doi:10.1016/j.jmb.2008.08.062 J. Mol. Biol. (2008) 383, 652–666

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Crystal Structure of the *Escherichia coli* 23S rRNA:m^{5}C Methyltransferase RlmI (YccW) Reveals Evolutionary Links between RNA Modification Enzymes

S. Sunita[1], Karolina L. Tkaczuk[2,3], Elzbieta Purta[2,4,5], Joanna M. Kasprzak[6], Stephen Douthwaite[5], Janusz M. Bujnicki[2,6] and J. Sivaraman[1]*

[1]*Department of Biological Sciences, National University of Singapore, 14 Science Drive, Singapore 117543*

[2]*Laboratory of Bioinformatics and Protein Engineering, International Institute of Molecular and Cell Biology in Warsaw, ul. Ks. Trojdena 4, 02-109 Warsaw, Poland*

[3]*Institute of Technical Biochemistry, Technical University of Lodz, B. Stefanowskiego 4/10, 90-924 Lodz, Poland*

[4]*Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Pawinskiego 5a, 02-106 Warsaw, Poland*

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

[6]*Bioinformatics Laboratory, Institute of Molecular Biology and Biotechnology, Faculty of Biology, Adam Mickiewicz University, ul. Umultowska 98, 61-614 Poznan, Poland*

Received 8 May 2008; received in revised form 19 August 2008; accepted 21 August 2008
Available online 29 August 2008
Edited by J. Doudna

Methylation is the most common RNA modification in the three domains of life. Transfer of the methyl group from *S*-adenosyl-L-methionine (AdoMet) to specific atoms of RNA nucleotides is catalyzed by methyltransferase (MTase) enzymes. The rRNA MTase RlmI (*r*RNA *l*arge subunit *m*ethyltransferase gene *I*; previously known as YccW) specifically modifies *Escherichia coli* 23S rRNA at nucleotide C1962 to form 5-methylcytosine. Here, we report the crystal structure of RlmI refined at 2 Å to a final *R*-factor of 0.194 (R_{free} = 0.242). The RlmI molecule comprises three domains: the N-terminal PUA domain; the central domain, which resembles a domain previously found in RNA:5-methyluridine MTases; and the C-terminal catalytic domain, which contains the AdoMet-binding site. The central and C-terminal domains are linked by a β-hairpin structure that has previously been observed in several MTases acting on nucleic acids or proteins. Based on bioinformatics analyses, we propose a model for the RlmI–AdoMet–RNA complex. Comparative structural analyses of RlmI and its homologs provide insight into the potential function of several structures that have been solved by structural genomics groups and furthermore indicate that the evolutionary paths of RNA and DNA 5-methyluridine and 5-methylcytosine MTases have been closely intertwined.

Keywords: crystal structure; rRNA modification; methyltransferase; YccW; RlmI

*Corresponding author. E-mail address: dbsjayar@nus.edu.sg.

Abbreviations used: AdoMet, *S*-adenosyl-L-methionine; MTase, methyltransferase; RFM, Rossmann-fold methyltransferases; m^{5}C, 5-methylcytosine; SelMet, L-selenomethionine; AUC, analytical ultracentrifugation; PDB, Protein Data Bank; r-protein, ribosomal protein; AdoHcy, *S*-adenosylhomocysteine; ITC, isothermal titration calorimetry; m^{5}U, 5-methyluridine.

Introduction

S-Adenosyl-L-methionine (AdoMet) is a ubiquitous methyl donor in biological systems, and AdoMet-dependent methyltransferases (MTases) represent a large group of enzymes that show considerable diversity with respect to structure and mechanism of action. Functions mediated by these enzymes include signal transduction, metabolism, biosynthesis, gene expression, and protein trafficking and sorting;[1,2] substrates for these enzymes include various small molecules, proteins, lipids, and nucleic acids.[3] At least seven different three-dimensional folds are exhibited by AdoMet-dependent MTases.[4] The largest superfamily—the Rossmann-fold methyltransferases (RFM)—is characterized by the presence of an AdoMet-binding/catalytic domain that resembles that of Rossmann-fold oxidoreductases.[5] MTases often contain additional domains, sometimes with elaborations of the common fold, and these are often involved in the recognition and binding of substrates or in mediation of oligomerization.[1]

RNA molecules exhibit a large variety of more than 100 different types of nucleoside modification†—more than half of which involve methylation.[6–8] Most of the modifications are found in tRNAs,[9] and many of these have been connected to roles that include maintenance of translational efficiency and fidelity, regulation of cell cycle transitions, and tRNA–protein interaction.[10] In rRNAs, nucleoside methylations, together with their MTases, play crucial roles in the assembly, maturation, and regulation of the protein synthetic cellular machinery,[11–13] and can additionally confer antibiotic resistance.[14] The types of modified nucleosides found in RNA are, by now, fairly comprehensively charted and, in many cases, the enzymes that are responsible for these modifications are also known. However, our understanding of the structures and catalytic mechanisms of these RNA-specific MTases has conspicuously lagged behind.

Here we report the crystal structure and bioinformatics/phylogenetic study of the recently characterized 5-methylcytosine (m^5C) RNA MTase RlmI (*r*RNA *l*arge subunit *m*ethyltransferase gene *I*; formerly known as YccW). RlmI specifically methylates nucleotide C1962 in *Escherichia coli* 23S rRNA.[15] RlmI belongs to the COG1092 family, which comprises a number of functionally uncharacterized putative MTases with similar domain compositions and also includes longer proteins such as the 23S rRNA: m^2G2445 MTase RlmL (formerly known as YcbY).[16] A comparative structural analysis of RlmI enables us to predict the potential function of several presently uncharacterized MTases. The relationship between the higher-order structure and catalytic function of RlmI further reveals important clues about the evolution of MTases that modify the C5 atom of pyrimidines in DNA and RNA.

† http://www.modomics.genesilico.pl/modification_list

Results and Discussion

Structure of RlmI

The structure of recombinant *E. coli* RlmI was solved by the multiwavelength anomalous dispersion method[17] from synchrotron data using L-selenomethionine (SelMet)-labeled protein. The model was refined to a final *R*-factor of 0.194 (R_{free} = 0.242) at 2.0 Å resolution, with good stereochemical parameters (Fig. 1a, Table 1). The RlmI protein consists of 396 amino acids and an N-terminal $(His)_6$ tag, and the crystallographic model clearly shows the path of

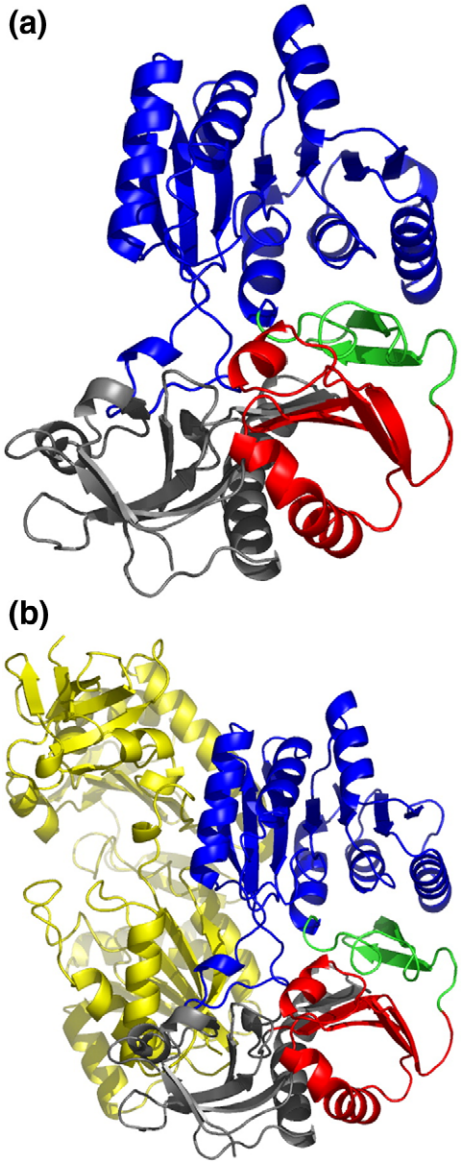

Fig. 1. Structure of RlmI. (a) Ribbon diagram of the RlmI monomer. (b) Ribbon diagram of the RlmI dimer. The catalytic RFM domain is shown in blue, the β-hairpin structure is shown in green, the EEHEE domain is shown in red, and other nonconserved structural elements are shown in gray. The second subunit in the dimer is shown in yellow.