein is essential for structure and function. *Structure*, **15**, 111–122.
29. Kiss,T. (2002) Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular functions. *Cell*, **109**, 145–148.
30. Tollervey,D., Lehtonen,H., Jansen,R., Kern,H. and Hurt,E.C. (1993) Temperature-sensitive mutations demonstrate roles for yeast fibrillarin in pre-rRNA processing, pre-rRNA methylation, and ribosome assembly. *Cell*, **72**, 443–457.
31. Mayrose,I., Graur,D., Ben-Tal,N. and Pupko,T. (2004) Comparison of site-specific rate-inference methods for protein sequences: empirical Bayesian methods are superior. *Mol. Biol. Evol.*, **21**, 1781–1791.
32. Bashford,D. (1997) In Yutaka Ishikawa,R.R.O., John,V.~W. Reynders and Marydell,T. (eds), *Lecture Notes in Computer Science,* Springer, Berlin, Vol. 1343, pp. 233–240.
33. Gouet,P., Courcelle,E., Stuart,D.I. and Metoz,F. (1999) ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics*, **15**, 305–308.