doi:10.1016/j.jmb.2006.05.047 J. Mol. Biol. (2006) 360, 774–787

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

The Structure of the RNA m^{5}C Methyltransferase YebU from *Escherichia coli* Reveals a C-terminal RNA-recruiting PUA Domain

B. Martin Hallberg[1]†, Ulrika B. Ericsson[1]†, Kenneth A. Johnson[1,2] Niels Møller Andersen[3], Stephen Douthwaite[3], Pär Nordlund[1] Albert E. Beuscher IV[1,2] and Heidi Erlandsen[1,2]*

[1]*Department of Medical Biochemistry and Biophysics Karolinska Institute 17177 Stockholm, Sweden*

[2]*Department of Biochemistry and Biophysics, The Arrhenius Laboratories for Natural Sciences, Stockholm University 10691 Stockholm, Sweden*

[3]*Department of Biochemistry and Molecular Biology University of Southern Denmark Campusvej 55 DK-5230 Odense M, Denmark*

**Corresponding author*

Nucleotide methylations are the most common type of rRNA modification in bacteria, and are introduced post-transcriptionally by a wide variety of site-specific enzymes. Three 5-methylcytidine (m^{5}C) bases are found in the rRNAs of *Escherichia coli* and one of these, at nucleotide 1407 in 16 S rRNA, is the modification product of the methyltransferase (MTase) YebU (also called RsmF). YebU requires *S*-adenosyl-L-methionine (SAM) and methylates C1407 within assembled 30 S subunits, but not in naked 16 S rRNA or within tight-couple 70 S ribosomes. Here, we describe the three-dimensional structure of YebU determined by X-ray crystallography, and we present a molecular model for how YebU specifically recognizes, binds and methylates its ribosomal substrate. The YebU protein has an N-terminal SAM-binding catalytic domain with structural similarity to the equivalent domains in several other m^{5}C RNA MTases including RsmB and PH1374. The C-terminal one-third of YebU contains a domain similar to that in *p*seudo*u*ridine synthases and *a*rchaeosine-specific transglycosylases (PUA-domain), which was not predicted by sequence alignments. Furthermore, YebU is predicted to contain extended regions of positive electrostatic potential that differ from other RNA-MTase structures, suggesting that YebU interacts with its RNA target in a different manner. Docking of YebU onto the 30 S subunit indicates that the PUA and MTase domains make several contacts with 16 S rRNA as well as with the ribosomal protein S12. The ribosomal protein interactions would explain why the assembled 30 S subunit, and not naked 16 S rRNA, is the preferred substrate for YebU.

Keywords: m^{5}C RNA methyltransferase; structure; SAD phasing; rRNA modification; PUA domain

Introduction

The positions and types of modification in bacterial rRNAs have been well characterized, although our knowledge regarding their function and the enzymes that catalyze the modifications is less complete. The most common bacterial rRNA modification is nucleotide methylation, and in *Escherichia coli* somewhat less than half of the enzymes responsible for these site-specific modifications have been identified.[1] There are 24 methylation sites in the rRNAs within mature *E. coli* ribosomes, and three of these are 5-methylcytidine (m^{5}C) located at m^{5}C967 and m^{5}C1407 in 16 S rRNA and at m^{5}C1962 in 23 S rRNA. The m^{5}C967 modification is catalyzed specifically by the *S*-adenosyl-L-methionine (SAM)-dependent methyltransferase (MTase) RsmB, which was the first rRNA MTase of this type to be identified.[2]

† B.M.H. and U.B.E. contributed equally to this work.

Abbreviations used: MTase, methyltransferase; SAM, *S*-adenosyl L-methionine; SAH, *S*-adenosyl homocysteine; m^{5}C, 5-methylcytidine; m^{5}U, 5-methyl uridine; SAD, single anomalous dispersion; rmsd, root-mean-square deviation; CYT, cytosine; PUA, *p*seudo*u*ridine synthases and *a*rchaeosine-specific transglycosylases; r-protein, ribosomal protein.

E-mail address of the corresponding author: erlandsen@dbb.su.se

Using RsmB as a search-probe led to the identification of more than 55 putative RNA m^5C MTases.[3] The RsmB homologs were divided into eight subfamilies that are distinguished by their N-terminal and/or C-terminal extensions. The variable N-terminal and/or C-terminal domains in the different subfamilies were proposed to serve as RNA-binding modules, although the mechanisms by which this could be achieved were not known.[3] Subfamily I, of which RsmB is the founding member, is the largest subfamily, presently with over 20 bacterial enzymes identified. Subfamily II contains Nop2p of *Saccharomyces cerevisiae* and the human proliferation-associated nucleolar antigen p120; both of these proteins are necessary for large ribosomal subunit assembly.[4] Subfamily III contains the *E. coli* protein YebU, recently renamed RsmF, and orthologs of this MTase are evident in a wide range of bacteria.[1] The wide phylogenetic distribution of the enzymes in subfamily III, together with the loss of fitness associated with YebU inactivation in *E. coli*,[1] suggest that these enzymes have important conserved functions.

Two RNA m^5C MTase structures have been solved: RsmB from *E. coli*,[5] and PH1374 from *Pyrococcus horikoshii*.[6] The structure of RsmB has been determined both in the unliganded and SAM-bound forms; however, no m^5C MTase has been solved in a complex with its RNA substrate. It would be of particular interest to determine how the highly conserved MTase domains recognize and methylate specific sites in the bacterial rRNA. Here, we present the structure of YebU at 2.9 Å resolution. We use the YebU structure to propose the most likely mechanism for its substrate specificity and recognition, as well as how the enzyme catalyses m^5C formation by a SAM-dependent mechanism. The YebU structure contains a novel C-terminal PUA-like domain, reminiscent of the PUA domains found in the RNA-modifying pseudouridine synthases and archaeosine-specific transglycosylases. PUA domains have an overall dumb-bell shape of mixed α/β-secondary structure, and are thought to interact primarily with NCA trinucleotide sequences.[7] The closest structural similarity to the PUA-like domain of YebU is found in eukaryotic proteins that lack an MTase domain, and for which no function has been ascribed. The PUA-like domain of YebU is connected to its MTase domain by a stiff linker, and this unusual arrangement indicates a novel means of RNA binding.

Results and Discussion

Structure determination

The SAM-dependent m^5C MTase YebU from *E. coli* was cloned, expressed, purified and crystallized. The three-dimensional structure was determined in space group *P*1 with four molecules in the asymmetric unit. The single-wavelength anomalous dispersion (SAD) method was used with selenomethionine-incorporated protein and the structure was built manually in XtalView.[8] The structures of the RsmB m^5C MTase (PDB id code *1SQF*),[5] and the human p120 homologue protein PH1374 from *P. horikoshii* (PDB ID code *1IXK*),[6] which are currently the only other RNA m^5C MTases with known 3D structures,[3] were used to aid in building the initial trace of YebU. A sequence alignment of the three RNA m^5C MTases is shown in Figure 1.

The final YebU structure (Figure 2) was refined to a resolution of 2.9 Å (Table 1) and has a final *R*-factor of 23.1% and R_{free} of 28.2%. The model includes most of the residues in the four molecules that are found in the asymmetric unit, and has good stereochemistry, with no residue in the disallowed regions of the Ramachandran plot, and with 1.5% (six residues in both molecules A and B), 2.1% (eight residues in molecule C) and 1.3% (five residues in molecule D) in the generously allowed regions (Table 2). The final model comprises residues 7–474 (in molecules A and D) and residues 8–474 (molecules B and C). Weak or absent electron density is observed in two regions, residues 21–26 and 75–85 in all four molecules, possibly due to flexibility or disorder; thus, these regions are not included in the structure.

Overall fold of YebU

YebU contains two distinct structural domains with a total of 18 α-helices and 21 β-strands (Figures 1 and 2(b)). The N-terminal MTase domain (catalytic domain) of YebU is located between residues 7 and 310, and is very similar to the MTase domains of RsmB (residues 136–429) and PH1374 (residues 7–315), as well as to other canonical SAM-dependent MTases. The overall root-mean-square deviations (rmsd) for the structural superpositions are 1.3 Å using 228 C^α atoms in RsmB and 1.6 Å using 223 C^α atoms in PH1374. The four molecules of YebU in the asymmetric unit are identical in the structure of their MTase domains, due to the NCS constraints used throughout refinement.

The N-terminal portion of YebU (residues 7–102) consists of a so-called N1 sequence motif of βαββαβ topology, which has been implicated in RNA recognition and binding.[5] The N1-domain structure in YebU is almost identical with that in RsmB. However, the two regions in this domain that could not be seen in YebU (residues 21–26 and 75–85) were clearly observed in the RsmB structure, and this was presumably due to stabilization by the RNA-binding domain at the N terminus of RsmB, which is absent from YebU (Figure 1).

The topology of the MTase domain is a mixed, seven-stranded twisted β-sheet structure arranged in the canonical order of 3-2-1-4-5-7-6 (β7- β6- β5- β8- β9- β12- β10/ β11 (Figure 1) found in most of the SAM-dependent MTases.[9] The β-sheet is sandwiched on one side by helices α5, α6 and α7, and on the other side by helices α8, α9 and α10 in the same manner as that seen in RsmB. However, a few

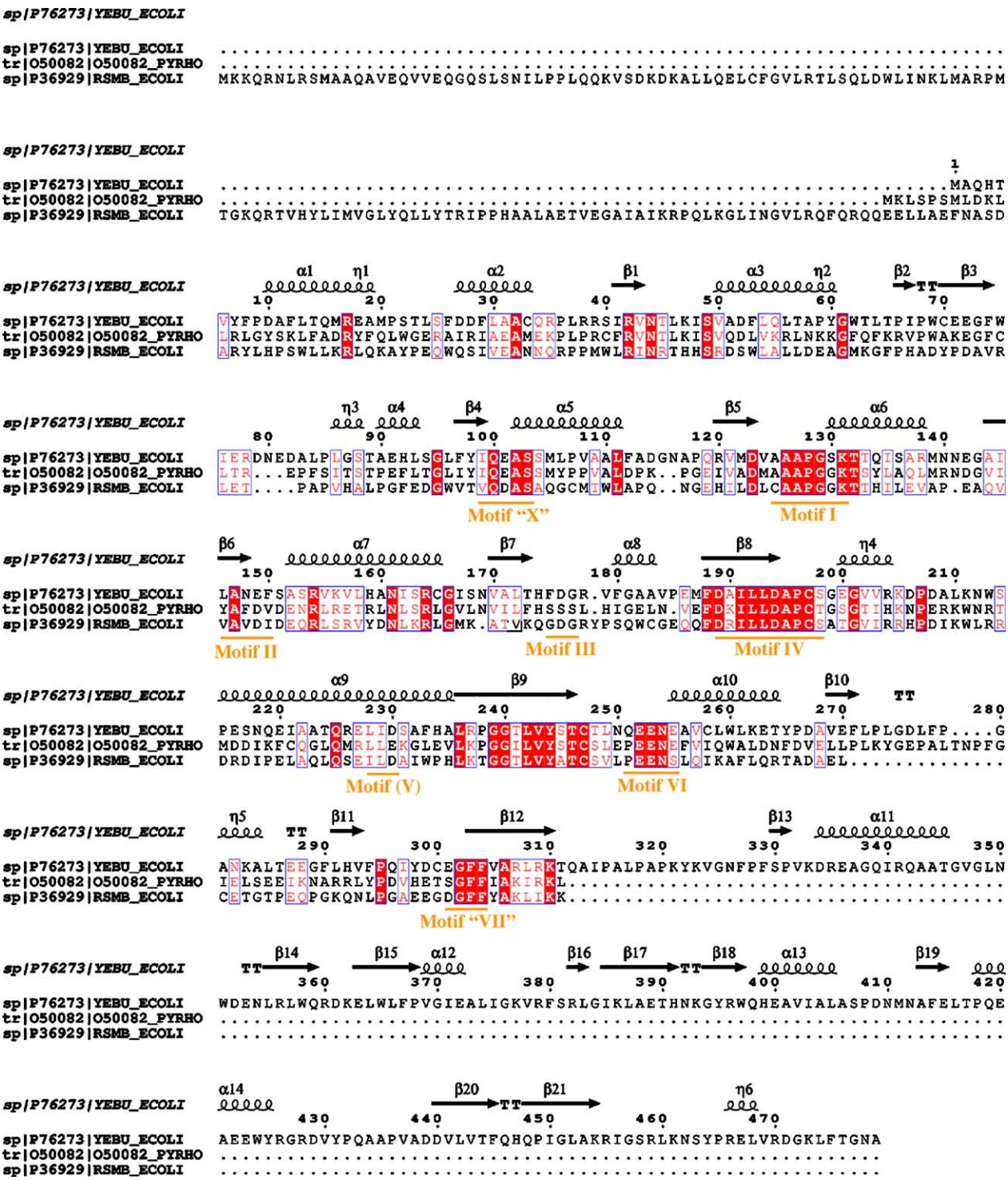

Figure 1. Sequence alignment with secondary structure assignment of YebU (sp|P76273|YEBU_ECOLI) with the homologous SAM-dependent RNA m^5C MTases: human p120 homologue protein PH1374 from *Pyrococcus horikoshii* (tr|O50082|O50082_PYRHO) and *E. coli* RsmB m^5C MTase (sp | P36929 | RSMB_ECOLI). The alignment was performed using the server CLUSTALW (http://www.ebi.ac.uk/clustalw/),[40] and the Figure was made using ESPript.[41]

regions differ between YebU and RsmB: the loop-region between YebU residues 112–119 is much longer than in RsmB. This results in a protrusion of the main chain in this region on the back of the YebU structure (i. e. on the opposite side to the active site). Also, the regions 179–187 and 273–301 show significant differences in the main chain between the two MTase domains of YebU and RsmB, reflecting the lack of sequence similarity here.

The C-terminal region of YebU (residues 311–474) consists of a long, rigid proline-rich loop/belt (residues 312–332) that reaches over the MTase domain to start the novel C-terminal domain. The C-terminal domain consists of two subdomains, the

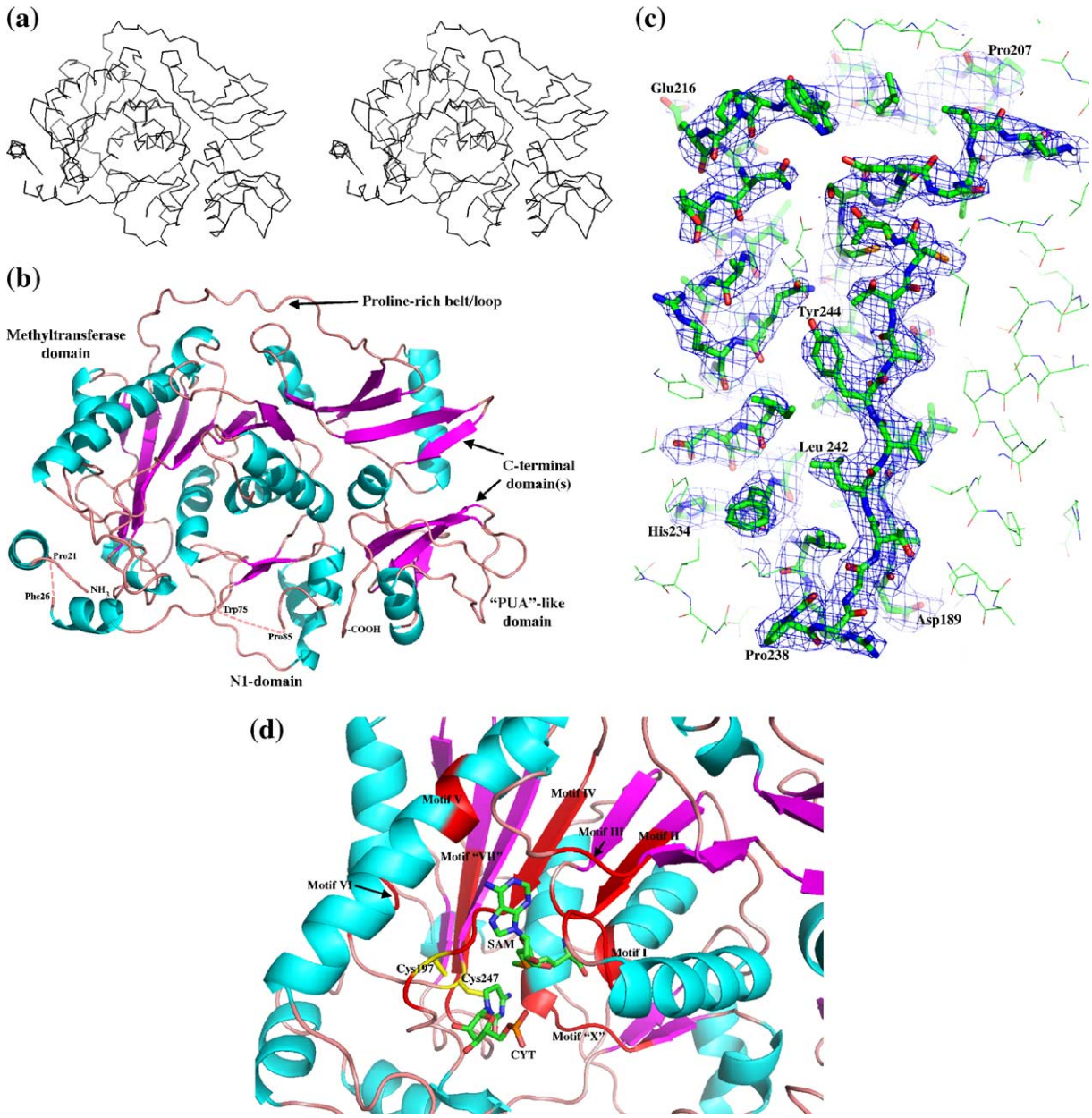

Figure 2. Structure of YebU. (a) A C^{α} trace in wall-eyed stereo. (b) YebU secondary structural assignment. The α-helices are colored cyan, the β strands are colored magenta and loop/coiled regions are colored salmon. The same orientation as in (a) is used. (c) Representative $2F_{obs}–F_{calc}$ electron density contoured at 1.0σ in the MTase region 189–248. (d) Zoom-in of the active site of YebU. SAM was modeled into the active site by superimposing the RsmB and YebU structures; the cytosine substrate (CYT) was modeled from the target U1939 base of the RumA-RNA-SAH complex[21] with some modifications. SAM-binding motifs (red), and the two cysteine residues putatively active during catalysis (yellow) are shown.

first of which has six β-strands and three α-helices, where the β-strands form a central sheet in an anti-parallel fashion, and the α-helices pack against the edges of the sheet. The second subdomain is smaller, with three β-strands and one α-helix and, compared to the first subdomain, contains more coil regions that lack detectable secondary structure. The core of the second subdomain is also formed from the β-strands, although here the first two strands are parallel and bracket the α-helix with the third strand running anti-parallel. The two C-terminal subdomains are held together by a short linker region of approximately six amino acid residues. As discussed below, the C-terminal domain of YebU shows structural similarity to the PUA domains found in other RNA-binding proteins.

RNA m^5C MTase domain architecture

RNA m^5C MTases have a conserved MTase core region consisting of a small N-terminal region (N1 domain) of 45–50 amino acid residues and a larger C-terminal region of 220–230 amino acid residues that contains several sequence motifs (Figure 1). The

Table 1. Data collection, and substructure solution statistics

	Se-Met (P1)
Wavelength (Å)	0.9682
Resolution (Å)	30–2.9 (3.0–2.9)
No. unique reflections	45,870
No. observed reflections	148,085
Completeness (%)	97.9 (97.5)
$<I/\sigma(I)>$	9.13 (2.14)
R_{sym}	10.3 (57.2)
Anomalous signal-to-noise ratio[a]	1.31 (1.21)
CC (all/weak)[b]	20.4/4.5

Values in parentheses are for the outermost shell.
[a] $<|F^+-F^-|/\sigma(F^+-F^-)>$ as calculated by XPREP (Bruker AXS).
[b] Correlation coefficent as a measure for the agreement between E_o^2 and E_c^2; expressed as a percentage for the best solution in SHELXD.[29]

MTase region is responsible for binding the SAM cofactor,[5] and contains the residues necessary for catalysis. However, an isolated MTase domain does not seem to be sufficient for the site-specific RNA recognition necessary for cell viability. The RNA substrate recognition has been suggested to be promoted by the non-core N-terminal and C-terminal extensions, or possibly by additional proteins that serve as RNA-binding modules or subunits.

RsmB, the best-studied member of subfamily I, contains an N-terminal extension of approximately 220 amino acid residues (Figure 3). The target of RsmB is in *E. coli* 16 S rRNA at nucleotide C967, which is within a stem–loop that is highly conserved in all small subunit RNAs.[10] Members of subfamily II, which includes the human p120 protein, have a large N-terminal extension that contains nuclear and nucleolar localization motifs,[11] as well as a C-terminal extension. A homologue of the human p120 protein found in *P. horikoshii* (PH1374, PDB ID code *1IXK*) is the only three-dimensional structure that has been solved for this subfamily, and reveals only the MTase domain, and not the N or the C-terminal domain.[6] The biological substrate for this subgroup is unknown.

The *E. coli* protein YebU (RsmF) is the founding member of subfamily III and contains a conserved C-terminal extension in addition to the MTase core (Figures 1 and 3). YebU has recently been reported to be responsible for methylation of C1407 of 16 S rRNA.[1] The remaining subfamilies contain proteins with different types of N-terminal extensions, or with the core MTase region alone or with an insert; their substrates are, for the most part, unknown.

Functional annotation of hypothetical proteins based on the YebU structure

Sequence comparisons of the C-terminal domain of YebU show that it has little obvious similarity to other proteins of known structure. However, comparison of higher-order structural motifs was more productive, and a DALI[12] search using both C-terminal subdomains of the YebU structure gave 18 hits with a *Z*-score higher than 2, where most of the hits were RNA-interacting proteins. The level of sequence identity in the hits, at 11–16%, is low. The best structural alignment was found for a human protein (PDB ID code *1T5Y*; Northeast Structural Genomics Consortium Target Hr2118) that has no known function and is a homologue of the yeast protein NIP7p (Figure 4); this protein gave a *Z*-score of 10.7 and rmsd of 3.4 Å for 130 C^{α} atoms with a sequence identity of only 13%. A BLAST[13] search using the sequence of yeast NIP7p revealed one further structure, the human protein KD93 that is expressed in hematopoietic stem/progenitor cells. The structure of this latter protein (PDB ID code *1SQW*)[14] is identical in sequence with 1T5Y, with an rmsd of 0.4 Å; these structures should thus be regarded as different annotations of the same protein. The Gene Ontology functional annotation at the European Bioinformatics Institute for 1T5Y suggests that this protein is involved in RNA binding, although no experimental verification of this function has been provided.

The *S. cerevisiae* nucleolar protein Nip7p has been studied in more detail. This protein is conserved among eukaryotes and archaea, and has been shown in yeast to be required for accurate processing of the 27 S precursor of the 25 S and 5.8 S ribosomal RNAs. Depletion of Nip7 leads to defective pre-rRNA processing, and consequently reduced levels of the 60 S subunit and decreased rates of protein synthesis.[15] A good structural match is found between the Nip7p homologues and the first C-terminal subdomain of YebU (Figure 4), suggesting that both have RNA-recognition functions. However, an equivalent of the MTase domain is lacking in Nip7p, suggesting that another protein might have to be recruited in order to bind the putative RNA target. This idea is supported by two-hybrid screens, where Nip7p was found to interact with the nucleolar protein Nop8p, which is also essential for 60 S subunit biogenesis, as well as with the exosome subunit Rrp43p.[16]

Table 2. Refinement statistics and quality of structure

Resolution (Å)	29.03–2.90 (2.975–2.90)			
R_{cryst}[a] (%)	23.14 (29.8)			
R_{free}[a] (%)	28.19 (36.1)			
RMS deviation from ideal geometry				
Bond lengths (Å)	0.022			
Bond angles (deg.)	1.75			
Estimated coordinate error (Å)	0.415			
Average *B*-factor (Å^2)	65.5			
Ramachandran plot[b]:				
Molecule	A	B	C	D
Most favored (%)	85.9	86.1	86.6	86.4
Allowed (%)	12.6	12.4	11.3	12.3
Generously allowed (%)	1.5	1.5	2.1	1.3
Disallowed (%)	0	0	0	0

Values in parentheses are for the outermost shell.
[a] $R=\Sigma|F_{obs}-F_{calc}|/\Sigma|F_{obs}|$.
[b] According to the definition used in PROCHECK.[39]

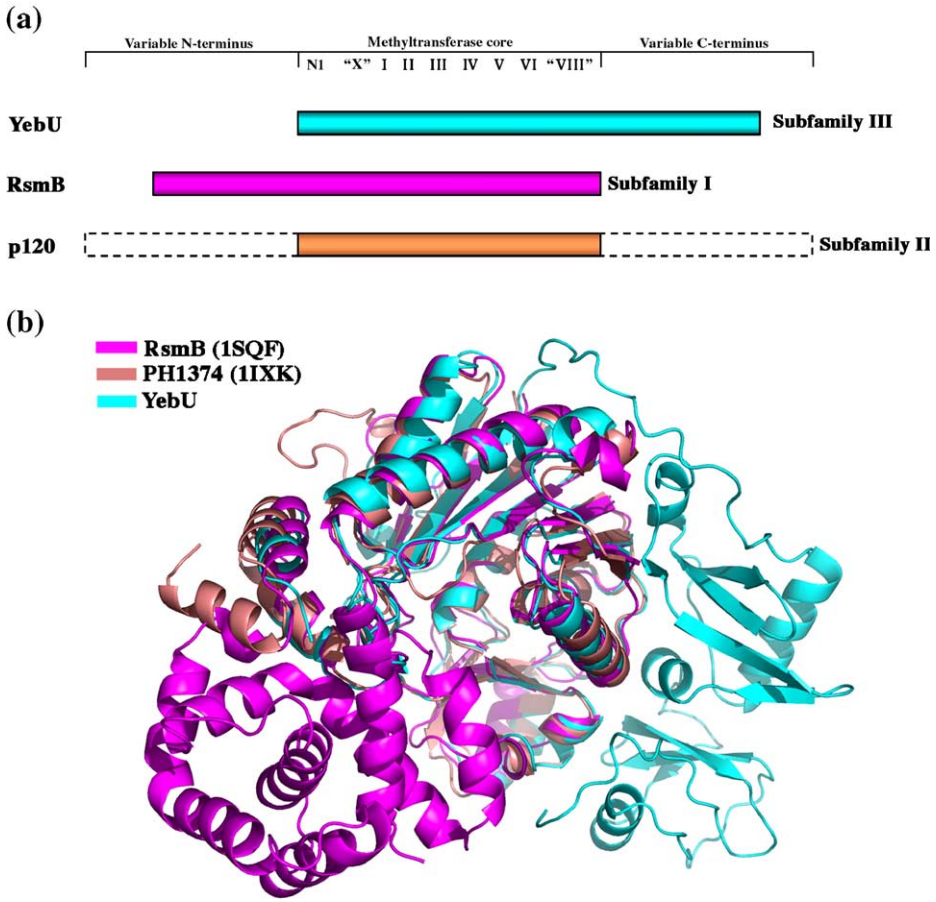

Figure 3. (a) Conserved regions in the putative RNA m^5C MTase family. The representative members shown are: subfamily III, YebU (in cyan); subfamily I, RsmB (in magenta); and subfamily II, human p120 (in salmon). The figure is adapted from Reid et al.[3] (b) Superposition of the three known RNA m^5C MTase structures: *E. coli* YebU, *E. coli* RsmB and *Pyrococcus horikoshii* human p120 homologue protein PH1374. The color scheme used is the same as that in (a).

The DALI search revealed another conserved hypothetical protein of unknown function from *Thermoplasma acidophilum* as the second best match to the YebU C-terminal domain (with a *Z*-score of 8.5 and rmsd of 3.1 Å for 115 C^{α} atoms, and a sequence identity of only 11%) (PDB ID code *1Q7H*). In addition, the DALI search disclosed some of the same structures that were reported in the Nip7p homologue KD93 study:[14] the tRNA pseudouridine synthase b fragment (PDB ID code *1k8w*)[7] and archaeosine tRNA-guanine transglycosylase (PDB ID code *1iq8*)[17] with reasonable *Z*-scores of 6.4 and 5.7, respectively. Both of these are RNA-modifying enzymes and contain a PUA domain, which is widely distributed in eukaryotic and archaeal enzymes of this type. The PUA domain is thought to bind to RNA molecules with complex folded structures.[18] The superposition of archaeosine tRNA-guanine transglycosylase complexed with valine tRNA (PDB ID code *1J2B*)[19] shows that the C-terminal domain of YebU is structurally similar to the PUA domain (Figure 5). The archaeal PUA domain has been suggested to form a positive electrostatic patch that can bind to the acceptor base-pairs of tRNA.[19] An electrostatic surface representation of YebU reveals a region of positive potential on one side of the PUA domain (Figure 6(a), right arrow), with a second positive patch closer to the MTase domain (Figure 6(a), left arrow).

Substrate recognition by YebU

Our knowledge about how MTases recognize and bind nucleic acid target sequences is rather sparse, and is based presently on the HhaI and RumA structures. The HhaI structure was solved in complex with the cofactor analogue *S*-adenosyl-L-homocysteine (SAH) and a duplex 13-mer DNA oligonucleotide containing 5-fluorocytosine at the target site (Figure 7(a));[20] and the structure of RumA, an m^5U RNA MTase, was solved in complex with RNA and SAH (Figure 7 (b)).[21] Both the HhaI and RumA MTase domains are closely similar to that in YebU, with an overall rmsd of 1.6 Å. Both structures feature the target nucleotide in a flipped-out conformation, where it is swung out of its stacked conformation in the free substrate and towards the active site of the enzyme.

The electrostatic potential on the surface of the MTase structures is a good indicator of where the RNA or DNA substrates would bind. Both the HhaI

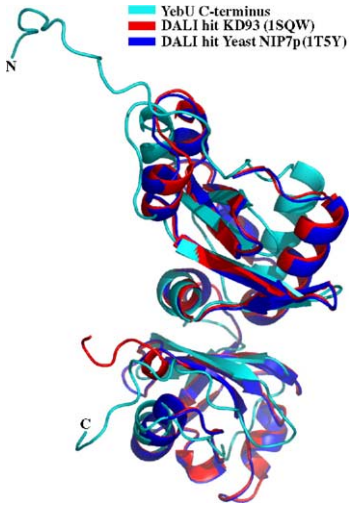

Figure 4. Superposition of the C-terminal domain of YebU with the best hit from the DALI search: Northeast Structural Genomics Consortium Target HR2118: A human homolog of *Saccharomyces cerevisiae* Nip7p, a protein of unknown function (PDB ID code *1T5Y*) and the structure of KD93, a protein expressed in human hematopoietic stem/progenitor cells (PDB ID code *1SQW*).[14] The N and the C termini of the c-terminal domain of YebU are indicated. YebU is colored cyan, 1T5Y is colored blue and 1SQW is colored red.

and RumA structures show strong positively charged electrostatic potential at exactly the surface regions that interact with the negatively charged backbone of the nucleic acid within the complex (Figure 7(a) and (b)). In YebU, there are two contiguous regions of favorable electrostatic potential (Figure 7(d)), one from the MTase domain and another from the C-terminal domain. These regions are positioned differently than in Hha1, RumA and RsmB (Figure 7(c)), which suggests a different RNA-binding mode for YebU.

The two spatially-separated regions of electrostatic potential in YebU suggest that it recognizes an additional section of rRNA, distinct from the methylation site. In order for the target cytosine to reach the active site from the binding groove between the MTase domain and the PUA domain, it must be flipped out of a helical structure, similar to mechanisms observed for the other MTases mentioned above. The actual distance between a modeled cytosine at the active site and the closest patch of positive electrostatic potential is approximately 9 Å, which fits a structural model in which a flipped out target cytosine binds at the active site next to SAM, while other regions of the rRNA bind to the electropositive patches in the MTase domain and the PUA-like C-terminal domain.

Docking of rRNA to YebU

As an initial step in constructing a molecular model of how the rRNA might interact with YebU, the relevant region of the RumA-RNA structure with the flipped-out conformation of the target nucleotide (PDB ID code *2BH2*) was modeled onto the YebU structure. The RumA RNA was fit into the elongated groove, situated near the YebU active site that begins with two β-strands at residues 171–176 and 342–345, and ends at residue 403 near the YebU C terminus. This groove has the longest contiguous region of high electropositive potential on the YebU surface (Figure 7(d)), and readily accommodates the RNA substrate with space to rotate the nucleotides so that the methylation target is positioned within 4 Å of the catalytic Cys197 residue. This model fits well with the YebU crystal structure and is supported by AutoDock[22] calculations, where fragments of the RumA RNA were found to dock to positions similar to those found with manual fitting. However, although the RumA-RNA model supports the idea that YebU might bind the rRNA target in a flipped-out conformation, this form of modeling cannot conclusively deduce the specific conformation that the C1407 rRNA site assumes in the catalytic complex.

Further molecular models of YebU were built to examine how YebU might bind with the C1407 methylation site in the context of the *E. coli* 30 S ribosomal subunit. One model includes the part of the 16 S rRNA helix 44 that contains C1407 (nucleotides C1402–G1416 and A1483–U1500), as well as three RNA strands (A790–A794, G887–A892 and A908–C912) that are near the 1402/1483 nucleotides. As expected, docking this RNA model with AutoDock to the YebU structure did not place the C1407 methylation site in a reasonable orientation or within a reasonable distance from the YebU active site and, clearly, conformational changes within the RNA must occur to facilitate binding. These were achieved by manually aligning the region of 16 S rRNA helix 44 containing C1407 to fit into the electropositive surface groove of YebU; this places the C1407 base at the entrance to the catalytic cavity (Figure 8(b)). Further conformational changes in the rRNA are necessary before C1407 can become methylated.

The remainder of the YebU MTase domain makes very few RNA interactions, but instead, is predicted to bind with two β-strands within the 30 S subunit ribosomal protein (r-protein) S12 (Figure 8). This model fits well with the observation that YebU is unable to methylate protein-free 16 S rRNA,[1] and suggests that the S12 interaction is an important factor for enzyme-target recognition. The docking model additionally predicts that the PUA domain of YebU interacts with other regions of the rRNA, in particular at nucleotides A695–C699, A780–A802, A900–U904 and G1514–U1522. The predicted structure of these nucleotides complexed with the PUA domain is reminiscent of the tRNA-bound archaeosine tRNA-guanine transglycosylase structure (PDB ID code *1J2B*), and these additional rRNA interactions may

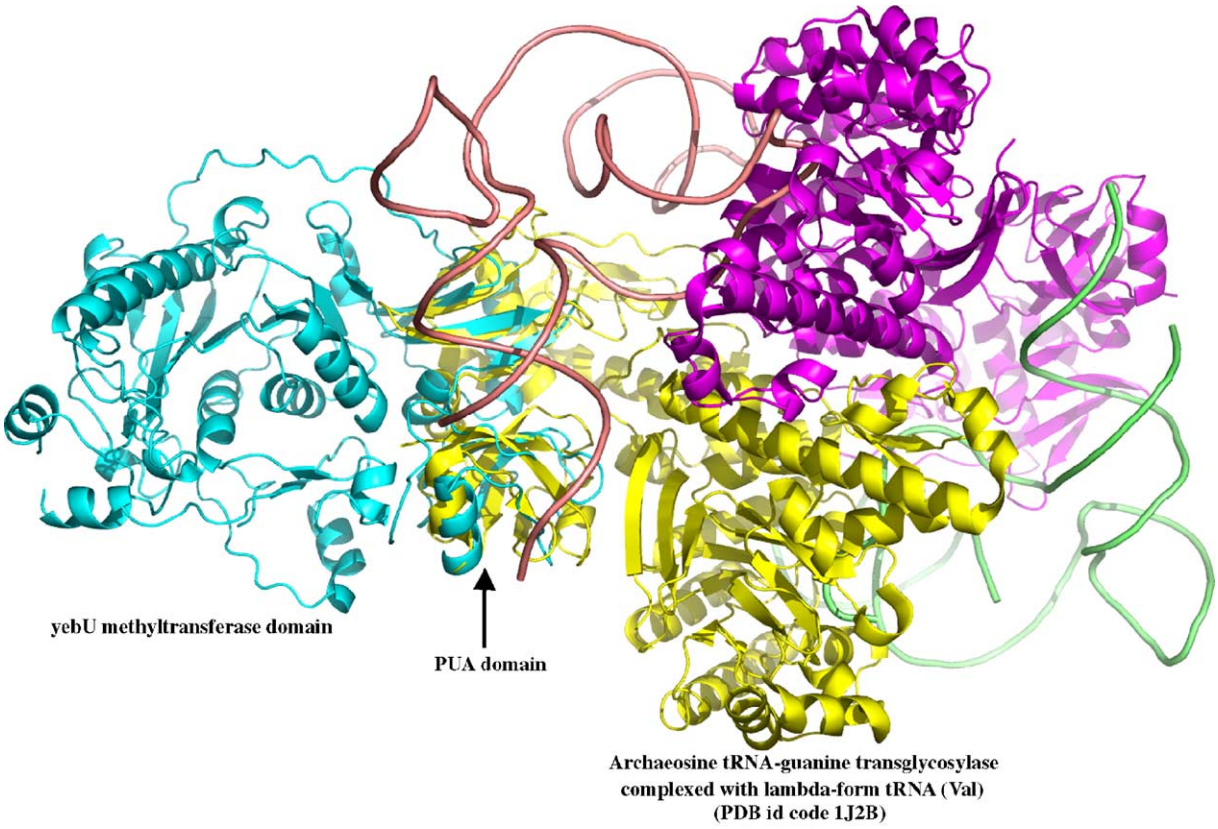

Figure 5. Structural superposition of YebU (cyan) with archaeosine tRNA-guanine transglycosylase (chain A, yellow; chain B, magenta) in its complex with valine tRNA (lime-green and salmon) (PDB ID code *1J2B*).[19] The position of the PUA domain is indicated.

enhance the specificity of YebU in targeting the correct methylation site.

In principle, two putative roles in rRNA recognition can be envisioned for the PUA domain of YebU: either the PUA domain binds to a portion of the rRNA that is adjacent to the methylation site within the primary structure, or it binds an rRNA structure that, while close in the higher-order structure, is further away in the primary structure. The first case requires a conformational change to occur upon rRNA binding that subsequently brings the PUA domain closer to the MTase domain, and improves the RNA binding. A comparison of MTases in RNA-bound and unbound structures shows only minor conformational changes in the MTase domain. Instead, these conformational

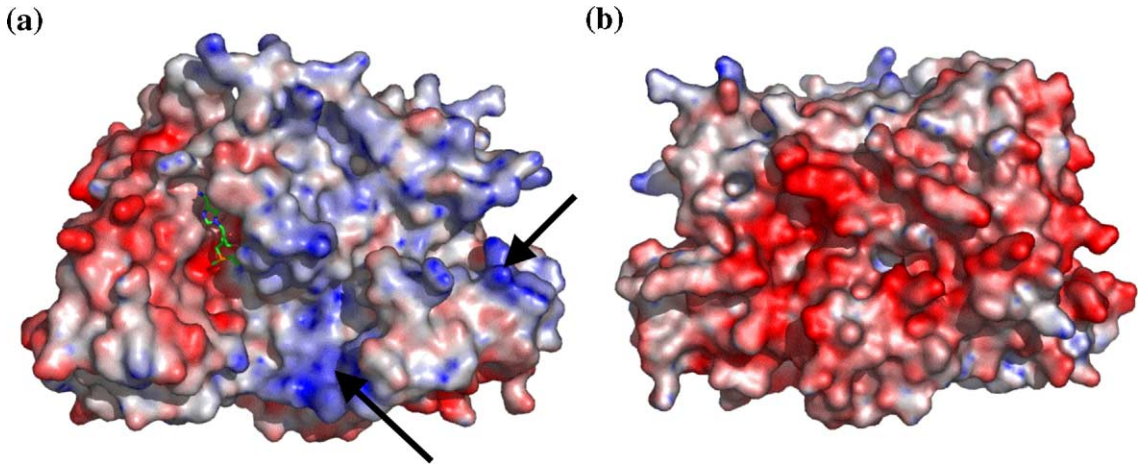

Figure 6. Molecular surface representations of YebU. The two views are (a) front and (b) back. The electrostatic surface potential for YebU is shown; blue and red represent regions of positive and negative electrostatic potential, respectively; contoured at $\pm 8k_B/T$. The MTase active site is highlighted by the presence of a stick model of SAM colored green (carbon), blue (nitrogen), red (oxygen) and orange (sulfur). Two regions of clear positive potential discussed in the text are marked with arrows for clarity.

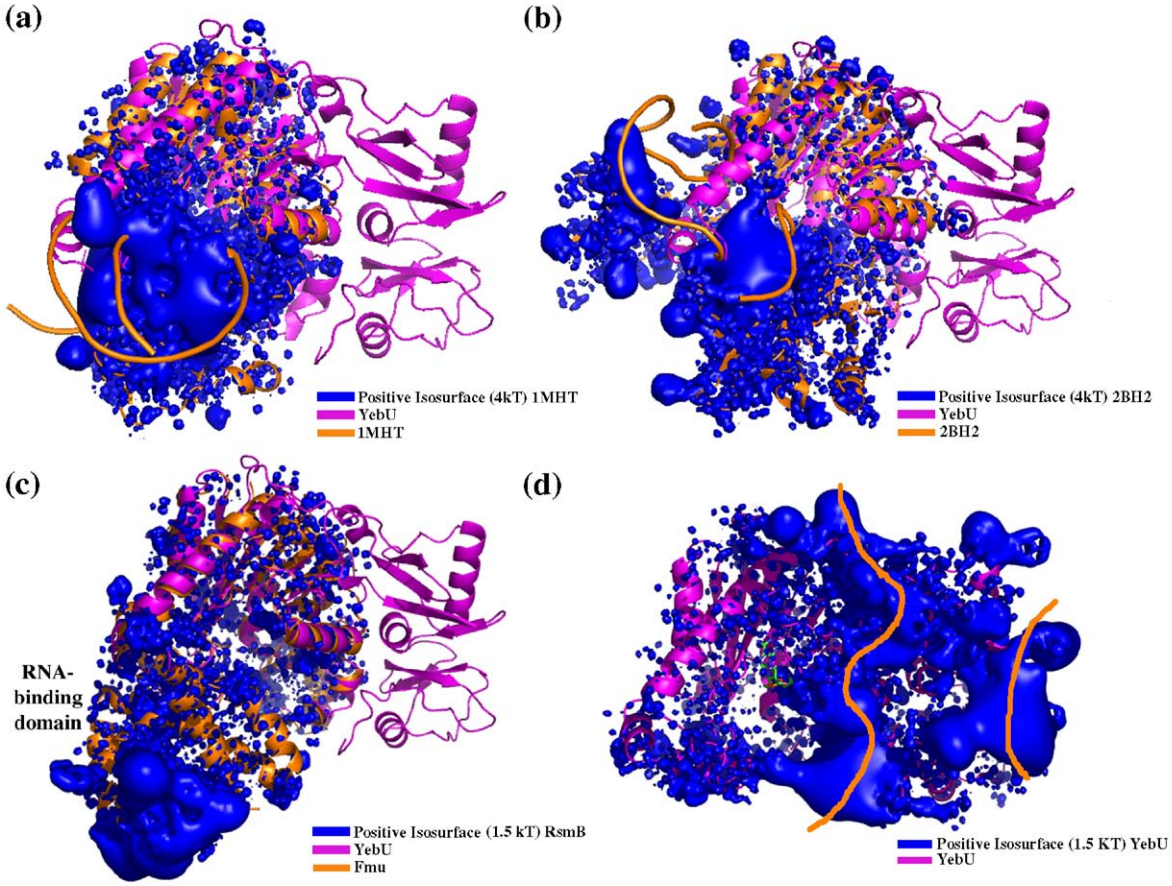

Figure 7. Electrostatic properties of MTases shown with their nucleic acid substrates. (a) DNA-MTases; (b) RNA-MTases; (c) m^5C RNA MTases RsmB; and (d) YebU. The positive potential isosurfaces (4kT/e in (a) and (b); 1.5kT/e in (c) and (d)) are shown in blue at various contour levels. The same orientation of YebU (magenta ribbon) is used in all four parts of the Figure to facilitate comparison of RNA-binding. The conserved SAM-binding MTase domains were superimposed onto YebU in (a)–(c) using the program TOP,[38] and the electrostatic potentials were calculated using APBS[37] and visualized in Pymol. The nucleic acid substrates are shown as orange coils in (a) and (b); the position of the RNA-binding site in YebU is suggested (orange coil in (d)).

changes could potentially occur in the long proline-rich loop/belt found in the sequence (residues 312–332) between the MTase domain and the C-terminal PUA domain (Figures 1 and 2). However, these domains form a tight interface and this long connector sequence may be relatively rigid due to the five proline residues found in this 20 residue connection between the MTase and PUA domains, which would argue against this RNA-binding model. Rather, the proline-rich connector between the MTase domain and the C-terminal PUA-like domain may simply function as a distance spacer to aid YebU in recognizing the substrate rRNA properly.

A likely scenario is that the PUA domain binds to a second ribosomal region, which is separate from the MTase-binding region, and thereby acts as an RNA-binding cofactor to assist recognition of the target cytosine.[18] This model requires that the ribosomal regions involved are in a relatively stable configuration, which would be achieved by assembly of the 16 S rRNA with a set (or at least a subset) of r-protein to attain a conformation that is recognized by YebU.

Surface recognition of rRNA by conserved residues in the MTase and PUA domains of YebU

To identify the amino acid residues necessary for RNA interaction, we compared the MTase and the PUA domains of YebU with the corresponding domains in other known protein structures. In the MTase domains of the three known m^5C MTases (YebU, RsmB and PH1374), the residues Ala344, Gly384, Leu387, Ala388, Glu400, Leu405, Pro408, Asp429, Val430, Ala435, Pro436, Val441, Leu442, Val443 and Lys461 (YebU numbering throughout) are partially or completely conserved, and the majority of these residues are either located at the active site SAM-binding motifs or are putatively involved in substrate binding (Figure 8(a)).

Comparison of the PUA domains in Nip7p, KD93 and the archaeosine tRNA-guanine transglycosylase with that in YebU revealed that only four residues (Leu405, Val430, Val441, and Val443) are completely conserved (identical) in the four structures. Partially conserved PUA residues are located in the hydrophobic core of the PUA domain (these are particularly

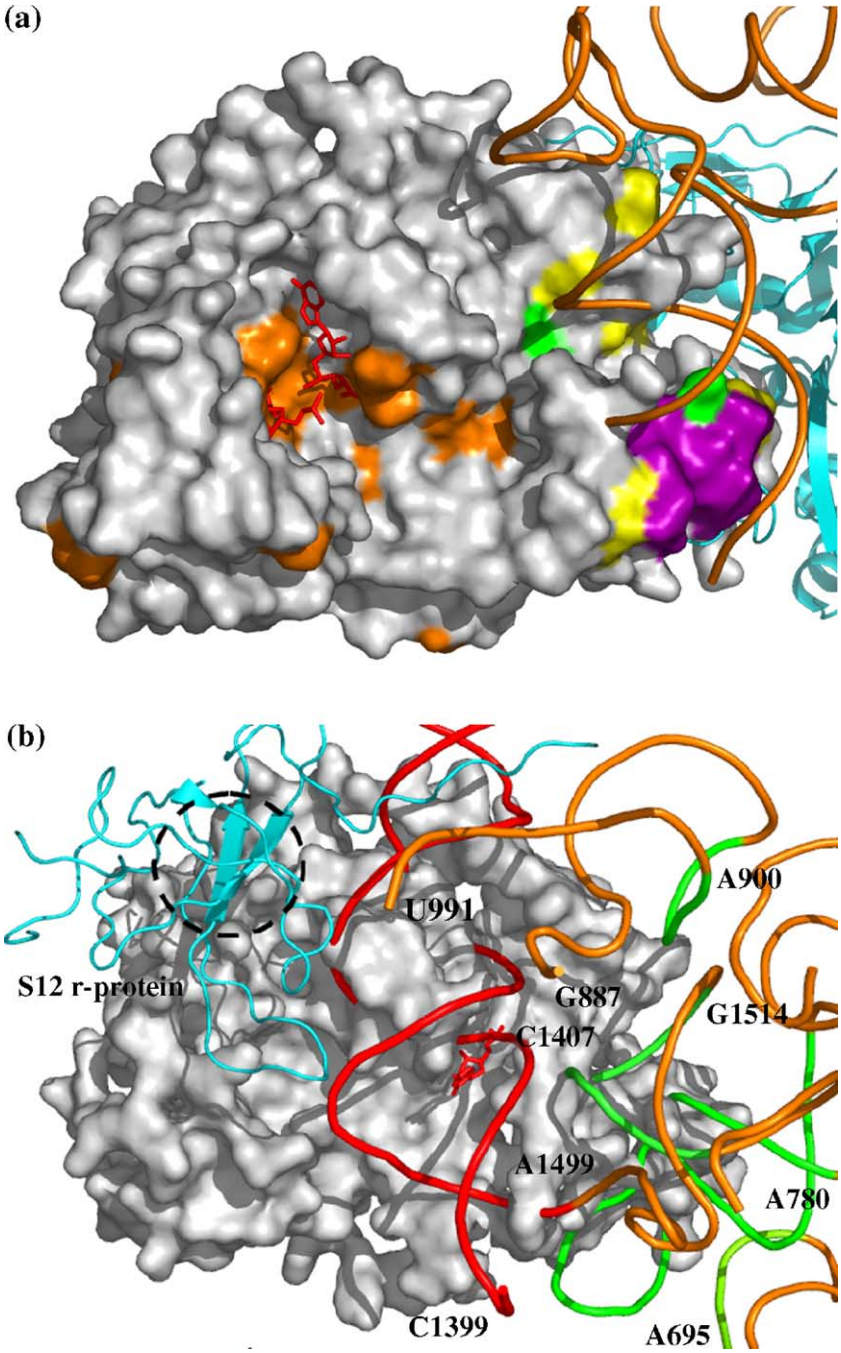

Figure 8. (a) A surface representation of YebU showing the conserved residues. Residues Leu405, Val430, Val441 and Val443 (green) are conserved in the four PUA domain structures. Also indicated are residues that are similar in YebU and archaeosine tRNA-guanine transglycosylase (1J2B) (yellow), RNA-interacting residues of the PUA domain (purple) (see Table 3), and residues conserved in the RsmB, PH1374 and YebU m^5C RNA MTase domains (orange). The modeled cofactor SAM and cytosine substrate are shown as red sticks in the MTase active site. The RNA from the superimposed archaeosine tRNA-guanine transglycosylase (1J2B) structure is shown in orange coil together with the archaeosine tRNA-guanine transglycosylase structure (cyan). (b) Modeled interactions between YebU and the 30 S ribosome. YebU is shown as a molecular surface along with rRNA and r-proteins (cyan ribbon) within 20 Å. The nucleotide C1407 methylation target (stick figure) is within helix 44 of 16 S rRNA (red). The S12 r-protein is shown in cyan with a broken-line circle showing the β-sheets predicted to interact with YebU. Regions of rRNA predicted to interact with the PUA domain (green) are shown, as are other rRNA regions within 20 Å of YebU (orange); specific nucleotides are numbered to indicate the position and polarity of the sequence.

apparent when comparing YebU and archaeosine tRNA-guanine transglycosylase), and are therefore more likely necessary for the domain architecture than for RNA interactions. Figure 8(a) shows all the conserved residues mapped onto the surface of the YebU structure together with the RNA from the superimposed archaeosine tRNA-guanine transglycosylase structure.

Eight residues in the C3 region of the PUA domain were previously proposed to be involved in RNA interactions (Table 3),[14] and are shown on the surface of YebU (purple residues in Figure 8(a)). The variation here in the identities of the amino acid residues indicates that the putative RNA-binding sites in YebU and Nip7p/KD93 are possibly different from that of archaeosine tRNA-guanine transglycosylase. However, it cannot be ruled out that differences in the RNA targets of YebU and the archaeosine transglycosylase might be complemented by the differences in charge characteristics of the enzyme PUA domains.

The active site of YebU

Enzymes that catalyze the 5-methylation of pyrimidines generally use the thiol group of a cysteine residue to perform a nucleophilic attack on the 6-position in order to activate the 5-position towards the one-carbon transfer from the SAM cofactor. Subsequently, a proton is removed from

Table 3. Residues in the PUA domain proposed to be involved in RNA interactions in archaeosine tRNA-guanine transglycosylase (1J2B)[19] and KD93[14] and the YebU and Nip7p counterparts

YebU	1J2B	KD93 (1SQW)	Nip7p (1T5Y)
Glu423	Phe519	Ser106	Ser106
Gly427	Lys522	Tyr109	Tyr109
Arg428	Gly523	Gly110	Gly110
Asp429	Asp525	Asn111	Asn111
Tyr431	Phe527	Leu114	Leu114
Arg459	Arg573	Ile161	Ile161
Leu460	Lys576	Phe164	Phe164
Asn462	Arg578	Gln166	Gln166

the 5-position of a 5, 6-dihydropyrimidine intermediate, and β-elimination results in the 5-methylated product.

Members of the RNA m^5C MTase family contain six signature motifs (Figures 1 and 2) that also are homologous to the motifs found in DNA m^5C MTases.[23] One of the best-conserved motifs (motif IV) in MTases is the completely conserved Pro-Cys dipeptide sequence that was previously believed to contain the cysteine catalytic residue (Cys197 in YebU, Cys325 in RsmB), but is now more generally accepted to be involved in the β-elimination product release step. The RNA m^5C MTases (but not the DNA m^5C MTases) also have a second conserved cysteine residue in motif VI (Cys247 in YebU, Cys375 in RsmB), which has been confirmed as the catalytic cysteine nucleophile.[23] Both Cys197 and Cys247 in YebU are similarly positioned in RsmB, and are effectively identical within the X-ray resolution limits of this study. In a docked model, the sulfur groups of Cys197 and Cys247 are located close to the C5 atom of the cytosine substrate on the same side of the pyrimidine ring, and are thus reasonably positioned for catalysis. The methyl group of SAM is positioned pointing towards the C5 atom of cytosine where it is to be donated. Thus, the positions of these catalytically important residues allow us to speculate that the catalytic reaction in YebU will be very similar to that of the other m^5C RNA MTases.

Bujnicki *et al*. used site-directed mutagenesis to show that two additional residues, Lys in motif I and Asp in motif IV, are essential for the activity of the *S. cerevisiae* MTase Trm4p, and were proposed to be involved in SAM-binding and recognition.[24] The two residues are conserved across the RNA m^5C MTases, and are evident in YebU at positions Lys131 and Asp194, in RsmB at Lys260 and Asp322, and in the human p120 homologue protein PH1374 at Lys132 and Asp194. The YebU residue Lys131 is located in a similar position and orientation in RsmB and PH1374, where it can make a hydrogen bond with the carboxyl moiety of the SAM cofactor. Asp194 in YebU is located within hydrogen bonding distance of Lys131 and the amine moiety of SAM. In our model, Asp194 is located close to N4 of the target cytosine, with its position and orientation closely resembling that of Asp322 in RsmB and Asp194 in PH1374. We envisage that Asp194 in YebU forms a hydrogen bond with the target cytosine to position it for catalysis, and suggest this residue may be of vital importance in the active site.

Sequence alignment indicates that there is a difference in motif III of the active sites in RsmB and YebU (Figure 1). Asp303 in the RsmB structure forms an important hydrogen bond with SAM (Asp303 $O^{\delta 1}$–SAM N6; 3.0 Å), and the corresponding interaction for YebU can be created *via* Asp176, which is one residue further along in the sequence. Placement of Asp176 in the active site forms a bulge further back in structure at residue Asn169, causing the YebU backbone to diverge from the conformation seen in RsmB. The YebU and RsmB backbones are eventually brought back into alignment by another frame-shift after residue Val179 (Tyr306 in RsmB). A second backbone difference also allows Phe175 to fit into the YebU structure where Gly302 is found in the RsmB (Figure 1). In motif III of the p120 structure, no SAM-binding Asp residue is found,[6] although Ser177 points towards the nucleotide moiety of SAM and thus could serve as a SAM-binding residue instead of Asp176 (in YebU) or Asp303 (in RsmB).

In conclusion, we show in this study that while the *E. coli* protein YebU has many structural characteristics in common with other RNA MTases, it also embodies features found in other types of proteins. The structural features are brought together in a novel combination in YebU, and are used to recognize and methylate the 16 S rRNA target at nucleotide C1407 in the 30 S ribosomal subunit. The key regions at the N terminus of YebU include the SAM-binding catalytic domain that has structural similarity to the equivalent domains in several other m^5C RNA MTases including RsmB and PH1374. The C-terminal one-third of YebU contains a PUA-domain, which was not predicted from sequence comparisons. In eukaryotes, the PUA-like domains are separated from the catalytic domains, functioning in complex with RNA-modifying and processing enzymes. This suggests that the PUA domains have evolved in eukaryotes to become general interaction modules for recognition of specific RNA sites. In YebU, a stiff linker between the MTase domain and the PUA domain suggest a mode of RNA binding that is novel for bacterial enzymes. Our docking model onto the 30 S subunit indicates that these regions interact with the 16 S rRNA, and with r-protein S12, and explains the substrate preference of YebU. The structure of YebU presented here provides a structural basis for designing further biochemical work on RNA m^5C MTases.

Materials and Methods

Cloning, protein production and purification

The *E. coli* YebU gene was cloned into the PT73.3HisGW vector.[25] The resulting construct encodes a polypeptide with the YebU gene and both N- and C-terminal hexa-histidine tails. BL21(DE3) cells were

transformed with the expression plasmid, and were grown at 37 °C in Luria-Bertani (Miller) medium supplemented with tetracycline. Protein overproduction was induced at an absorbance at 600 nm of 0.6 by the addition of 0.25 mM isopropyl-β-D-thiogalactopyranoside and was continued for 4 h at 25 °C. The methionine-pathway inhibition method[26] was used to obtain selenomethionine-substituted protein. Harvested cells were resuspended and sonicated in lysis buffer (20 mM sodium phosphate buffer (pH 7.4), 0.5 M NaCl, 20 mM imidazole, 10 mM β-mercaptoethanol, 10% (v/v) glycerol, 0.1 mg ml^{-1} lysozyme, 0.5 mg ml^{-1} DNAse I and Complete EDTA-free Protease Inhibitor mixture (Roche Biosciences)). The cell lysate obtained by centrifugation at 50,000***g*** for 20 min was loaded onto a Ni^{2+}-loaded Hi-Trap Chelating column (GE Healthcare) equilibrated with wash buffer (lysis buffer without lysozyme or DNAse I). His-tagged YebU was eluted with elution buffer (wash buffer containing 200 mM imidazole) and the fractions containing YebU were pooled and applied onto a HiPrep 26/60 Superdex 200 gel-filtration column (GE Healthcare) equilibrated in GF buffer (20 mM Hepes (pH 7.4), 150 mM NaCl, 10 mM β-mercaptoethanol, 10% (v/v) glycerol, 2 mM EDTA). After purification, YebU was concentrated to 17 mg ml^{-1} using an Amicon Ultra device (Millipore).

Crystallization

YebU crystals were grown at room temperature by the hanging-drop, vapour-diffusion method by mixing 1 μl of protein sample in GF buffer (20 mM Hepes (pH 7.4), 150 mM NaCl, 10 mM β-mercaptoethanol, 10% (v/v) glycerol, 2 mM EDTA) and 1 μl of reservoir solution consisting of sodium acetate (pH 4.6), 15% (w/v) PEG 8000, 50 mM mono-potassium dihydrogen phosphate, 10 mM $BaCl_2$ or 15% PEG 5000 MME, and 0.2 M ammonium sulphate. The crystals were harvested and soaked for 1–2 s in cryoprotectant (20% PEG400 in mother liquor) and flash-frozen in liquid nitrogen before data collection.

Data collection and structure determination

A relatively complete P1 SAD dataset was collected on the high-energy side of the selenium K-edge at Beamline i711 (Max-lab II, Lund, Sweden). The data were processed by Mosflm,[27] and scaled by SCALA.[28] Intensities were further analyzed in XPREP (Bruker AXS) and the ΔF_{ano} values were exported to SHELXD[29] for selenomethionine substructure solution. SHELXD found 26 of the possible 32 sites in the asymmetric unit. These sites were thereafter used to calculate phases in the SOLVE package,[30] using locally scaled unmerged intensities. Further, RESOLVE[31] was used for phase extension and 4-fold averaging.

Model building and refinement

The averaged and extended map in space group *P*1 was used for initial tracing and model building of the four molecules in the asymmetric unit. The structures of the RsmB m^5C MTase (PDB ID code *1SQF*),[5] and the human p120 homologue protein PH1374 from *P. horikoshii* (PDB ID code *1IXK*),[6] currently the only other family members of the RNA m^5C MTase[3] with previously known 3D structures, were used to build the initial trace of the MTase domain of molecule A in YebU using the crystallographic rebuilding program O.[32] The remaining portion of the C terminus was built by hand one residue at a time. The remaining three molecules were then placed in density by hand, using the structure of molecule A, using the program Xtalview,[8] and the real-space refinement routine in Xfit was used for adjusting the initial (manual) fit of the B,C, and D molecules. After all four molecules had been placed in density, overall refinement was done in the CCP4 program Refmac (v.5.2).[33] Several rounds of side-chain and main chain adjustments and refinement were carried out, in which the four molecules were superimposed and the best-resolved structures were used to improve regions of weak density in the other structures. Non-crystallographic symmetry and TLS refinement was used throughout.[34,35]

Molecular modeling and docking

An initial model of the YebU-rRNA complex was constructed by manually orienting the flipped-out 36 nt RNA conformation from the RumA-23 S rRNA crystal structure (PDB ID code *2BH2*) onto the YebU structure and positioning the flipped out nucleotide (U1939) close to the Cys197 and Cys247 YebU catalytic residues while aligning the remainder of the RNA backbone with the neighboring binding pockets with high positive electrostatic potential near β-strands 171–176, and 342–345, and near the C terminus.

Docking calculations were performed with AutoDock 3.0.5, using a three-dimensional grid of dimensions 94.1 Å×94.1 Å×94.1 Å around the YebU structure, to predict receptor–ligand binding energies and conformations. The RNA ligands, which included the tRNA-bound archaeosine tRNA-guanine transglycosylase (PDB ID code *1J2B*) and the RNA-bound RumA (PDB ID code *2BH2*), were docked to YebU as rigid molecules and were allowed no flexible bonds upon preparation of the models. The genetic algorithm local search (GALS) algorithm was used for searching the ligand Cartesian and conformation space. For each ligand, 20 docking runs of 10^7 energy evaluations were calculated. Each generation of the genetic algorithm (GA) had a population size of 150 individuals. Other GA parameters, such as crossover rate, mutation rate, local search probability, and others, were set to standard values.

Sequence searches and surface potential calculations

The BLAST searches were performed using a Web-server†. The DALI[12] searches were also performed using a Web-server‡. The coordinates of the YebU structure were submitted to the DALI server, whereupon the structure was compared against a representative subset of structures from the Protein Data Bank. The outputs returned to the user are, among others, the *Z*-score (statistical significance) of the best domain–domain alignment, rmsd of C^{α} atoms in rigid-body superimposition and the number of structurally equivalent residues (sequence identity). For the analysis of conserved residues, the structures were superimposed using the Protein structure comparison service SSM[36] at the European Bioinformatics Institute§.

Surface potentials were calculated using APBS,[37] and the structural figures were generated in PyMOL‖. The

† http://www.ncbi.nlm.nih.gov/blast/
‡ http://www.ebi.ac.uk/dali/
§ http://www.ebi.ac.uk/msd-srv/ssm
‖ http://www.pymol.sourceforge.net/

superpositions were done using the program TOP,[38] which is incorporated into the CCP4 suite of programs.[33]

Protein Data Bank accession code

The atomic coordinates and structure factors of YebU have been deposited with the RCSB Protein Data Bank with the identifier code 2FRX.

Acknowledgements

We thank Benita Engvall and Martin Andersson for performing the cloning of YebU and Yngve Cerenius, Max-lab, Sweden, for help during data collection. B.M.H. acknowledges the support from the Swedish Research Council (grant number 2005–5028). S.D. thanks the Danish Research Agency (FNU-grant #21-04-0520) and the Nucleic Acid Center of the Danish Grundforskningsfond for support.

References

1. Andersen, N. M. & Douthwaite, S. (2006). YebU is a m^5C methyltransferase specific for 16 S rRNA nucleotide 1407. *J. Mol. Biol.* **359**, 777–786.
2. Gu, X. R., Gustafsson, C., Ku, J., Yu, M. & Santi, D. V. (1999). Identification of the 16S rRNA m^5C967 methyltransferase from *Escherichia coli*. *Biochemistry*, **38**, 4053–4057.
3. Reid, R., Greene, P. J. & Santi, D. V. (1999). Exposition of a family of RNA m(5)C methyltransferases from searching genomic and proteomic sequences. *Nucl. Acids Res.* **27**, 3138–3145.
4. King, M., Ton, D. & Redman, K. L. (1999). A conserved motif in the yeast nucleolar protein Nop2p contains an essential cysteine residue. *Biochem. J.* **337**, 29–35.
5. Foster, P. G., Nunes, C. R., Greene, P., Moustakas, D. & Stroud, R. M. (2003). The first structure of an RNA m^5C methyltransferase, Fmu, provides insight into catalytic mechanism and specific binding of RNA substrate. *Structure (Camb.)*, **11**, 1609–1620.
6. Ishikawa, I., Sakai, N., Tamura, T., Yao, M., Watanabe, N. & Tanaka, I. (2004). Crystal structure of human p120 homologue protein PH1374 from *Pyrococcus horikoshii*. *Proteins: Struct. Funct. Genet.* **54**, 814–816.
7. Hoang, C. & Ferre-D'Amare, A. R. (2001). Cocrystal structure of a tRNA Psi55 pseudouridine synthase: nucleotide flipping by an RNA-modifying enzyme. *Cell*, **107**, 929–939.
8. McRee, D. E. (1999). XtalView/Xfit– a versatile program for manipulating atomic coordinates and electron density. *J. Struct. Biol.* **125**, 156–165.
9. Cheng, X. & Roberts, R. J. (2001). AdoMet-dependent methylation, DNA methyltransferases and base flipping. *Nucl. Acids Res.* **29**, 3784–3795.
10. Brimacombe, R. (1995). The structure of ribosomal RNA: a three-dimensional jigsaw puzzle. *Eur. J. Biochem.* **230**, 365–383.
11. Valdez, B. C., Perlaky, L., Henning, D., Saijo, Y., Chan, P. K. & Busch, H. (1994). Identification of the nuclear and nucleolar localization signals of the protein p120. Interaction with translocation protein B23. *J. Biol. Chem.* **269**, 23776–23783.
12. Holm, L. & Sander, C. (1996). Mapping the protein universe. *Science*, **273**, 595–603.
13. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410.
14. Liu, J. F., Wang, X. Q., Wang, Z. X., Chen, J. R., Jiang, T., An, X. M. *et al.* (2004). Crystal structure of KD93, a novel protein expressed in human hematopoietic stem/progenitor cells. *J. Struct. Biol.* **148**, 370–374.
15. Zanchin, N. I., Roberts, P., DeSilva, A., Sherman, F. & Goldfarb, D. S. (1997). Saccharomyces cerevisiae Nip7p is required for efficient 60S ribosome subunit biogenesis. *Mol. Cell Biol.* **17**, 5001–5015.
16. Zanchin, N. I. & Goldfarb, D. S. (1999). Nip7p interacts with Nop8p, an essential nucleolar protein required for 60S ribosome biogenesis, and the exosome subunit Rrp43p. *Mol. Cell Biol.* **19**, 1518–1525.
17. Ishitani, R., Nureki, O., Fukai, S., Kijimoto, T., Nameki, N., Watanabe, M. *et al.* (2002). Crystal structure of archaeosine tRNA-guanine transglycosylase. *J. Mol. Biol.* **318**, 665–677.
18. Aravind, L. & Koonin, E. V. (1999). Novel predicted RNA-binding domains associated with the translation machinery. *J. Mol. Evol.* **48**, 291–302.
19. Ishitani, R., Nureki, O., Nameki, N., Okada, N., Nishimura, S. & Yokoyama, S. (2003). Alternative tertiary structure of tRNA for recognition by a posttranscriptional modification enzyme. *Cell*, **113**, 383–394.
20. Klimasauskas, S., Kumar, S., Roberts, R. J. & Cheng, X. (1994). HhaI methyltransferase flips its target base out of the DNA helix. *Cell*, **76**, 357–369.
21. Lee, T. T., Agarwalla, S. & Stroud, R. M. (2005). A unique RNA Fold in the RumA-RNA-cofactor ternary complex contributes to substrate selectivity and enzymatic function. *Cell*, **120**, 599–611.
22. Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K. & Olson, A. J. (1998). Automated docking using a Lamarckian genetic algorithm and and empirical binding free energy function. *J. Comput. Chem.* **19**, 1639–1662.
23. Liu, Y. & Santi, D. V. (2000). m5C RNA and m5C DNA methyl transferases use different cysteine residues as catalysts. *Proc. Natl Acad. Sci USA*, **97**, 8263–8265.
24. Bujnicki, J. M., Feder, M., Ayres, C. L. & Redman, K. L. (2004). Sequence-structure-function studies of tRNA: m^5C methyltransferase Trm4p and its relationship to DNA:m^5C and RNA:m^5U methyltransferases. *Nucl. Acids Res.* **32**, 2453–2463.
25. Tobbell, D. A., Middleton, B. J., Raines, S., Needham, M. R., Taylor, I. W., Beveridge, J. Y. & Abbott, W. M. (2002). Identification of in vitro folding conditions for procathepsin S and cathepsin S using fractional factorial screens. *Protein Expr. Purif.* **24**, 242–254.
26. Van Duyne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L. & Clardy, J. (1993). Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. *J. Mol. Biol.* **229**, 105–124.
27. Leslie, A. G. W. (1992). *Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography, No. 26,* Daresbury Laboratory, Warrington, UK.
28. Evans, P. R. (1993). *Proceedings of CCP4 Study Weekend on Data Collection and Processing,* Daresbury Laboratory, Warrington, UK.
29. Schneider, T. R. & Sheldrick, G. M. (2002). Substructure solution with SHELXD. *Acta Crystallog. sect. D*, **58**, 1772–1779.

30. Terwilliger, T. C. & Berendzen, J. (1999). Automated MAD and MIR structure solution. *Acta Crystallog. sect. D*, **55**, 849–861.
31. Terwilliger, T. C. (2002). Automated structure solution, density modification and model building. *Acta Crystallog. sect. D*, **58**, 1937–1940.
32. 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 Crystallog. sect. A*, **4**, 110–119.
33. Collaborative Computational Project Number 4. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallog. sect. D*, **50**, 760–763.
34. Winn, M. D., Murshudov, G. N. & Papiz, M. Z. (2003). Macromolecular TLS refinement in REFMAC at moderate resolutions. *Methods Enzymol.* **374**, 300–321.
35. Winn, M. D., Isupov, M. N. & Murshudov, G. N. (2001). Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallog. sect. D*, **57**, 122–133.
36. Krissinel, E. & Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallog. sect. D*, **60**, 2256–2268.
37. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl Acad. Sci. USA*, **98**, 10037–11041.
38. Lu, G. (2000). TOP: a new method for protein structure comparisons and similarity searches. *J. Appl. Crystallog.* **33**, 176–183.
39. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283–291.
40. Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucl. Acids Res.* **22**, 4673–4680.
41. Gouet, P., Courcelle, E., Stuart, D. I. & Metoz, F. (1999). ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics*, **15**, 305–308.

Edited by J. Doudna

(Received 20 January 2006; received in revised form 13 May 2006; accepted 18 May 2006)
Available online 6 June 2006