Crystal Structure of ErmC′, an rRNA Methyltransferase Which Mediates Antibiotic Resistance in Bacteria‡

Dirksen E. Bussiere,[§,||] Steven W. Muchmore,[||,⊥] Christopher G. Dealwis,[⊥,#] Gerd Schluckebier,[⊥] Vicki L. Nienaber,[⊥] Rohinton P. Edalji,[@] Karl A. Walter,[@] Uri S. Ladror,[@] Thomas F. Holzman,[@] and Cele Abad-Zapatero*[,⊥]

Laboratory of Protein Crystallography, Department of Scientific Information, Analysis and Management, and Protein Biochemistry Group, Abbott Laboratories, Abbott Park, Illinois 60064-3500

Received December 18, 1997; Revised Manuscript Received March 16, 1998

ABSTRACT: The prevalent mechanism of bacterial resistance to erythromycin and other antibiotics of the macrolide−lincosamide−streptogramin B group (MLS) is methylation of the 23S rRNA component of the 50S subunit in bacterial ribosomes. This sequence-specific methylation is catalyzed by the Erm group of methyltransferases (MTases). They are found in several strains of pathogenic bacteria, and ErmC is the most studied member of this class. The crystal structure of ErmC′ (a naturally occurring variant of ErmC) from *Bacillus subtilis* has been determined at 3.0 Å resolution by multiple anomalous diffraction phasing methods. The structure consists of a conserved α/β amino-terminal domain which binds the cofactor *S*-adenosyl-L-methionine (SAM), followed by a smaller, α-helical RNA-recognition domain. The β-sheet structure of the SAM-binding domain is well-conserved between the DNA, RNA, and small-molecule MTases. However, the C-terminal nucleic acid binding domain differs from the DNA-binding domains of other MTases and is unlike any previously reported RNA-recognition fold. A large, positively charged, concave surface is found at the interface of the N- and C-terminal domains and is proposed to form part of the protein−RNA interaction surface. ErmC′ exhibits the conserved structural motifs previously found in the SAM-binding domain of other methyltransferases. A model of SAM bound to ErmC′ is presented which is consistent with the motif conservation among MTases.

Erythromycin, a member of the macrolide−lincosamide−streptogramin B group (MLS[1]), is one of the most common antibiotics used against Gram-positive bacteria. These antibiotics inhibit bacterial protein synthesis by altering ribosome function (*1*). The organism that produces this antibiotic (*Streptomyces erythreus*) possesses an erythromycin resistance methylase gene (designated *erm*) which is involved in the protection of the bacterial ribosome by methylating the 23S rRNA of the 50S ribosomal subunit. To date, more than 30 *erm*-related genes have been identified from bacterial sources and either the *erm*C or *erm*A genes have been shown to be present in up to 98% of the erythromycin-resistant strains of *Staphylococcous aureus* that were isolated clinically (*2*). Other members of the *erm* family include *erm*E, *erm*K, and *erm*AM with amino acid sequence identities of 20, 27, and 49%, respectively, when they are compared with *erm*C.

*erm*C is the best characterized member of the *erm* gene family. It was originally identified in *S. aureus* but has been subsequently transferred naturally to *Bacillus subtilis* (*3*). This gene encodes the 244-residue ErmC methyltransferase (MTase) within plasmid pE194 (*4*). ErmC′ is a variant encoded by plasmid pIM13 (*5*) from *B. subtilis* and differs from ErmC in only five amino acids (*6*). ErmC′ is more biochemically stable than ErmC and can be readily cloned and expressed in *Escherichia coli* (*7*). All members of the Erm group of MTases mono- and dimethylate the exocyclic N6 position of a highly conserved adenine nucleotide (A2058, *E. coli* numbering, Figure 1) (*8*, *9*) within the peptidyltransferase loop of the 23S rRNA (*10*).

MTases are enzymes which methylate a wide variety of substrates using *S*-adenosyl-L-methionine (SAM) as the universal methyl donor and releasing *S*-adenosyl-L-homocysteine (SAH) as a reaction product. The crystal structures of a number of MTases have been determined. Two DNA

‡ Refined coordinates of ErmC′ have been deposited in the Brookhaven Protein Data Bank under accession code 2ERC. Until their release, requests should be sent to the corresponding author.

* To whom correspondence should be addressed: Department 46Y, Building AP-10, Abbott Laboratories, 100 Abbott Park Rd., Abbott Park, IL 60064-3500. Phone: (847) 937-0294. Fax: (847) 937-2625. E-mail: abad@abbott.com or abad@mozart.pprd.abbott.com.

§ Department of Scientific Information, Analysis and Management.

|| The contributions of these authors to this work should be considered equivalent.

⊥ Laboratory of Protein Crystallography.

Present address: Department of Pharmacology, B367, Yale University Medical School, New Haven, CT 06510.

@ Protein Biochemistry Group.

[1] Abbreviations: C5-MTases, C5 methyltransferases; COMT, catechol *O*-methyltransferase; CCD, charged-coupled detectors; DTT, dithiothreitol; EDTA, ethylenediaminetetraacetic acid; *erm*, erythromycin resistance methylase gene; GNMT, glycine *N*-methyltransferase; IPTG, isopropyl β-D-thiogalactopyranoside; N6-MTases, N6 methyltransferases; N4-MTases, N4 methyltransferases; $NH_4(SO_4)_2$, ammonium sulfate; $MgCl_2$, magnesium chloride; MLS, macrolide−lincosamide−streptogramin B group; MTases, methyltransferases; NCS, noncrystallographic symmetry; NaCl, sodium chloride; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; PCR, polymerase chain reaction; PDB, Protein Data Bank; PMSF, phenylmethanesulfonate fluoride; rRNA, ribosomal RNA; SAH, *S*-adenosyl-L-homocysteine; SAM, *S*-adenosyl-L-methionine; Se-Met, selenomethionine; TDGM, Tris/DTT/glycerol/magnesium.

S0006-2960(97)03113-9 CCC: $15.00
Published on Web 04/28/1998

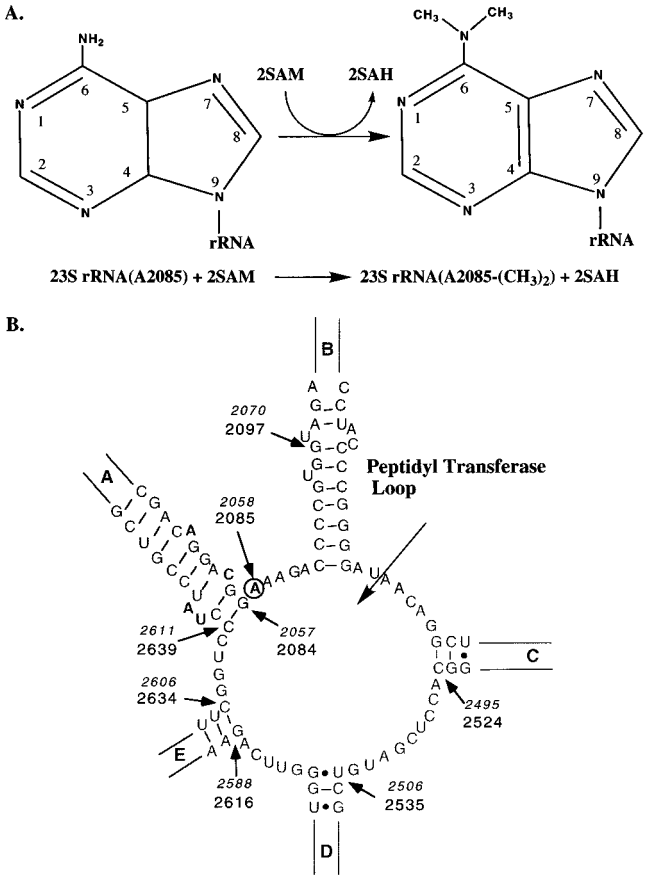

FIGURE 1: Reaction catalyzed by the Erm methyltransferases. (A) Schematic of the dimethylation of the N6 exocyclic nitrogen of the target adenine. The cofactor *S*-adenosyl-L-methionine (SAM) is the methyl donor, and *S*-adenosyl-L-homocysteine (SAH) is the demethylated product. (B) The proposed secondary structure for the conserved peptidyl transferase loop within domain V of the 23S rRNA in *B. subtilis* (*52*) is shown with each helical stem labeled (A−E). Arrows point to nucleotide markers along the sequence (upper, *E. coli* numbering; and lower, *B. subtilis* numbering). The target adenine is circled (adapted from ref *7*).

MTases which methylate C5 of the cytosine ring (C5-MTases) have been structurally characterized: M.*Hha*I from *Haemophilus haemolyticus* (*11*−*14*) and M.*Hae*III from *Haemophilus aegypticus* (*15*). These MTases methylate a cytosine base within two different specific recognition sequences (5′-GCGG-3′ for M.*Hha*I and 5′-GGCC-3′ for M*Hae*III) in double-stranded DNA. In addition, the structure of a DNA MTase which methylates the exocyclic N6 amino group of adenine (N6-MTase) from *Thermus aquaticus* (M.*Taq*I) has been described when it is complexed with the methyl donor SAM, with the reaction product SAH, and also with the antibiotic inhibitor sinefungin (adenosylornithine) (*16*). Finally, the structure of M.*Pvu*II, a DNA MTase which modifies the exocyclic N4 of cytosine (N4-MTase), has revealed strong structural similarities among the active sites of the three DNA Mtases (*17*).

As part of the restriction−modification system in organisms, DNA MTases recognize and methylate a specific nucleotide within a particular DNA sequence (*18*). Structurally, all DNA MTases consist of two separate domains: a common catalytic domain and a variable DNA-recognition domain, which is responsible for the interaction with the various DNA sequences. In contrast, the three-dimensional structure of *Vaccinia* virus protein VP39 (*19*) is a single-domain structure, and the substrate-recognition module is closely associated with the cofactor domain. VP39 is an mRNA methyltransferase that methylates both ends of *Vaccinia* virus messenger RNA at the O2′ positions.

The structures of two MTases capable of methylating small molecules have also been reported: catechol *O*-methyltransferase (COMT), a Mg^{2+}-dependent enzyme which attaches a methyl group to the exocyclic oxygen of catechol (*20*), and the tetrameric glycine *N*-methyltransferase (GNMT) from rat liver (*21*), which catalyzes the methylation of glycine to sarcosine. Protein MTases have also been characterized structurally by the crystal structure of CheR (*22*), a protein MTase involved in the chemotactic response in *Salmonella typhimurium*.

Although substrate recognition is accomplished by a variety of structural motifs in the different MTases, all of them share a structurally homologous α/β catalytic domain which contains the SAM-binding site. This module consists of a central β-sheet of mixed strand polarity. It contains five parallel β-strands with a topology analogous to that of a dinucleotide-binding fold (5−4−1−2−3) and is extended after β5 by two additional antiparallel strands (6−7) (*23*).

In this work, we present the crystal structure of ErmC′, a variant of ErmC, refined to 3.0 Å resolution. Recently, the solution structure of another member of the Erm family (ErmAM) has been determined by NMR methods (*24*). The amino acid sequences of ErmAM and ErmC′ are 49% identical, and the folds of both proteins are similar. The structures of ErmC′ and ErmAM provide a three-dimensional fingerprint for other Erm-like proteins, relate this class of enzymes to other MTases, and provide a framework for a better understanding of the mechanisms of rRNA recognition by bacterial proteins. In addition, the three-dimensional structure of ErmC′ is a valuable aid in the design of potent ErmC inhibitors. Given the importance of the Erm family of MTases in bacterial resistance to the MLS group of antibiotics, it is likely that inhibition of this group of enzymes will prolong the effective life span of current antibiotics.

MATERIALS AND METHODS

Protein Expression and Purification. The expression vector pTERM31 was constructed by PCR amplification of the ermC′ gene and the upstream kdsB cistron from pERM-1 (*7*). Subcloning the PCR product into pET24+ (Novagen, Madison, WI) was performed using the *Bam*HI and *Hin*dIII sites included in the "tailed" PCR primers. This new construct allowed the expression of ErmC′ by translational coupling to kdsB, under the control of the T7*lac* promoter. The pTERM31 plasmid was transformed into *E. coli* strain BL219(DE3)/pLysS (Novagen), and the resulting strain was used for production of ErmC′. Transformed cells were grown at 27.5 °C in a New Brunswick Scientific (Edison, NJ) Micros fermentor containing 10 L of Superbroth (BIO 101, La Jolla, CA), supplemented with kanamycin, chloramphenicol, and glucose. When the culture optical density reached 1.10, ErmC′ expression was induced by the addition of 1 mM isopropyl β-D-thiogalactopyranoside (IPTG). Cells were harvested 400 min postinduction.

Se-Met-labeled ErmC′ was produced by introducing the expression vector pTERM31 into the *E. coli* met-auxotrophic strain B834(DE3). Fermentations were performed with this

Table 1: Summary of the Data Reduction and Structure Determination Statistics of ErmC′ (Space Group *P*622)

wavelength (Å)	no. of reflections	resolution (Å)	R_{merge}[a]	$f'(e)$	$f''(e)$
		Data Reduction Statistics			
		Beamline X8C (NSLS Brookhaven National Laboratory)			
$\lambda_1 = 0.907\ 79$	4643 (85.4%)	3.41	6.1 (12.8)	−1.642	3.329
$\lambda_2 = 0.979\ 35$	3617 (83.4%)	3.69	5.5 (4.2)	−12.576	3.803
$\lambda_3 = 0.979\ 19$	3283 (74.7%)	3.67	6.6 (7.3)	−12.944	6.555
		Beamline F1 (CHESS, Cornell University)			
0.92178	8016 (93.1%)	2.9	7.5 (39.8)		

	MAD Phasing Statistics Using Maximum Likelihood[b] resolution (Å)					
	42.9−9.4	9.4−6.8	6.8−5.3	5.5−4.8	4.8−4.3	3.9−3.6
figure of merit[c]	0.92 (0.88)	0.91 (0.79)	0.87 (0.74)	0.85 (0.75)	0.82 (0.65)	0.65 (0.48)
phasing power[d]						
λ_1	4.19	5.01	4.02	3.40	2.77	1.50
λ_2	8.03	8.24	5.14	4.11	3.05	1.76
λ_3	9.73	9.25	6.07	4.83	4.27	2.49

[a] $R_{merge} = \sum_h\sum_i|I_{hi} - \langle I_h\rangle|/\sum_h\sum_i I_{hi}$, where $\langle I_h\rangle$ is the mean intensity of the I_{hi} measurements for reflection h with indices hkl. The first figure corresponds to the overall value; in parentheses is the one for the last shell. [b] Reference *34*. [c] Overall, the last shell in parentheses. The figure of merit for each reflection h of indices hkl is calculated as FOM $= \sqrt{(\langle A_h\rangle^2 + \langle B_h\rangle^2)/(\langle A_h^2\rangle + \langle B_h^2\rangle)}$, where A_h and B_h denote the unknown real and imaginary parts of the complex native structure factor for reflection h, respectively. The averages $\langle A_h\rangle$, $\langle B_h\rangle$, $\langle A_h^2\rangle$, and $\langle B_h^2\rangle$ are taken with respect to its two-dimensional probability distribution (eq 23 in ref *34*). The overall FOM in each resolution shell is calculated as the unweigthed average of the individual FOMs of all the unique reflections in that shell. [d] Phasing power $= (1/N_{refl})\sum_h[|\Delta^{ANO}_{obs}(h,i)|/\int_0^{2\pi}[|\Delta^{ANO}_{obs}(h,i) - \Delta^{ANO}_{calc}(h,i)|]p(\varphi_h)d(\varphi_h)]$ for wavelength i and reflection h of indices hkl in the corresponding resolution shell. $|\Delta^{ANO}_{obs}(h,i)|$ and $|\Delta^{ANO}_{calc}(h,i)|$ represent the observed and calculated values for the i wavelength, respectively; $p(\varphi_h)$ is the probability of a phase value of φ for reflection h.

strain as described elsewhere (*25*). Briefly, cells were grown at 37 °C until the culture reached an optical density of 0.85. At this point, IPTG was added to induce expression of ErmC′ and the culture temperature was shifted to 29 °C. Cells were harvested 13.5 h postinduction. Incorporation of Se-Met into the protein preparations was assessed by mass spectrometry. The data indicated the presence of three or four Se-Met amino acid residues per molecule.

Frozen cell paste (200−250 g) was thawed at room temperature and resuspended in 5−10 volumes of cold lysis buffer [50 mM Tris, 5 mM DTT, 1 mM PMSF, 2 mM EDTA, and 0.2% Triton X-100 (pH 7.8)]. The cells were lysed with a French press, and cell debris was removed by centrifugation. The supernatant was dialyzed overnight against 20 L of Tris/DTT/glycerol/magnesium (TDGM) buffer at pH 7.8 (50 mM Tris, 5 mM DTT, 10% glycerol, and 10 mM $MgCl_2$). The dialysate was then applied to a Sepharose Fast Flow column (Pharmacia) that had been pre-equilibrated in TDGM buffer. Fractions were assayed for methyltransferase activity, and those containing ErmC′ were pooled, applied to a TSK SP-5PW column (TosoHaas, Montgomeryville, PA), and eluted with an NaCl gradient. The purified protein was then concentrated on a YM-10 (Amicon) membrane.

Crystallographic Methods. Crystals of ErmC′ were obtained by the hanging-drop vapor diffusion method. Protein solutions at concentrations in the range of 8−10 mg/mL contained 0.1 M Tris-HCl buffered at pH 7.5, 2 mM $MgCl_2$, 10% (v/v) glycerol, 5 mM DTT, and an approximately 40-fold molar excess of SAH. Equilibrating solutions in the reservoir contained 3.2 M $NH_4(SO_4)_2$ in 0.1 M citrate buffered at pH 4.0. Crystallization conditions for both the native and Se-Met forms of the protein are similar, although additional DTT (up to 20 mM) was present in the Se-Met protein solutions. Native crystals grew as hexagonal plates with the dimensions $0.250 \times 0.250 \times 0.005$ mm^3. Se-Met crystals were thicker with a maximum of 0.050 mm in the third dimension. Crystals were prepared for low-temperature data collection (−170 °C) by transfer to a cryoprotective solution containing 25% (v/v) glycerol, 0.1 M citrate (pH 4.0), and 3.2 M $NH_4(SO_4)_2$. Crystals diffracted to approximately 3.6 Å on a Rigaku RU200 rotating anode equipped with Yale focusing mirrors using an RAXIS-IIC image plate. The space group was determined as *P*622, with one ErmC′ molecule per asymmetric unit, with the following cell dimensions: $a = b = 147.4$ and $c = 58.1$ Å. Multiple anomalous diffraction (MAD) (*26*) data to 3.5 Å resolution were collected from a single frozen crystal at the National Synchrotron Light Source (Beamline X8C, Brookhaven National Laboratory, Upton, NY). The data were recorded using a CCD-based detector (*27*) using 0.2° oscillations per frame and a 30 s exposure. Data sets for each wavelength consisted of two separate 90° sweeps separated by 180° in ω. The data were integrated using the MADNESS (*28*) data reduction package and were scaled and merged with programs from the XDS suite (*29*) (Table 1). An additional data set was collected at the Cornell High Energy Synchrotron Source (Cornell University, Ithaca, NY) (Beamline F1, 2.9 Å data, $\lambda = 0.921\ 78$ Å). The higher-resolution data set was also recorded on a CCD detector (*30*) in oscillation frames of 0.2° each. Data reduction was carried out with the HKL suite of programs (*31*) (Table 1). Data from the MAD phasing sets were imported into both the CCP4 (*32*) and PHASES (*33*) crystallographic program suites. The positions of three of the five expected selenium atoms were derived by interpretation of both anomalous difference and dispersive difference Patterson maps. Once these positions were determined, heavy atom positional refinement and phase calculations were undertaken in the SHARP (*34*) maximum-likelihood phasing program. A log-likelihood gradient map (E. d. La Fortelle, unpublished results) suggested the presence of a fourth Se atom, and examination of residual difference Fourier maps confirmed its position. This site was then incorporated into the phase calculations (Table 1). Phase

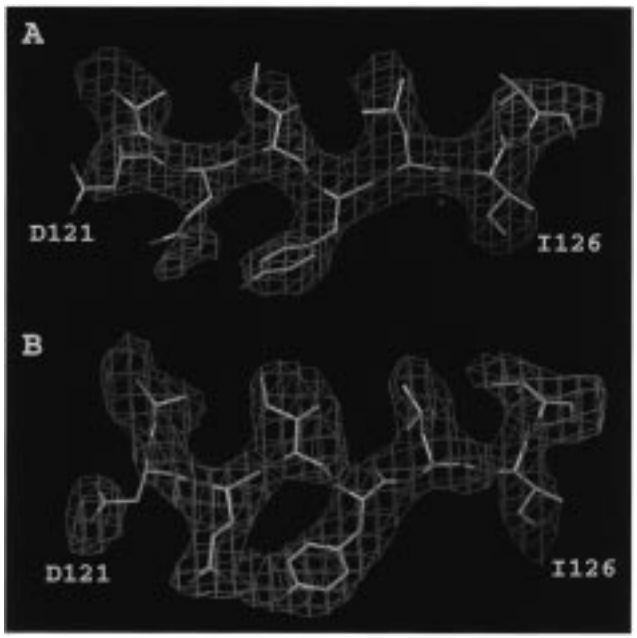

FIGURE 2: Illustrative electron density maps. (A) Portion of the original MAD phased electron density map using the maximum-likelihood phasing algorithm supplemented with the solvent-flattening protocol. The portion corresponds to the central β-sheet (β5) of the SAM-binding domain. (B) Corresponding section of the final $2F_o - F_c$ map. The contour level is 1σ.

improvement via iterative density modification was undertaken using a modified solvent-flattening protocol (E. d. La Fortelle, unpublished results) with the SOLOMON program (*35*). A sample of the density-modified electron density map is presented in Figure 2A. Examination of electron density maps calculated using data from wavelength 1 ($\lambda = 0.90779$ Å) revealed readily interpretable features. Atomic models were built using the graphics programs O (*36*) and QUANTA (*37*). Refinement of the initial model was accomplished with iterative cycles of model building and refinement against a 3.5 Å data set using simulated annealing protocols in the XPLOR (*38*) software package to a crystallographic *R*-factor, where $R_{cryst} = 26.3\%$ and $R_{free} = 43.5\%$ (10% of the data) (*39*). Detailed analysis of the second and higher-resolution data set for the resolution range from 3.6 to 2.9 Å revealed significant deviations for symmetry-related reflections in *P*622. Therefore, the space group was reassigned as *P*6 with two molecules in the asymmetric unit, and the stronger data were reprocessed to a 3.0 Å cutoff for subsequent refinement (Table 2). Models derived from the initial phasing experiments and refinement were used to calculate both $2F_o - F_c$ and $F_o - F_c$ electron density maps in the new space group. Noncrystallographic symmetry (NCS) restraints were applied to both monomers at a level of 175 and 250 kcal/mol, for both side and main chain atoms, respectively. The NCS term was kept in all refinements.

After the first round of refinement in *P*6, R_{cryst} reached approximately 34%, and more importantly, the R_{free} had fallen to 37%. Subsequent refinement cycles were repeated using the standard residual target function and simulated annealing−torsion angle dynamic refinement followed by 100 cycles of Powell minimization; an isotropic temperature factor correction and bulk solvent correction were also applied. Five cycles of this protocol improved the model/data agreement to an R_{cryst} of 24.1% and an R_{free} of 31.8%.

Table 2: Data Reduction and Refinement Statistics for ErmC′ (Space Group *P*6)[a]

number of observations	138 382
number of unique reflections	14 568
% of data $>1\sigma$	94.0 (84.3)[b]
% of data $>3\sigma$	86.6 (62.9)[b]
R_{merge} (%)	6.8 (27.1)[b]
number of protein atoms	3934
number of reflections used (8−3 Å)	11 972
R_{free} (%) (8−3 Å)	31.3
R_{cryst} (%) (8−3 Å)	23.2
rmsd[c] on bonds (Å)	0.014
rmsd[c] on angles (deg)	2.12
rmsd[c] on improper angles (deg)	1.0
rmsd[c] on dihedral angles (deg)	21.8
G-factor[d]	−0.12

[a] $R_{cryst} = \sum_h ||F_o| - |F_c|| / \sum_h |F_{obs}|$, where F_o and F_c are the observed and calculated structure factor amplitudes for reflection *h* with indices *hkl*, respectively. R_{free} was calculated against 8.8% of the reflections randomly removed from the 3.0 Å data set (1052 reflections). R_{mer} definition as in Table 1. [b] Overall; last shell in parentheses at the corresponding σ cutoff. [c] Root-mean square deviations from ideality. [d] Reference *42*.

Finally, to improve the quality of this medium-resolution structure, the atomic model was subjected to 100 rounds of Powell minimization which included a conformation database potential (D. E. Bussiere, J. Kuszewski, A. M. Gronenborg, and G. M. Clore, unpublished results). After the use of this conformational potential, the final R_{cryst} and R_{free} values were 23.2 and 31.3%, respectively (Table 2). A portion of the refined $2F_o - F_c$ electron density map is illustrated in Figure 2B. Brookhaven Protein Data Bank (PDB) (*40*) searches for similarity of folds were conducted using the program SARF (*41*). Three-dimensional structure superpositions were performed using the least-squares superposition routines of the package O (*36*) using a cutoff of 3.8 Å for improvement and a consecutive stretch of three amino acid residues for the smallest aligned fragment size.

RESULTS AND DISCUSSION

ErmC′ Structure. The refined model of the ErmC′ MTase contains residues Gln10−Lys244 of the entire ErmC′ molecule (residues 1−244). The density for residues Met1−Ser9 was not interpretable; therefore, this portion was considered to be disordered and was not included in the final model. The Ramachandran plot (not shown) and model evaluation statistics are reasonable for 3.0 Å resolution (*42*) (Table 2). No solvent molecules have been included.

The three-dimensional appearance of the ErmC′ molecule can be described as bilobed with two domains of uneven size (Figure 3). The polypeptide fold of the larger catalytic domain is similar to the one described for other MTases (*19*−*23*) and consists of an α/β topology formed around a central β-sheet containing six parallel β−strands (labeled β1−β6) and an antiparallel strand (β7) inserted between β5 and β6 (topology 6−7−5−4−1−2−3) (*23*) (Figure 3A,B). This conserved module binds the SAM cofactor. The loop connecting strands β6 and β7 (Pro159−Val169) exhibits poor electron density and also shows the largest deviations between the two molecules in the asymmetric unit. After β7, the polypeptide continues into the smaller RNA-binding domain (Figures 3 and 4). The connection between β5 and β6 is made up of two helices (αE and αF) which form an

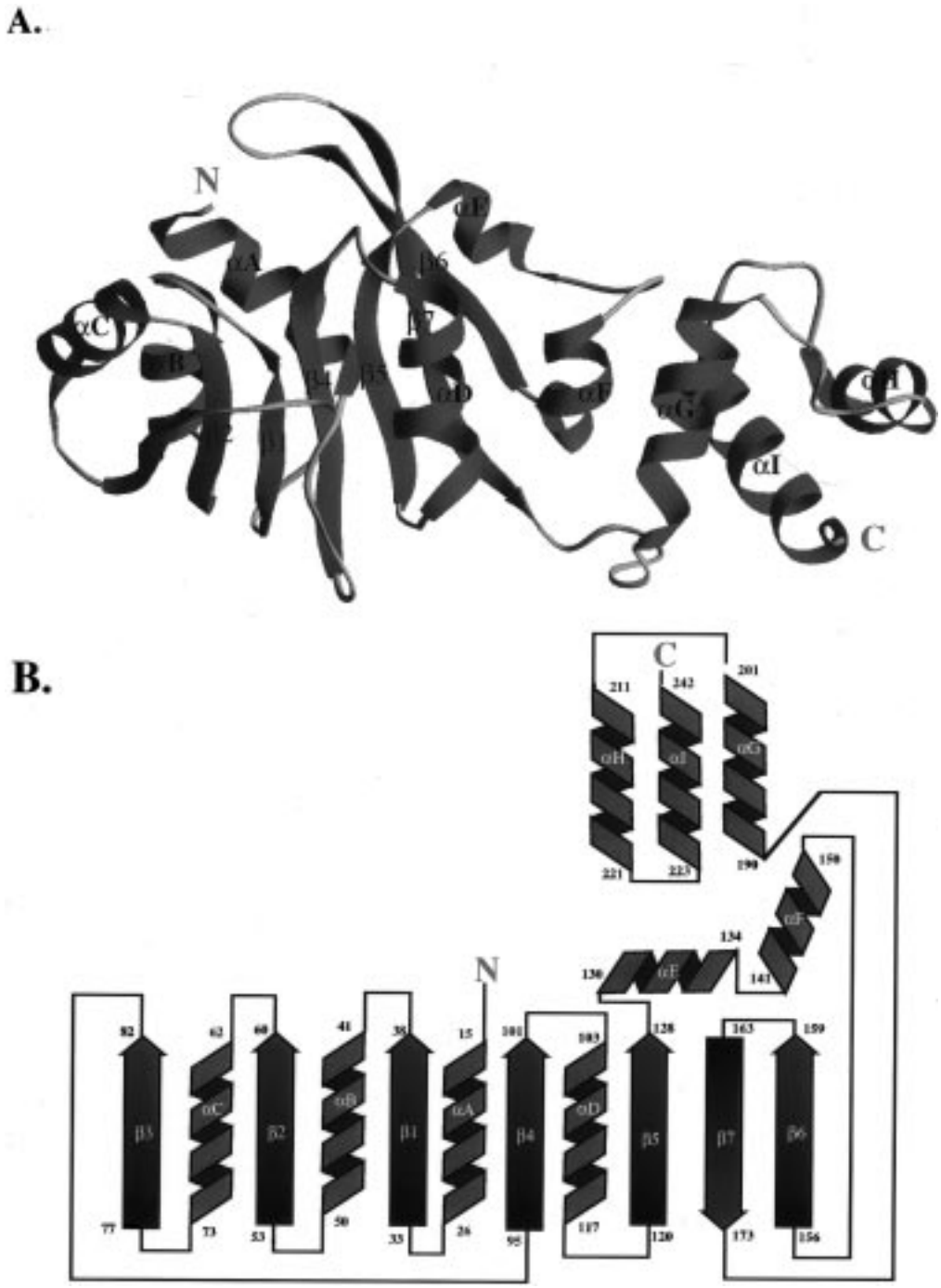

FIGURE 3: Ribbon representation of the structure of ErmC′. (A) The backbone structure with its two separate domains is presented. This figure was prepared with the RIBBONS (*53*) software package. (B) Schematic representation of the topology of the structure of ErmC′. The strands are labeled $\beta1-\beta7$ and the helices αA−αI to relate to the nomenclature of other MTases (*23*).

L-like protrusion from the catalytic domain that extends toward the smaller helical domain. Helices αD, αF, and αG form a concave interface in the interdomain region, with αF forming the back of the surface (Figure 3).

The RNA binding domain is completely α-helical, consisting of three helices (αG−αI). The first and longest of them has four turns and is arranged perpendicular to the remaining two (Figures 3A and 4). Searches through the PDB failed to detect any structural homology between the α-helical domain of ErmC′ and other protein folds. This domain is topologically different from the three commonly found RNA-binding domains previously described (*43*, *44*). The RNA-binding domain of the ribosomal protein L11 has also been found to contain three α-helices, forming an α-helical bundle with a topology strikingly similar to the homeodomain class of eukaryotic DNA binding proteins (*45*). Attempts to sumperimpose the helical RNA-binding domains of L11 and ErmC′ have failed to reveal a common motif.

Comparison with ErmAM. The fold of the recently reported solution structure of ErmAM (*24*) is the same as the one observed for ErmC′. Coordinate errors for ErmAM and ErmC′ can be estimated as 1.4 and 0.5 Å, respectively (*24*, *46*). Overall superposition between all the Cα atoms of ErmAM and ErmC′ results in 165 equivalent Cα pairs with an rms deviation of 2.1 Å. About two-thirds of the equivalent Cα atoms from the overall superposition cor-

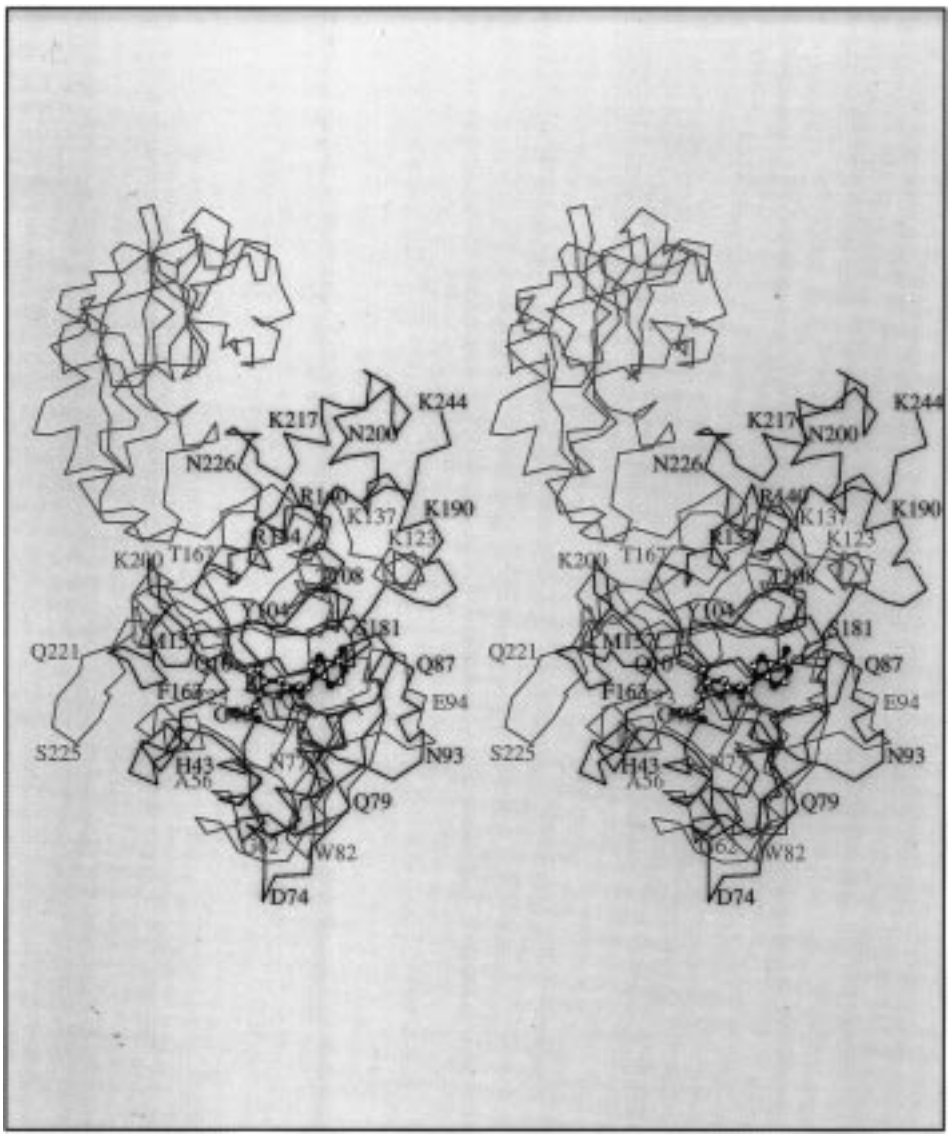

FIGURE 4: Stereodiagram of the polypeptide backbone fold of M.*Taq*I MTase (red) superimposed on the refined structure of ErmC′ (blue). The view is chosen to illustrate the similarity of the SAM-binding domains and the differences in the fold and disposition of the recognition domains for DNA (M.*Taq*I) and RNA (ErmC′). For reference, the position of SAM as found in M.*Taq*I (PDB set 2ADM) is shown as a ball-and-stick model (*16*). Blue, yellow, and red denote the nitrogen, sulfur, and oxygen atoms, respectively, in the SAM. Sequence numbers label the path of the SAM-binding domain of M.*Taq*I (red). The entire polypeptide chain of ErmC′ (blue) is labeled with markers approximately every 10 residues. The two structures were superimposed using the program suite O (*36*) to the rms deviation listed in Table 3. This figure was prepared with MOLSCRIPT (*54*).

respond to residues within the catalytic domain (105 pairs) which upon separate optimization gives 123 pairs and an rms of 1.9 Å. Independent superposition of the RNA-recognition domain yields 50 Cα pairs with an rms deviation of 1.9 Å (Table 3). A rigid body motion (∼10° rotation and ∼2 Å translation) is required to optimize the superposition of the smaller domains after the overall aligment. This systematic shift could be the result of the large structural differences currently observed between the two structures at the end of β7, at the interdomain linking residues (Arg178−Arg182 in ErmC′ and Arg177−Asp181 in ErmAM) (Figure 5). However, the possibility that this domain displacement reflects a significant difference between the structures cannot be ruled out.

Maximum flexibility is observed in ErmC′ at the amino terminus and in the loop between β6 and β7 (vide supra). Considerable flexibility is also observed in ErmAM at the amino terminus and at the conection between β6 and β7. In addition, the NMR structure is poorly defined at the loop linking β1 and αB and on both sides of αC (β2, α2, and α3, respectively) (*24*). The uncertainty in these regions is most likely due to missing NOE resonances (*24*). The increased rigidity of the residues of ErmC′ in the proximity of αC is probably due to the effect of intermolecular contacts. In solution, it has been shown by light scattering (E. Matayoshi and J. Huffaker, unpublished results) that ErmC′ forms oligomeric aggregates. Moreover, intermolecular contacts in the crystal near αC facilitate the formation of a ring-like dodecamer of ErmC′ molecules displaying approximate 622 symmetry. Both structures have a β-bulge at the conserved residue Leu155 even though the following residues are different (Leu155-Ser-Met and Leu154-Leu-Lys for ErmC′ and ErmAM, respectively). The final refined electron density map could not confirm the presence of a cis Pro at position 165 in ErmC′ corresponding to the observed cis Pro (Pro164) in ErmAM (*24*). Within the

Table 3: Comparison of ErmC′ with Other Methyltransferases[a]

	*Hha*I 327	TaqI 421	COMT 221	GNMT 292	ErmAM[d] 243	VP39 348	CheR 288
rms[b] (Å)	2.1	1.9	1.8	2.0	1.9 (2.1)	1.9	1.9
no. of pairs	110	115	113	118	123 (165)	92	89
% identity[c]	15	21	11	12	40 (45)	8	11

[a] *Hha*I, *H. haemolyticus* methyltransferase I (*11*), PDB entry 3MHT. TaqI, *T. aquaticus* methyltransferase I (*16*), PDB entry 2ADM. COMT, cathecol *O*-methyltransferase (*20*), PDB entry 1VID. GNMT, glycine *N*-methyltransferase (*21*), PDB entry 1XVA. ErmAM, rRNA methyltransferase (*24*) solution structure determined by NMR, PDB entry 1YUB. VP39, *Vaccinia* protein, mRNA methyltransferase (*19*), PDB entry 1VPT. CheR, chemotactic receptor protein methyltransferase (*22*), PDB entry 1AF7. [b] Root-mean-square deviation in angströms between the corresponding number of equivalent amino acid residues after the overall least-squares superposition as described in Materials and Methods. [c] Percent identity of the equivalent amino acid residues after superposition. [d] Catalytic domain (Asn11−Lys180) superposition; in parentheses are the corresponding values for the overall least-squares superposition.

respective limitations of the experimental techniques employed (*46*), the structural comparison between ErmC′ and ErmAM suggests that the structural differences between these two rRNA MTases are probably not significant at this level of resolution.

Relation to Other MTases. Overall superposition of several MTases onto ErmC′ reveals that M.*Taq*I is the closest structural homologue with an rms deviation of 1.9 Å for 115 Cα pairs and 21% sequence identity within the SAM-binding domain (Table 3). This is consistent with the observation that both enzymes are N6-MTases, catalyzing the methylation of the N6 exocyclic nitrogen of adenine. In addition, both M.*Taq*I and ErmC′ belong to the γ group of MTases (*18*) because of the similar arrangement of two highly conserved sequence motifs (I and IV) first characterized in DNA amino methyltransferases (*47*). These motifs are part of the 10 sequence motifs initially identified in cytosine methyltransferases (motifs I−X) (*48*). A structure-based sequence comparison of more than 40 amino MTases (N6-adenine and N4-cytosine) has validated nine (I−VIII and X) of the original motifs found in C5-cytosine MTases (*49*). In ErmC′ as in M.*Taq*I, motif X precedes the remaining eight motifs which occur later in strict sequence (X and I−VIII), prior to the nucleic acid-recognition domain (*49*). Figure 4 shows the structural superposition of the catalytic domains between M.*Taq*I and ErmC′, and Figure 5 depicts the amino acid sequence alignment of ErmC′ with M.*Taq*I based on the three-dimensional superposition of the two structures. The structures of ErmC′ and M.*Taq*I superimpose well within the SAM-binding domain and at the regions of the conserved sequence motifs, which include those residues that are thought to bind the target adenine (*50*). At the C terminus of β1, motif I is centered around residues Gly-X-Gly and is typically preceded by Asp or Glu four residues toward the N terminus. Motif I forms the turn connecting β1 to αB and creates part of the binding pocket for the methionine and ribose regions of SAM (*23*). The MTase consensus sequence at motif II contains a negatively charged residue at the end of β2 (Glu59 in ErmC′; Figure 5) which hydrogen bonds to the ribose hydroxyls of SAM and is followed by a hydrophobic residue (Ile60) which is within van der Waals contact of the adenine ring of SAM (*49*). Within motif III, Asp84 in ErmC′ corresponds to Asp89 in M.*Taq*I which forms a hydrogen bond with the exocyclic amino group of the adenine moiety in SAM. The sequence GNIPY (motif IV; Figures 4 and 5) corresponds to the catalytic sequence GNPPY in M.*Taq*I (*49*). The departure from the PP sequence conserved among nearly all DNA amino MTases is worth noting (*49*). The consensus sequence (N/S)IP(Y/F) has been observed in the analysis of 12 rRNA N6-MTases sequences by Cheng (*49*). Whether the composite sequence (G/N/S)IP(Y/F) can be considered a fingerprint for rRNA MTases or even di-MTases requires further investigation.

The residues comprising motif V form αD in ErmC′ and do not superimpose well with the corresponding αD in M.*Taq*I due to a rigid body shift. This is caused by the deletion of 34 residues (Gly109−Leu142) in M.*Taq*I immediately after the catalytic sequence G104NPPY108 (motif IV), which results in a displacement of the two helices relative to each other (Figures 4 and 5). Motif V begins with Asn105 (N105ISTDIIRKIV), and although the amino acid sequence in this conserved helix is similar to the one found in M.*Taq*I (N141LYGAFLEKAV), the negatively charged glutamate (Glu148 in M.*Taq*I) has been replaced by an arginine (Arg112 in ErmC′). Residue Phe146 in motif V of M.*Taq*I is within van der Waals distance of the adenine fragment of SAM; Phe is replaced by Ile110 in ErmC′. Motif VI consists of a cluster of three hydrophobic residues at the C terminus of β5 (FVV and LIV in M.*Taq*I and ErmC′, respectively) which have been suggested to be involved in placing the target (substrate) adenine ring on the side opposite motif IV (*49*, *51*). Motif VII is weakly conserved even among the class of DNA C5-MTases and corresponds in M.*Hha*I to residues in the loop connecting αE and β6 (*49*). The corresponding region in the two N6-MTases superimposes residues Glu149-Val150 in ErmC′ (between αF and β6) with Glu184-Gly185 in M.*Taq*I (Figures 4 and 5).

The loop between β6 and β7 on ErmC′ is topologically equivalent to motif VIII in other MTases (Figure 5) and contains a phenylalanine residue (Phe196 M.*Taq*I) that has been proposed to interact with the target adenine in DNA (*51*) and to play a role in catalysis via cation−π interactions (*50*). This loop forms a prominent arched feature over the catalytic domain comprising residues immediately following Val158 (Pro159−Val169). It hangs over the active site and will be referred to as the "adenine-binding loop". It contains Phe163 which corresponds to the conserved Phe or Tyr found in other MTases within the sequence F163PXPXVXS171 which appears to be a consensus sequence among the *erm* gene family (K. Stewart, private communication). A search of the SWISSPROT database against the sequence FXPXPXVXS found that, of the 50 top-ranked protein sequences containing this motif, over 90% of them were involved in the binding of adenine or adenine analogues. The identified proteins ranged from DNA-binding proteins to several viral coat proteins and included over 25 methyltransferases. Therefore, it is likely that this sequence functions in the recognition of adenine in the single-stranded portion of the target RNA. Presumably, the conformation of this region of the chain could be altered by the binding of the 23S rRNA or appropriate substrates or inhibitors.

Cofactor Binding. Although SAH was added to the crystallization solutions, no satisfactory electron density corresponding to this reaction product was found either during the refinement or in the final maps. However, the

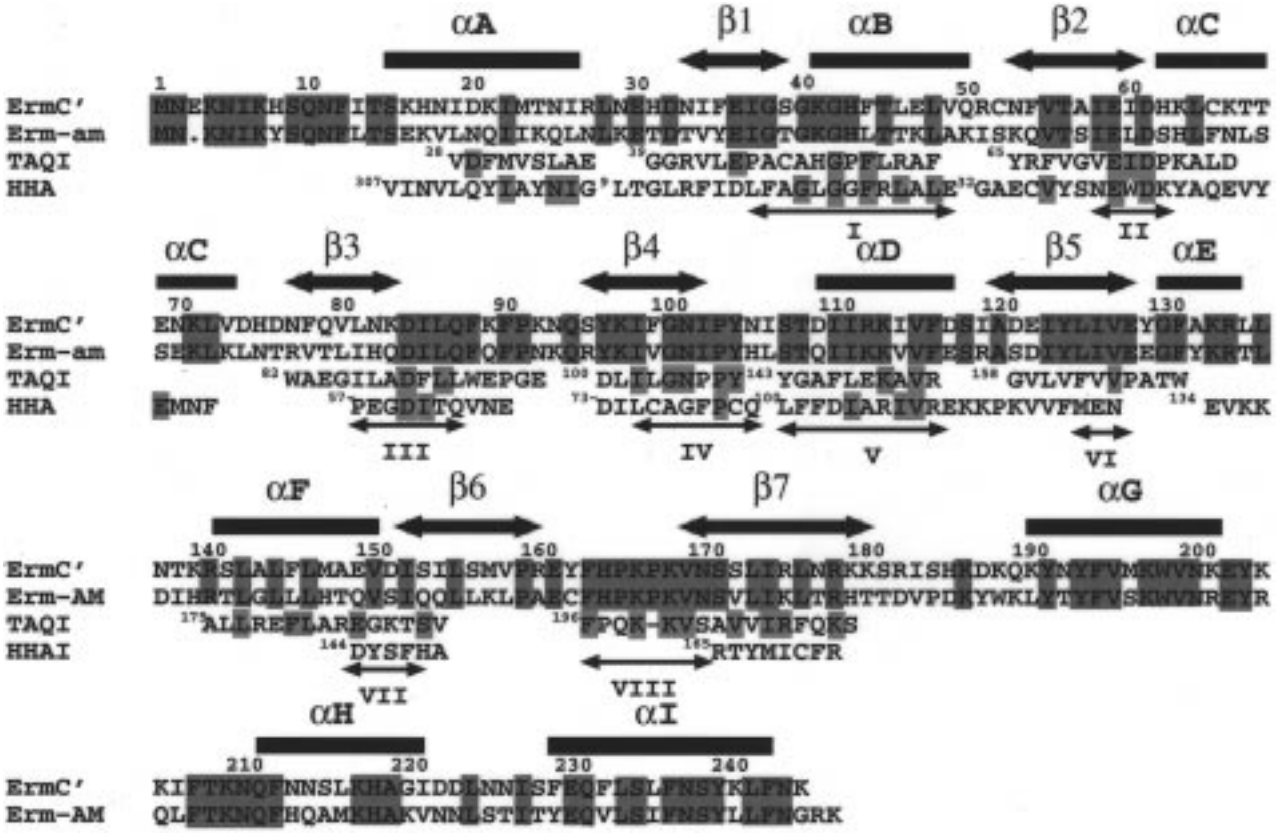

FIGURE 5: Structure-based sequence alignment of ErmC′, ErmAM, M.*Taq*I, and M.*Hha*I with the secondary structure elements of ErmC′ indicated by rectangles for α-helices and double-headed arrows for *β*-strands. Noted are the position and approximate extent of the nine conserved motifs present in the currently analyzed nucleic acid MTases. Regions of extensive sequence identity between ErmC′ and ErmAM are green; regions of identity between M.*Taq*I and M.*Hha*I are magenta. The superscript number to the left of the start of each M.*Taq*I or M.*Hha*I sequence indicates the starting residue number.

approximate location and orientation of the adenosine moiety of the cofactor can be modeled with confidence due to the conservation of the SAM-binding motifs in the structure of ErmC′ (vide supra) and the structural superposition of the cofactor binding domains of ErmC′ and M.*Taq*I (rms deviation of 1.9 Å for 115 Cα pairs, Table 3) (*16*). The adenine ring is most likely contained within a hydrophobic cavity lined by Ile37 (Pro45 in M.*Taq*I, motif I) at the floor and with Ile60 (Ile72, motif II), Ile85-Leu86 (Phe90-Ile91, motif III), and Ile110 (Phe146, motif V) forming the walls (Figure 6A). The side chain of Glu59 (Glu71, motif II) would provide the dominant interactions for the ribose hydroxyls, and Asp84 (Asp89, motif III) would interact with the exocyclic nitrogen of the adenine group of the cofactor. This suggested mode of cofactor binding is consistent with the amide proton shifts observed in the NMR structure of ErmAM upon the addition of SAH (*24*).

rRNA Recognition. Figure 6B depicts views of the electrostatic potential on the surface of the ErmC′ structure. Of note is an extensive concave interdomain surface with a large area of positive electrostatic potential. This region is approximately 20 × 25 Å^2 and could readily bind a helical section of rRNA such as helix B (Figure 1), which is distal to the target adenine located at the peptidyl transferase loop of domain V of the 23S rRNA (Figures 1 and 6B). Exposed, but deep within this surface of ErmC′, are hydrophobic residues such as Phe116, Phe145, and Met196, which could interact with several of the RNA bases. At the top of this positively charged surface, there is a ridge of basic residues (Lys133, Arg134, Lys139, Arg140, and Lys201, Figure 6B) which extends from Tyr104 within the NIPY active site sequence. All these basic residues are solvent-exposed and form a line of positive charges that extends from Arg160, within the active site loop, to Lys201 at the C terminus of αG in the RNA-binding domain (Figures 3 and 6B). The single-stranded rRNA region containing the target adenine of domain V (A2085, Figure 1) would be able to interact with some of these charged residues. A similar mode of rRNA binding is also suggested by the structure of ErmAM (*24*). Cocrystallization of the ErmC′ in the presence of SAH and selected rRNA fragments from domain V will be necessary to establish the binding mode between ErmC′ and rRNA.

ErmC′ is another example of the modular architecture of MTases whereby the catalytic domain combines with various substrate-recognition modules to methylate a large variety of substrates in the cellular milieu. In view of the universal nature of SAM as a methyl donor, further structural studies of ErmC′ in complex with cofactor analogues, with rRNA fragments, and with specifically designed inhibitors will be necessary to design potent drugs which will aid in overcoming bacterial resistance to the MLS group of antibiotics.

ACKNOWLEDGMENT

We thank Drs. J. Greer and D. Norbeck of Advanced Technology for support and Drs. L. Katz, P. Zhong, and K. Stewart for discussions. We also thank Drs. S. Ginell, R. Alkire (NSLS, line X8C), and J. Pflugrath for help in data

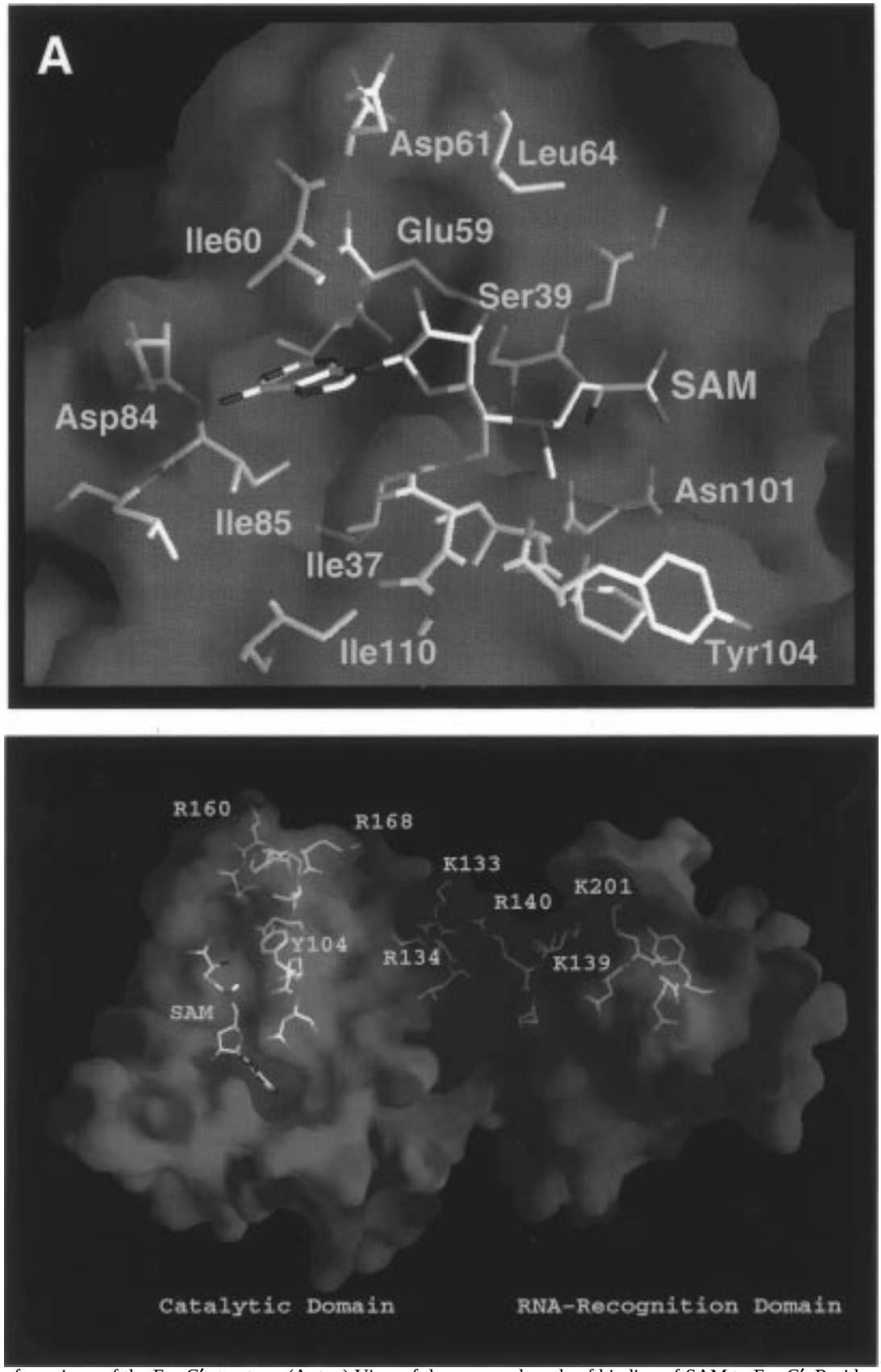

FIGURE 6: Surface views of the ErmC′ structure. (A, top) View of the proposed mode of binding of SAM to ErmC′. Residues making van der Waals contact with SAM (<3.7 Å) have been labeled. The side chain of Glu59 is not well defined in the refined structure and is positioned pointing away from the ribose hydroxyls. (B, bottom) View of the electrostatic potential on the surface of the structure of ErmC′ with two areas highlighted. The negatively charged region on the amino-terminal domain (red, potential $< -10k_bT$), showing the position and orientation of the SAM model based on binding to other MTases. An extensive area of positive charge (blue, potential $> 10k_bT$) in the concave interdomain surface proposed to be the surface of interaction for the helical regions of RNA. Some of the residues proposed to be involved in RNA binding are shown. The chemical structure of the modeled reaction product is represented as a ball-and-stick model. This figure was prepared with the program GRASP (*55*).

collection and reduction and Dr. E. Westbrook for synchrotron access. The assistance of Drs. E. de La Fortelle, J. Irwin, and G. Bricogne in using the SHARP phasing package is greatly appreciated. We acknowledge the efforts of Dr. E. Matayoshi and J. Huffaker in performing light scattering measurements with ErmC′ samples. We also thank Dr. A. Brünger and his research group for advice during the course of the refinement. The constructive criticism of the anonymous referees is appreciated.

REFERENCES

1. Weisblum, B. (1995) *Antimicrob. Agents Chemother. 39*, 577−585.
2. Westh, H., Hougaard, D. M., Vuust, J., and Rosdahl, V. T. (1995) *Antimicrob. Agents Chemother. 39*, 369−373.
3. Shivakumar, A. G., and Dubnau, D. (1981) *Nucleic Acids Res. 9*, 2549−2562.
4. Iordanescu, S. (1976) *Arch. Roum. Pathol. Exp. Microbiol. 35*, 111−118.
5. Monod, M., Denoya, C., and Dubnau, D. (1986) *J. Bacteriol. 167*, 138−147.
6. Su, S. L., and Dubnau, D. (1990) *Biochemistry 29*, 6033−6042.
7. Zhong, P., Pratt, S. D., Edalji, R. P., Walter, K. A., Holzman, T. F., Shivakumar, A. G., and Katz, L. (1995) *J. Bacteriol. 177*, 4327−4332.
8. Denoya, C. D., and Dubnau, D. (1987) *J. Bacteriol. 169*, 3857−3860.
9. Denoya, C. D., and Dubnau, D. (1989) *J. Biol. Chem. 264*, 2615−2624.
10. Skinner, R., Cundliffe, E., and Schmidt, F. J. (1983) *J. Biol. Chem. 258*, 12702−12706.
11. Cheng, X., Kumar, S., Posfai, J., Pflugrath, J. W., and Roberts, R. (1993) *Cell 74*, 299−307.
12. Klimasauskas, S. K., Roberts, R. J., and Cheng, X. (1994) *Cell 76*, 357−369.
13. O'Gara, M., Klimasauskas, S., Roberts, R. J., and Cheng, X. (1996) *J. Mol. Biol. 261*, 634−645.
14. O'Gara, M., Roberts, R. J., and Cheng, X. (1996) *J. Mol. Biol. 263*, 597−606.
15. Reinisch, K. M., Chen, L., Verdine, G. L., and Lipscomb, W. N. (1995) *Cell 82*, 143−153.
16. Schluckebier, G., Kozak, M., Bleimling, N., Weinhold, E., and Saenger, W. (1997) *J. Mol. Biol. 265*, 56−67.
17. Gong, W., O'Gara, M., Blumenthal, M., and Cheng, X. (1997) *Nucleic Acids Res. 25*, 2702−2715.
18. Wilson, G. G., and Murray, N. E. (1991) *Annu. Rev. Genet. 25*, 585−627.
19. Hodel, A. E., Gershon, P. D., Shi, X., and Quiocho, F. A. (1996) *Cell 85*, 247−256.
20. Vidgren, J., Svensson, L. A., and Liljas, A. (1994) *Nature 368*, 354−357.
21. Fu, Z., Hu, Y., Konishi, K., Takata, Y., Ogawa, H., Gomi, T., Fujioka, M., and Takusagawa, F. (1996) *Biochemistry 35*, 11985−11993.
22. Djordjevic, S., and Stock, A. M. (1997) *Structure 5*, 545−558.
23. Cheng, X. (1995) *Annu. Rev. Biophys. Biomol. Struct. 24*, 293−318.
24. Yu, L., Petros, A. M., Schnuchel, A., Zhong, P., Severin, J. M., Walter, K., Holzman, T. F., and Fesik, S. W. (1997) *Nat. Struct. Biol. 4*, 483−489.
25. Bussiere, D. E., Bastia, D., and White, S. W. (1995) *Cell 80*, 651−660.
26. Hendrickson, W. A. (1991) *Science 254*, 51−58.
27. Strauss, M. G., Westbrook, E. M., Naday, I., Coleman, T. A., Westbrook, M. L., Travis, D. J., Sweet, R. M., Pflugrath, J. W., and Stanton, M. (1990) *Nucl. Instrum. Methods Phys. Res., Sect. A 297*, 275−295.
28. Messerchmidt, A., and Pflugrath, J. W. (1987) *J. Appl. Crystallogr. 20*, 306−315.
29. Kabsch, W. (1988) *J. Appl. Crystallogr. 21*, 916−924.
30. Tate, M. W., Eikenberry, E. F., Barna, M. E., Wall, J. L., Lowrance, J. L., and Grunner, S. M. (1995) *J. Appl. Crystallogr. 28*, 196−205.
31. Gewirth, D. (1996) *The HKL Manual*, 5th ed., Yale University Press, New Haven, CT.
32. CCP4 (1994) *Acta Crystallogr. D50*, 760−763.
33. Furey, W., and Swaminathan, S. (1998) *Methods Enzymol. 277* (in press).
34. La Fortelle, E. d., and Bricogne, G. (1997) *Methods Enzymol. 276*, 472−494.
35. Abrahams, J. P., and Leslie, A. G. W. (1996) *Acta Crystallogr. D52*, 30−42.
36. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallogr. A47*, 110−119.
37. Biosym, Inc. (1994) *Quanta Manual*.
38. Brünger, A. T., Kuriyan, J., and Karplus, M. (1987) *Science 235*, 458−460.
39. Brünger, A. T. (1992) *Nature 355*, 472−474.
40. Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Brice, M. D., Rodgers, J. R., Kennard, O., and Shimanouchi, T. (1977) *J. Mol. Biol. 112*, 532−543.
41. Alexandrov, N. N. (1996) *Protein Eng. 9*, 727−732.
42. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr. 26*, 283−291.
43. Burd, C., and Dreyfuss, G. (1994) *Science 265*, 615−620.
44. Nagai, K. (1996) *Curr. Opin. Struct. Biol. 6*, 53−61.
45. Xing, Y., GuhaThakurta, D., and Draper, D. E. (1997) *Nat. Struct. Biol. 4*, 24−27.
46. Brünger, A. T. (1997) *Nat. Struct. Biol. 4* (Suppl.), 862−865.
47. Wilson, G. G. (1992) *Methods Enzymol. 216*, 259−279.
48. Posfai, J., Bhagwat, A. S., Posfai, G., and Roberts, R. J. (1989) *Nucleic Acids Res. 17*, 2421−2435.
49. Malone, T., Blumenthal, R. M., and Cheng, X. (1995) *J. Mol. Biol. 253*, 618−632.
50. Schluckebier, G., Labahn, J., Granzin, J., and Saenger, W. (1998) *Biol. Chem.* (in press).
51. Schluckebier, G., O'Gara, M., Saenger, W., and Cheng, X. (1995) *J. Mol. Biol. 247*, 16−20.
52. Noller, H. F. (1984) *Annu. Rev. Biochem. 53*, 119−162.
53. Carson, M. (1987) *J. Mol. Graphics 5*, 103−106.
54. Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946−950.
55. Nicholls, A. J., Sharp, K. A., and Honing, B. (1991) *Proteins: Struct., Funct., Genet. 11*, 281−296.

BI973113C