Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of the methyltransferase domain from the Modoc virus, a flavivirus with no known vector

Anna M. Jansson,[a] Emma Jakobsson,[a]‡ Patrik Johansson,[a]§ Violaine Lantez,[b] Bruno Coutard,[b] Xavier de Lamballerie,[c] Torsten Unge[a] and T. Alwyn Jones[a]*

[a]Department of Cell and Molecular Biology, Uppsala University, Biomedical Center, Box 596, SE 751 24 Uppsala, Sweden, [b]Laboratoire Architecture et Fonction des Macromolécules Biologiques, UMR 6098, AFMB–CNRS–ESIL, Case 925, 163 Avenue de Luminy, 13288 Marseille, France, and [c]Unité des Virus Emergents, Faculté de Médecine, 27 Boulevard Jean Moulin, 13005 Marseille, France

‡ Current address: CIC bioGUNE, Bizkaia Technology Park, Edificio 801A, 48160 Derio, Spain.
§ Current address: Cell Protein and Structural Sciences, AstraZeneca Research and Development, SE 431 83 Mölndal, Sweden.

Correspondence e-mail: alwyn@xray.bmc.uu.se

Received 4 February 2009
Accepted 7 May 2009

PDB References: Modoc virus methyltransferase domain, 2wa1, r2wa1sf; SAM complex, 2wa2, r2wa2sf.

The Modoc virus (MODV) is a flavivirus with no known vector (NKV). Evolutionary studies have shown that the viruses in the MODV group have evolved in association with mammals (bats, rodents) without transmission by an arthropod vector. MODV methyltransferase is the first enzyme from this evolutionary branch to be structurally characterized. The high-resolution structure of the methyltransferase domain of the MODV NS5 protein ($MTase_{MODV}$) was determined. The protein structure was solved in the apo form and in complex with its cofactor *S*-adenosyl-L-methionine (SAM). Although it belongs to a separate evolutionary branch, $MTase_{MODV}$ shares structural characteristics with flaviviral MTases from the other branches. Its capping machinery is a relatively new target in flaviviral drug development and the observed structural conservation between the three flaviviral branches indicates that it may be possible to identify a drug that targets a range of flaviviruses. The structural conservation also supports the choice of MODV as a possible model for flavivirus studies.

1. Introduction

The genus *Flavivirus*, belonging to the family *Flaviviridae*, comprises nearly 80 different viruses and a large number of arboviruses that are pathogenic to humans, such as Japanese encephalitis virus (JEV), tick-borne encephalitis virus (TBEV), West Nile virus (WNV) and Murray Valley encephalitis virus (MVEV), which are responsible for neurological diseases, and dengue virus (DENV) and yellow fever virus (YFV), which cause haemorrhagic diseases. Flaviviruses are enveloped and carry their genetic information in a positive-strand RNA genome of ~11 kb. The genus *Flavivirus* is divided into three subgroups depending on the vector with which the virus is associated: tick-borne, mosquito-borne and a third subgroup that contains viruses with no known vector (NKV). NKVs constitute an original evolutionary branch associated with the infection of bats or rodents, presumably in the absence of any vector transmission. However, it is important to notice that besides the 'classical' group of NKVs, other flaviviruses with no identified vector exist embedded within the large evolutionary group of mosquito-borne viruses (*e.g.* Sokuluk, Tokose and Entebbe bat virus) or outside the evolutionary lineage of canonical flaviviruses (*e.g.* Tamana bat virus). The murine flavivirus Modoc (MODV) was first isolated in 1958 from white-footed deer mice in Modoc County, California, USA (Johnson, 1967). This virus was subsequently classified as a flavivirus (Calisher *et al.*, 1989; Varelas-Wesley & Calisher, 1982) and phylogenetic analysis placed MODV in the 'classical' NKV group (Kuno *et al.*, 1998). MODV is neuroinvasive and has a similar pathology to flaviviral encephalitis in humans

(Leyssen, Croes *et al.*, 2003; Leyssen, Paeshuyse *et al.*, 2003; Leyssen *et al.*, 2001). This makes MODV a potential model virus for studying flaviviral infections.

The co-transcriptional modification of the 5′-end of viral and eukaryotic mRNA is known as mRNA capping. Capping of mRNA is a multistep reaction in which the reactions are catalyzed by the same or different enzymes depending on the organism (Furuichi & Shatkin, 2000). The general features of mRNA capping involve an initial step in which the 5′-phosphate group of the mRNA is removed by RNA triphosphatase, an activity carried out by the NS3 protein, as exemplified for WNV (Bartelma & Padmanabhan, 2002). This is followed by the addition of a GMP to the remaining 5′-diphosphate by RNA guanylyltransferase. At present, it is unclear which protein carries this activity, although preliminary results may suggest that NS5A is involved in this step (Bollati *et al.*, 2009). In a third step, N7 of the guanine base is methylated by RNA guanine-N7-methyltransferase. The final step is a second methylation, catalyzed by a nucleoside-2′-*O*-methyltransferase, on the first ribose in the 3′-direction of the three phosphates linking the guanosine. The flaviviral NS5MTase has been assigned both N7-methylation and 2′-*O*-methylation activity (Egloff *et al.*, 2002; Ray *et al.*, 2006). Both methylations employ *S*-adenosyl-L-methionine (SAM) as the methyl donor, producing *S*-adenosyl-L-homocysteine (SAH). Capping of the mRNA is essential for the stability of the mRNA, as well as for ribosome binding.

The replicative machinery is a relatively new target for drug design against emerging viruses. In flaviviruses, lack of N7 methylation renders the virus nonreplicative (Dong, Ren *et al.*, 2008) and WNV deficient in 2′-*O*-methylation activity resulted in nonvirulent infections in mice (Dong, Zhang *et al.*, 2008). Thus, the flaviviral MTase is an attractive drug target. Here, we report the structure of the Modoc virus methyltransferase domain ($MTase_{MODV}$). The structure is presented both in the light of evolution and in the context of being a suitable target for antiviral drugs against flaviviruses.

2. Materials and methods

2.1. Cloning, protein expression and purification of $MTase_{MODV}$

Modoc virus strain 3321 (from the personal collection of XdL) was propagated onto BHK21 cells under standard conditions (310 K, MEM medium, 5% CO_2). At gross cytopathic effect, the supernatant was collected, clarified by centrifugation and 200 µl was submitted to nucleic acid extraction using the Virus minikit and the EZ1 Biorobot (both from Qiagen). Reverse transcription was realised using random hexaprimers (Taqman RT reagent, Applied Biosystems). An 879-nucleotide segment including the complete methyltransferase domain of the NS5 gene from Modoc virus (293 amino-acid residues) was initially amplified using primers 5′-G GGG ACA AGT TTG TAC AAA AAA GCA GGC TTC GAA GGA GAT AGA ACC ATG CAT CAT CAT CAT CAT CAT GGC ATT TGT TCG AGT GCC CCC ACA CTG-3′ (forward) and 5′-GGG GAC CAC TTT GTA CAA GAA AGC TGG GTA TTA GGT GCC TGA TTT CTC AGC CTT GAT-3′ (reverse) and the high-fidelity Triplemaster PCR system kit (Eppendorf) and cloned into a pDEST14 vector (Invitrogen). A truncated methyltransferase gene fragment coding for only the first 268 residues was isolated by PCR amplification using the primers 5′-ATG GCT CAT CAT CAT CAT CAT GGC ATT TGT TCG AGT GC-3′ (forward) and 5′-CTA AAG ATT GGA TCT TGT TCC TG-3′ (reverse) and *Pfu* Ultra DNA polymerase (Stratagene). This added an N-terminal pentahistidine tag to the construct. DNA sequencing revealed four mutations in the genomic template: S77T, G87S, R134T and R202A. The PCR product was purified on a 1% agarose gel and ligated into a pCR T7/CT-TOPO vector (Invitrogen). The plasmids were transformed into *Escherichia coli* Top10F′ cells (Invitrogen) and six clones were selected for plasmid preparation with the QIAprep Spin Miniprep Kit (Qiagen). Correct directionality of the inserted PCR fragment was determined by analytical PCR with the same forward primer as for amplification and the 5′-TAT GCT AGT TAT TGC TCA G-3′ reverse primer annealed to the vector. The construct was further verified by nucleotide sequencing (Uppsala Genome Center, Rudbeck Laboratory, Uppsala, Sweden). The plasmid from a correct clone was transformed into *E. coli* Rosetta cells for large-scale expression. The cells were cultured at 310 K in LB medium with 50 µg ml^{-1} ampicillin to a density corresponding to an OD_{600} of 0.6–1. Protein production was induced with 0.1 m*M* isopropyl β-D-1-thiogalactopyranoside (Calbiochem) and continued for 3 h at 298 K. The cells were harvested by centrifugation, washed with 1× SSP buffer (150 m*M* NaCl, 250 m*M* NaH_2PO_4 pH 7.4) and stored at 253 K.

For protein purification, cells were suspended in a buffer (50 m*M* Na_2HPO_4/NaH_2PO_4 pH 8.0, 50 m*M* Na_2SO_4, 200 m*M* NaCl, 10 m*M* imidazole, 10% glycerol, 0.5% Triton X-100) supplemented with RNase, DNase, lysozyme and β-mercaptoethanol and lysed in a Constant Cell Disruptor (Constant Systems Ltd) operated at 200 MPa. Cleared cell lysate was incubated with Ni–NTA agarose slurry (Qiagen) for 1 h at 281 K. The slurry was washed with the same buffer as above without Triton X-100 and with 20 m*M* imidazole and elution was performed in the same buffer with 200 m*M* imidazole. The protein was further purified by size-exclusion chromatography (HiLoad 16/60 Superdex-200; GE Healthcare) in a buffer containing 50 m*M* HEPES–NaOH pH 7.0 and 200 m*M* $(NH_4)_2SO_4$. The protein was >95% pure as determined visually by SDS–PAGE analysis (PhastSystem, GE Healthcare). The protein was concentrated to 10 mg ml^{-1} in 50 m*M* HEPES–NaOH pH 7.0, 50 m*M* $(NH_4)_2SO_4$ buffer in a Vivaspin concentrator with a molecular-weight cutoff of 10 kDa (Vivascience). The typical protein yield was 8 mg per litre of culture.

2.2. Crystallization

$MTase_{MODV}$ was crystallized both in the apo form and in complex with its cofactor *S*-adenosyl-L-methionine (SAM).

An initial crystal hit was obtained in condition B12 (12% PEG 4000, 100 m*M* unbuffered sodium acetate, 100 m*M* HEPES–NaOH pH 7.5) of the JBScreen Classic HTS I (Jena Bioscience) using the vapour-diffusion technique. The sitting drops consisted of 0.5 µl reservoir solution added to 0.5 µl protein solution [10 mg ml^{-1} in 50 m*M* HEPES–NaOH pH 7.0 and 50 m*M* $(NH_4)_2SO_4$] at 293 K. The reservoir composition was optimized with the vapour-diffusion technique and hanging drops. Large crystals were obtained overnight in drops containing 1 µl protein solution mixed with 1 µl reservoir solution containing 8% PEG 4000, 100 m*M* unbuffered sodium acetate and 100 m*M* HEPES–NaOH pH 7.5. These crystallization conditions were further optimized with the batch method under constant conditions. Hanging drops of 1 µl protein solution [10 mg ml^{-1} in 50 m*M* HEPES pH 7.0 and 50 m*M* $(NH_4)_2SO_4$] and 1 µl reservoir solution were set up against a batch reservoir [composed of 4% PEG 4000, 50 m*M* unbuffered sodium acetate and 75 m*M* HEPES–NaOH, 25 m*M* $(NH_4)_2SO_4$]. The drops were immediately streak-seeded with the initial apo crystals. For cocrystallization experiments with the SAM cofactor, the protein solution was mixed with a 15-fold molar excess of the ligand and incubated for 30 min on ice. The cocrystallization was set up using the same conditions as for the native crystals. Native crystals were cryoprotected by a quick dip in reservoir solution supplemented with 30% glycerol and flash-cooled in liquid nitrogen. The crystals of the protein–ligand complex were cryoprotected similarly but with reservoir solution supplemented with 30% PEG 400 and 5 m*M* SAM.

Table 1
Data-collection and refinement statistics.

Values in parentheses are for the outer shell.

	Apo	SAM complex
Data collection		
Resolution (Å)	2.0 (2.11–2.0)	1.8 (1.9–1.8)
Space group	*I*4	$P2_12_12_1$
Unit-cell parameters (Å, °)	$a = b = 180.7$, $c = 52.4$, $\alpha = \beta = \gamma = 90$	$a = 52.5$, $b = 62.2$, $c = 160.0$, $\alpha = \beta = \gamma = 90$
No. of unique reflections	54461 (6673)	45634 (4527)
Average multiplicity	5.2 (3.3)	6.7 (4.6)
Completeness (%)	94.9 (80.6)	92.0 (63.7)
$R_{\rm merge}$†	0.090 (0.48)	0.070 (0.315)
$\langle I/\sigma(I)\rangle$‡	14.1 (3.0)	21.5 (4.1)
Refinement		
Resolution range (Å)	45–2.0	40–1.8
No. of reflections used	51678	43257
R value§ (%)	19.4	18.6
$R_{\rm free}$¶ (%)	22.2	23.5
No. of atoms	4108	4152
No. of solvent molecules	296	305
Mean *B* factors (Å^2)		
Protein *A* chain	31.6	19.7
Protein *B* chain	23.3	15.54
Ligand	—	17.35
Solvent H_2O	31.6	24.2
Solvent SO_4	39.2	43.4
Ramachandran outliers†† (%)	0.7	0.7
R.m.s.d. from ideal bond length‡‡ (Å)	0.016	0.015
R.m.s.d. from ideal bond angle‡‡ (°)	1.4	1.4

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, as obtained by *SCALA* (Evans, 1993). ‡ $\langle I/\sigma(I)\rangle$ indicates the average of the intensity divided by its standard deviation, as obtained by *SCALA* (Evans, 1993). § $R = 100 \times \sum_{hkl} |F_{\rm obs} - F_{\rm calc}| / \sum_{hkl} |F_{\rm obs}|$. ¶ $R_{\rm free}$ is the same as *R*, but for a subset (5%) of the reflections that were excluded from the refinement process (Brünger, 1992). †† Calculated using a strict-boundary Ramachandran plot (Kleywegt & Jones, 1996). ‡‡ Calculated using small-molecule-based parameters (Engh & Huber, 1991).

2.3. Data collection, structure determination and refinement

Data were collected on beamline ID14-2 at the European Synchrotron Research Facility (ESRF) in Grenoble using a MAR CCD detector (MAR USA Inc). The data sets were indexed and integrated with *MOSFLM* (Leslie, 1999) and the intensities were merged with *SCALA* (Evans, 1993) as implemented in the *CCP*4 program suite (Collaborative Computational Project, Number 4, 1994). The crystals of the native protein belonged to space group *I*4, with unit-cell parameters $a = b = 180.7$, $c = 52.4$ Å. Crystals of the protein complexed with the SAM cofactor belonged to space group $P2_12_12_1$, with unit-cell parameters $a = 52.5$, $b = 62.2$, $c = 160.0$ Å. The structure with the SAM cofactor was solved by molecular replacement with *Phaser* (McCoy *et al.*, 2007) using the $\rm MTase_{DENV}$ structure as a search model (PDB code 1l9k; Egloff *et al.*, 2002). The first model was built using *ARP*/*wARP* (Lamzin *et al.*, 2001). Further rebuilding and refinement was performed using *O* (Jones *et al.*, 1991) and *REFMAC*5 (Murshudov *et al.*, 1997; Pannu *et al.*, 1998). The structure of the apo protein was solved in the same manner, now using the SAM-complex structure as the search model. 296 and 305 solvent molecules were added to the apo and SAM-complex models, respectively. Data-collection and refinement statistics are presented in Table 1.

2.4. Alignments

Structural alignments of the flavivirus MTases were made with the lsq commands as implemented in *O* (Kleywegt & Jones, 1997). Structural alignments between the $\rm MTase_{MODV}$ and the more distantly related MTases from the *E. coli* FtsJ heat-shock protein, M.*Taq*I, vaccinia virus VP39 and catechol *O*-MTase were identified with the *DALI* server (Holm *et al.*, 2008; Holm & Sander, 1993) and optimized using *SUPERPOSE* (Krissinel & Henrick, 2004), part of the *CCP*4 program suite. The dimer interface was calculated with *AREAIMOL* (Lee & Richards, 1971), also part of the *CCP*4 program suite.

Secondary-structure boundaries were determined by visual inspection. Figures were prepared using *O* and rendered with *MOLRAY* (Harris & Jones, 2001).

3. Results

3.1. Structure determination

The NS5 gene product in flaviviruses has been predicted to contain two domains: an N-terminal methyltransferase domain and a C-terminal polymerase domain (Egloff *et al.*, 2002). In the Modoc virus genome, the MTase region corresponds to residues 1–293 of the NS5 gene product. Initially, a full-length construct of the MTase domain was expressed and purified but did not yield any crystals. A sequence alignment revealed that several of the MTases with known structures had shorter C-termini than $MTase_{MODV}$. In addition, the only structure of a flavivirus MTase at the time was only modelled to residue 267. Therefore, we designed a new construct truncated at the C-terminus after Leu268. Hence, this construct was 25 amino acids shorter than the predicted domain for $MTase_{MODV}$.

This truncated form of $MTase_{MODV}$ crystallized in space group $I4$, with a solvent content of 67%, whereas $MTase_{MODV}$ in complex with the cofactor SAM crystallized in space group $P2_12_12_1$, with a solvent content of 45%. The structures of the native protein and of the cofactor complex were solved to resolutions of 2.0 and 1.8 Å, respectively. The $MTase_{MODV}$ model had a final R and R_{free} of 19.4% and 22.2%, respectively. The corresponding values for the $MTase_{MODV}$ SAM model were 18.6% and 23.5%, respectively. Data-collection and refinement statistics are presented in Table 1.

Both the native protein and the cofactor complex crystallized with two molecules in the asymmetric unit. The two monomers are related by a nearly perfect twofold noncrystallographic symmetry operator. A superposition of 219 Cα atoms in the *A* and *B* chains of the native protein resulted in an r.m.s.d. of 0.6 Å. A similar comparison for the SAM complex resulted in an r.m.s.d. of 0.4 Å over the same range. When superimposing the *A* chain and *B* chain of the native protein with the *A* and *B* chain of the cofactor complex, the r.m.s.d. was 0.8 Å for 480 Cα atoms. The interface between the chains has a surface area of about 800 Å^2, which is in the lower range for homodimers (Jones & Thornton, 1996). The protein elutes as a monomer in size-exclusion chromatography (data not shown). Hence, dimerization seems to be a crystallographic artefact. For the structures of both the native protein and the SAM complex, density was interpretable for almost the entire *B* chain. In the *A* chain there was no density for the last 19 residues at the C-terminus nor for residues 35–56 (part of the B1 β-strand, the A3 α-helix and the following loop). Since the *A* chain is less complete than the *B* chain, all analysis and pictures are based on the *B* chain.

3.2. Overview of the structure

The $MTase_{MODV}$ structure is characterized by a mixed α/β-fold composed of a central seven-stranded β-sheet surrounded by mainly helical formations (Fig. 1). The twisted β-sheet is flanked by two helices on each side, also adjacent in sequence. N- and C-terminal extensions protrude from this central core. The N-terminal structure forms a helix–turn–helix motif followed by a strand and another helix, whereas the C-terminal portion forms an additional helix and a strand. These extensions are aligned 'head-to-tail', which allows the two strands to form a parallel β-sheet perpendicular to the main sheet. The seven-stranded β-sheet together with the flanking helices αA and αX on one side and αD and αE on the other overlap with the consensus core domain that characterizes most SAM-dependent methyltransferases (Fauman *et al.*, 1999). $MTase_{MODV}$, however, lacks the two helices of the core domain that are normally inserted between strands $\beta2/\beta3$ and $\beta3/\beta4$. This seems to be a typical characteristic, as observed in

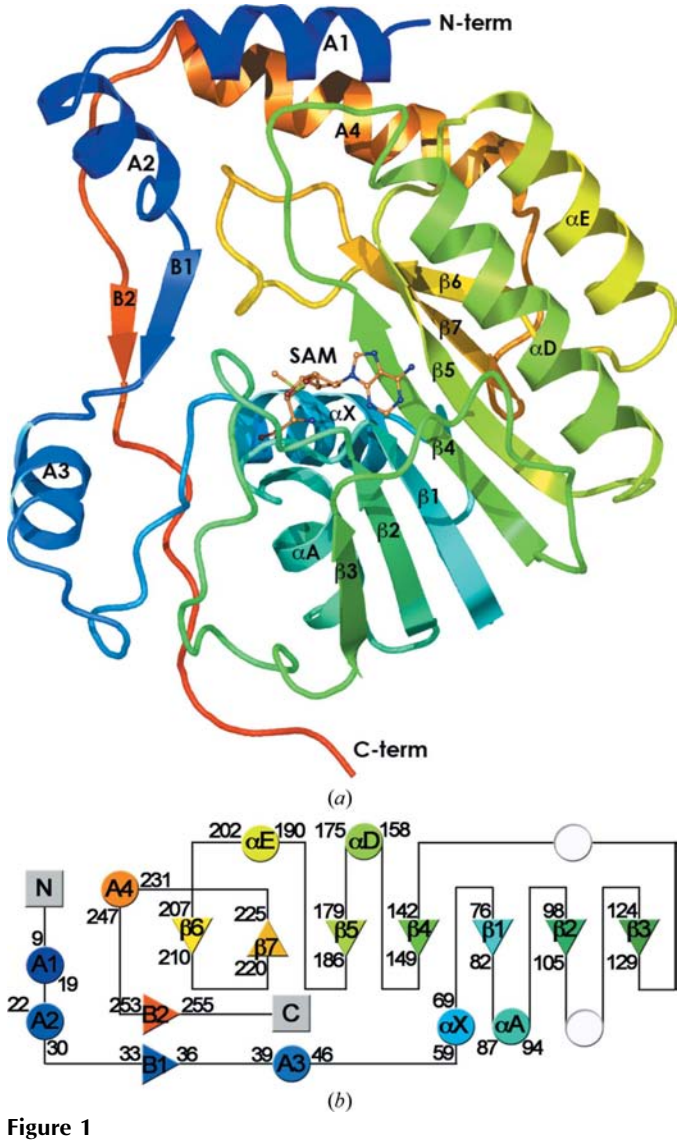

Figure 1
(*a*) The crystal structure of $MTase_{MODV}$ in complex with the cofactor SAM. Cartoon representation of the *B* chain in rainbow colouring from blue to red from the N-terminus to the C-terminus. The naming of the secondary-structure elements corresponds to that of Egloff *et al.* (2002), where Greek letters are used within the core fold common to all SAM-dependent MTases and roman letters are used for the N-terminal and C-terminal extensions. (*b*) Topology diagram of $MTase_{MODV}$. α-Helices are represented by circles and β-strands are represented by triangles. The colours are coherent with the above three-dimensional representation of the structure. The empty circles between $\beta2$ and $\beta3$ and between $\beta3$ and $\beta4$ are the two helices that deviate from the consensus fold for SAM-dependent methyltransferases.

the previously determined structures of MTases from other flaviviruses. These also display a SAM-dependent MTase core truncated by two helices (Assenberg *et al.*, 2007; Egloff *et al.*, 2002; Mastrangelo *et al.*, 2007; Zhou *et al.*, 2007).

3.3. Binding of the cofactor *S*-adenosyl-L-methionine

Inspection of the electron-density maps of the SAM cocrystallization data revealed strong additional electron density in the vicinity of the $MTase_{MODV}$ active site. The density found in the σ_A-weighted $2|F_o| - |F_c|$ map (Read, 1986) allowed unambiguous modelling of all SAM non-H atoms (Fig. 2*a*). The cofactor is bound in a cavity lined by the $\beta4$ strand and the loop connecting $\beta2$ and $\beta3$ on one side and αX and the loop connecting $\beta4$ and $\beta5$ on the other, with the ribose positioned above $\beta1$. Hydrophobic interactions with the adenine together with hydrogen bonds and electrostatic interactions between the protein and the SAM molecule stabilize the cofactor in the binding site. Specifically, the adenosine ring is accommodated in a hydrophobic pocket made up by the side chains of Thr104, Leu106, Val133 and Ile148, and is further positioned by hydrogen bonds between N1 and the main-chain N atom of Val133 and between N3 and the main-chain N atom of Leu106, as well as a hydrogen bond between N6 of the adenine and the carboxylate group of Asp132 (Fig. 2*b*). Ribose binding is stabilized by hydrogen-bond interactions between O2′ and the main-chain N atom of Gly107 *via* a bridging water molecule and the imidazole ring of His111. The methionine tail of the SAM molecule is positioned by hydrogen bonds to both carboxylate O atoms O and OXT from Ser57 and Ser87 (mutated residue; Gly in wild type), together with a hydrogen bond between the N atom and the carboxylate of Asp147. A comparison of the active-site cavity before and after binding of the SAM cofactor reveals that Leu106 is clearly shifted towards the adenine of the SAM, hence tightening the hydrophobic pocket around it.

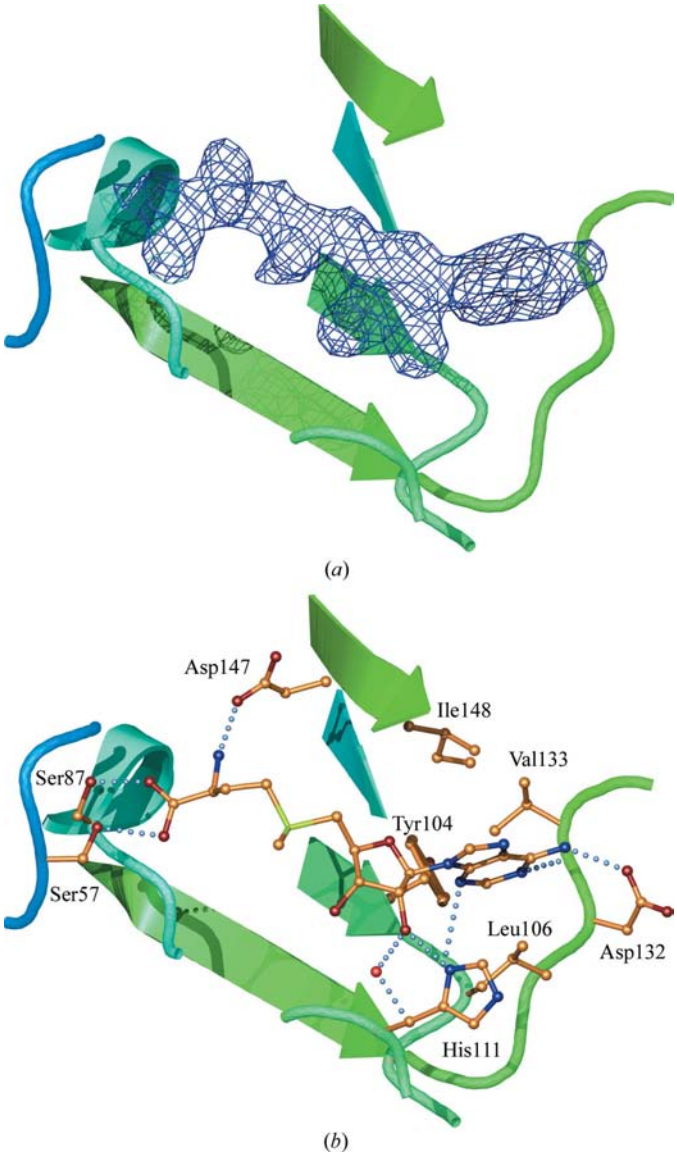

Figure 2
(*a*) Unbiased density found in the active site after initial refinement of a model without the SAM cofactor. The density was calculated as a $2|F_o| - |F_c|$ map using only phases from the protein. (*b*) SAM is stabilized in the active site by both hydrophobic and electrostatic interactions.

3.4. Comparison of overall structure between $MTase_{MODV}$ and related MTases

In the following comparisons of the structure of $MTase_{MODV}$ with related protein structures, we use the secondary-structure nomenclature described in Fig. 1.

Superimposing the Cα atoms of other flaviviral MTases onto the $MTase_{MODV}$ structure reveals a high level of conservation of the backbone (Table 2). A comparison of the Cα atoms of MTases from DENV, Meaban virus, WNV and MVEV gives an r.m.s.d. of around 1 Å for an alignment of ~220 Cα atoms. A closer look at the residues in the active site shows that many of the side chains involved in stabilization of the SAM cofactor are conserved. In all five viruses, the adenine is accommodated in a hydrophobic pocket and the key residues in the interaction are conserved, including Ser57, His111, Asp147 and Ile148. Other vicinal residues, such as Asp132 and Val133, are conserved in four of the structures but

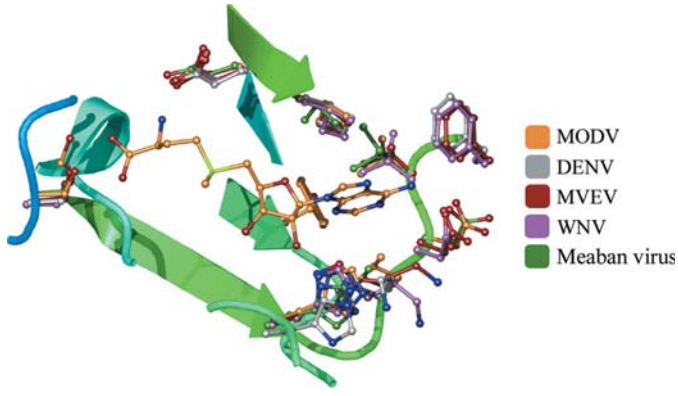

Figure 3
A superimposition of $MTase_{MODV}$ with MTases from DENV, MVEV, WNV (belonging to the mosquito-borne branch) and Meaban virus (from the tick-borne branch). Most stabilizing residues are conserved and the active sites from all three lineages align very well.

Table 2
Structure alignment.

Statistics are from *DALI* (Holm *et al.*, 2008; Holm & Sander, 1993). *Z* score, measure of the quality of the alignment. R.m.s.d., average deviation in distance between aligned Cα atoms; L.ALI, length of the alignment; %IDE, percentage sequence identity between the compared chains.

PDB code	Name	*Z* score	R.m.s.d.	L.ALI	%IDE
2oxt	Meaban virus methyltransferase	34.9	0.8	223	54
2px2, 2pxa, 2px5, 2px4, 2px8, 2pxc	Murray Valley virus methyltransferase	33.6†	1.1†	220†	55†
2p3l, 1r6a, 1l9k, 2p41, 2p40, 2p1d, 2p3q, 2p3o	Dengue virus methyltransferase	33.1†	1.0†	218†	54†
2oy0	West Nile virus methyltransferase	32.3	1.0	219	53
1eiz, 1ej0	FtsJ heat-shock protein, *E. coli*	16.2	2.0	153	26
2adm	M.*Taq*I	11.4	3.2	160	10
1av6	Vaccinia virus VP39	9.6	3.4	165	10
1vid	Catechol-*O*-MTase	9.1	2.9	143	13

† An average of the values for the relevant PDB entries.

are exchanged for residues with similar side chains in the dengue virus enzyme (Fig. 3).

In the more distantly related methyltransferases there is still some level of conservation at the core. A structural alignment of $MTase_{MODV}$, *E. coli* MTase FtsJ (Bügl *et al.*, 2000), the DNA MTase M.*Taq*I (Schluckebier *et al.*, 1997), vaccinia virus MTase VP39 (Hodel *et al.*, 1996) and catechol *O*-MTase (Vidgren *et al.*, 1994) was made. Catechol *O*-MTase and *E. coli* MTase FtsJ share most of the typical core-domain features. Both harbour several more extended loops compared with $MTase_{MODV}$. The structure of catechol *O*-MTase displays a significantly longer loop than the equivalent stretch between β4 and αD in $MTase_{MODV}$. FtsJ has an extended loop structure between β2 and β3 and an additional α-helix in the corresponding coil structure between β3 and β4 of $MTase_{MODV}$. The FtsJ structure only consists of this compact core domain, whereas catechol *O*-MTase also has N- and C-terminal extensions. However, these are not located like those in $MTase_{MODV}$. VP39 from vaccinia virus has the elements of the typical core, but with some shifts in secondary-structure positioning. For example, the αX helix of the vaccinia virus MTase is at an angle compared with that of $MTase_{MODV}$ and the αA helix is shifted in parallel outwards from the central β-sheet. The loop between β2 and β3 is extended in $MTase_{MODV}$, forming a lid that assists in the positioning of the SAM cofactor. Furthermore, helix αE is replaced by a two-stranded β-sheet in the vaccinia virus structure: β6 and β9. The αE helix is the only part that the vaccinia virus MTase lacks in the consensus fold (Egloff *et al.*, 2002). The DNA MTase M.*Taq*I is a much larger protein and has two domains, whereas the other methyltransferases discussed here have only one. One of these domains contains the MTase core, which is largely conserved, although most of the secondary-structure elements are shifted and have different boundaries.

The cofactor-binding site is also mostly conserved in these structures. Catechol *O*-MTase has an active site that deviates from the others in that the adenine is stabilized solely by hydrogen bonds and is not accommodated in a hydrophobic pocket. The active-site Asp is conserved in the FtsJ, vaccinia virus VP39 and catechol *O*-MTase structures, but in the DNA MTase M.*Taq*I it is exchanged for an Asn. Some of these observed structural differences between these proteins can be explained by their different substrate specificities.

4. Discussion

Flaviviruses with no known arthropod vector are mainly associated with rodents and bats. The history of NKV infection in their natural hosts is poorly known but may be characterized by a prolonged asymptomatic infection. Accidental infection of humans with NKVs has been reported and has led to neurological disease (Billoir *et al.*, 2000). Studies have shown that the NKV group has evolved separately from the tick-borne and the mosquito-borne groups (Gould *et al.*, 2003). In particular, the chronic infection of mice and hamsters (in which the pathophysiology of infection of the central nervous system is evocative of that observed with other neurovirulent flaviviruses) constitutes an attractive model for the study of antivirals (Charlier *et al.*, 2004).

$MTase_{MODV}$ is the first MTase structure to be determined from the third lineage of flaviviruses, the NKV branch. Structural comparisons between $MTase_{MODV}$ and flaviviral MTases from the distinct mosquito-borne and tick-borne lineages show a high similarity. Cα alignments of $MTase_{MODV}$ and $MTase_{DENV}$, $MTase_{MVEV}$, $MTase_{Meaban\ virus}$ and $MTase_{WNV}$ have r.m.s.d.s of $\sim$1 Å for $\sim$220 Cα atoms (82% of the $MTase_{MODV}$ structure) and the residues important for SAM binding and accommodation of the RNA substrate are highly conserved. From the $MTase_{MODV}$ structure we can therefore see that the structures have remained closely related despite divergent evolution.

Initially, the flaviviral NS5 MTase was found to have nucleoside-2′-*O*-methylation activity (Egloff *et al.*, 2002). Recently, the MTases from MVEV and WNV ($MTase_{WNV}$) have been shown to have nucleoside-2′-*O*-methylation activity as well as N7-methylation activity (Dong, Ren *et al.*, 2008; Ray *et al.*, 2006; Assenberg *et al.*, 2007). These activities are also seen in the MTases of DENV and YFV, both from the mosquito-borne group, as well as the MTase from Powassan encephalitis virus from the tick-borne group (Dong, Ren *et al.*, 2008). On the basis of the structural similarities that we observe here, we expect $MTase_{MODV}$ to also exhibit these activities. Biochemical studies of $MTase_{WNV}$ demonstrate a unique methylation pattern not found in, for example, cellular methyltransferases (Dong *et al.*, 2007). These studies also show that N7-methylation by the flaviviral MTase is essential for the propagation of the virus, making the flaviviral MTase a very attractive drug target. The common methylation pattern found in flaviviral MTases also offers the possibility of a drug that can target a range of flaviviruses.

In flaviviral MTases, three specific binding sites have been identified: SAM/SAH binding, GTP binding and RNA binding

(Egloff *et al.*, 2002). Although the positions for methylation on the RNA are located on different nucleotides and experiments have shown that SAM is the methyl donor in both methylation reactions, only one SAM-binding site has been identified (Dong, Ren *et al.*, 2008). Cocrystal structures of $MTase_{DENV}$ and of $MTase_{MVEV}$ show that mRNA cap analogues can bind in the GTP-binding site (Assenberg *et al.*, 2007; Egloff *et al.*, 2007). This binding site is mainly important during 2′-*O*-methylation (Dong, Ren *et al.*, 2008), positioning the 2′-*O* of the first ribose in proximity of the leaving methyl group of the SAM cofactor. A possible RNA-binding site was first identified in $MTase_{DENV}$ and this site is conserved among flaviviral MTases (Dong, Ren *et al.*, 2008; Egloff *et al.*, 2002). Mutation studies in $MTase_{WNV}$ verified this positively charged surface to be the RNA-binding site. The studies also identified the residues critical for RNA binding and for N7- and 2′-*O*-methylation. These results further suggest that the position of RNA binding is shifted between the two methylation steps. Mutations that interfere with N7-methylation lead to lowered or abandoned viral replication (Dong, Ren *et al.*, 2008). A structural comparison between $MTase_{MODV}$ and $MTase_{WNV}$ identified these critical residues and they were found to be fully conserved in $MTase_{MODV}$.

Experiments show that the Modoc virus is neuroinvasive in immunodeficient mice and in immunocompetent hamsters, causing a lethal degeneration of neural tissue (Leyssen, Paeshuyse *et al.*, 2003; Leyssen *et al.*, 2001). The observed pathological effect is highly similar to that of flavivirus encephalitis in humans (Leyssen, Paeshuyse *et al.*, 2003; Leyssen *et al.*, 2001). These results also support the Modoc virus as a suitable model virus for the evaluation of chemoprophylactic or chemotheraputic strategies against flavivirus infections. The structural results presented in this paper support the choice of Modoc virus as a model virus for studying flavivirus infections, as well as the choice of the flavivirus methyltransferase as a possible drug target.

We thank Yngve Cerenius at MAX-lab, Lund, where the initial data sets were collected. We thank Agata Naworyta, Mikael Nilsson, Annette Roos and Adrian Suárez Covarrubias for assistance in data collection at the ESRF. Terese Bergfors is thanked for critical reading of the manuscript. This work was supported by the EU IP Project VIZIER (CT 2004–511960).

References

Assenberg, R., Ren, J., Verma, A., Walter, T. S., Alderton, D., Hurrelbrink, R. J., Fuller, S. D., Bressanelli, S., Owens, R. J., Stuart, D. I. & Grimes, J. M. (2007). *J. Gen. Virol.* **88**, 2228–2236.

Bartelma, G. & Padmanabhan, R. (2002). *Virology*, **299**, 122–132.

Billoir, F., de Chesse, R., Tolou, H., de Micco, P., Gould, E. A. & de Lamballerie, X. (2000). *J. Gen. Virol.* **81**, 781–790.

Bollati, M., Milani, M., Mastrangelo, E., Ricagno, S., Tedeschi, G., Nonnis, S., Decroly, E., Selisko, B., de Lamballerie, X., Coutard, B., Canard, B. & Bolognesi, M. (2009). *J. Mol. Biol.* **385**, 140–152.

Brünger, A. T. (1992). *Nature (London)*, **355**, 472–475.

Bügl, H., Fauman, E. B., Staker, B. L., Zheng, F., Kushner, S. R., Saper, M. A., Bardwell, J. C. A. & Jakob, U. (2000). *Mol. Cell*, **6**, 349–360.

Calisher, C. H., Karabatsos, N., Dalrymple, J. M., Shope, R. E., Porterfield, J. S., Westaway, E. G. & Brandt, W. E. (1989). *J. Gen. Virol.* **70**, 37–43.

Charlier, N., Leyssen, P., De Clercq, E. & Neyts, J. (2004). *Antivir. Res.* **63**, 67–77.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Dong, H., Ray, D., Ren, S., Zhang, B., Puig-Basagoiti, F., Takagi, Y., Ho, C. K., Li, H. & Shi, P.-Y. (2007). *J. Virol.* **81**, 4412–4421.

Dong, H., Ren, S., Zhang, B., Zhou, Y., Puig-Basagoiti, F., Li, H. & Shi, P.-Y. (2008). *J. Virol.* **82**, 4295–4307.

Dong, H., Zhang, B. & Shi, P.-Y. (2008). *Antivir. Res.* **80**, 1–10.

Egloff, M.-P., Benarroch, D., Selisko, B., Romette, J. L. & Canard, B. (2002). *EMBO J.* **21**, 2757–2768.

Egloff, M.-P., Decroly, E., Malet, H., Selisko, B., Benarroch, D., Ferron, F. & Canard, B. (2007). *J. Mol. Biol.* **372**, 723–736.

Engh, R. A. & Huber, R. (1991). *Acta Cryst.* A**47**, 392–400.

Evans, P. R. (1993). *Proceedings of the CCP4 Study Weekend. Data Collection and Processing*, edited by L. Sawyer, N. Isaacs & S. Bailey, pp. 114–122. Warrington: Daresbury Laboratory.

Fauman, E. B., Blumenthal, R. M. & Cheng, X. (1999). *S-Adenosylmethionine-Dependent Methyltransferases*, edited by X. Cheng & R. M. Blumenthal, pp. 1–7. Singapore: World Scientific.

Furuichi, Y. & Shatkin, A. J. (2000). *Adv. Virus Res.* **55**, 135–184.

Gould, E. A., de Lamballerie, X., Zanotto, P. M. & Holmes, E. C. (2003). *Adv. Virus Res.* **59**, 277–314.

Harris, M. & Jones, T. A. (2001). *Acta Cryst.* D**57**, 1201–1203.

Hodel, A. E., Gershon, P. D., Shi, X. & Quiocho, F. A. (1996). *Cell*, **85**, 247–256.

Holm, L., Kääriäinen, S., Rosenström, P. & Schenkel, A. (2008). *Bioinformatics*, **24**, 2780–2781.

Holm, L. & Sander, C. (1993). *J. Mol. Biol.* **233**, 123–138.

Johnson, H. N. (1967). *Jpn. J. Med. Sci. Biol.* **20**, 160–166.

Jones, S. & Thornton, J. M. (1996). *Proc. Natl Acad. Sci. USA*, **93**, 13–20.

Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.

Kleywegt, G. J. & Jones, T. A. (1996). *Structure*, **4**, 1395–1400.

Kleywegt, G. J. & Jones, T. A. (1997). *Methods Enzymol.* **277**, 525–545.

Krissinel, E. & Henrick, K. (2004). *Acta Cryst.* D**60**, 2256–2268.

Kuno, G., Chang, G.-J. J., Tsushiya, K. R., Karabatsos, N. & Cropp, C. B. (1998). *J. Virol.* **72**, 73–83.

Lamzin, V. S., Perrakis, A. & Wilson, K. S. (2001). *International Tables for Crystallography*, Vol. *F*, edited by M. G. Rossmann & E. Arnold, pp. 720–722. Dordrecht: Kluwer Academic Publishers.

Lee, B. & Richards, F. M. (1971). *J. Mol. Biol.* **55**, 379–400.

Leslie, A. G. W. (1999). *Acta Cryst.* D**55**, 1696–1702.

Leyssen, P., Croes, R., Rau, P., Heiland, S., Verbeken, E., Sciot, R., Paeshuyse, J., Charlier, N., De Clercq, E., Meyding-Lamadé, U. & Neyts, J. (2003). *Brain Pathol.* **13**, 279–290.

Leyssen, P., Paeshuyse, J., Charlier, N., Van Lommel, A., Drosten, C., De Clercq, E. & Neyts, J. (2003). *J. Neurovirol.* **9**, 69–78.

Leyssen, P., Van Lommel, A., Drosten, C., Schmitz, H., De Clercq, E. & Neyts, J. (2001). *Virology*, **279**, 27–37.

Mastrangelo, E., Bollati, M., Milani, M., Selisko, B., Peyrane, F., Canard, B., Grard, G., de Lamballerie, X. & Bolognesi, M. (2007). *Protein Sci.* **16**, 1133–1145.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.

Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.

Pannu, N. S., Murshudov, G. N., Dodson, E. J. & Read, R. J. (1998). *Acta Cryst.* D**54**, 1285–1294.

Ray, D., Shah, A., Tilgner, M., Guo, Y., Zhao, Y., Dong, H., Deas, T. S., Zhou, Y., Li, H. & Shi, P.-Y. (2006). *J. Virol.* **80**, 8362–8370.
Read, R. J. (1986). *Acta Cryst.* A**42**, 140–149.
Schluckebier, G., Kozak, M., Bleimling, N., Weinhold, E. & Saenger, W. (1997). *J. Mol. Biol.* **265**, 56–67.
Varelas-Wesley, I. & Calisher, C. H. (1982). *Am. J. Trop. Med. Hyg.* **31**, 1273–1284.
Vidgren, J., Svensson, L. A. & Liljas, A. (1994). *Nature (London)*, **368**, 354–358.
Zhou, Y., Ray, D., Zhao, Y., Dong, H., Ren, S., Li, Z., Guo, Y., Bernard, K. A., Shi, P.-Y. & Li, H. (2007). *J. Virol.* **81**, 3891–3903.